NeurIPS 2021 Workshop
Tackling Climate Change with Machine Learning

Learned Benchmarks for Subseasonal Forecasting

Soukayna Mouatadid[1], Paulo Orenstein[2], Genevieve Flaspohler[3], Miruna Oprescu[4], Judah Cohen[5], Franklyn Wang[6], Sean Knight[3], Maria Geogdzhayeva[3], Sam Levang[7], Ernest Fraenkel[3], Lester Mackey[4]

1 University of Toronto
2 impa Instituto de Matemática Pura e Aplicada
3
4 Microsoft
5 aer
6 Veritas Harvard
7 salient Predictions

Introduction

Subseasonal Forecasting (3-6 weeks ahead) is a crucial pre-requisite for:

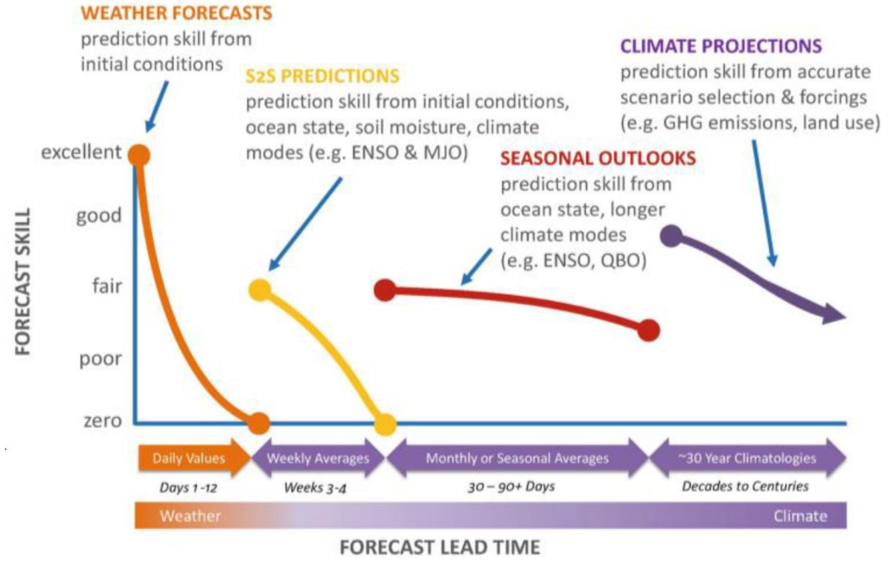
- Allocating water resources
- Preparing for droughts and floods
- Managing wildfires
- Agriculture planning

But…

It is a challenging forecast horizon for both meteorological and ML models

Objective:

- We develop a toolkit of subseasonal models that outperform operational weather models as well as state-of-the-art learning methods from the literature.



Source: Pechlivanidis and Crochemore, 2020

Forecasting Tasks

- Target variables:
  - Average temperature (°C)
  - Accumulated precipitation (mm)

- Lead times:
  - weeks 3-4 ahead
  - weeks 5-6 ahead

- Geographical region:
  - Contiguous U.S., on a 1° x 1° grid (G = 862 grid points)

- Evaluation metrics:
  - Root mean squared error: $\mathrm{RMSE}(\hat{\mathbf{y}}_t, \mathbf{y}_t) = \sqrt{\frac{1}{G}\sum_{g=1}^{G}(\hat{y}_{t,g} - y_{t,g})^2}$
  - Skill: $\mathrm{skill}(\hat{\mathbf{y}}_t, \mathbf{y}_t) = \frac{\langle \hat{\mathbf{y}}_t - \mathbf{c}_t, \mathbf{y}_t - \mathbf{c}_t \rangle}{\|\hat{\mathbf{y}}_t - \mathbf{c}_t\|_2 \cdot \|\mathbf{y}_t - \mathbf{c}_T\|_2} \in [-1, 1]$

Dataset

- *SubseasonalClimateUSA* dataset:
  - Regularly updated collection of ground-truth measurements and model forecasts.
  - Publicly accessible through the *subseasonal_data* Python package
- Variables include:
  - Temperature
  - Precipitation
  - CFSv2
  - Stratospheric geopotential height
  - Madden-Julian Oscillation
  - Multivariate ENSO index
  - Pressure
  - Relative humidity
  - Sea surface temperature
  - Sea ice concentration

Baseline Models

- Climatology
  - Standard baseline for subseasonal forecasting
  - Average temperature or precipitation for specific day and month over 1981-2010
- CFSv2
  - Operational U.S. physics-based model from NCEP
  - Main NWP baseline deployed in the U.S.
- Persistence
  - Today equals tomorrow

Learning Models

- **AutoKNN**, introduced in (Hwang et al., 2019)
- **Informer**, introduced in (Zhou, 2021)
- **LocalBoosting**, introduced in (Prokhorenkova et al., 2018)
- **MultiLLR**, introduced in (Hwang et al., 2019)
- **N-BEATS**, introduced in (Orenshkin, 2020)
- **Prophet**, introduced in (Taylor and Letham, 2018)
- **Salient 2.0**, introduced in (Schmitt, 2019)

Our Toolkit

- **Climatology++**
  - Use adaptively selected window around target day for averaging
- **CFSv2++**
  - Average over range of issuance date and lead times
  - Adaptively debiasing using selected window
- **Persistence++**
  - Learned combination of lagged measurements with NWP

Ensembling

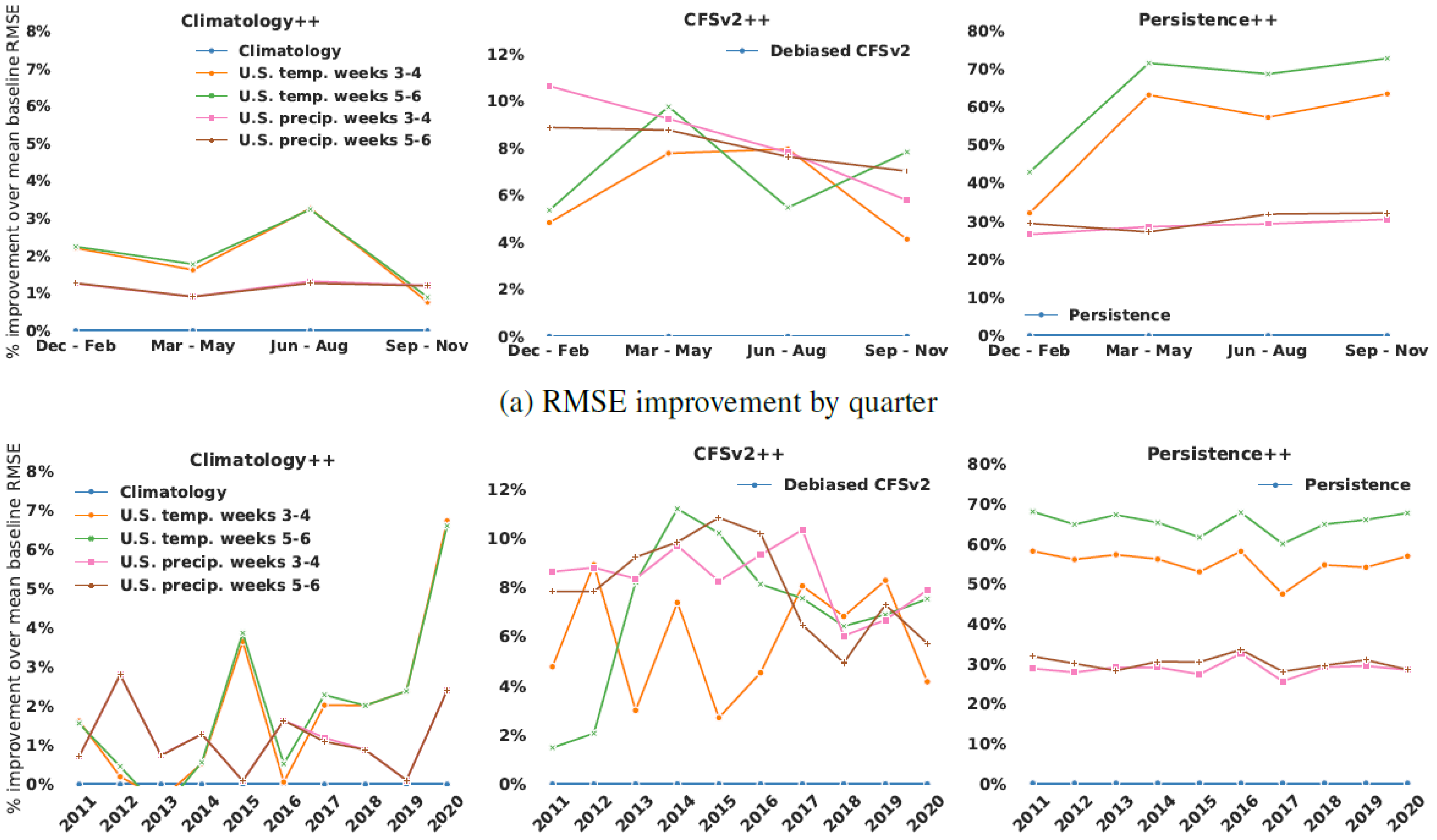
- **Uniform ensemble**
  - Equal-weighted average of the toolkit model forecasts
  - Standard solution in the weather community
- **Online ensemble**
  - Based on the AdaHedgeD online learning algorithm (Flaspohler et al., 2021)
  - Results in an adaptive convex combination of base models
- **Base models**
  - Climatology++, CFSv2++, Persistence++, LocalBoosting, MultiLLR and Salient 2.0

Results

Table 1. Average percentage skill and percentage improvement over mean debiased CFSv2 RMSE across 2011-2020 in the contiguous U.S. The best performing model in each model group is bolded, and the best performing model overall is shown in green.

		% Improvement over Mean Deb. CFSv2 RMSE				Average % Skill			
		Temperature		Precipitation		Temperature		Precipitation	
Group	Model	weeks 3-4	weeks 5-6	weeks 3-4	weeks 5-6	weeks 3-4	weeks 5-6	weeks 3-4	weeks 5-6
Baselines	Climatology	0.13	**2.93**	**7.79**	**7.51**	–	–	–	–
	Deb. CFSv2	–	–	–	–	**23.07**	**15.98**	4.79	3.01
	Persistence	−109.94	−170.10	−28.27	−31.92	9.40	5.77	**7.84**	**7.31**
Toolkit	Climatology++	2.01	4.83	**8.86**	**8.57**	19.84	20.23	**15.44**	15.23
	CFSv2++	5.89	**7.08**	8.36	8.06	**31.40**	**27.88**	15.38	**15.29**
	Persistence++	**6.00**	6.43	8.61	7.89	30.19	24.91	12.39	8.88
Learning	AutoKNN	0.93	3.22	7.73	7.33	12.41	9.63	5.76	5.06
	Informer	−39.99	−63.66	0.65	0.19	−5.17	−1.46	5.70	5.16
	LocalBoosting	−0.76	−0.29	7.36	6.89	14.67	12.29	11.11	9.58
	MultiLLR	**2.45**	2.21	7.12	6.65	**22.37**	15.62	9.62	7.52
	N–BEATS	−46.71	−52.05	−19.19	−21.32	7.95	2.79	5.14	4.18
	Prophet	1.13	**3.78**	**8.42**	**8.12**	21.13	**20.55**	**13.41**	**13.26**
	Salient 2.0	−6.84	−3.95	2.99	2.66	12.46	13.45	9.24	8.92
Ensembles	Uniform Toolkit	6.47	7.55	9.47	**9.05**	31.96	**28.93**	18.05	**17.54**
	Online Toolkit	**6.71**	**7.67**	**9.51**	9.04	**32.07**	28.63	**18.19**	17.30

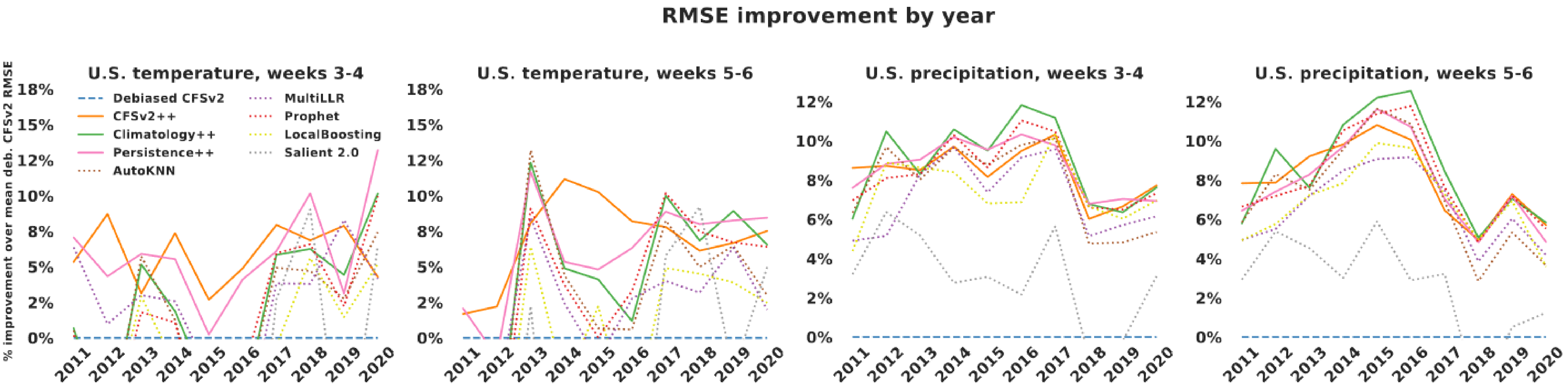
Results: toolkit vs baselines



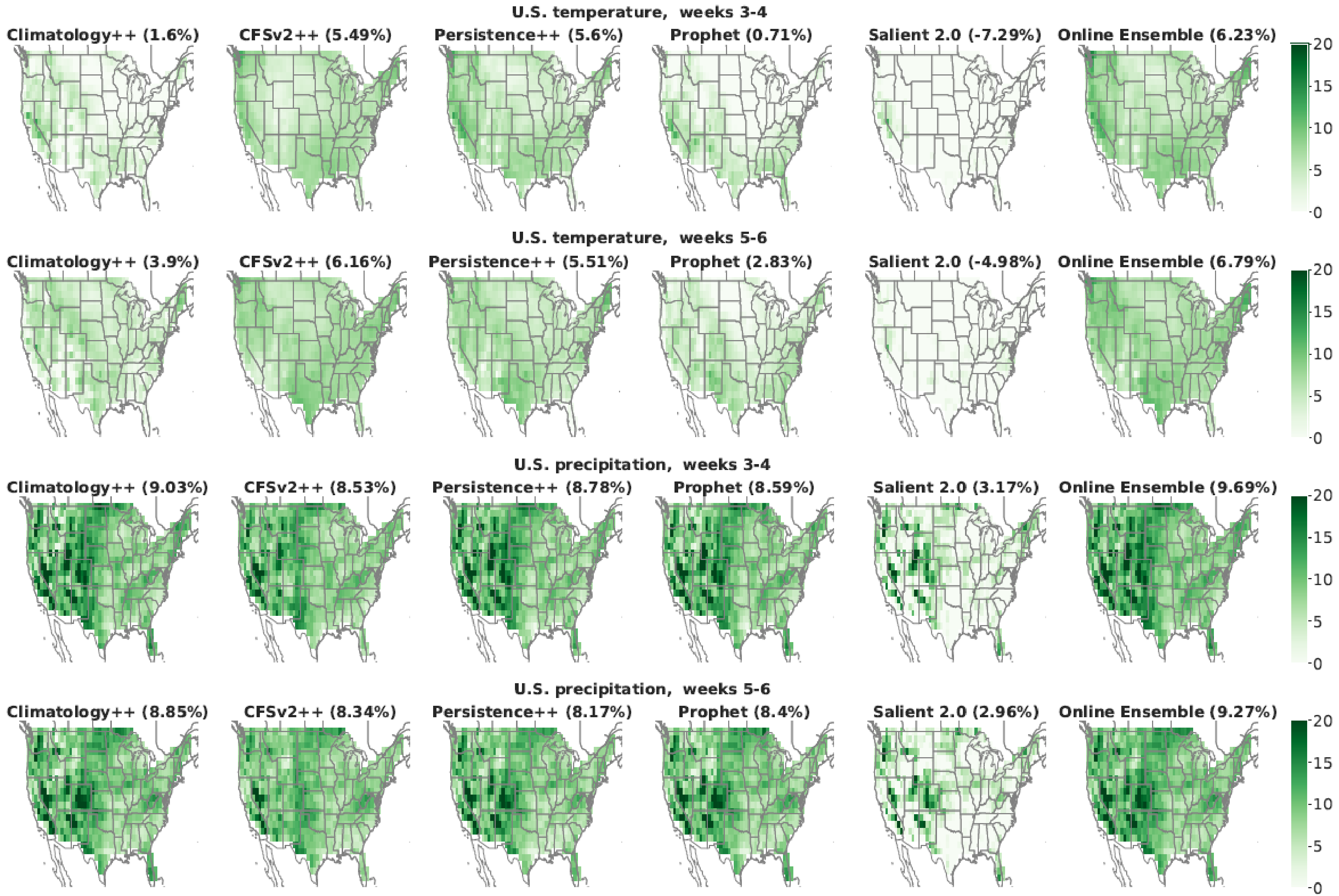
(a) RMSE improvement by quarter

(b) RMSE improvement by year

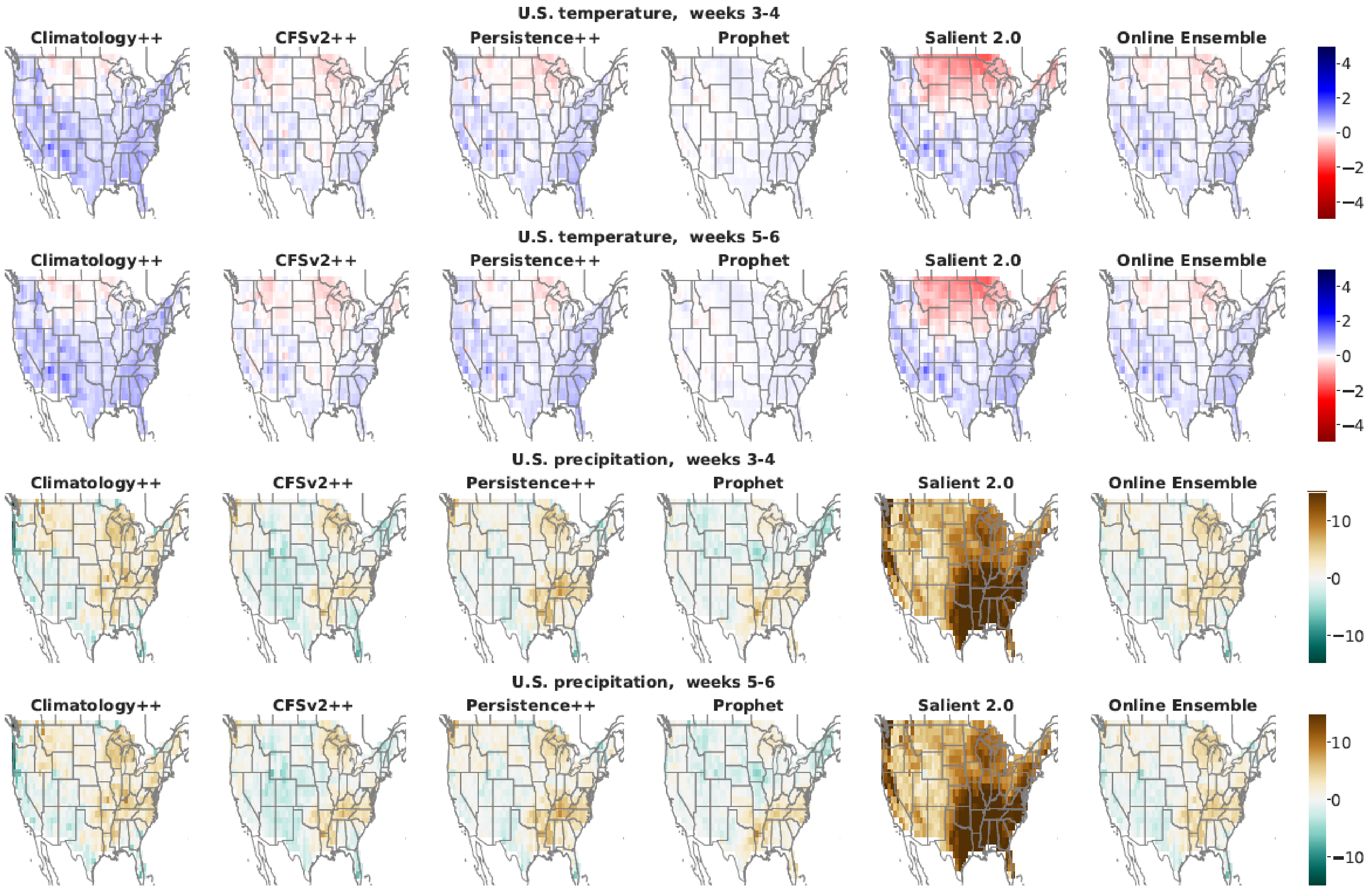
Results: toolkit vs learning



Results: Percentage improvement over mean deb. CFSv2 RMSE



Results: Mean model bias



Results: Comparing to ECMWF

Group	Model	% Improvement over Mean Deb. CFSv2 RMSE				Average % Skill			
		Temperature		Precipitation		Temperature		Precipitation	
		weeks 3-4	weeks 5-6	weeks 3-4	weeks 5-6	weeks 3-4	weeks 5-6	weeks 3-4	weeks 5-6
Baselines	Climatology	**1.56**	**3.92**	**8.7**	**7.56**	–	–	–	–
	Debiased CFSv2	–	–	–	–	**22.64**	**15.71**	2.84	1.68
	Persistence	−105.57	−169.22	−28.05	−33.43	9.12	2.27	**8.11**	**6.21**
Toolkit	Climatology++	3.88	6.44	**9.79**	**8.61**	22.09	23.2	**15.34**	**15.06**
	CFSv2++	5.65	6.65	8.94	7.6	30.91	26.87	14.6	13.85
	Persistence++	**7.06**	**7.86**	9.06	7.57	**31.46**	**28.04**	10.03	6.61
ECMWF	Debiased Control	−29.05	−33.25	−30.81	−31.84	18.52	13.71	0.82	3.17
	Debiased Ensemble	**4.62**	**3.69**	**7.90**	**6.41**	**32.27**	**26.61**	**13.12**	**9.10**
Ensembles	Uniform Toolkit	**7.43**	**8.27**	10.04	**8.77**	**32.77**	**29.75**	16.53	**15.71**
	Online Toolkit	7.2	7.96	**10.08**	8.62	32.22	28.38	**17.19**	15.42

Conclusion

- Subseasonal forecasting is a hard but fundamental problem
- Adaptive de-biasing of classical benchmarks yields sizable improvement
- Toolkit models are not only accurate, but highly scalable
- Online ensembling is highly advantageous
- Combining NWP and ML models is a powerful strategy
- All data and code are open source