Applying Machine Learning to Improve Subseasonal to Seasonal (S2S) Forecasts

Judah Cohen, **Sonja Totz**, Genevieve Flaspohler, Soukaya Mouatadid, Paulo Orenstein, Dara Entekhabi and Lester Mackey

11/7/2022

UNIVERSITY OF TORONTO

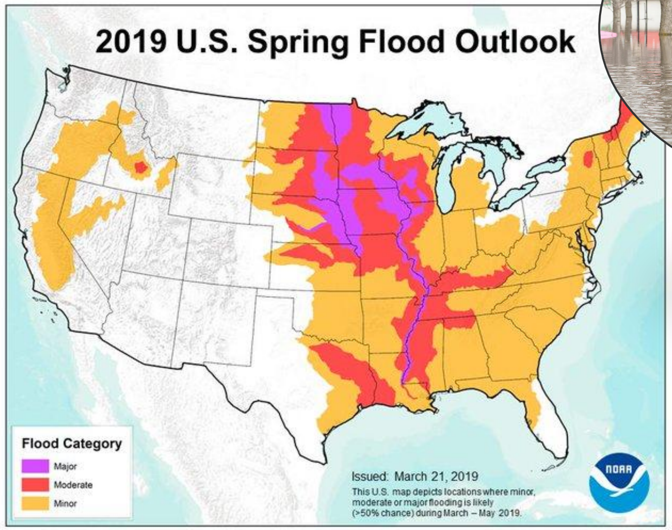
impa
Instituto de Matemática Pura e Aplicada

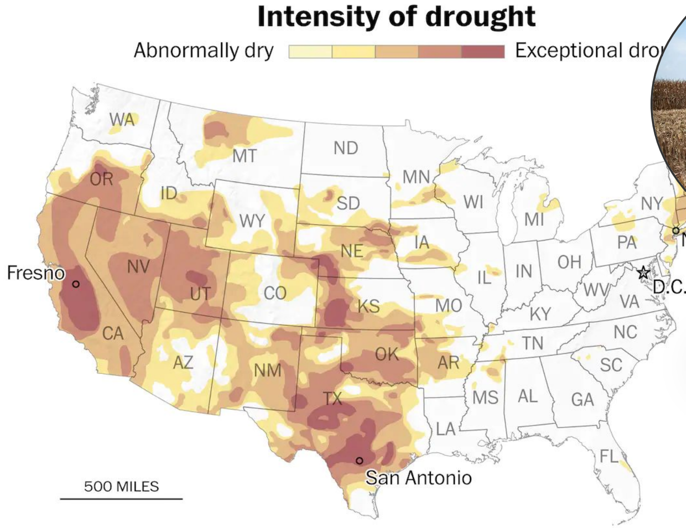
Microsoft

Verisk

Subseasonal Forecasting: precipitation & temperature

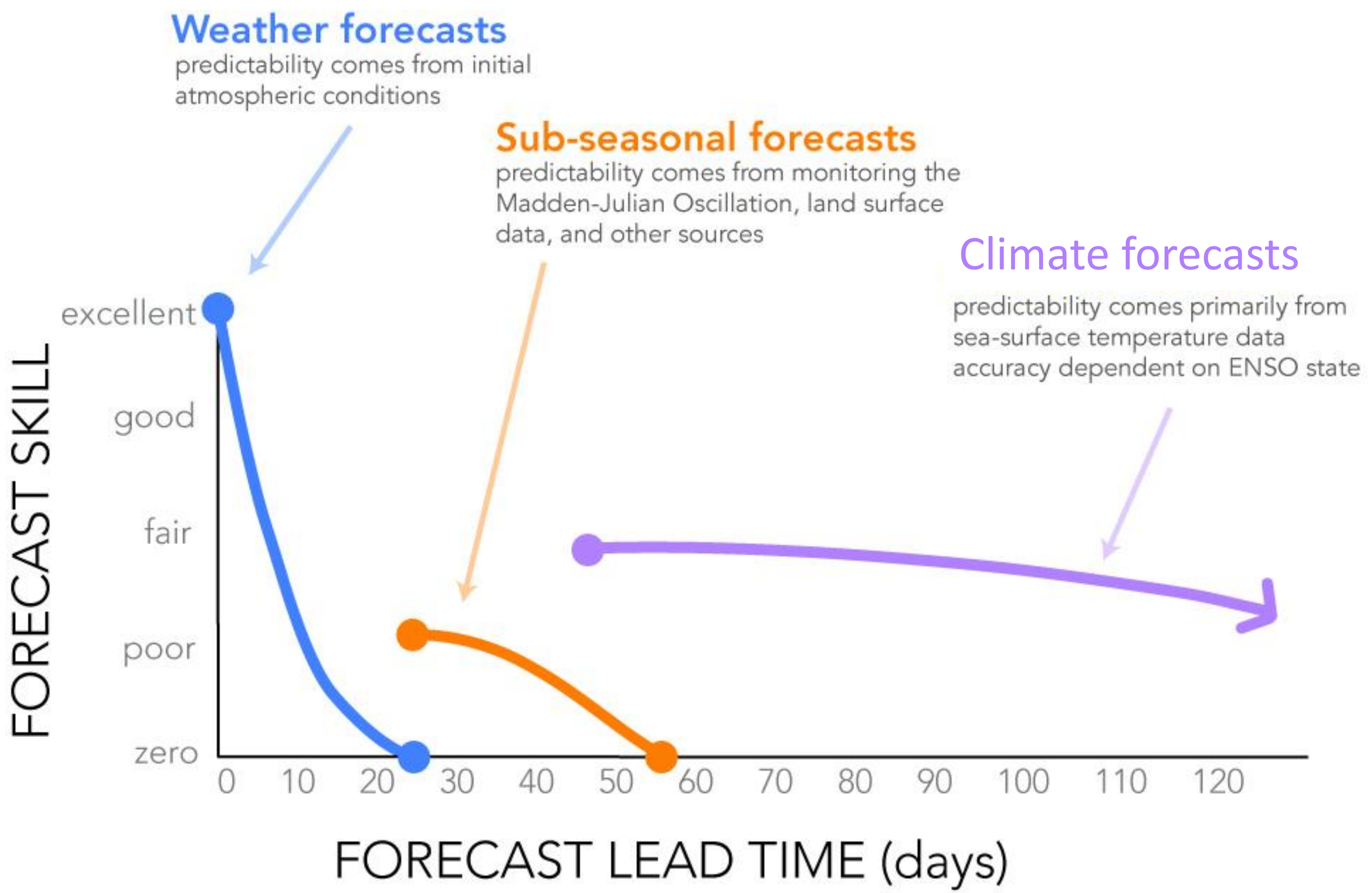
- Allocating water resources
- Managing wildfires
- Preparing for weather extremes e.g., droughts, heavy rainfall, and flooding
- Crop planting, irrigation scheduling, and fertilizer application
- Energy pricing




Data as of Aug. 31
Source: North American Drought Monitor
HANNAH DORMIDO/THE WASHINGTON POST

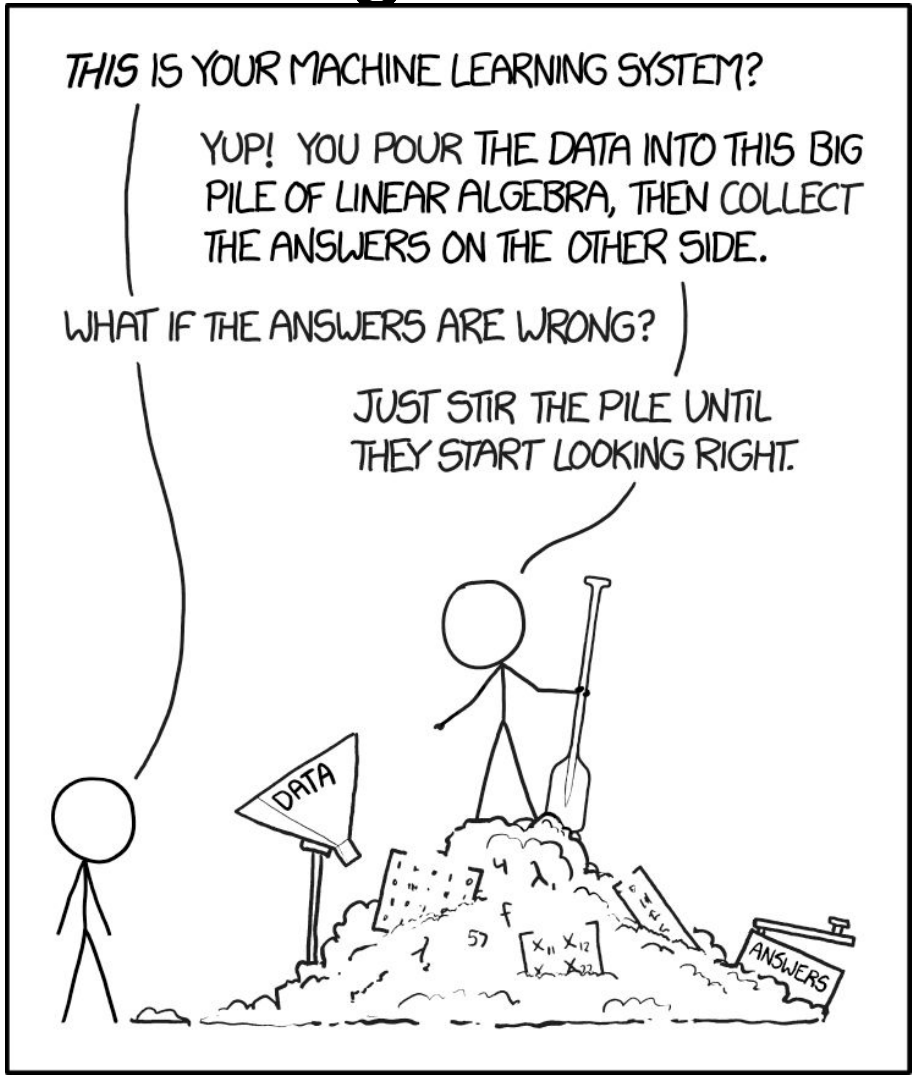
https://www.washingtonpost.com/business/2022/09/05/crops-climate-drought-food/
https://www.nytimes.com/2019/03/21/climate/climate-change-flooding.html
White et al., 2017



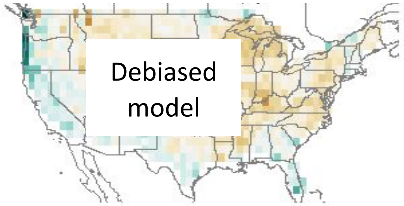
Source: https://iri.columbia.edu/news/qa-subseasonal-prediction-project/

Reasons for incorporating machine learning

- Data Science and Machine Learning (ML) have a long history in atmospheric science.
- Processing much larger datasets with new ML techniques could improve models and forecasts.
- Still need to understand the underlying physics and chemistry to select the data and methods: *the machine student needs an intelligent teacher.*
- While ML can be computationally expensive and hard to interpret, it can also find new features and generate new research questions.
- **Dynamical models have many shortcomings that can be revealed and improved with the help of data analysis.**



Adaptive Bias Correction (ABC)



Dynamical Model

Dy

Averages dy
over a range
and lead tim

Ensembled
dates and w
learned ada

y++

nean or
n in
lay of

and
n
d tuning
procedure.

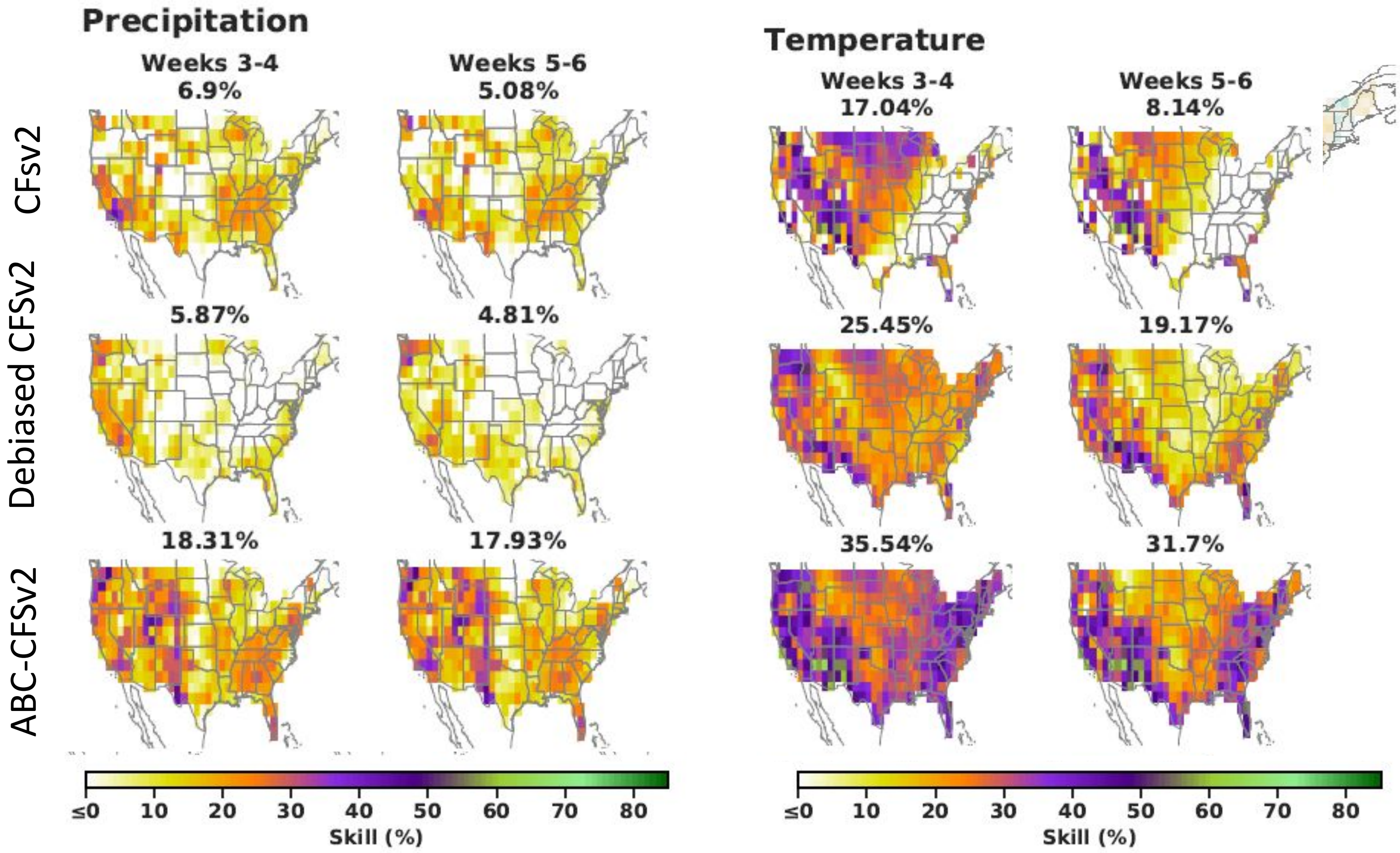
Adaptive Bias Correction (ABC)

Takes any dynamical model as input and produces a corrected model as output.

Efficient, explainable, and adaptive.

Corrected Dynamical Model

Adaptive Bias Correction: Physics + Hybrid Learning Model



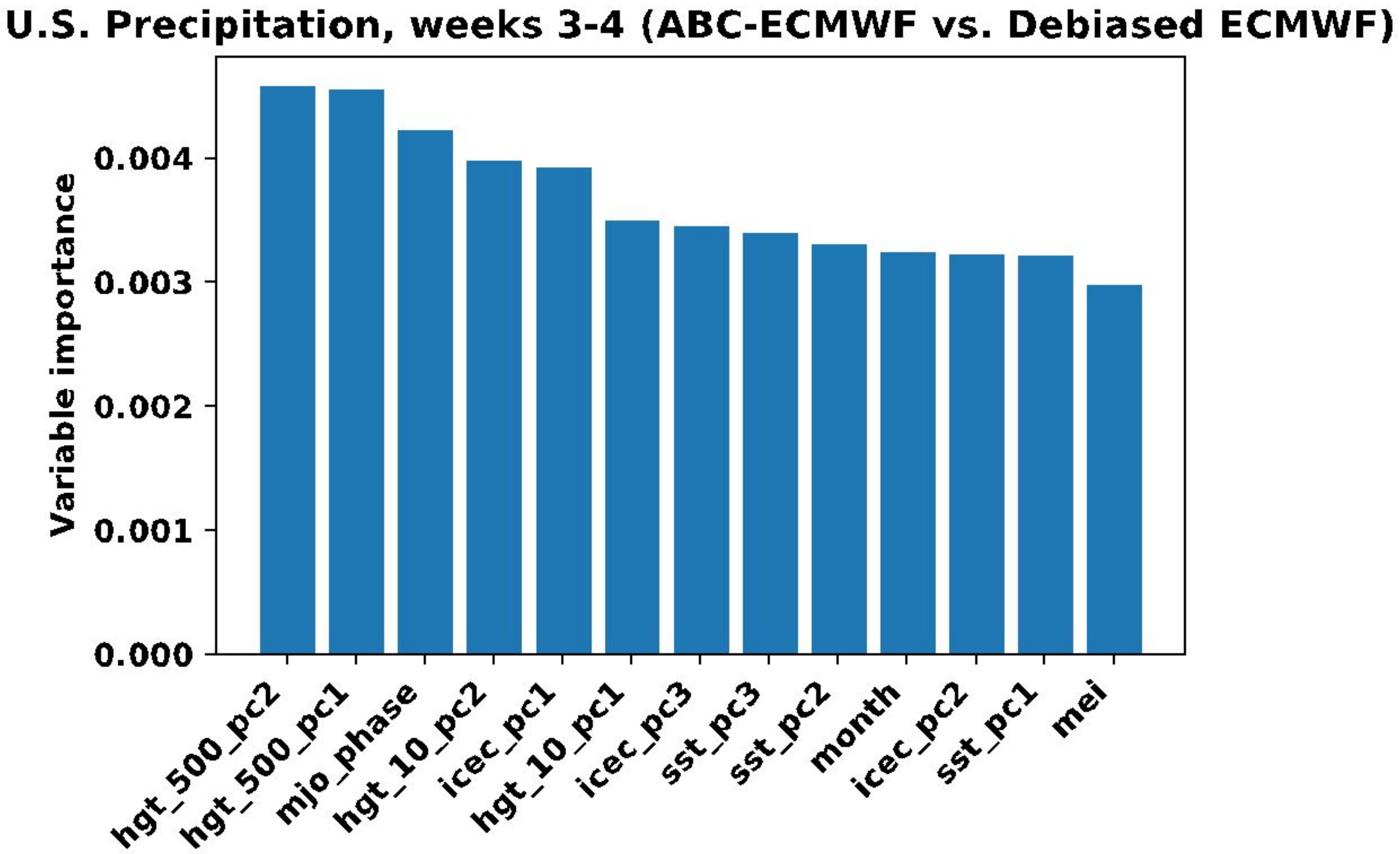
- Doubles or triples the forecasting skill of US operational dynamical model (CFSv2)

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

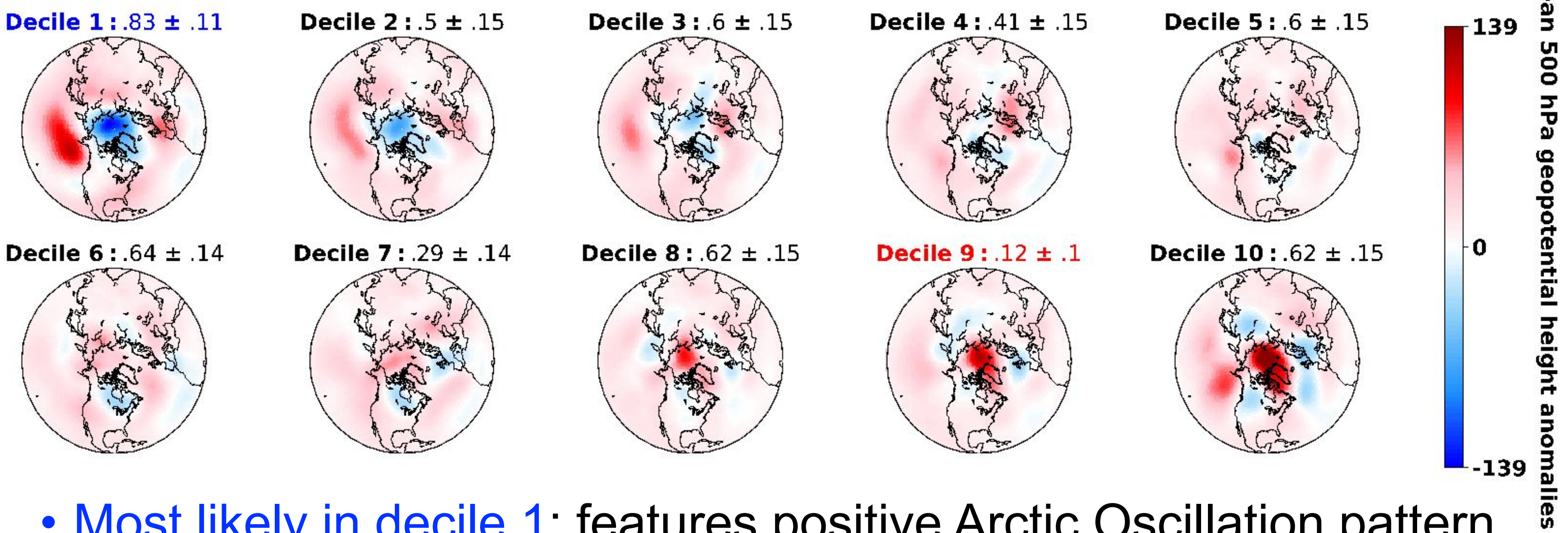
- **Takeaway:** ABC outperforms operational US model (CFSv2) and 7 state-of-the-art machine learning and deep learning methods from the

Explaining ABC Improvements



Global importance of each variable in explaining skill improvement

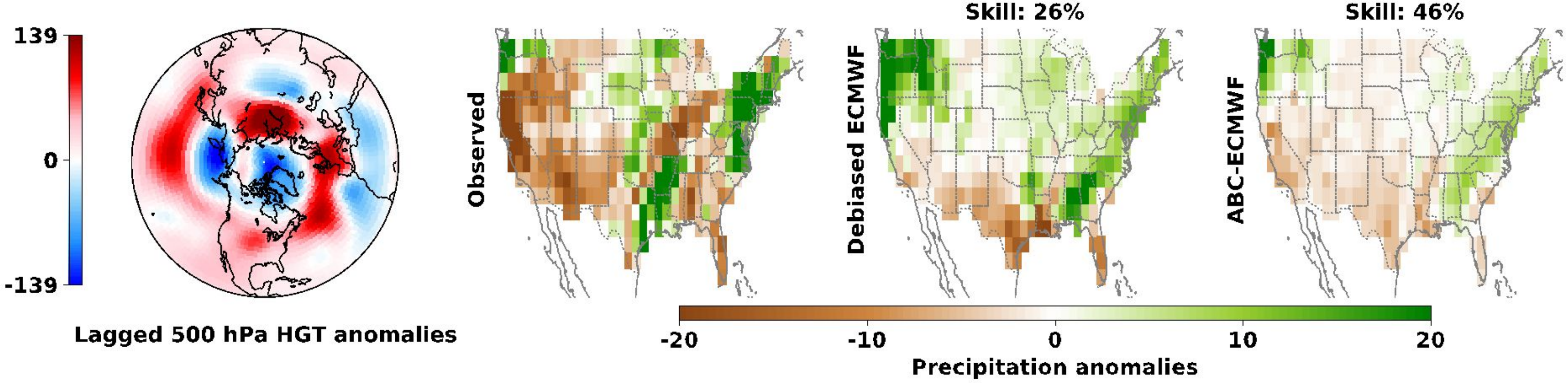
Positive impact of HGT 500 PC1 on ABC skill improvement



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

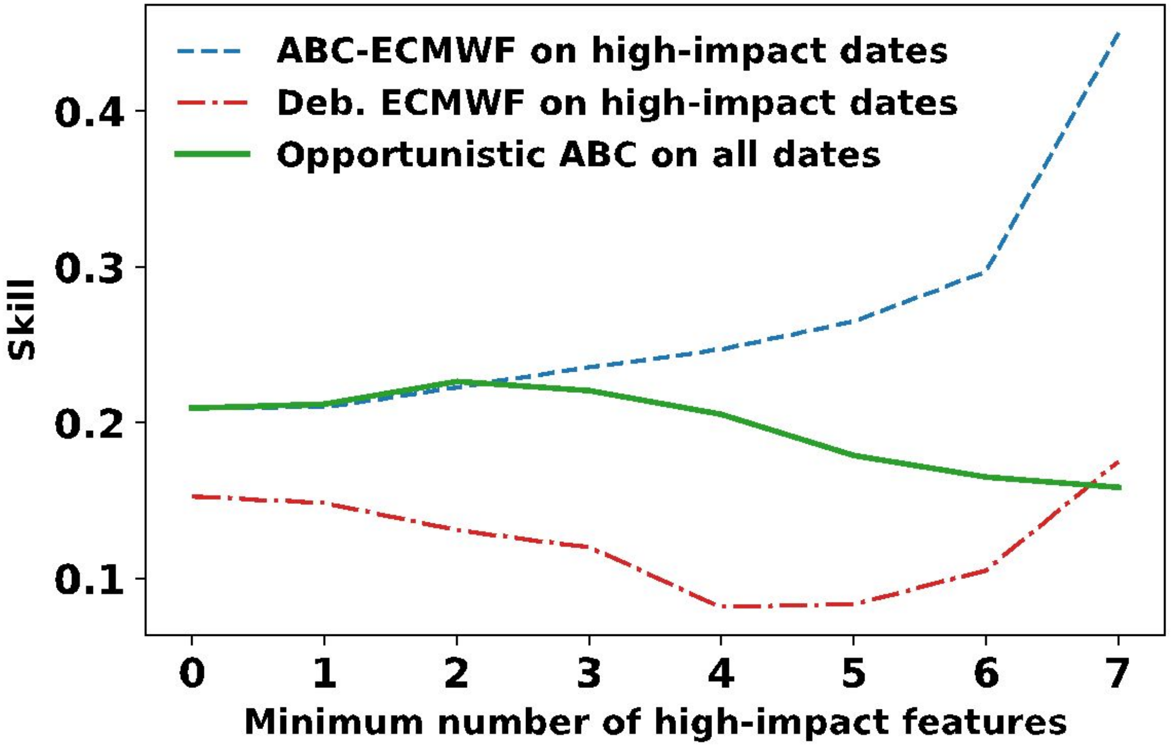
---

Forecast with largest HGT 500 PC1 impact in decile 1



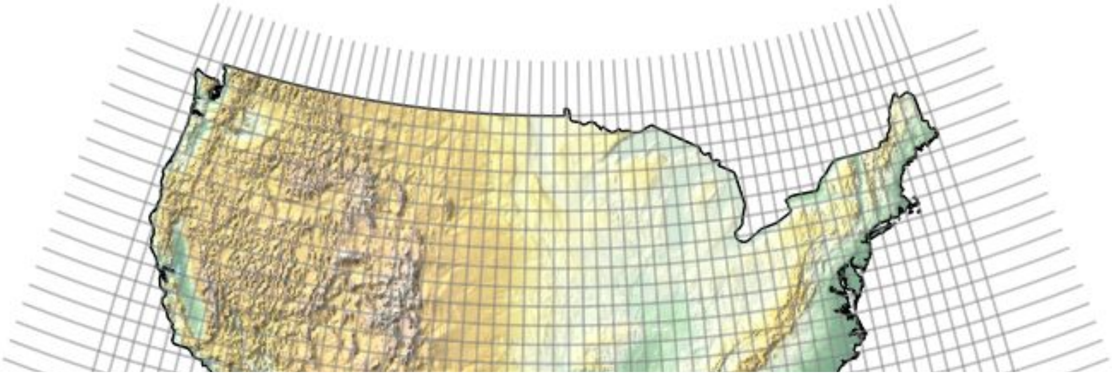
Forecasts of Opportunity

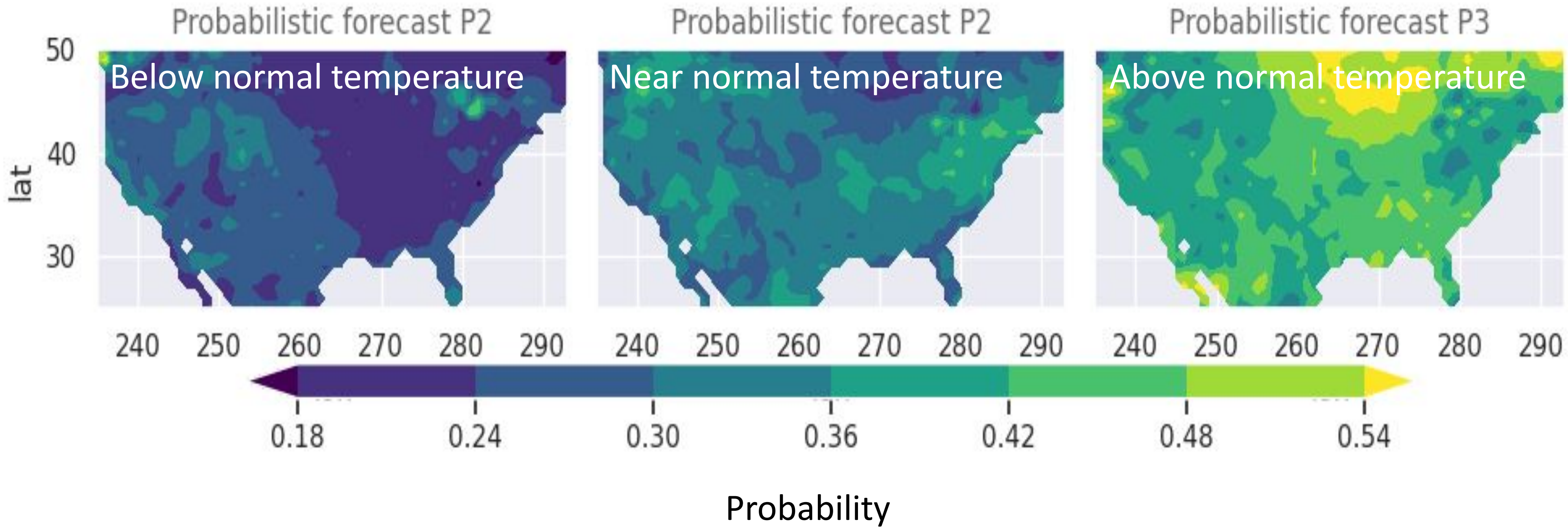
# High-impact variables	% Forecasts using ABC	High-impact skill (%) ABC	Debiased
0 or more	100.00	20.94	15.28
1 or more	95.93	20.99	14.84
2 or more	80.62	22.29	13.12
3 or more	58.61	23.56	12.00
4 or more	31.82	24.72	8.18
5 or more	14.59	26.51	8.35
6 or more	6.46	29.72	10.55
7 or more	2.15	45.00	17.53



- **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

Probabilistic forecasts





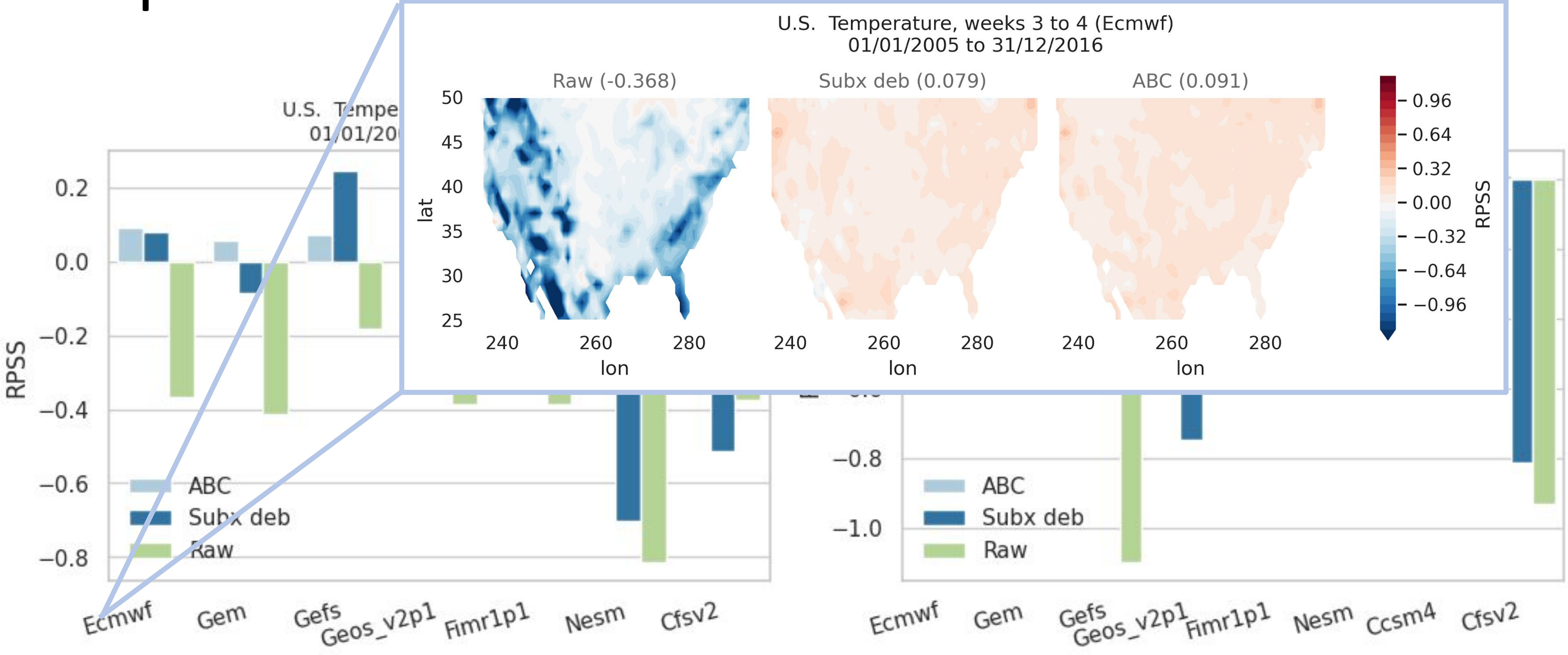
Ranked probability skill score

$$RPS = \frac{1}{ncat-1}\sum_{icat=1}^{ncat}(Pcumfct_{icat} - Pcumobs_{icat})^2$$

$$RPSS = 1 - \frac{RPS_{fct}}{RPS_{cli}}$$

$$\boxed{RPSS = 1 - \frac{\frac{1}{ncat-1}\sum_{icat=1}^{ncat}(Pcumfct_{icat} - Pcumobs_{icat})^2}{\frac{1}{ncat-1}\sum_{icat=1}^{ncat}(Pcumcli_{icat} - Pcumobs_{icat})^2}}$$

ABC probabilistic forecasts



Online Learning

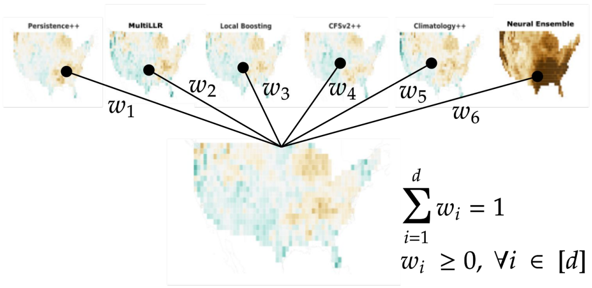
Regret measures the ensemble performance.

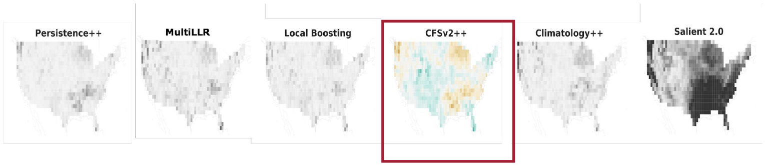
Regret measures the gap between the accuracy of the real-time ensemble and the model that would have been best in retrospect.

$$\text{Regret}_T = \underbrace{\sum_{t=1}^{T} \ell_t(\mathbf{w}_t)}_{\text{Real-time forecasting loss}} - \underbrace{\inf_{\mathbf{u} \in \mathbf{W}} \sum_{t=1}^{T} \ell_t(\mathbf{u})}_{\text{Total loss of best single model}}$$

Real-time forecasting loss

Total loss of best single model





We want ensembling strategies that have provably small regret.

New algorithms with optimal regret guarantees

We introduce DORM+: Delayed Optimistic Online Mirror Descent

Last observed loss subgradient $\mathbf{g}_{t-D} \in \partial \ell_{t-D}(\mathbf{w}_{t-D})$

Bregman divergence

$$\mathbf{w}_{t+1} = \operatorname*{argmin}_{\mathbf{w} \in \mathbf{W}} \langle \mathbf{g}_{t-D} + \mathbf{h}_{t+1} - \mathbf{h}_t, \mathbf{w} \rangle + \mathcal{B}_{\boxed{\lambda}\psi}(\mathbf{w}, \mathbf{w}_t)$$

$\mathcal{B}_\psi(\mathbf{w}, \mathbf{u}) =$
$\psi(\mathbf{w}) - \psi(\mathbf{u}) - \langle \nabla \psi(\mathbf{u}), \mathbf{w} - \mathbf{u} \rangle$

Ensemble weights

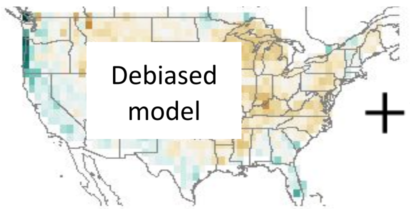
Hint vector: Estimate of future **and missed** feedback $\mathbf{g}_{t-D+1:t+1}$

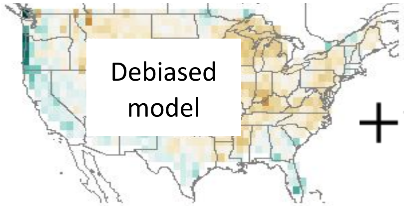
The regularization strength λ controls the balance of exploration and exploitation.

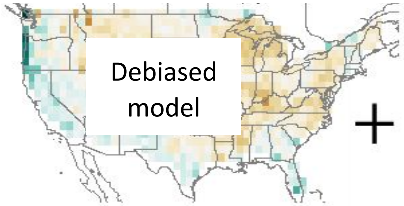
DORM+ optimally tunes λ to achieve task-targeted exploration.

The first algorithm to achieve optimal $\mathcal{O}(\sqrt{(D+1)T})$ regret growth while adaptively tuning λ for online learning with optimism and delay.

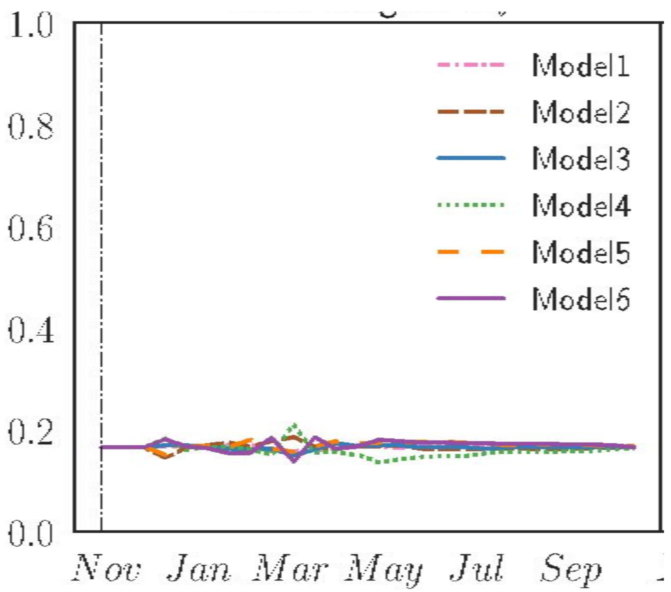
Ensembling methods

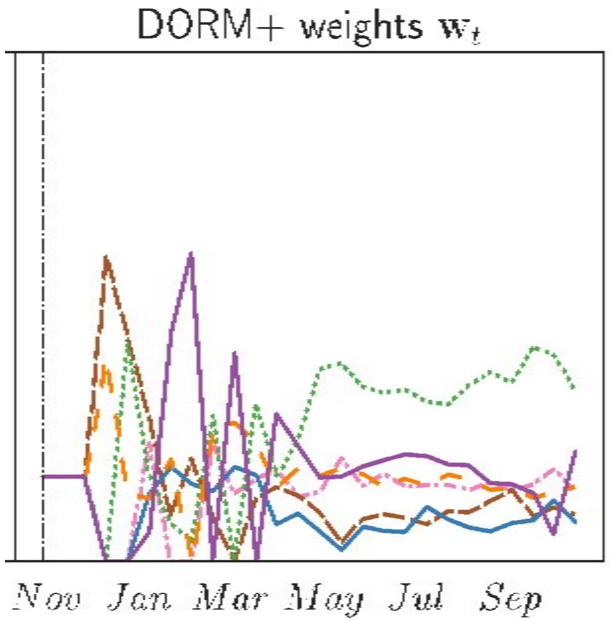
$\hat{Y}_t = w_{OL1}$ 

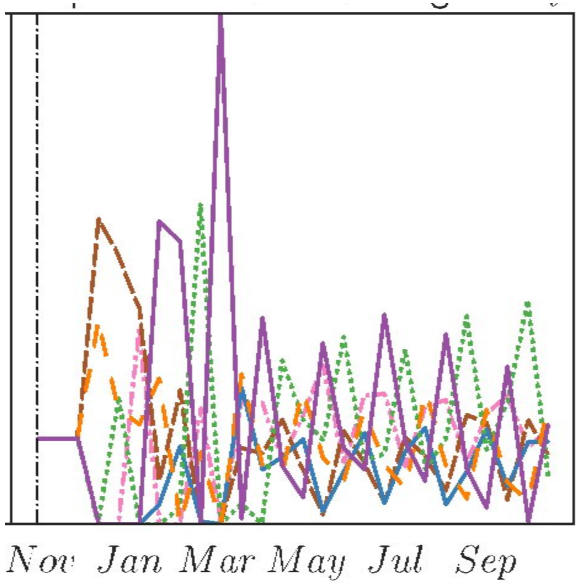
$+w_C$ 

$+w_O$ 

$+\cdots$





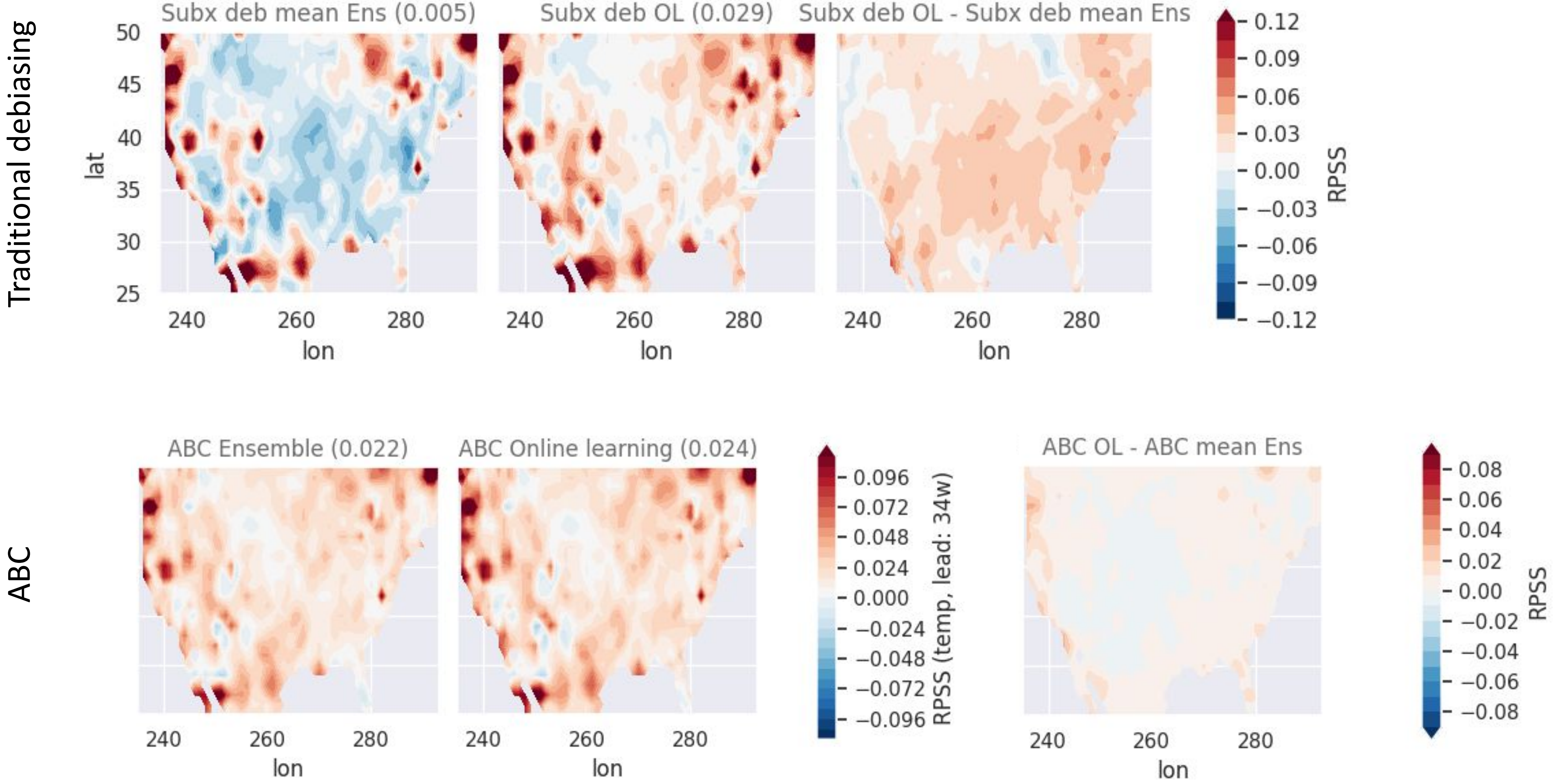


Conservative choice
State-of-the-art approaches

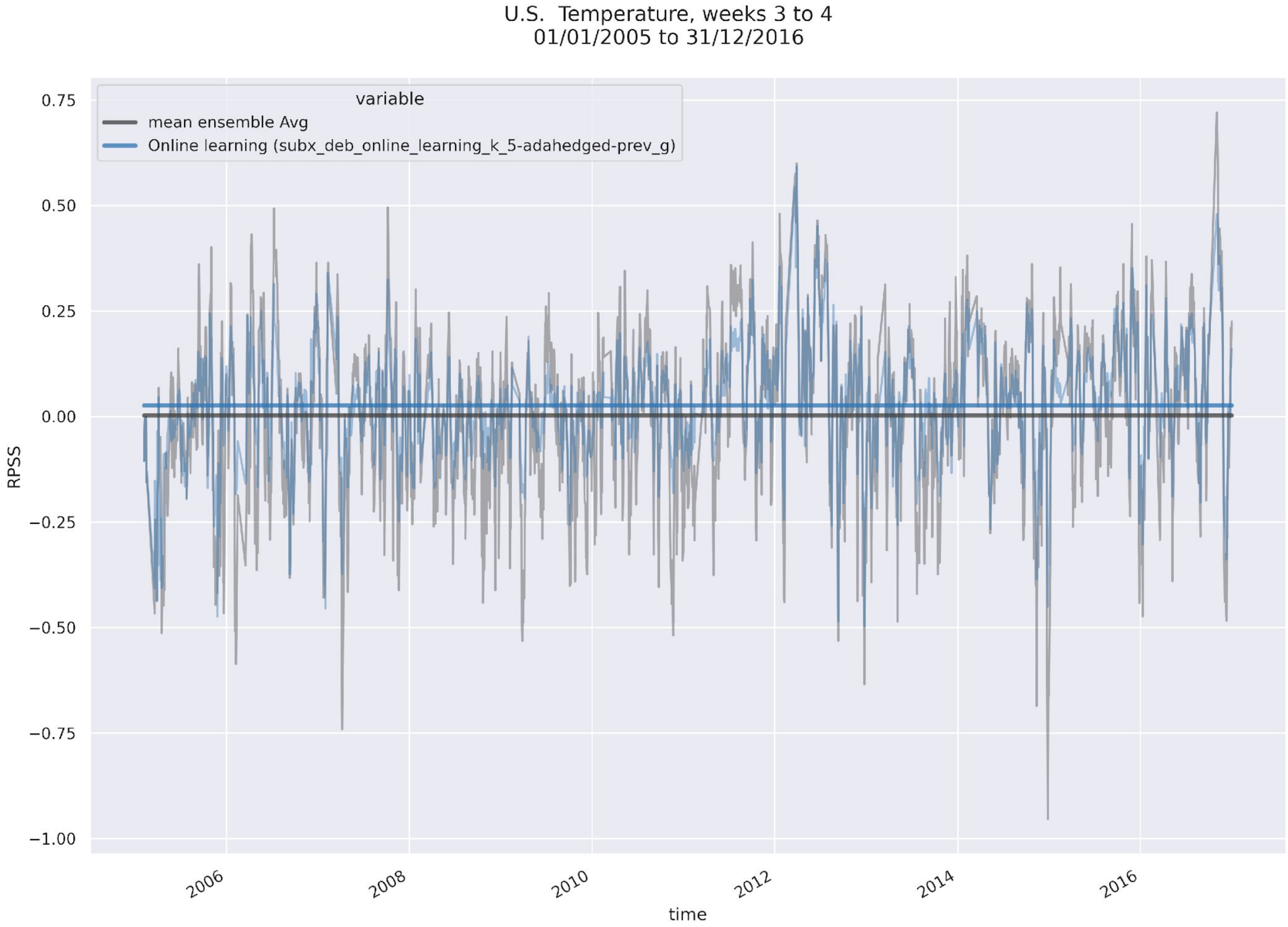
Online learning
Adapt weights online.

Aggressive choice
Human-chosen model weights

Ranked probability score - temperature

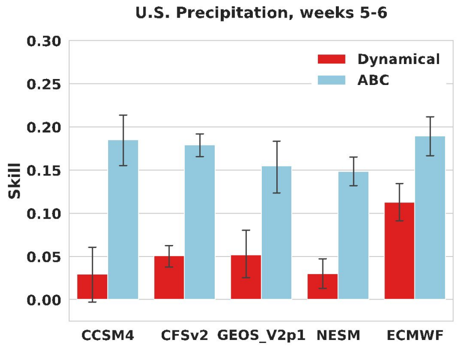


Online learning - temperature

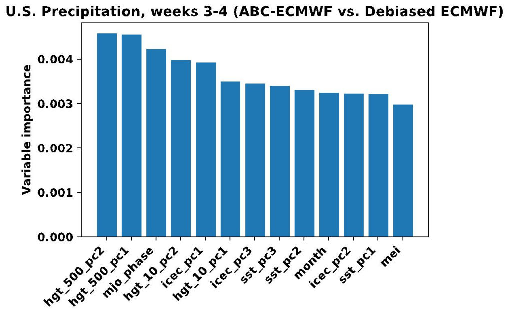


Summary

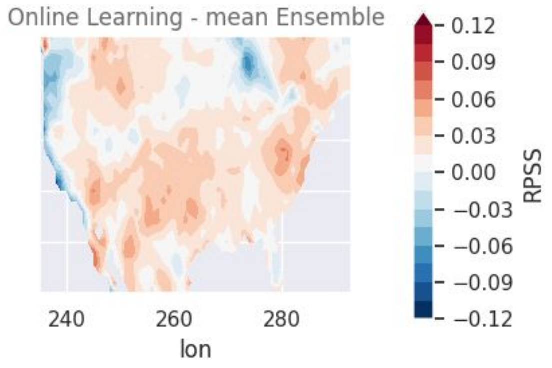
- ABC can double or triple forecasting accuracy of operational models



- Forecast improvements can be explained by common variables using Cohort Shapley, which can also be used to decide whether to use ABC or traditional debiasing



- Adaptive ensembling is only successful, when skills of model predictions differ sufficiently from each other



Thank you!

Support from:

NOAA
WEATHER
PROGRAM OFFICE

Online learning with optimism and delay

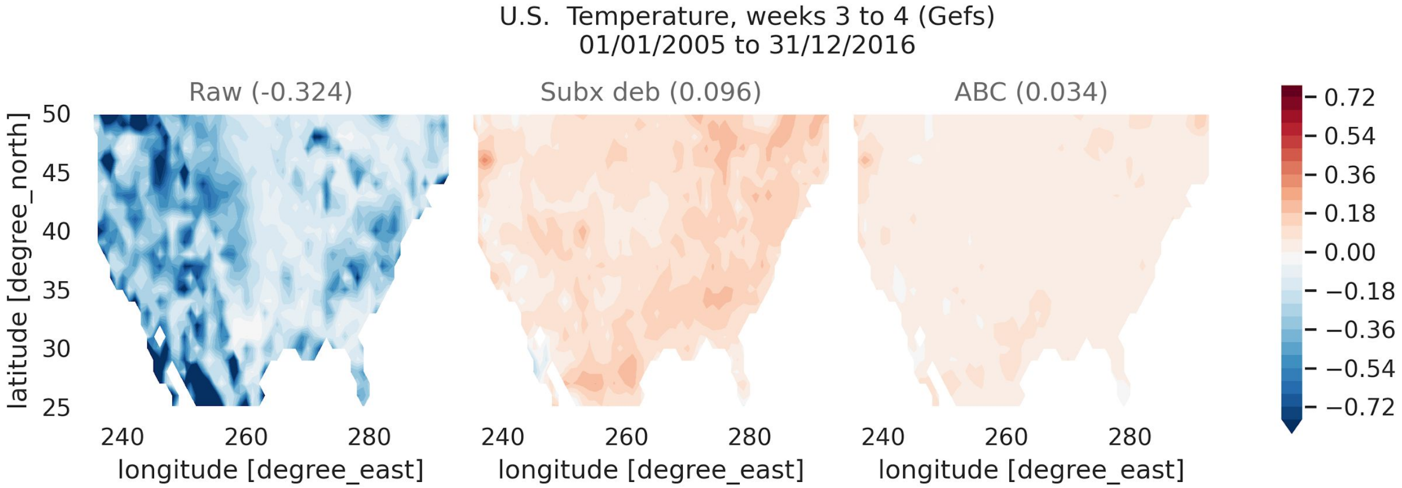
A general model for decision-making under uncertainty, composed of:

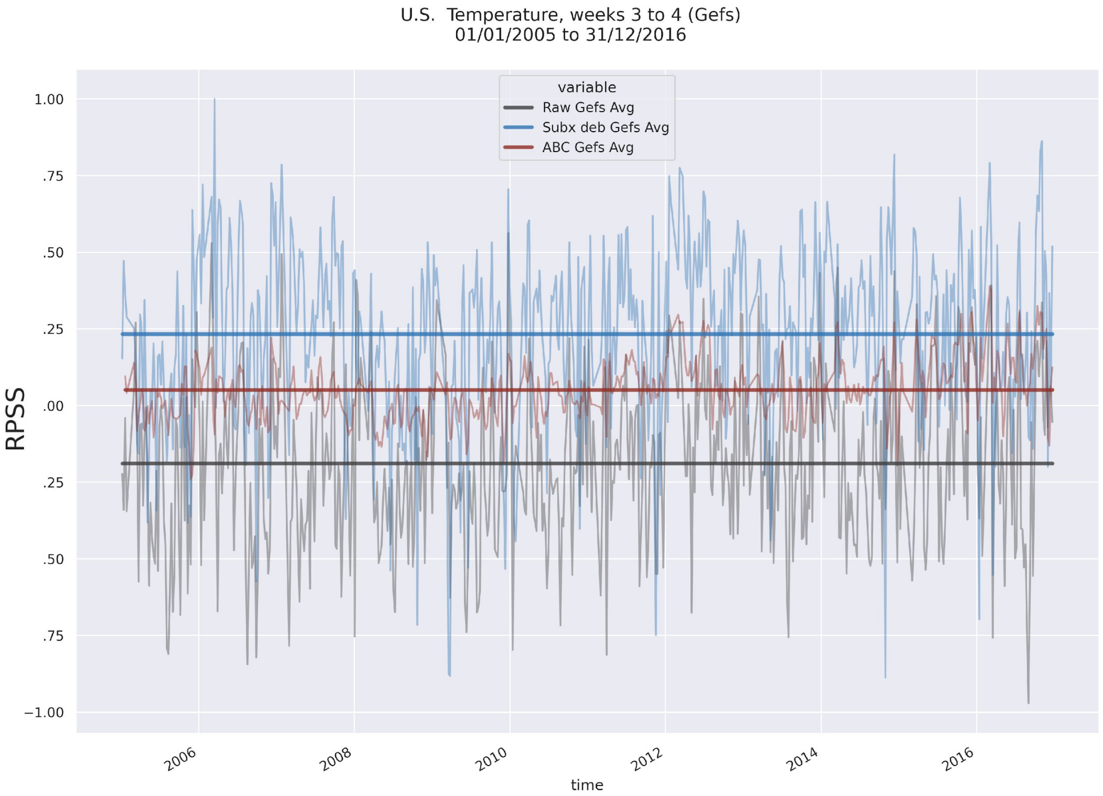
$$\Big(\overset{\text{Set of Plays}}{\mathbf{W}},\ \underset{\text{Set of Regret Competitors}}{\mathbf{U}},\ \overset{\text{Loss Generator}}{\mathcal{L}},\ \underset{\text{Optimistic Hint Generator}}{\tilde{\mathcal{L}}},\ \overset{\text{Feedback Delay}}{D}\Big)$$

The decision-making objective is to control the growth of regret (versus strategies in the set $\boldsymbol{U}$) by making plays from $\boldsymbol{W}$ that have low loss.

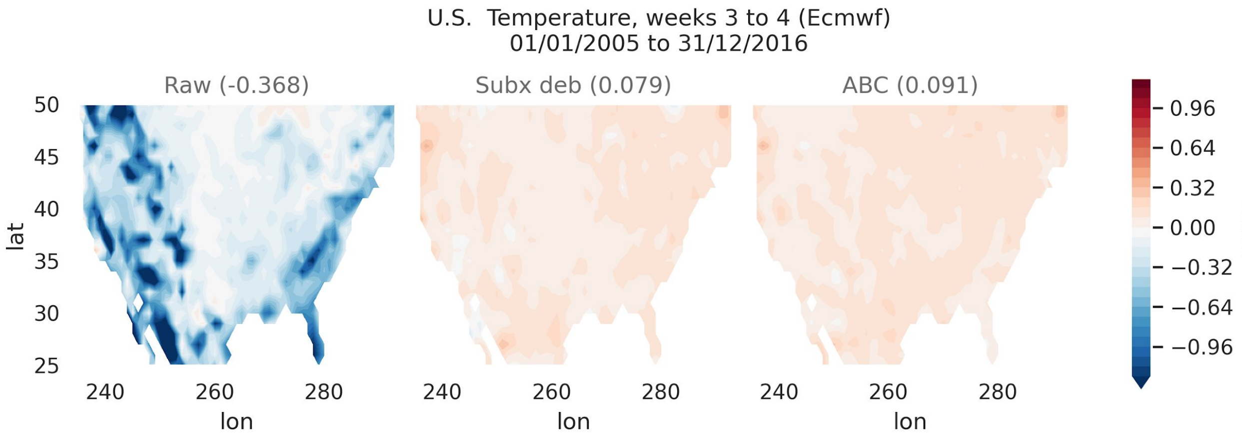
c.f. POMDP models: Actions = Plays. Rewards = Losses. Black-box loss generator instead of belief and transition function T. Provides robust, worst-case guarantees.

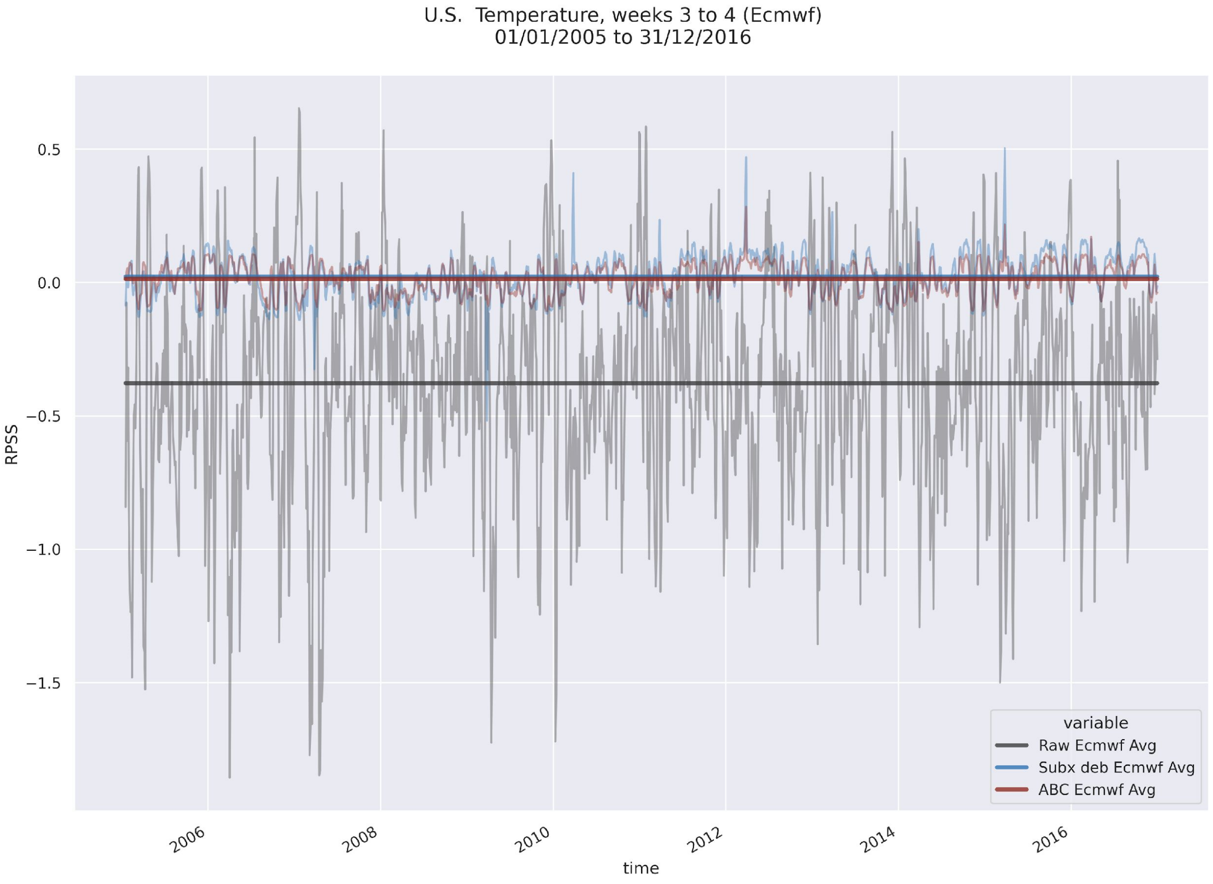
Gefs temp


U.S. Temperature, weeks 3 to 4 (Gefs)
01/01/2005 to 31/12/2016
Raw (-0.324)
Subx deb (0.096)
ABC (0.034)
latitude [degree_north]
50
45
40
35
30
25
240
260
280
longitude [degree_east]
0.72
0.54
0.36
0.18
0.00
−0.18
−0.36
−0.54
−0.72

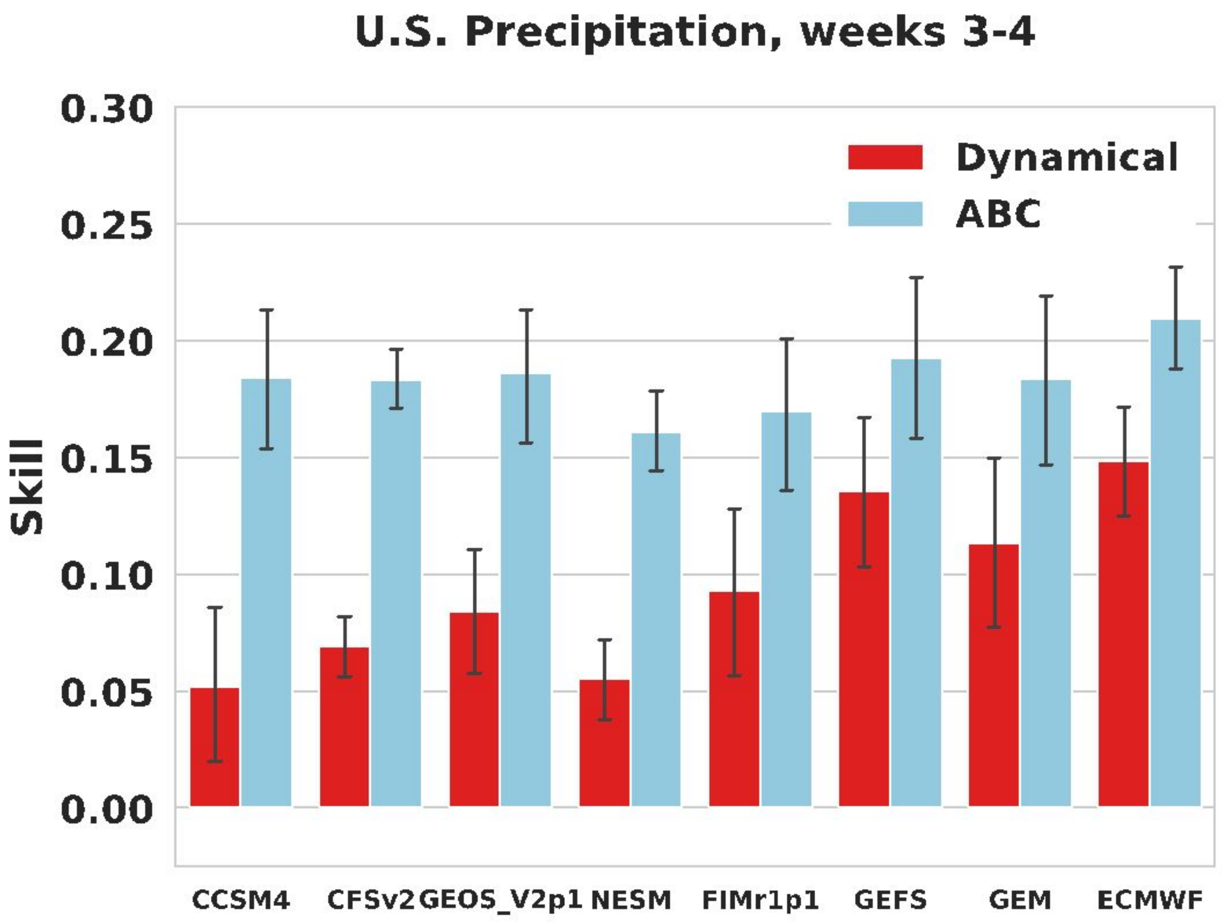

U.S. Temperature, weeks 3 to 4 (Gefs)
01/01/2005 to 31/12/2016
variable
Raw Gefs Avg
Subx deb Gefs Avg
ABC Gefs Avg
RPSS
1.00
.75
.50
.25
.00
.25
.50
.75
−1.00
2006
2008
2010
2012
2014
2016
time

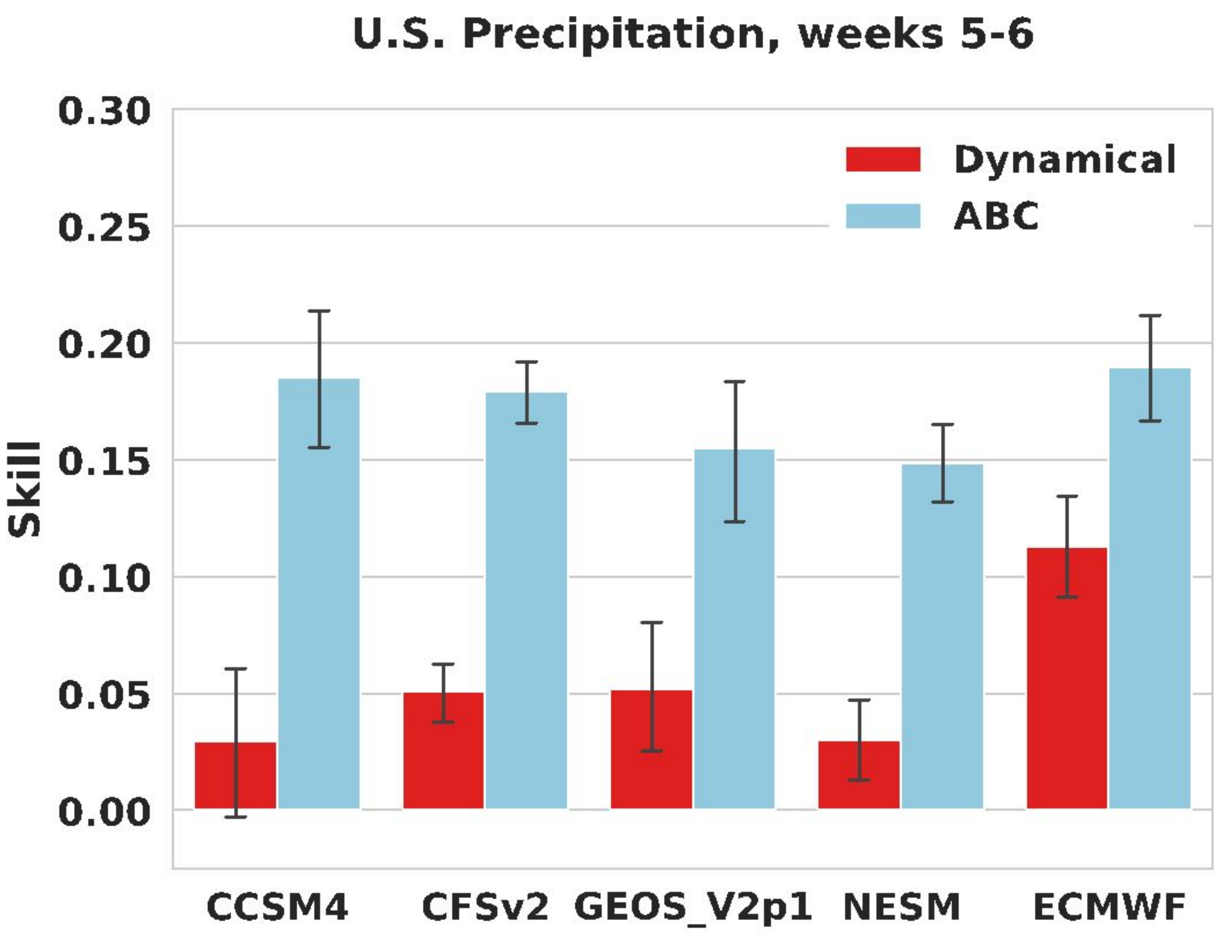
ECMWF - temp


U.S. Temperature, weeks 3 to 4 (Ecmwf)
01/01/2005 to 31/12/2016
Raw (-0.368)
Subx deb (0.079)
ABC (0.091)
lat
lon
0.96
0.64
0.32
0.00
−0.32
−0.64
−0.96


U.S. Temperature, weeks 3 to 4 (Ecmwf)
01/01/2005 to 31/12/2016
RPSS
time
variable
Raw Ecmwf Avg
Subx deb Ecmwf Avg
ABC Ecmwf Avg

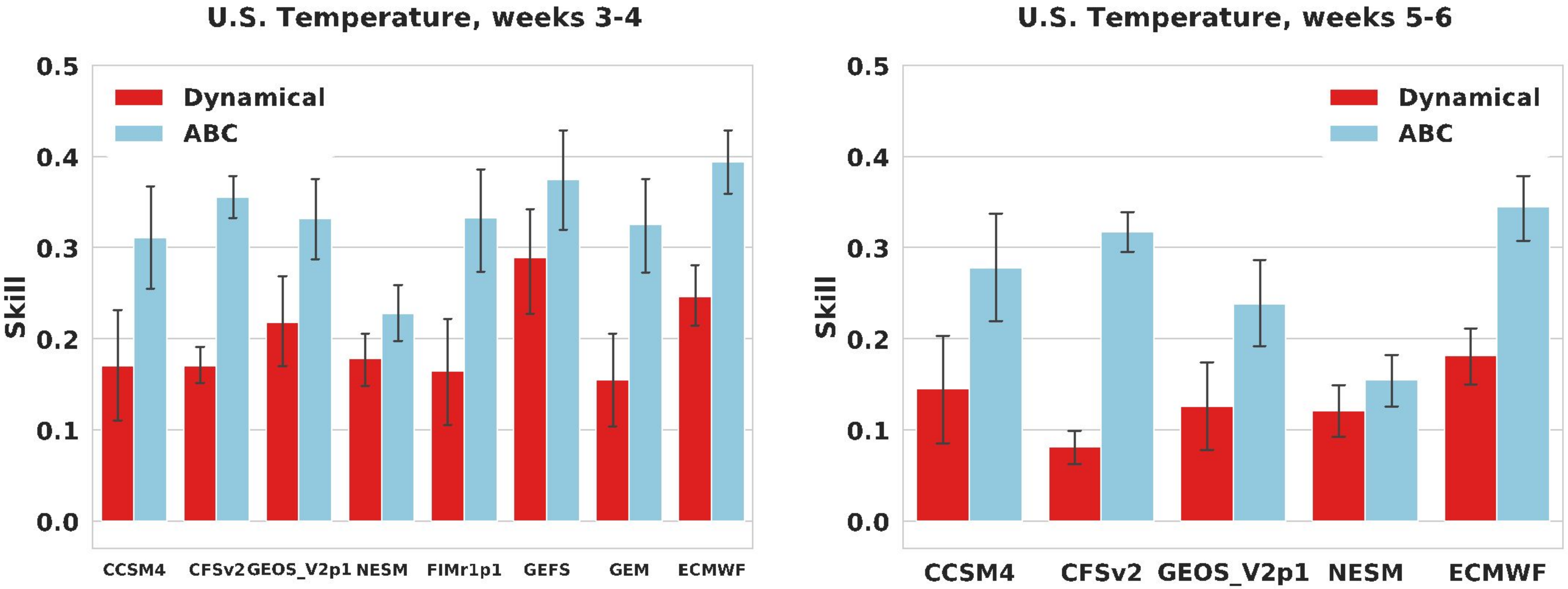
Adaptive Bias Correction (ABC): Hybrid Physics + Learning Model





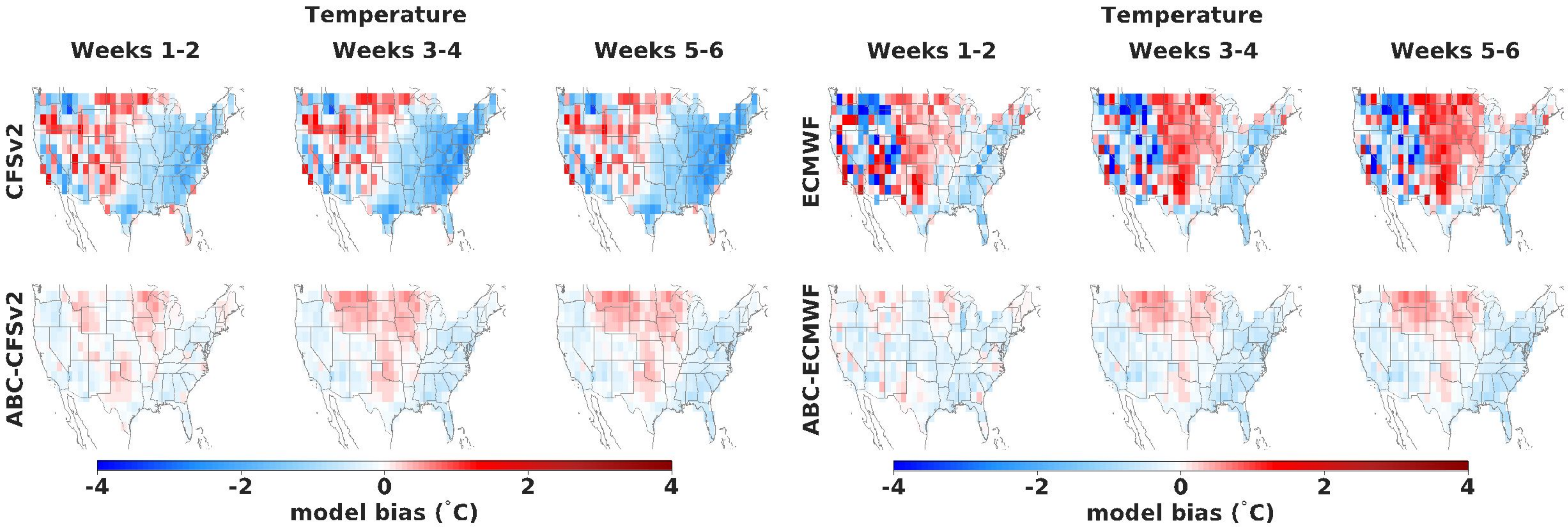
- Can be used to correct any dynamical model
- Including leading model from European Centre for Medium-Range Weather Forecasts

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 Reduces Systematic Model Bias

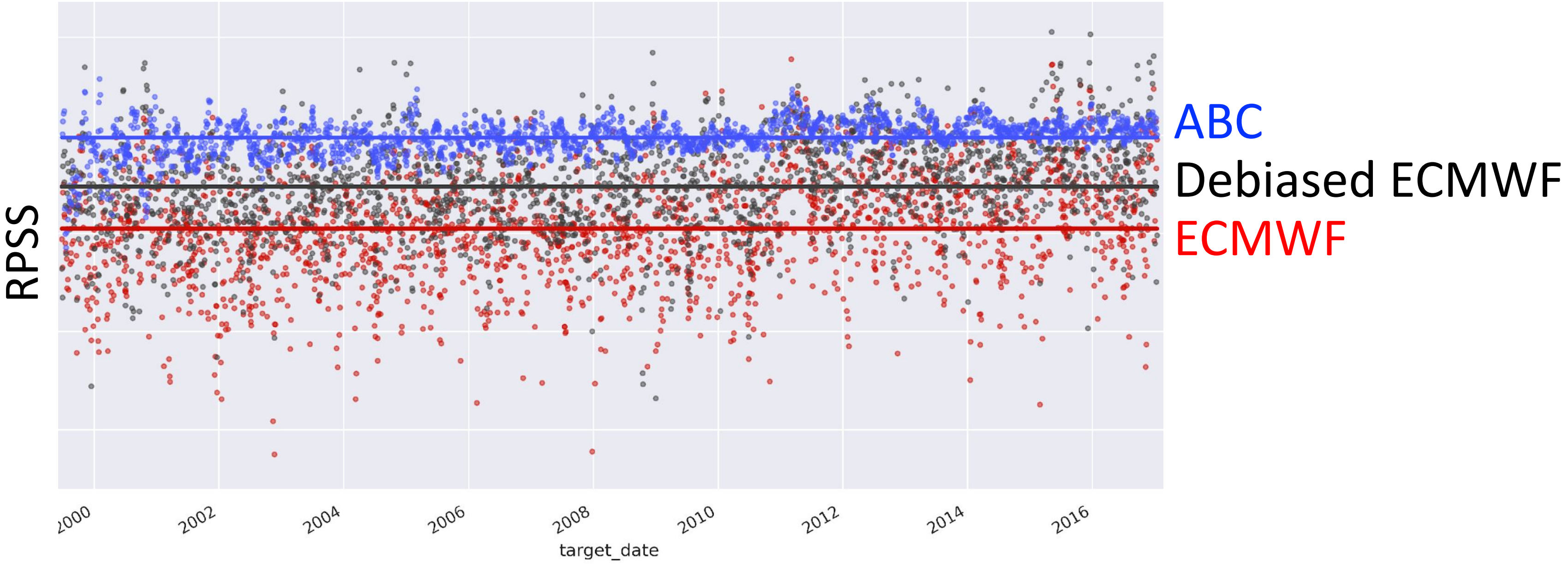


- Spatial distribution of model bias over the years 2018–2021
- CFSv2 = Climate Forecasting System v2, US operational dynamical model
- ECMWF = European Centre for Medium-Range Weather Forecasts, leading subseasonal model

Advancing Ensemble Subseasonal Forecasting with Machine Learning

Probabilistic ABC

- Forecast tercile probabilities
- Evaluate using Ranked Probability Skill Score (RPSS)

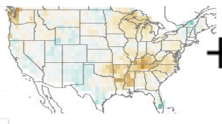
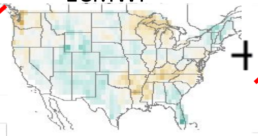
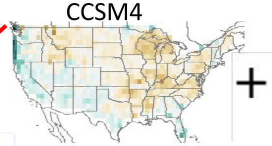


Optimistic online learning for adaptive model ensembling

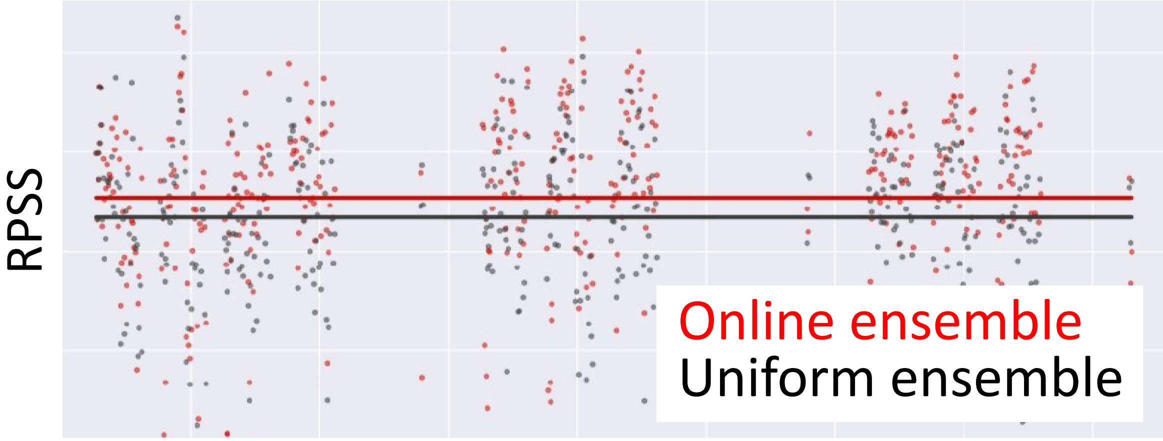
- Learns task- and time-dependent model weights w_i
- Accounts for delayed feedback (Flaspohler et al., 2021)
- No hyperparameters to tune

CFSv2 ECMWF CCSM4

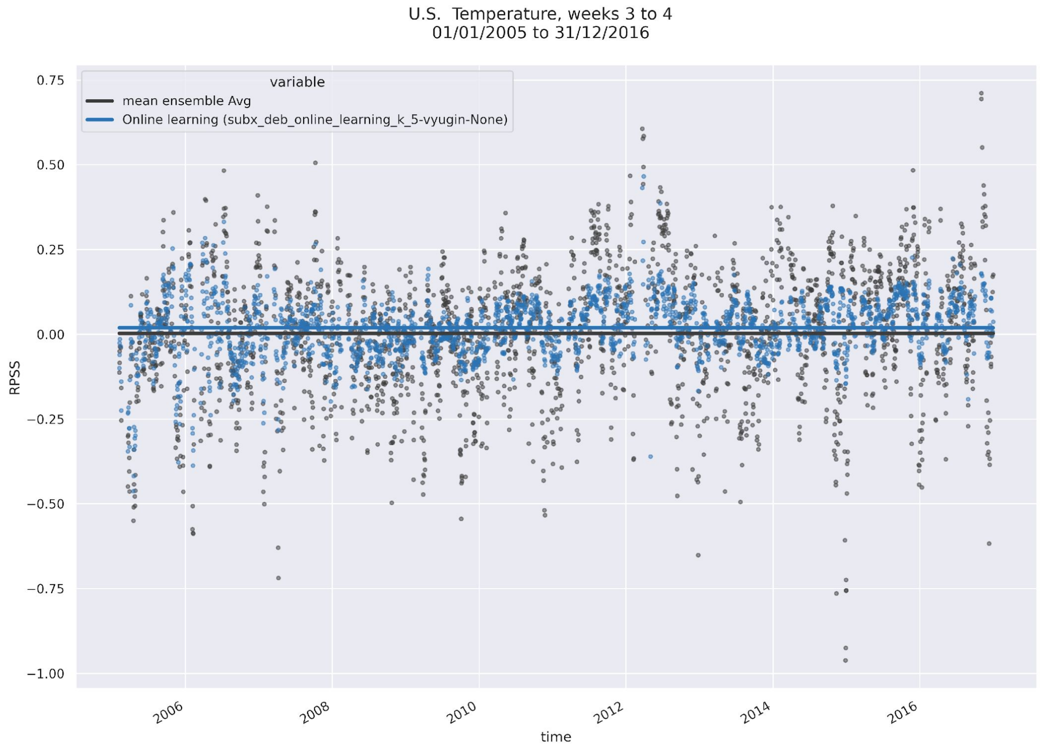
$$\hat{Y}_t = \cancel{\frac{1}{n}} \; + \; \cancel{\frac{1}{n}} \; + \; \cancel{\frac{1}{n}} \; + \cdots$$

0.1 0.6 0.1



With Sonja Totz, Genevieve Flaspohler, Judah Cohen, & Dara Entekhabi (MIT), Paulo Orenstein (IMPA), Arun Kumar & Yun Fan (CPC)



U.S. Temperature, weeks 3 to 4
01/01/2005 to 31/12/2016
variable
mean ensemble Avg
Online learning (subx_deb_online_learning_k_5-vyugin-None)
RPSS
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
2006
2008
2010
2012
2014
2016
time