Improving Subseasonal Forecasting with Machine Learning

Lester Mackey

Microsoft Research New England

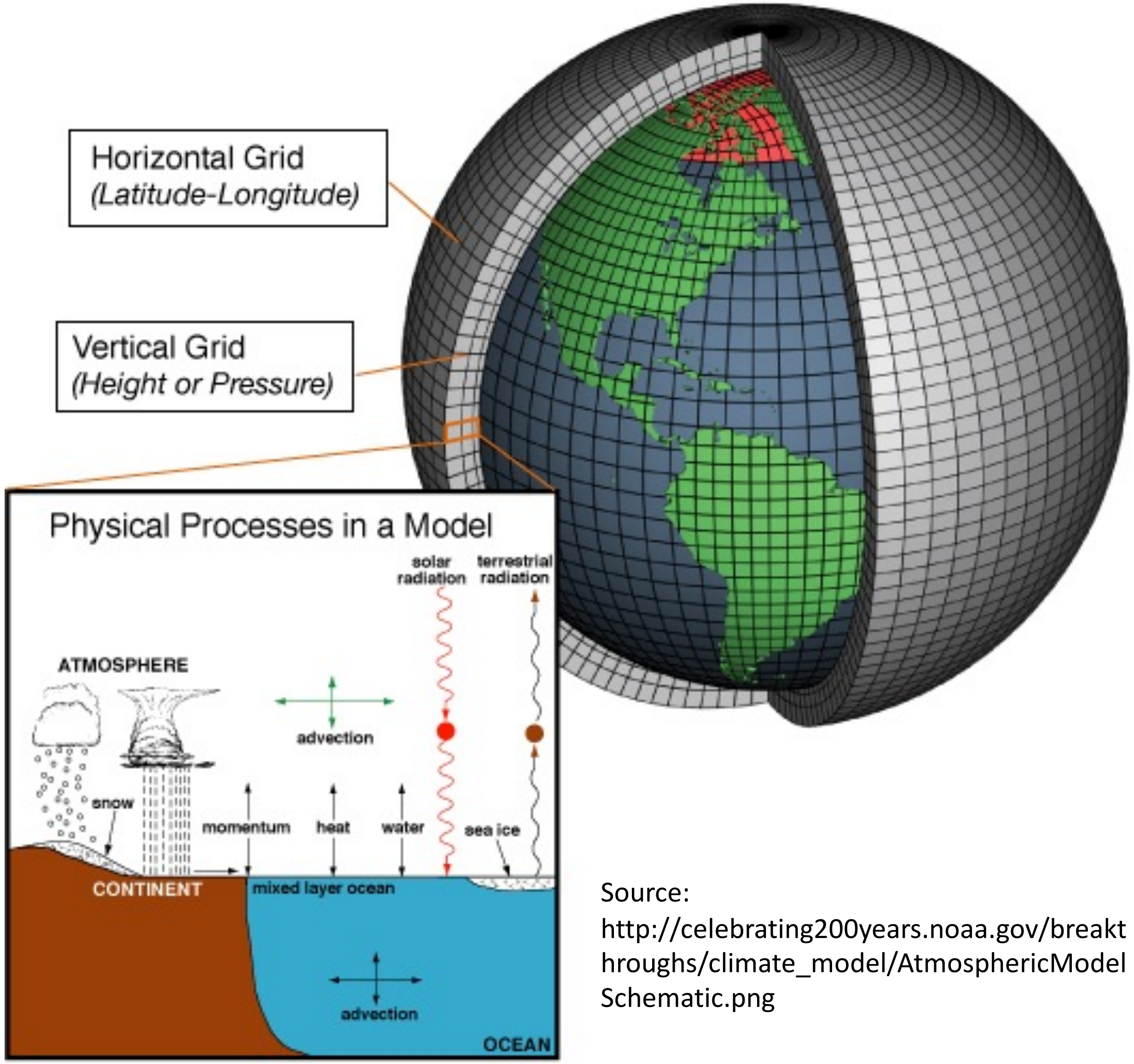
Joint work with **Judah Cohen, Jessica Hwang, Paulo Orenstein, Soukayna Mouatadid, Genevieve Flaspohler, Sonja Totz, Miruna Oprescu, Franklyn Wang, Sean Knight, Maria Geogdzhayeva, Sam Levang, Karl Pfeiffer, Ernest Fraenkel**

Judah Cohen

- Climatologist, director of seasonal forecasting at Atmospheric and Environmental Research
- Concern: Community not making the best use of historical data in weather / climate forecasting
  - Landscape dominated by **dynamical models**, purely physics-based models of atmospheric and oceanic evolution

Dynamical Models

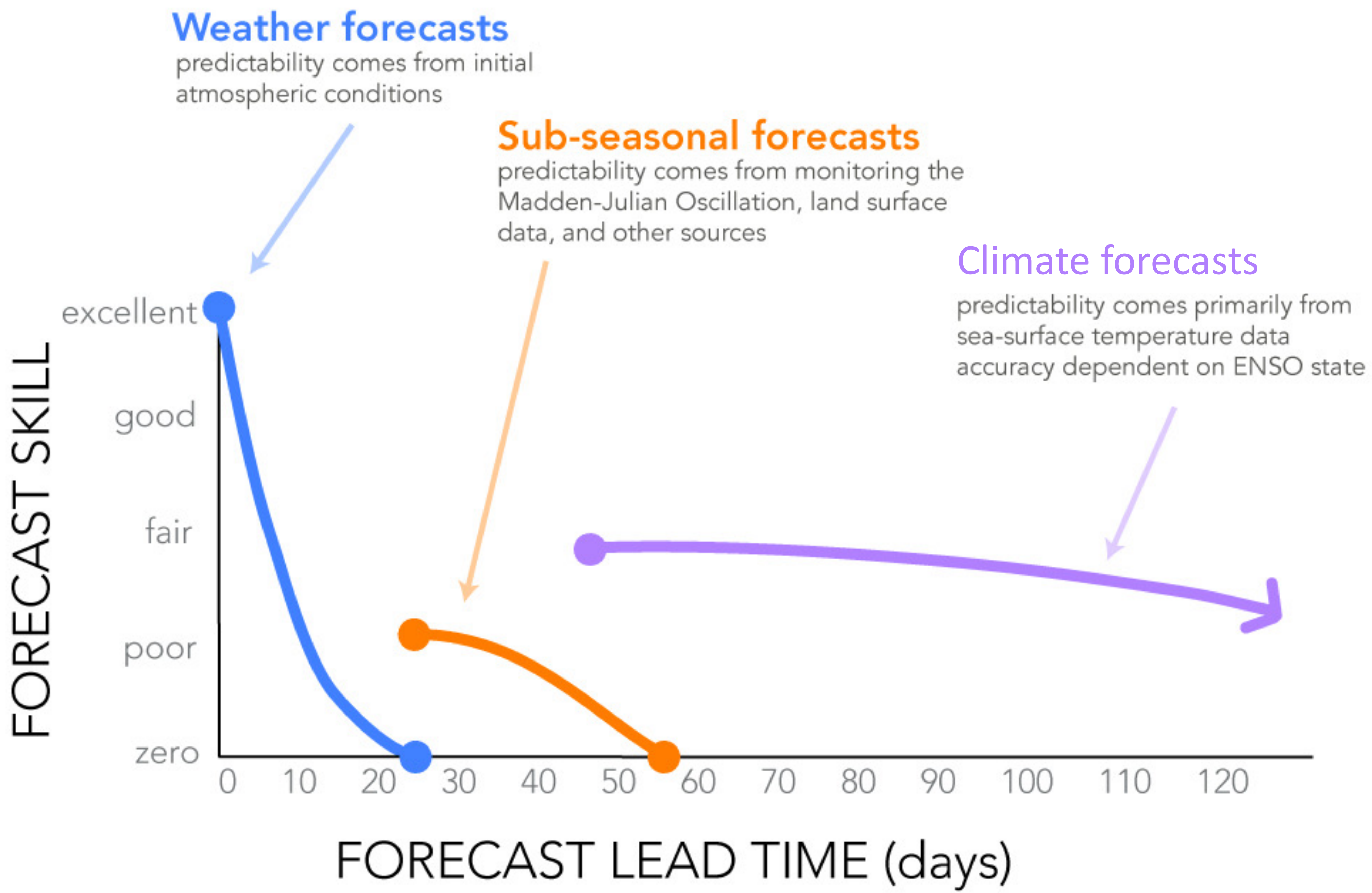
- Initialized with current weather conditions estimated from measurements
- Simulate future weather / climate by discretizing partial differential equations using supercomputers
- Accuracy limited by chaotic nature: errors in inputs rapidly amplified
- Ensembles with varying initial conditions / model parameters often formed to capture uncertainty
- Sometimes *debiased* by comparing predictions to truth over recent years



Source: http://celebrating200years.noaa.gov/breakthroughs/climate_model/AtmosphericModelSchematic.png

Judah Cohen

- Climatologist, director of seasonal forecasting at Atmospheric and Environmental Research
- Concern: Community not making the best use of historical data in weather / climate forecasting
  - Landscape dominated by **dynamical models**, purely physics-based models of atmospheric and oceanic evolution
- Concern: **Subseasonal forecasts** especially poor



Source: https://iri.columbia.edu/news/qa-subseasonal-prediction-project/

Subseasonal Forecasting: What and Why?

- **What:** Predicting temperature and precipitation 2 – 6 weeks out
- **Why:** (White et al., 2017, Meteorological Applications)
    - Allocating water resources
    - Managing wildfires
    - Preparing for weather extremes
        - e.g., droughts, heavy rainfall, and flooding
    - Crop planting, irrigation scheduling, and fertilizer application
    - Energy pricing

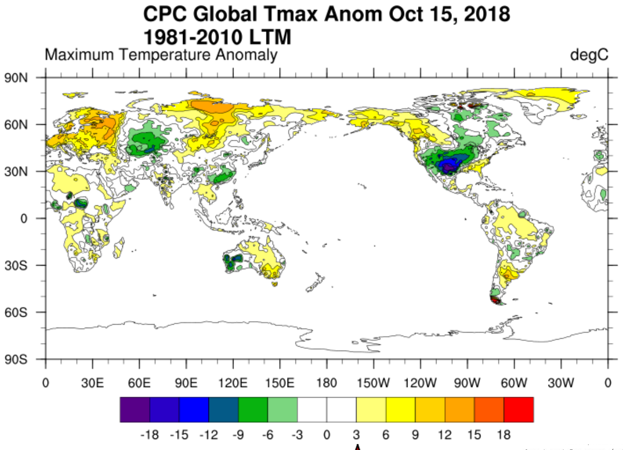
U.S. Bureau of Reclamation

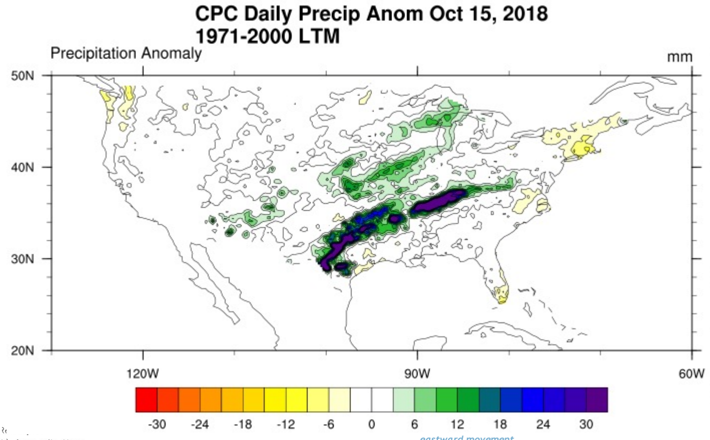
- "The mission of the [USBR] is to manage, develop, and protect water and related resources in an environmentally and economically sound manner in the interest of the American public."
- Manages water in 17 western states
  - Provides 1 out of 5 Western farmers with irrigation water for 10 million farmland acres
  - Generates enough electricity to power 3.5M U.S. homes
- "During the past eight years, every state in the Western United States has experienced drought that has affected the economy both locally and nationally through impacts to agricultural production, water supply, and energy."

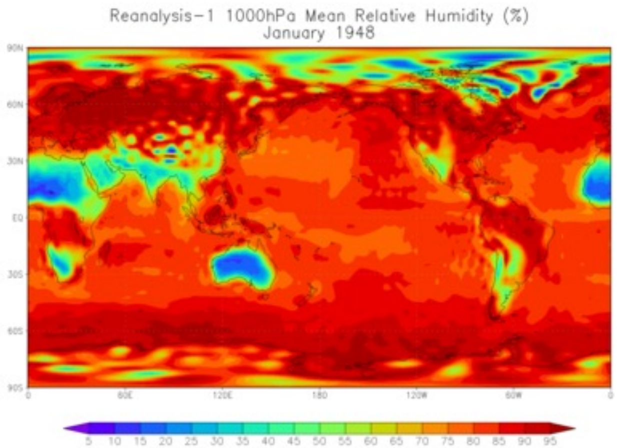
Credit: David Raff, USBR

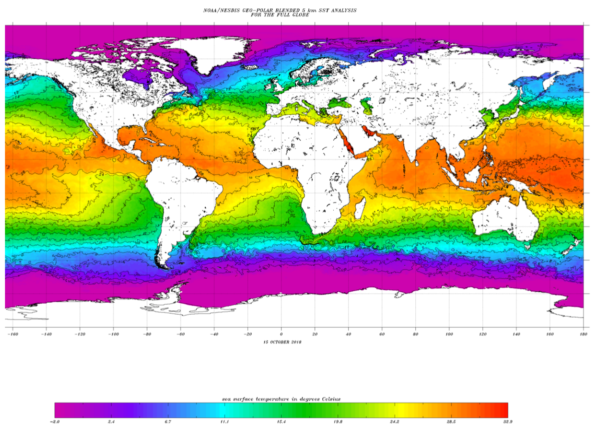
Our SubseasonalClimateUSA Dataset

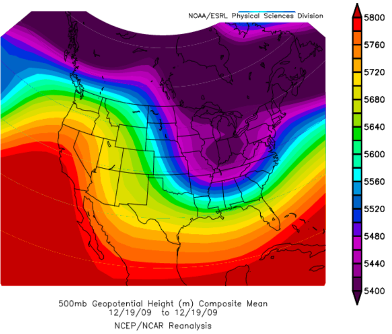
- To train and evaluate our models, we constructed a **SubseasonalClimateUSA dataset** from diverse data sources
- Updated daily + accessed via [subseasonal_data]() Python package

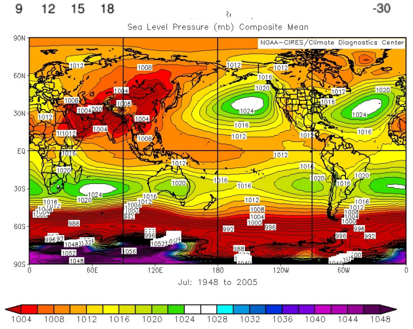


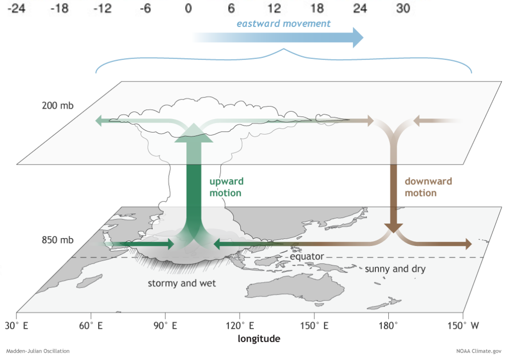


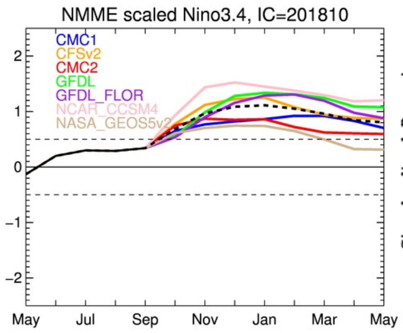


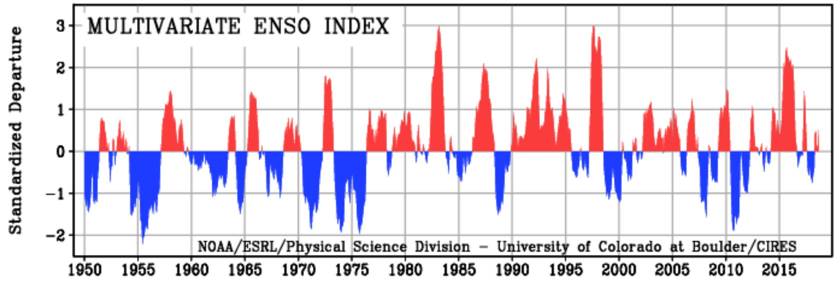




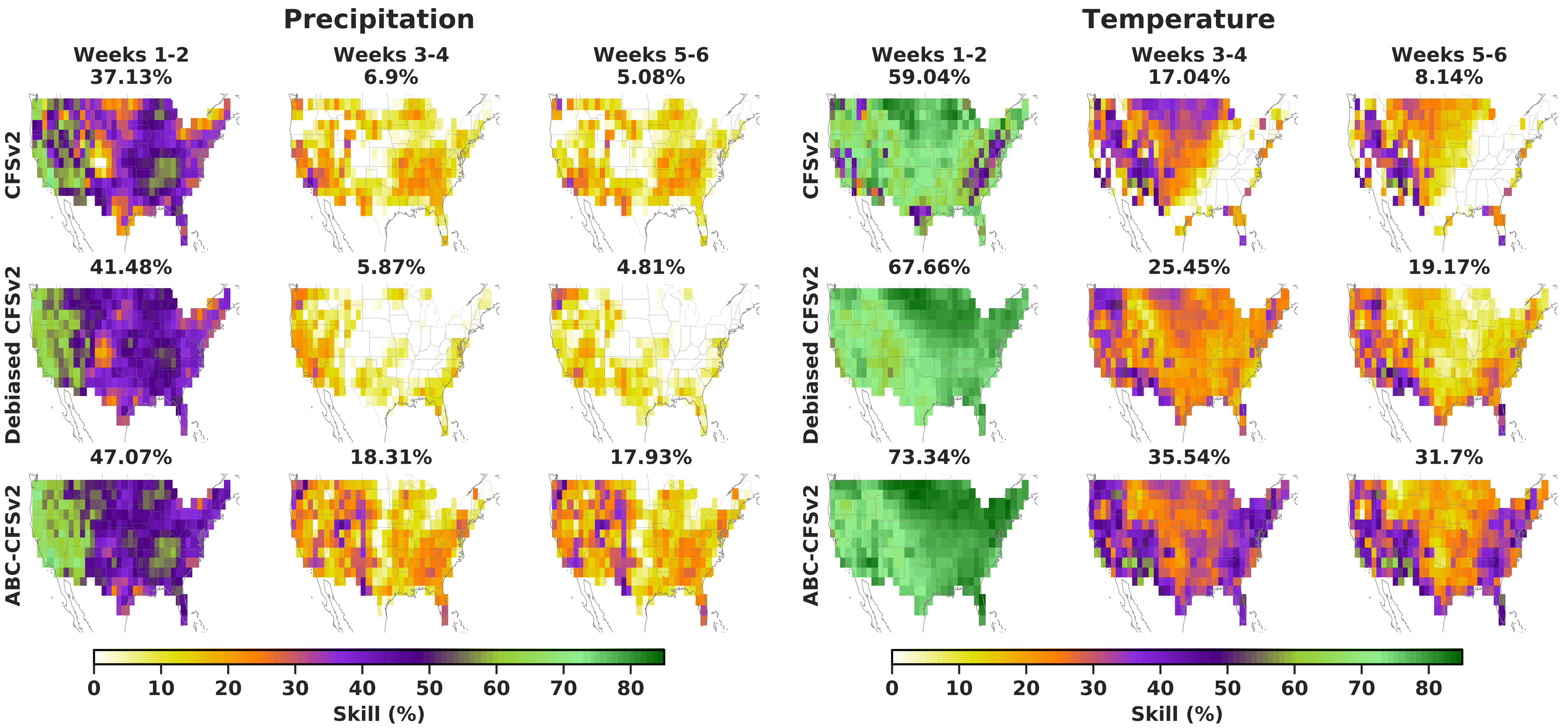






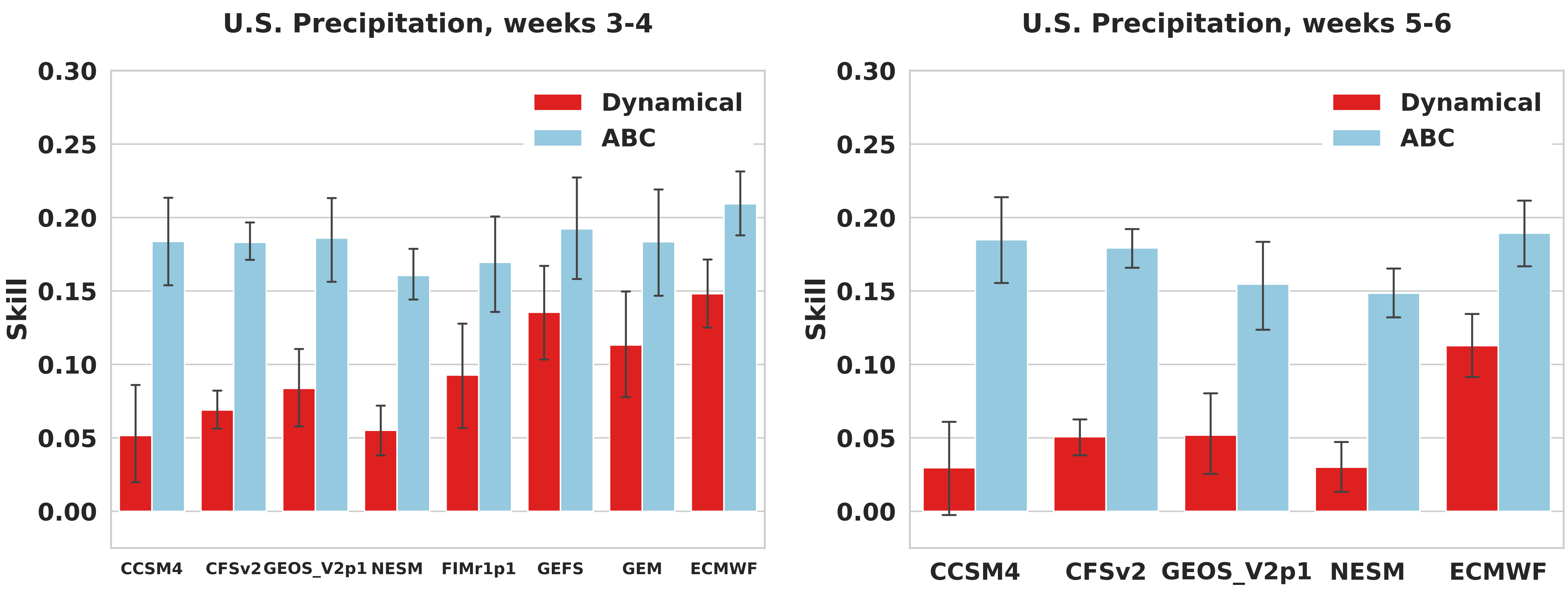


Adaptive Bias Correction (ABC): Hybrid Physics + Learning Model



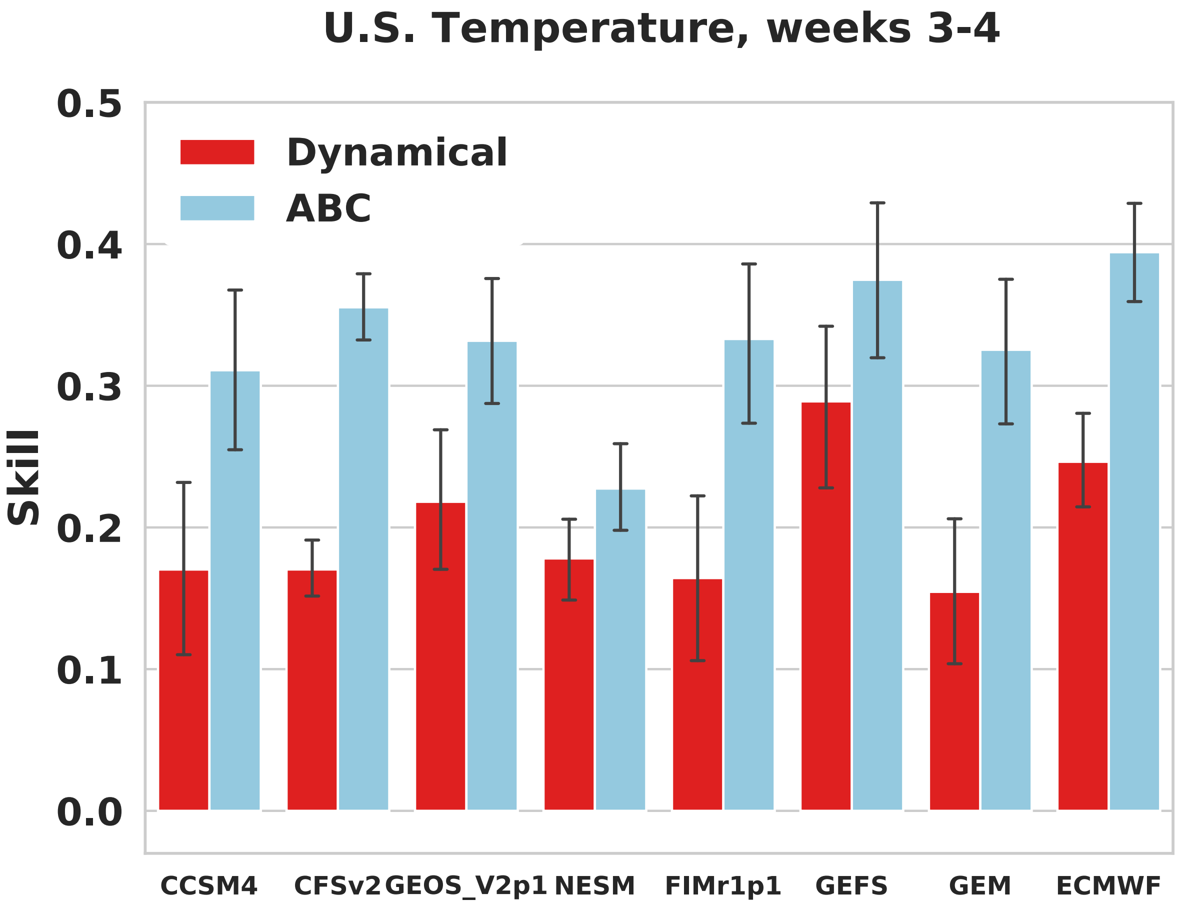
- Doubles or triples the forecasting skill of US operational dynamical model (CFSv2)

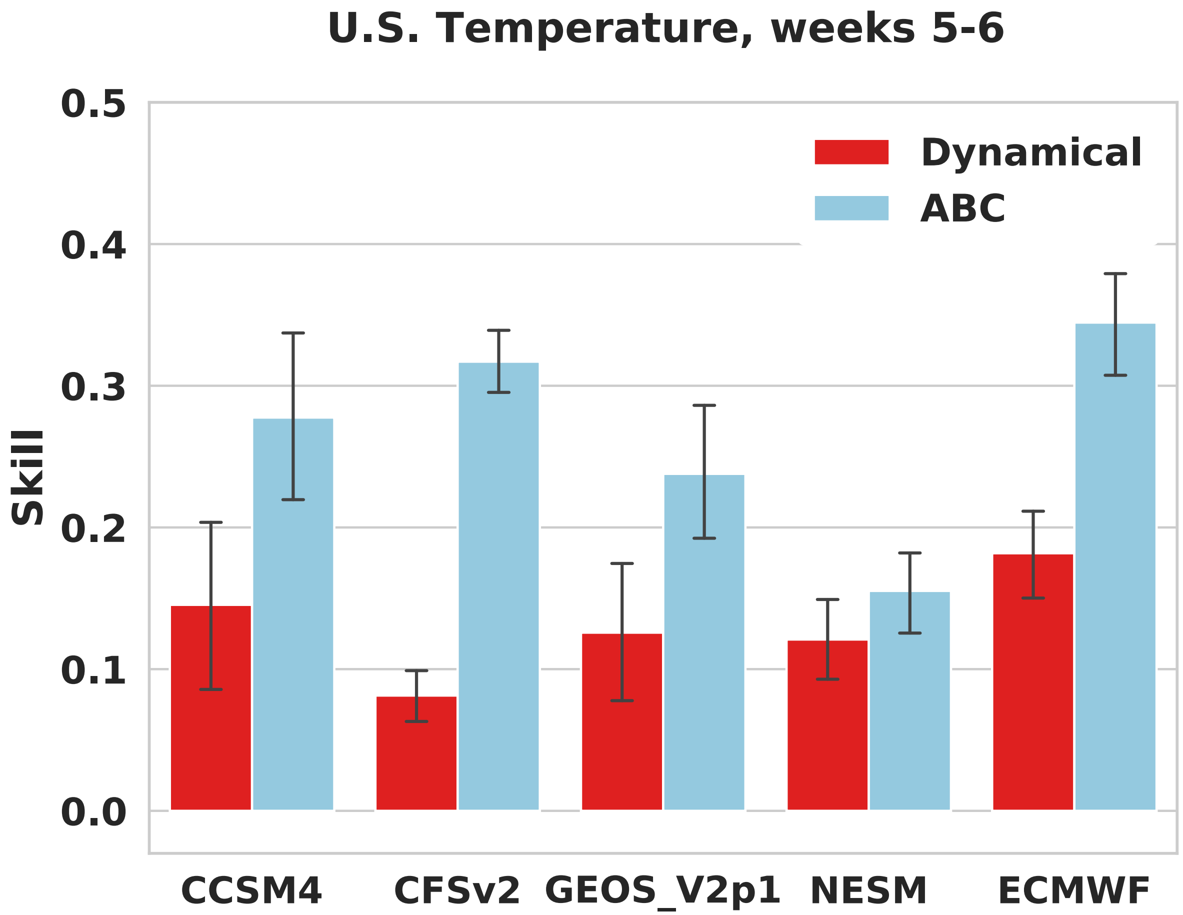
Adaptive Bias Correction (ABC): Hybrid Physics + Learning Model



- Can be used to correct any dynamical model
- Including leading model from European Centre for Medium-Range Weather Forecasts

Adaptive Bias Correction (ABC): Hybrid Physics + Learning Model





- Can be used to correct any dynamical model
- Including leading model from European Centre for Medium-Range Weather Forecasts

ABC: An Ensemble of 3 Learning Models

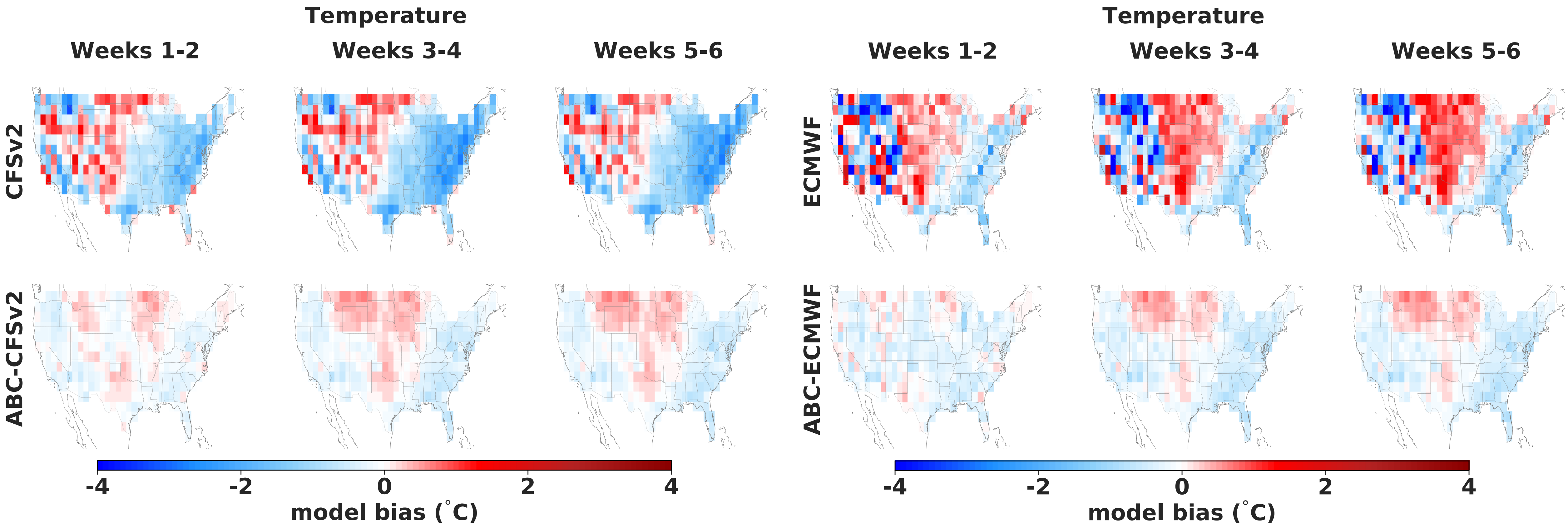
- **Climatology++**
  - Predicts historical mean or geographic median in window around target day of year
  - # of training years and window size chosen adaptively via an online tuning procedure
  - 250% more skillful than debiased CFSv2 for precipitation
- **Dynamical++**
  - Learned correction for raw dynamical model forecasts
  - Averages dynamical forecasts over a range of issuance dates and lead times, subtracts mean ensemble forecast, and adds mean ground-truth over a learned window
  - Ensembled lead times and issuance dates and window size chosen adaptively
  - Improves deb. CFSv2 temperature and precipitation skill by 50-275%
- **Persistence++**
  - Least squares regression per grid point to combine climatology, recent weather trends in the form of lagged temperature or precipitation measurements, and CFSv2 ensemble forecast
  - Improves deb. CFSv2 temperature and precipitation skill by 40-130%
- Also outperform 7 state-of-the-art machine learning and deep learning methods

Contiguous U.S. Performance (2010-2020)

		Average % Skill			
		Temperature		Precipitation	
Group	Model	weeks 3-4	weeks 5-6	weeks 3-4	weeks 5-6
Baselines	Debiased CFSv2	24.94	19.12	5.77	4.28
	Persistence	10.64	6.22	8.31	7.41
Learning	AutoKNN	12.43	8.56	6.66	5.93
	Informer	0.55	0.01	6.15	5.86
	LocalBoosting	14.44	12.69	10.82	9.72
	MultiLLR	24.5	16.68	9.49	7.97
	N−BEATS	9.21	4.16	5.48	4.46
	Prophet	20.21	19.78	13.51	13.41
	Salient 2.0	11.24	11.77	10.11	9.99
ABC	Climatology++	18.61	18.87	15.04	14.99
	CFSv2++	32.38	29.19	16.34	16.09
	Persistence++	32.4	26.73	13.38	9.77
	ABC	33.58	30.56	18.94	18.35

- **Takeaway:** ABC outperforms operational US model (CFSv2) and 7 state-of-the-art machine learning and deep learning methods from the literature

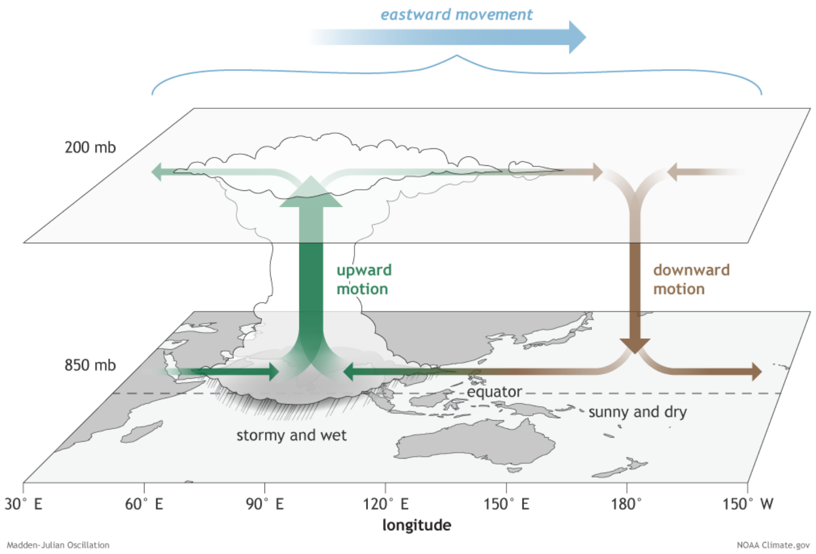
ABC Reduces Systematic Model Bias



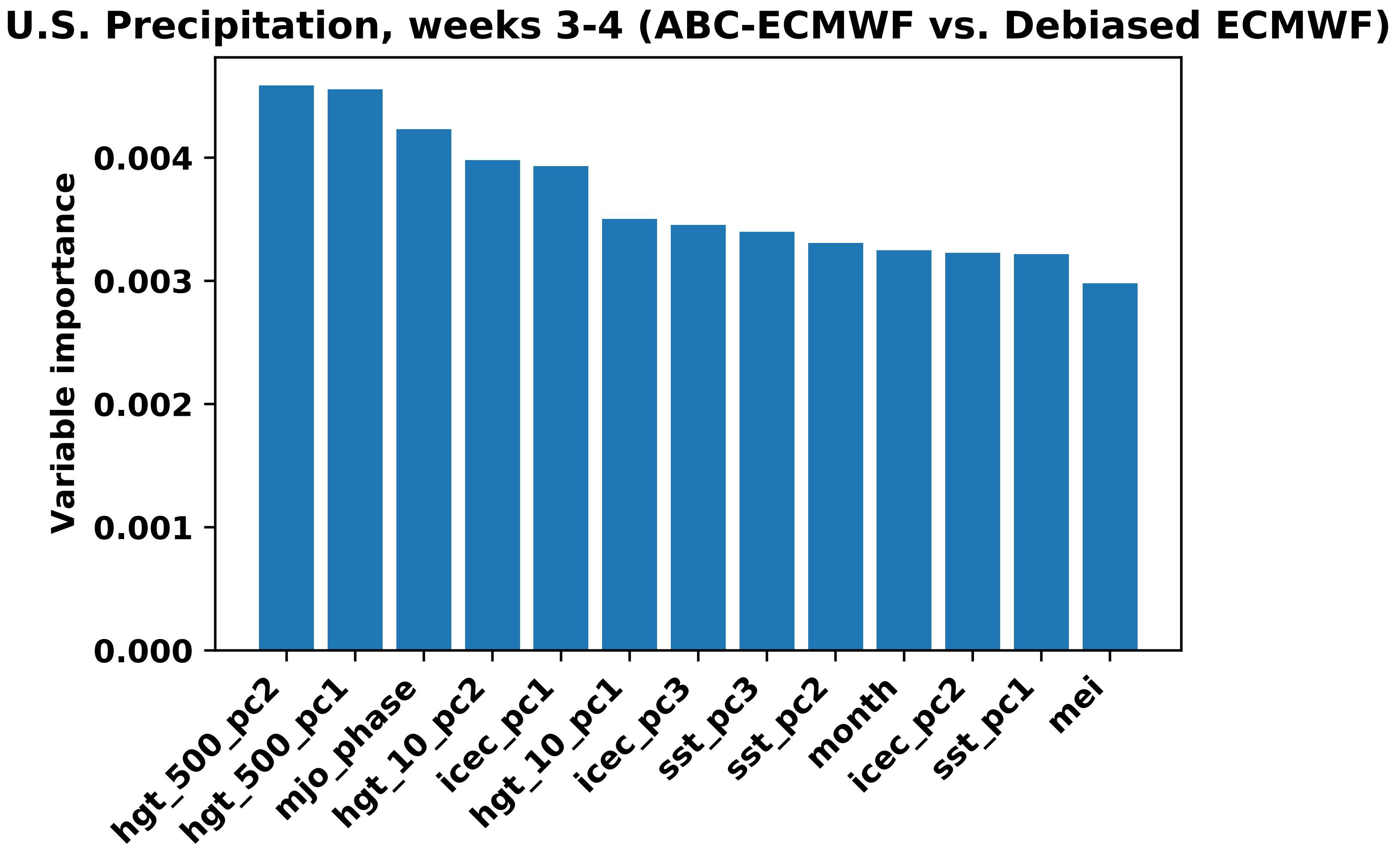
- Spatial distribution of model bias over the years 2018–2021
- CFSv2 = Climate Forecasting System v2, US operational dynamical model
- ECMWF = European Centre for Medium-Range Weather Forecasts, leading subseasonal model

Explaining ABC Improvements

- **Question:** When is ABC most likely to improve upon its input model?
- **Answer: Opportunistic ABC workflow**
  - Based on the optimal credit assignment principle of Shapley (1953)
  - Measures impact of explanatory variables on individual forecasts using Cohort Shapley (Mase et al., 2019) and overall using Shapley effects (Song et al., 2016)
- **Example:** Explain ABC improvements for weeks 3-4 precipitation using
  - **500 hPa geopotential height (HGT)**
    - Captures thermal structure, synoptic circulation
  - **Madden Julian Oscillation (MJO) phase**
    - 30-90 day oscillation in tropical atmosphere
  - **10 hPa geopotential height (HGT)**
    - Captures polar vortex variability
  - **Sea ice concentration (ICEC)**
    - Impacts near-surface temperatures
  - **Sea surface temperatures, multivariate ENSO index, target month, …**

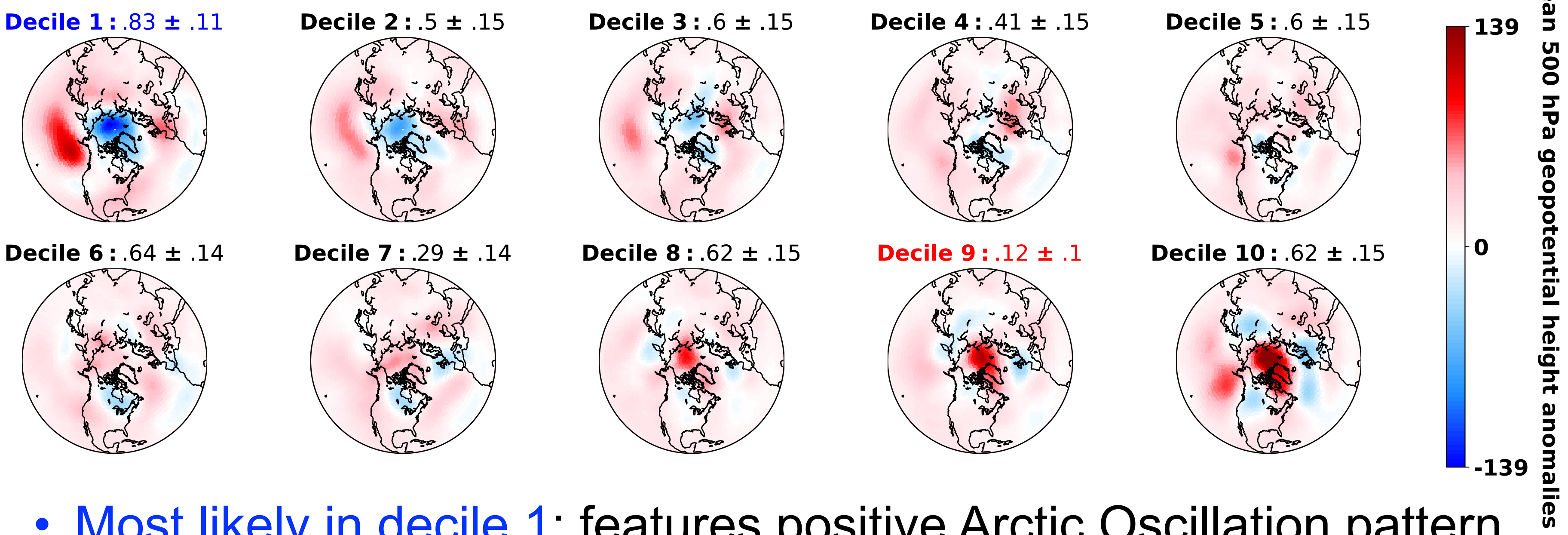


Explaining ABC Improvements



Global importance of each variable in explaining skill improvement

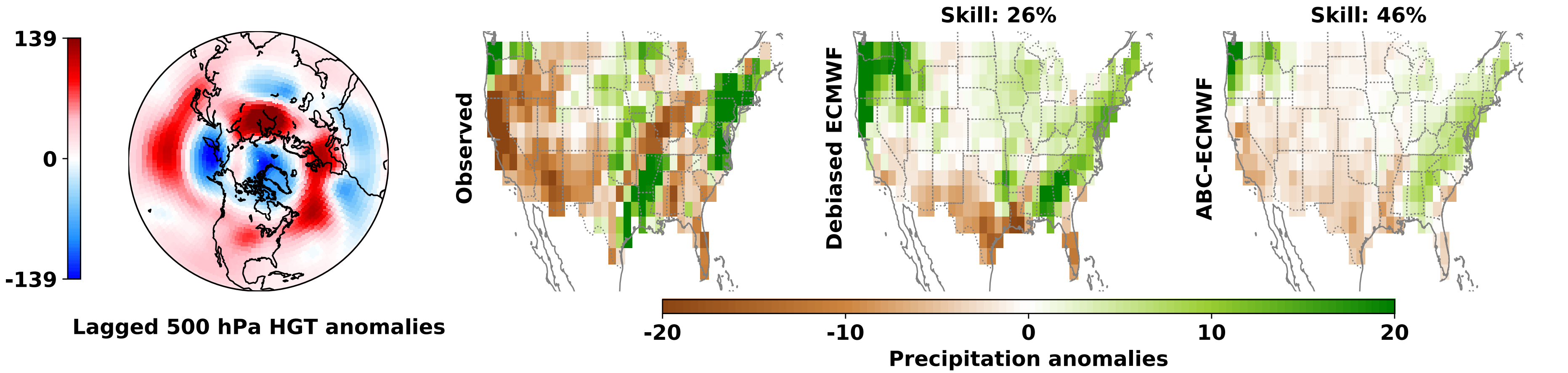
Positive impact of HGT 500 PC1 on ABC skill improvement



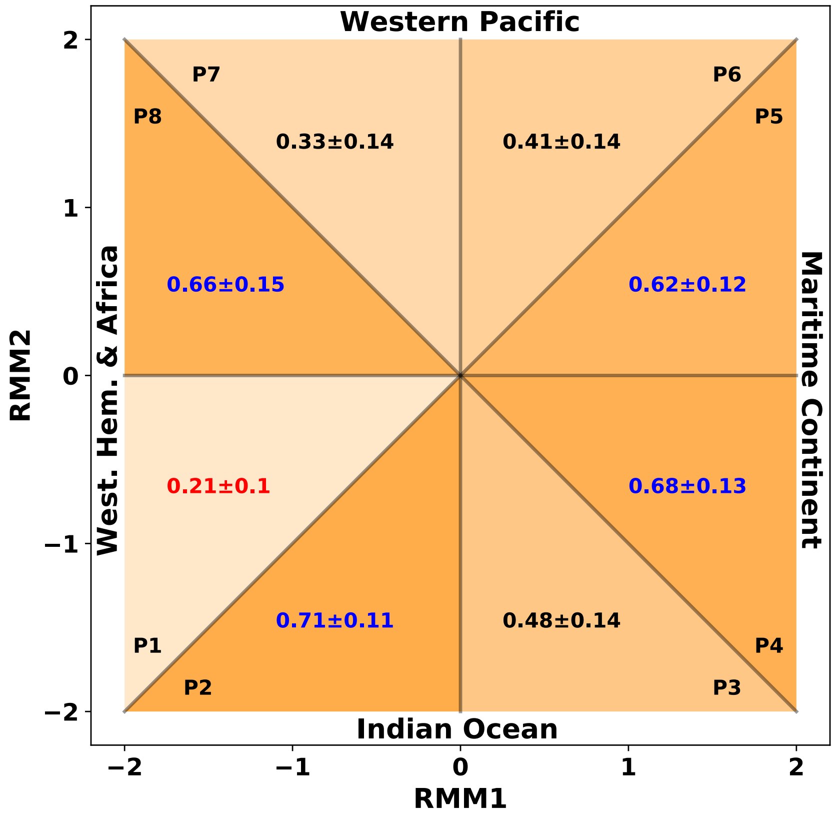
- Most likely in decile 1: features positive Arctic Oscillation pattern
- Least likely in decile 9: features opposite phase Arctic Oscillation

---

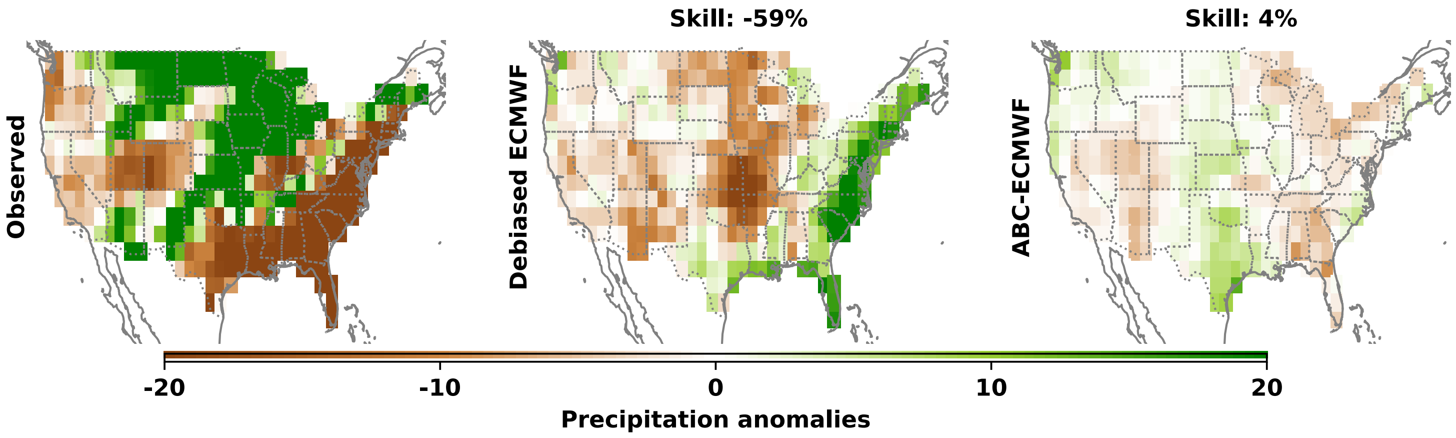
Forecast with largest HGT 500 PC1 impact in decile 1



Positive impact of MJO phase on ABC skill improvement

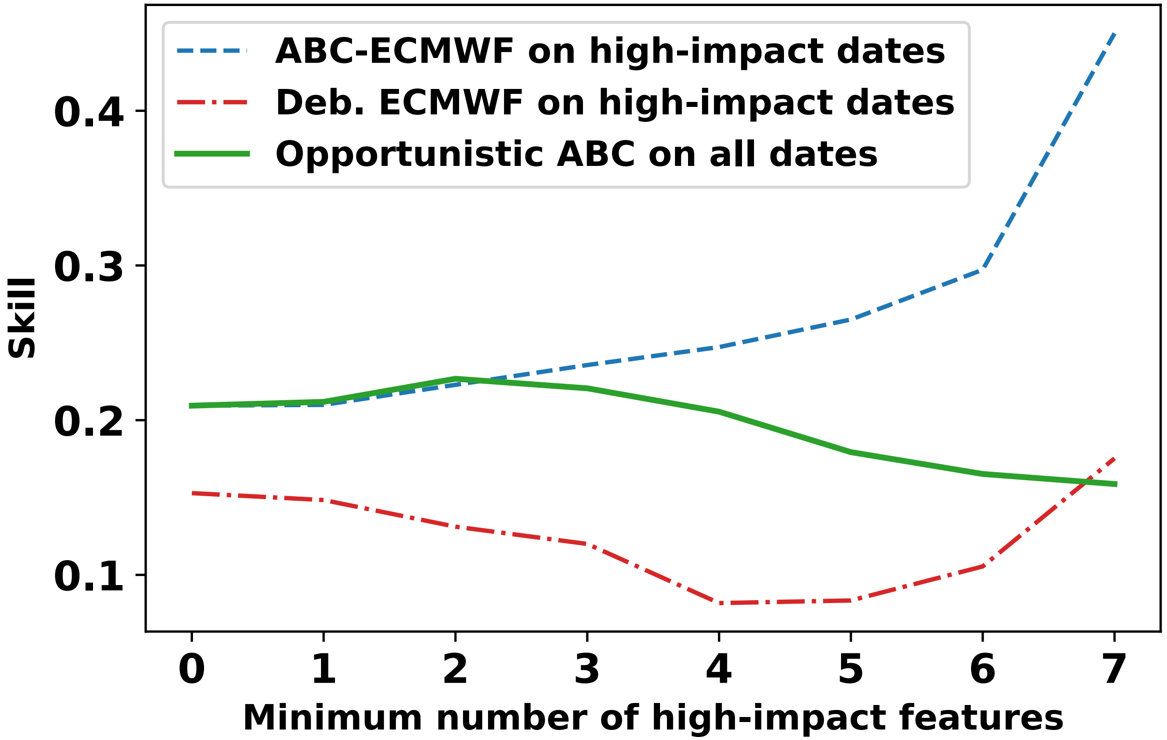


Forecast with largest MJO phase impact in deciles 2, 4, 5, 8



Forecasts of Opportunity

# High-impact variables	% Forecasts using ABC	High-impact skill (%) ABC	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



- **Idea:** Apply ABC opportunistically when multiple explanatory variables are in high-impact state and use baseline debiased dynamical model otherwise
- Effectively defining windows of opportunity based on variables observable at forecast issuance date

Next Steps and Open Questions

- **Extend forecasting region to the entire globe**
  - How should skill be measured? Overall? By region? Which regions?
- **Complement deterministic forecasts with probabilistic forecasts**
  - Forecast probability of each tercile (near normal, above normal, below normal)
  - Evaluate using Ranked Probability Skill Score

$$\mathrm{RPSS}(\hat{\boldsymbol{p}}, \boldsymbol{p}) = 1 - \frac{(\hat{p}_1 - p_1)^2 + (\hat{p}_3 - p_3)^2}{(\frac{1}{3} - p_1)^2 + (\frac{1}{3} - p_3)^2}$$

  - How well do deterministic forecasting techniques translate into this setting?
- **Improved multimodel ensembling**
  - Standard in the field is equal weighted averaging
  - But relative model performance varies over time and space
  - Monteleoni et al. (2011) use online learning to learn adaptive ensembling rules
  - Flaspohler et al. (2021) use optimistic online learning to deal with delayed feedback

Adaptive Bias Correction for Improved Subseasonal Forecasting
arxiv.org/abs/2209.10666

Learned Benchmarks for Subseasonal Forecasting
arxiv.org/abs/2109.10399

Online Learning with Optimism and Delay
arxiv.org/abs/2106.06885

Improving Subseasonal Forecasting in the Western U.S. with Machine Learning arxiv.org/abs/1809.07394

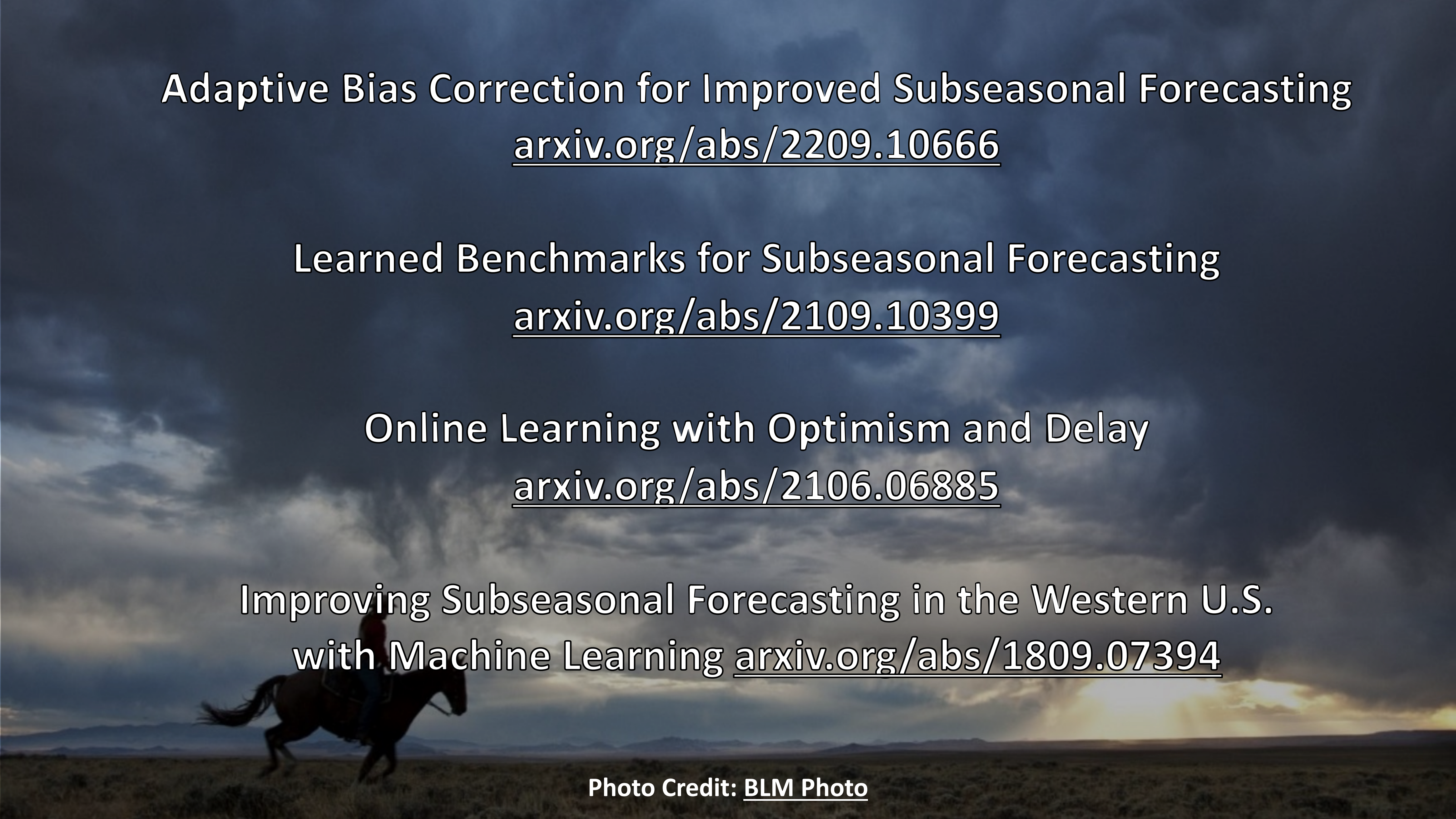


Photo Credit: BLM Photo

Online Learning with Optimism and Delay

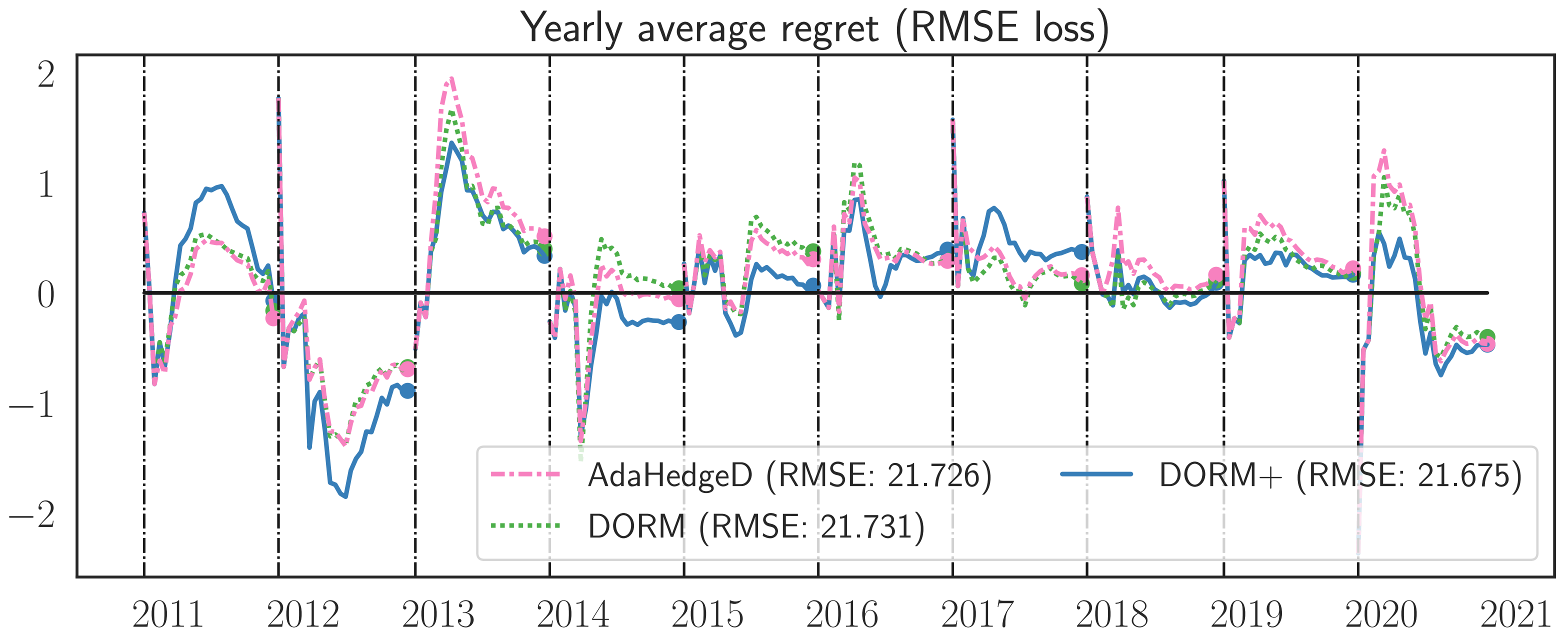
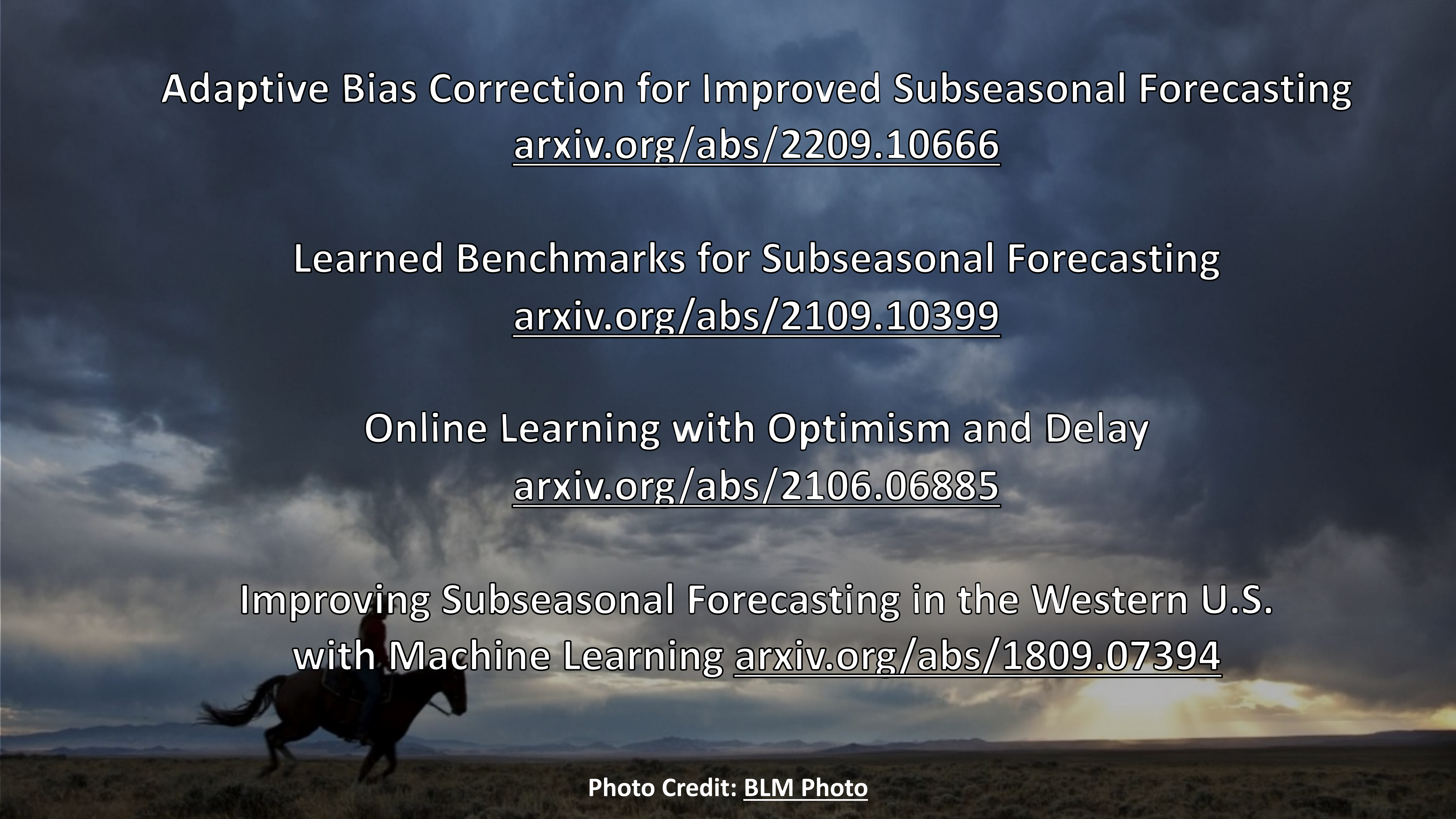


Table 1: **Average RMSE of 2011-2020 semimonthly forecasts**: The online learners compare favorably with the **best input models** and learn to downweight lower-performing candidates, like the *worst models*.

	AdaHedgeD	DORM	DORM+	Model1	Model2	Model3	Model4	Model5	Model6
P3-4	21.726	21.731	**21.675**	**21.973**	22.431	22.357	21.978	21.986	*23.344*
P5-6	21.868	21.957	**21.838**	22.030	22.570	22.383	22.004	**21.993**	*23.257*
T3-4	2.273	2.259	**2.247**	**2.253**	2.352	2.394	2.277	2.319	*2.508*
T5-6	2.316	2.316	**2.303**	**2.270**	2.368	2.459	2.278	2.317	*2.569*



Adaptive Bias Correction for Improved Subseasonal Forecasting
arxiv.org/abs/2209.10666

Learned Benchmarks for Subseasonal Forecasting
arxiv.org/abs/2109.10399

Online Learning with Optimism and Delay
arxiv.org/abs/2106.06885

Improving Subseasonal Forecasting in the Western U.S.
with Machine Learning arxiv.org/abs/1809.07394

Photo Credit: BLM Photo