Adaptive Bias Correction for Improved Subseasonal Forecasting

Soukayna Mouatadid [1] **Paulo Orenstein** [2] **Genevieve Flaspohler** [3 4 5] **Judah Cohen** [6 7] **Miruna Oprescu** [8] **Ernest Fraenkel** [9] **Lester Mackey** [10]

Abstract

Subseasonal forecasting—predicting temperature and precipitation 2 to 6 weeks ahead—is critical for effective water allocation, wildfire management, and drought and flood mitigation. Recent international research efforts have advanced the subseasonal capabilities of operational dynamical models, yet temperature and precipitation prediction skills remain poor, partly due to stubborn errors in representing atmospheric dynamics and physics inside dynamical models. Here, to counter these errors, we introduce an *adaptive bias correction* (ABC) method that combines state-of-the-art dynamical forecasts with observations using machine learning. We show that, when applied to the leading subseasonal model from the European Centre for Medium-Range Weather Forecasts (ECMWF), ABC improves temperature forecasting skill by 60-90% (over baseline skills of 0.18-0.25) and precipitation forecasting skill by 40-69% (over baseline skills of 0.11-0.15) in the contiguous U.S. We couple these performance improvements with a practical workflow to explain ABC skill gains and identify higher-skill windows of opportunity based on specific climate conditions.

[1]Department of Computer Science, University of Toronto, Toronto, ON, Canada [2]Instituto de Matemática Pura e Aplicada, Rio de Janeiro, Brazil [3]*n*Line Inc., Berkeley, CA, USA [4]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA [5]Department of Applied Ocean Science and Engineering, Woods Hole Oceanographic Institution, Falmouth, MA, USA [6]Atmospheric and Environmental Research, Lexington, MA, USA [7]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA [8]Department of Computer Science, Cornell University, Ithaca, NY, USA [9]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA [10]Microsoft Research New England, Cambridge, MA, USA. Correspondence to: Soukayna Mouatadid <soukayna@cs.toronto.edu>.

Accepted after peer-review at the 1st workshop on Synergy of Scientific and Machine Learning Modeling, SynS & ML ICML, Honolulu, Hawaii, USA. July, 2023.

Water and fire managers rely on subseasonal forecasts 2-6 weeks in advance to allocate water, manage wildfires, and prepare for droughts and other weather extremes. However, skillful forecasts for the subseasonal regime are lacking due to the complex dependence on both local weather and global climate variables and the chaotic nature of weather. Bridging the gap between short-term and seasonal forecasting has been the focus of several recent large-scale research efforts which have advanced the subseasonal capabilities of operational physics-based models (Vitart et al., 2017; Pegion et al., 2019; Lang et al., 2020). However, despite these advances, dynamical models still suffer from persistent systematic errors, which limit the skill of temperature and precipitation forecasts for longer lead times from 2 to 6 weeks ahead.

To overcome observed systematic errors of physics-based models on the subseasonal timescale, there have been parallel efforts in recent years to demonstrate the value of machine learning and deep learning methods in improving subseasonal forecasting (Li et al., 2016; Cohen et al., 2019; Hwang et al., 2019; Arcomano et al., 2020; He et al., 2020; Yamagami & Matsueda, 2020; Wang et al., 2021; Kim et al., 2021; Watson-Parris, 2021; Weyn et al., 2021; Srinivasan et al., 2021). While these works demonstrate the promise of statistical models for subseasonal forecasting, they also highlight the complementary strengths of physics- and learning-based approaches and the opportunity to combine those strengths to improve forecasting skill (Hwang et al., 2019; Kim et al., 2021).

To harness those complementary strengths, we introduce a hybrid dynamical-learning framework for improved subseasonal forecasting. In particular, we learn to adaptively correct the biases of dynamical models and apply our novel *adaptive bias correction* (ABC) to improve the skill of subseasonal temperature and precipitation forecasts. ABC can be applied operationally as a computationally inexpensive enhancement to any dynamical model forecast, and we use this property to substantially reduce the forecasting errors of eight operational dynamical models, including the state-of-the-art ECMWF model. We couple these performance improvements with a practical workflow for explaining ABC skill gains using Cohort Shapley (Mase et al., 2019) and identifying higher-skill windows of opportunity (Mariotti et al., 2020) based on relevant climate variables. To facilitate future deployment and benchmarking, we release our model and workflow code through the `Redacted` Python package.

1. Methods

We consider two prediction targets: average temperature (°C) and accumulated precipitation (mm) over a two-week period. These variables are forecasted at two time horizons: 15-28 days ahead (weeks 3-4) and 29-42 days ahead (weeks 5-6). We forecast each variable at $G = 376$ grid points on a $1.5° \times 1.5°$ grid across the contiguous U.S., bounded by latitudes 25N to 50N and longitudes 125W to 67W. To provide the most realistic assessment of forecasting skill (Risbey et al., 2021), all predictions in this study are formed in a real forecast manner that mimics operational use. In particular, to produce a forecast for a given target date, all learning-based models are trained and tuned only on data observable on the corresponding forecast issuance date. We evaluate each forecast according using uncentered anomaly correlation skill. For a collection of target dates, we report average skill using progressive validation (Blum et al., 1999) to mimic operational use. All data used in this work was obtained from the `Redacted` dataset (`Redacted`).

Our proposed adaptive bias correction (ABC) is a uniformly-weighted ensemble of our three machine learning models, Climatology++, Dynamical++, and Persistence++. Climatology++ predicts the historical mean or geographic median over all days in a window around the target day of year. The number of training years and the size of the observation window are determined adaptively using an automated tuning procedure. Dynamical++ is a learned correction for raw dynamical forecasts. After averaging dynamical forecasts over a range of issuance dates and lead times, Dynamical++ debiases the ensemble forecast by adding the mean value of the target variable and subtracting the mean forecast over a learned window of observations around the target day of year. Unlike standard debiasing strategies, which employ static ensembling and bias correction, Dynamical++ adaptively selects the range of ensembled lead times, the number of averaged issuance dates, and the size of the observation window using an automated tuning procedure. Persistence++ fits a least squares regression per grid point to optimally combine climatology, recent weather trends in the form of lagged temperature or precipitation measurements, and a dynamical ensemble forecast.

2. Results

Figure 1 highlights the advantage of ABC over raw dynamical models when forecasting accumulated precipitation and averaged temperature in the contiguous U.S. Here, ABC is

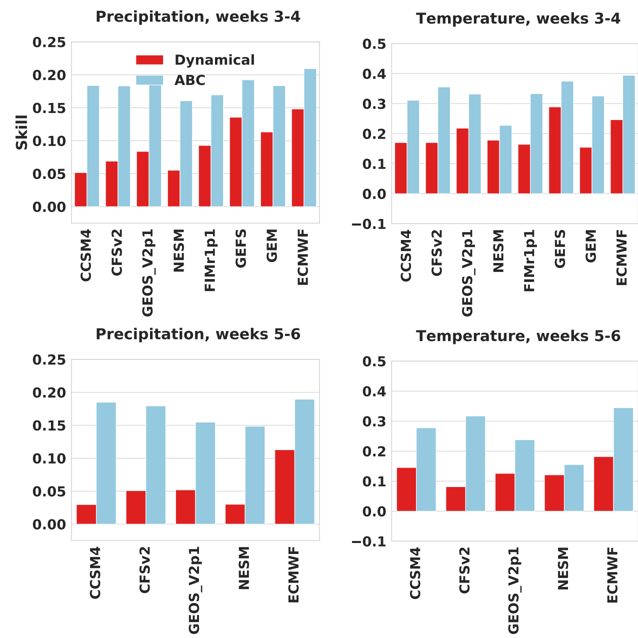

Figure 1. Average model skill for ECWMF and SubX dynamical models (red) and their ABC-corrected counterparts (blue) across the contiguous U.S. and the years 2018–2021. For each forecasting task and dynamical model input, ABC provides a pronounced improvement in skill.

applied to the leading subseasonal model, ECMWF, and to each of seven operational models participating in the Subseasonal Experiment (SubX, Pegion et al., 2019). Subseasonal forecasting skill, measured by uncentered anomaly correlation, is evaluated at two forecast horizons, weeks 3-4 and weeks 5-6, and averaged over all available forecast dates in the four-year span 2018–2021. We find that, for each dynamical model input and forecasting task, ABC leads to a pronounced improvement in skill. For example, when applied to the U.S. operational model CFSv2, ABC improves temperature forecasting skill by 109-289% and precipitation skill by 165-253%. When applied to the leading ECMWF model, ABC improves temperature skill by 60-90% and precipitation skill by 40-69%. Moreover, for precipitation, even lower-skill models like CCSM4 enjoy skill comparable to the best after the application of ABC. Overall and despite significant variability in dynamical model skill, ABC consistently reduces the systematic errors of its input model, bringing forecasts closer to observations for each target variable and time horizon.

The results presented so far assess overall model skill, averaged across all forecast dates. However, there is a growing appreciation that subseasonal forecasts can also benefit from selective deployment during "windows of opportunity," periods defined by observable climate conditions in which specific forecasters are likely to have higher skill (Mariotti et al., 2020). In this section, we propose a practical *opportunistic ABC workflow* that uses a candidate set of explanatory variables to identify windows in which ABC is especially likely to improve upon a baseline model. The same workflow can be used to explain the skill improvements achieved by ABC in terms of the explanatory variables.

The opportunistic ABC workflow is based on the optimal credit assignment principle (Shapley, 1953) and measures the impact of explanatory variables on individual forecasts using Cohort Shapley (Mase et al., 2019) and overall variable importance using Shapley effects (Song et al., 2016). We use these Shapley measures to interpret the contexts in which ABC offers improvements in terms of climate variables with known relevance for subseasonal forecasting skill.

As a running example, we use our workflow to explain the skill differences between ABC-ECMWF and debiased ECMWF when predicting precipitation in weeks 3-4. As our candidate explanatory variables we use Northern Hemisphere geopotential heights (HGT) at 500 and 10 hPa, the phase of the Madden-Julian Oscillation (MJO), Northern Hemisphere sea ice concentration (ICEC), global sea surface temperatures (SST), the multivariate El Niño-Southern Oscillation index (MEI.v2, Wolter & Timlin, 1993), and the target month. All variables are lagged appropriately to ensure that they are observable on the forecast issuance date.

We first use Shapley effects to determine the overall importance of each variable in explaining the precipitation skill improvements of ABC-ECMWF. We find the most important explanatory variables to be the first two principal components (PCs) of 500 hPa geopotential height, the MJO phase, the second PC of 10 hPa geopotential height, and the first PC of sea ice concentration. These variables are consistent with the literature exploring the dominant contributions to subseasonal precipitation (Chevallier et al., 2019).

We next use Cohort Shapley to identify the contexts in which each variable has the greatest impact on skill. For example, Figure 2 summarizes the impact of the first 500 hPa geopotential heights PC (`hgt_500_pc1`) on ABC-ECMWF skill improvement. This display divides our forecasts into 10 bins, determined by the deciles of `hgt_500_pc1`, and computes the probability of positive impact in each bin. We find that `hgt_500_pc1` is most likely to have a positive impact impact on skill improvement in decile 1, which features a positive Arctic Oscillation (AO) pattern, and least likely in decile 9, which features AO in the opposite phase. The ABC-ECMWF forecast most impacted by `hgt_500_pc1` in decile 1 is also preceded by a positive AO pattern and replaces the wet debiased ECMWF forecast with a more skillful dry pattern in the west.

Finally, we use the identified contexts to define windows of opportunity for operational deployment. Indeed, since all explanatory variables are observable on the forecast issuance date, one can selectively apply ABC when multiple variables are likely to have a positive impact on skill and otherwise issue a default, standard forecast (e.g., debiased ECMWF). We call this selective forecasting model *opportunistic ABC*. How many high-impact variables should we require when defining these windows of opportunity? Requiring a larger number of high-impact variables will tend to increase the skill gains of ABC but simultaneously reduce the number of dates on which ABC is deployed. Figure 3 illustrates this trade-off for ABC-ECMWF and shows that opportunistic ABC skill is maximized when two or more high-impact variables are required. With this choice, ABC is used for approximately 81% of forecasts and debiased ECMWF is used for the remainder.

3. Discussion and Conclusion

Dynamical models have shown increasing skill in accurately forecasting the weather (Bauer et al., 2015), but they still contain systematic biases that compound on subseasonal time scales and suppress forecast skill. ABC learns to correct these biases by adaptively integrating dynamical forecasts, historical observations, and recent weather trends. Our approach substantially reduces the forecasting errors of the leading subseasonal model from ECMWF and seven additional operational subseasonal forecasting models, with less skillful input models performing nearly as well as the ECMWF model after applying the ABC correction. This finding suggests that systematic errors in dynamical models are a primary contributor to observed skill differences and that ABC provides an effective mechanism for reducing these heterogeneous errors. Because ABC is also simple to implement and deploy in real-time operational settings, adaptive bias correction represents a computationally inexpensive strategy for upgrading operational models, while conserving valuable human resources.

While the learned correction of systematic errors can play an important role in skill improvement, it is no substitute for scientific improvements in our understanding and representation of the processes underlying subseasonal predictability. As such, we view ABC as a complement for improved dynamical model development. Fortunately, ABC is designed to be adaptive to model changes. As operational models are upgraded, process models improve, and systematic biases evolve, our ABC training protocol is designed to ingest the upgraded model forecasts and hindcasts reflecting those changes.

To capitalize on higher-skill forecasts of opportunity, we have also introduced an opportunistic ABC workflow that explains the skill improvements of ABC in terms of a candidate

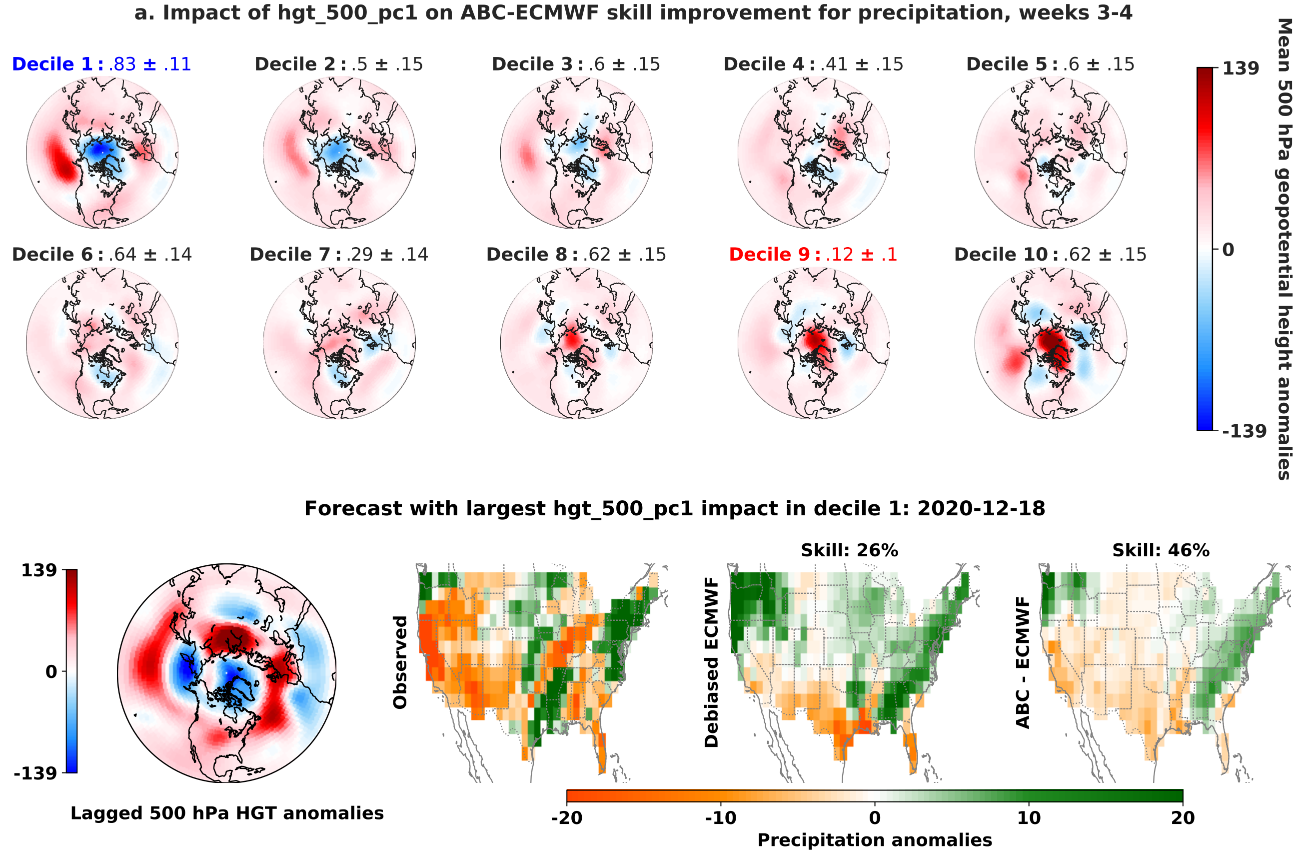

Figure 2. **Top:** To summarize the impact of `hgt_500_pc1`on ABC-ECMWF skill improvement for precipitation weeks 3-4, we divide our forecasts into 10 bins, determined by the deciles of `hgt_500_pc1`, and compute the probability of positive impact in each bin, as shown above each bin map. The highest probabilities of positive impact are shown in blue and the lowest probabilities of positive impact are shown in red. We find that `hgt_500_pc1`is most likely to have a positive impact on skill improvement in decile 1, which features a positive Arctic Oscillation (AO) pattern, and least likely in decile 9, which features AO in the opposite phase. **Bottom:** The forecast most impacted by `hgt_500_pc1`in decile 1 is also preceded by a positive AO pattern and replaces the wet debiased ECMWF forecast with a more skillful dry pattern in the west.

# High-impact variables	% Forecasts using ABC	High-impact skill (%) ABC	High-impact skill (%) Debiased
0 or more	100.00	20.94	15.28
1 or more	95.93	20.99	14.84
2 or more	80.62	22.29	13.12
3 or more	58.61	23.56	12.00
4 or more	31.82	24.72	8.18
5 or more	14.59	26.51	8.35
6 or more	6.46	29.72	10.55
7 or more	2.15	45.00	17.53

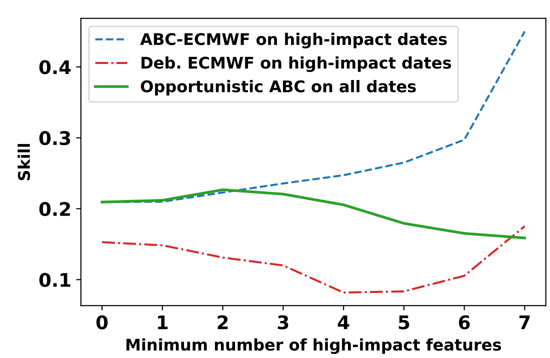

Figure 3. Defining windows of opportunity for opportunistic ABC forecasting of precipitation weeks 3-4. **Left:** When more explanatory variables fall into high-impact deciles or bins (e.g., the blue bins of Figure 2), the mean skill of ABC-ECMWF improves, but the percentage of forecasts using ABC declines. **Right:** The overall skill of opportunistic ABC is maximized when ABC-ECMWF is deployed for target dates with two or more high-impact variables and standard debiased ECMWF is deployed otherwise.

set of environmental variables, identifies high-probability windows of opportunity based on those variables, and selectively deploys either ABC or a baseline forecast to maximize expected skill. The same workflow can be applied to explain the skill improvements of any forecasting model and, unlike other popular explanation tools (e.g., Ribeiro et al., 2016; Lundberg & Lee, 2017), avoids expensive model retraining, requires no generation of additional forecasts beyond those routinely generated for operational or hindcast use, and allows for explanations in terms of variables that were not explicitly used in training the model.

Overall, we find that correcting dynamical forecasts using ABC yields an effective and scalable strategy to optimize the skill of the next generation of subseasonal forecasting models. We anticipate that our hybrid dynamical-learning framework will benefit both research and operations, and we release our open-source code to facilitate future adoption and development.

Broader impact

In the past decade, extreme weather events such as heatwaves, drought and floods, have affected millions and cost billions (Zhongming et al., 2021). Now, more than ever, improving our ability to forecast the weather and predict the climate is of major interest to all sectors of the economy and government agencies from the local to the national level. Weather forecasts span intervals 0-to-10 days ahead, and climate forecasts span intervals seasons-to-decades ahead; both are currently used operationally in decision-making, and their accuracy and reliability have improved consistently in recent decades. However, many critical applications, such as water allocation, wildfire management, and drought and flood mitigation, require subseasonal forecasts whose span intervals lie between these two extremes. Yet, skillful forecasts for the subseasonal regime are lacking due to the complex dependence on both local weather and global climate variables, as well as the chaotic nature of the weather. To address this need, we introduce the machine learning approach of *adaptive bias correction* (ABC) that combines state-of-the-art dynamical forecasts with observations. We anticipate that our hybrid dynamical-learning paradigm for subseasonal forecasting will benefit both research and operations. We additionally release our open-source ABC code to facilitate widespread adoption and future development.

References

Arcomano, T., Szunyogh, I., Pathak, J., Wikner, A., Hunt, B. R., and Ott, E. A machine learning-based global atmospheric forecast model. *Geophysical Research Letters*, 47(9):e2020GL087776, 2020.

Bauer, P., Thorpe, A., and Brunet, G. The quiet revolution of numerical weather prediction. *Nature*, 525(7567):47–55, 2015.

Blum, A., Kalai, A., and Langford, J. Beating the hold-out: Bounds for k-fold and progressive cross-validation. In *Proceedings of the twelfth annual conference on Computational learning theory*, pp. 203–208, 1999.

Chevallier, M., Massonnet, F., Goessling, H., Guémas, V., and Jung, T. Chapter 10 - the role of sea ice in sub-seasonal predictability. In Robertson, A. W. and Vitart, F. (eds.), *Sub-Seasonal to Seasonal Prediction*, pp. 201–221. Elsevier, 2019. ISBN 978-0-12-811714-9. doi: https://doi.org/10.1016/B978-0-12-811714-9.00010-3. URL https://www.sciencedirect.com/science/article/pii/B9780128117149000103.

Cohen, J., Coumou, D., Hwang, J., Mackey, L., Orenstein, P., Totz, S., and Tziperman, E. S2s reboot: An argument for greater inclusion of machine learning in subseasonal to seasonal forecasts. *WIREs Climate Change*, 10(2):e00567, 2019. doi: https://doi.org/10.1002/wcc.567.

He, S., Li, X., DelSole, T., Ravikumar, P., and Banerjee, A. Sub-seasonal climate forecasting via machine learning: Challenges, analysis, and advances. *arXiv preprint arXiv:2006.07972*, 2020.

Hwang, J., Orenstein, P., Cohen, J., Pfeiffer, K., and Mackey, L. Improving subseasonal forecasting in the western U.S. with machine learning. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '19, pp. 2325–2335, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450362016. doi: 10.1145/3292500.3330674. URL https://doi.org/10.1145/3292500.3330674.

Kim, M., Yoo, C., and Choi, J. Enhancing subseasonal temperature prediction by bridging a statistical model with dynamical arctic oscillation forecasting. *Geophysical Research Letters*, 48(15):e2021GL093447, 2021.

Lang, A. L., Pegion, K., and Barnes, E. A. Introduction to special collection:"bridging weather and climate: Subseasonal-to-seasonal (S2S) prediction". *Journal of Geophysical Research: Atmospheres*, 125(4):e2019JD031833, 2020.

Li, L., Schmitt, R. W., Ummenhofer, C. C., and Karnauskas, K. B. Implications of North Atlantic sea surface salinity for summer precipitation over the us midwest: Mechanisms and predictive value. *Journal of Climate*, 29(9):3143–3159, 2016.

Lundberg, S. M. and Lee, S.-I. A unified approach to interpreting model predictions. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

Mariotti, A., Baggett, C., Barnes, E. A., Becker, E., Butler, A., Collins, D. C., Dirmeyer, P. A., Ferranti, L., Johnson, N. C., Jones, J., Kirtman, B. P., Lang, A. L., Molod, A., Newman, M., Robertson, A. W., Schubert, S., Waliser, D. E., and Albers, J. Windows of opportunity for skillful forecasts subseasonal to seasonal and beyond. *Bulletin of the American Meteorological Society*, 101(5):E608 – E625, 2020. doi: 10.1175/BAMS-D-18-0326.1.

Mase, M., Owen, A. B., and Seiler, B. Explaining black box decisions by Shapley cohort refinement, 2019. URL https://arxiv.org/abs/1911.00467.

Pegion, K., Kirtman, B. P., Becker, E., Collins, D. C., LaJoie, E., Burgman, R., Bell, R., DelSole, T., Min, D., Zhu, Y., et al. The subseasonal experiment (subx): A multimodel subseasonal prediction experiment. *Bulletin of the American Meteorological Society*, 100(10):2043–2060, 2019.

Ribeiro, M. T., Singh, S., and Guestrin, C. "why should i trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pp. 1135–1144, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450342322. doi: 10.1145/2939672.2939778.

Risbey, J. S., Squire, D. T., Black, A. S., DelSole, T., Lepore, C., Matear, R. J., Monselesan, D. P., Moore, T. S., Richardson, D., Schepen, A., Tippett, M. K., and Tozer, C. R. Standard assessments of climate forecast skill can be misleading. *Nature Communications*, 12(1):4346, 2021. doi: 10.1038/s41467-021-23771-z.

Shapley, L. S. A value for n-person games. In Kuhn, H. W. and Tucker, A. W. (eds.), *Contributions to the Theory of Games (AM-28), Volume II*, pp. 307–318. Princeton University Press, Princeton, 1953. ISBN 9781400881970. doi: doi:10.1515/9781400881970-018.

Song, E., Nelson, B. L., and Staum, J. Shapley effects for global sensitivity analysis: Theory and computation. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1): 1060–1083, 2016.

Srinivasan, V., Khim, J., Banerjee, A., and Ravikumar, P. Subseasonal climate prediction in the western US using Bayesian spatial models. In de Campos, C. and Maathuis, M. H. (eds.), *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, volume 161 of *Proceedings of Machine Learning Research*, pp. 961–970. PMLR, 27–30 Jul 2021.

Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., Déqué, M., Ferranti, L., Fucile, E., Fuentes, M., et al. The subseasonal to seasonal (S2S) prediction project database. *Bulletin of the American Meteorological Society*, 98(1):163–173, 2017.

Wang, C., Jia, Z., Yin, Z., Liu, F., Lu, G., and Zheng, J. Improving the accuracy of subseasonal forecasting of china precipitation with a machine learning approach. front. *Earth Sci*, 9:659310, 2021.

Watson-Parris, D. Machine learning for weather and climate are worlds apart. *Philosophical Transactions of the Royal Society A*, 379(2194):20200098, 2021.

Weyn, J. A., Durran, D. R., Caruana, R., and Cresswell-Clay, N. Sub-seasonal forecasting with a large ensemble of deep-learning weather prediction models. *Journal of Advances in Modeling Earth Systems*, 13(7): e2021MS002502, 2021. doi: https://doi.org/10.1029/2021MS002502.

Wolter, K. and Timlin, M. S. Monitoring ENSO in COADS with a seasonally adjusted principal component index. In *Proc. of the 17th Climate Diagnostics Workshop*, pp. 52–57, 1993.

Yamagami, A. and Matsueda, M. Subseasonal forecast skill for weekly mean atmospheric variability over the northern hemisphere in winter and its relationship to midlatitude teleconnections. *Geophysical Research Letters*, 47(17): e2020GL088508, 2020.

Zhongming, Z., Linong, L., Xiaona, Y., Wangqiang, Z., Wei, L., et al. Wmo provisional report on the state of the global climate 2021. 2021.