AI4ES Journal Club

Presenter: Lluís Palma
Group: Computational Earth Science

BSC-CNS

2020/10/14

Article(s) of the day

Improving Subseasonal Forecasting in the Western U.S. with Machine Learning

Jessica Hwang
Department of Statistics, Stanford University
Stanford, California
jjhwang@stanford.edu

Paulo Orenstein
Department of Statistics, Stanford University
Stanford, California
pauloo@stanford.edu

Judah Cohen
Atmospheric and Environmental Research
Lexington, MA
jcohen@aer.com

Karl Pfeiffer
Atmospheric and Environmental Research
Lexington, MA
kpfeiffe@aer.com

Lester Mackey
Microsoft Research New England
Cambridge, MA
lmackey@microsoft.com

Link: https://arxiv.org/abs/1809.07394

Link: https://arxiv.org/abs/2006.07972

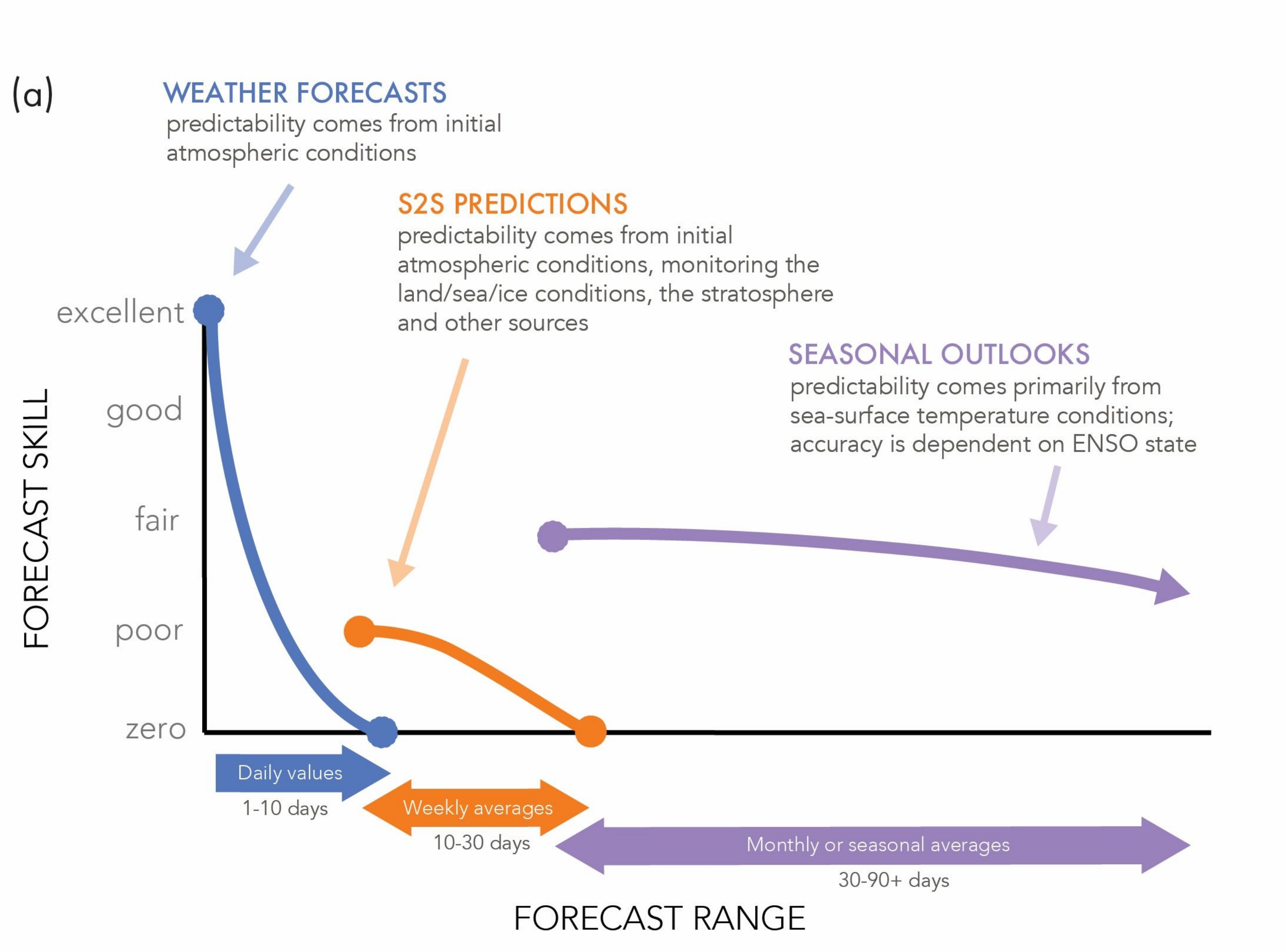
Sub-Seasonal Climate Forecasting via Machine Learning: Challenges, Analysis, and Advances

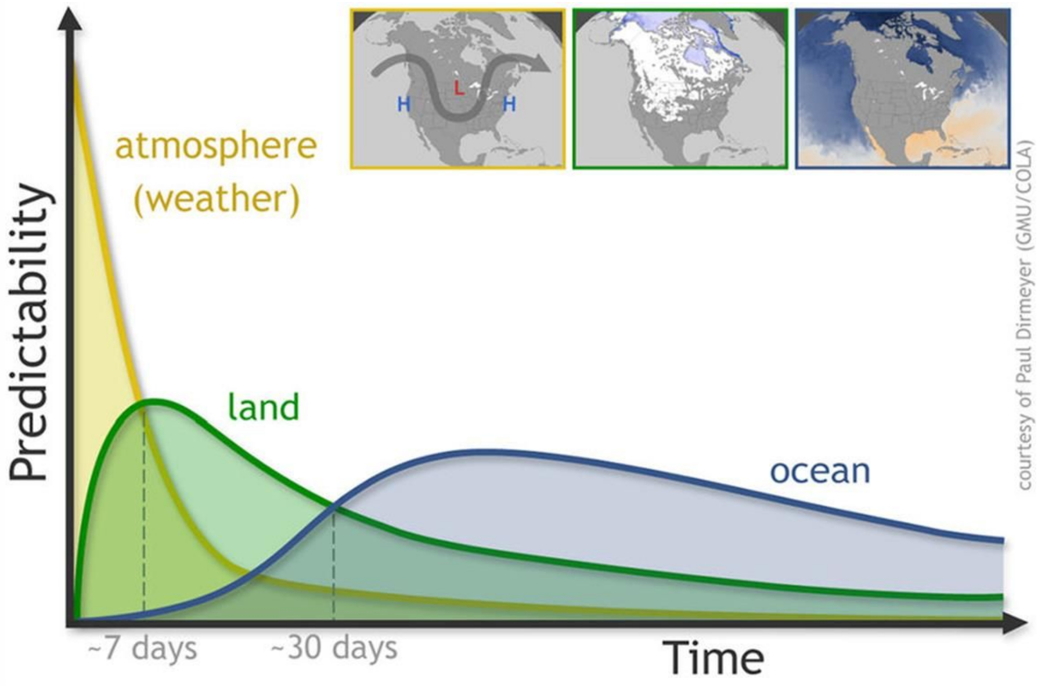
Sijie He *[1], Xinyan Li *[1], Timothy DelSole[2], Pradeep Ravikumar[3], and Arindam Banerjee[1]

[1]Department of Computer Science & Engineering, University of Minnesota, Twin Cities
[2]Department of Atmospheric, Oceanic, and Earth Science, George Mason University
[3]Machine Learning Department, Carnegie Mellon University
Email: {hexxx893@umn.edu, lixx1166@umn.edu, tdelsole@gmu.edu, pradeepr@cs.cmu.edu, banerjee@cs.umn.edu}

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

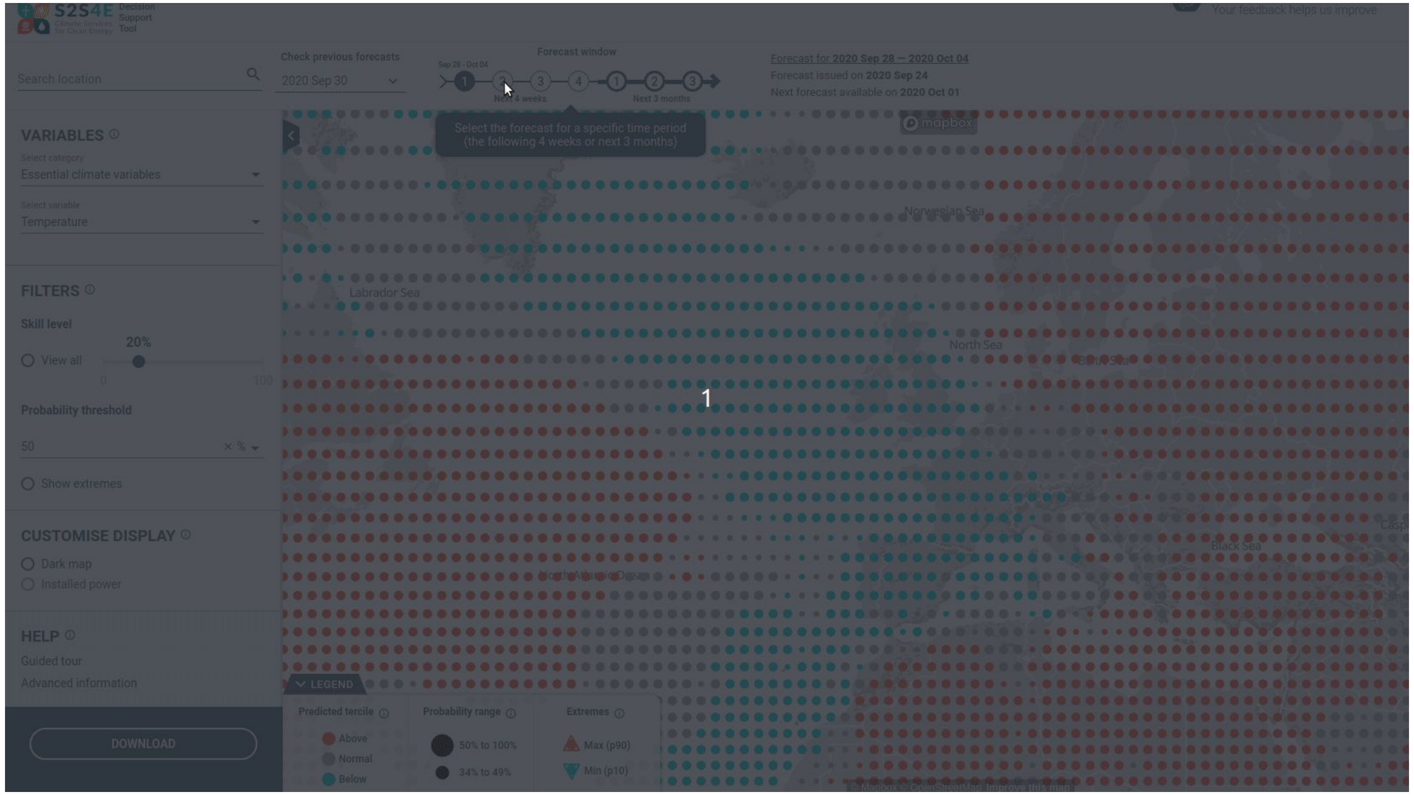
Context



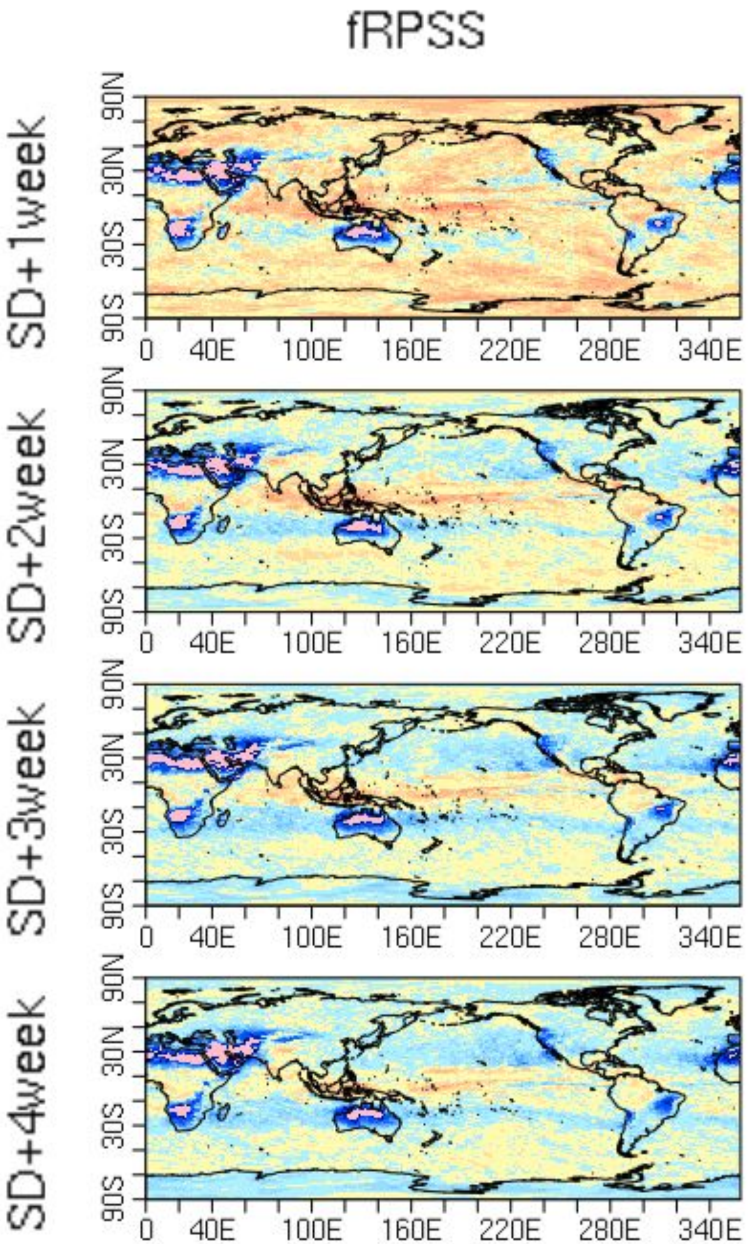


Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Context



https://s2s4e-dst.bsc.es/#/



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Context

Aim of the article:

- Propose an alternative ML model capable of improving subseasonal temperature and precipitation forecast on the western US region, in comparison to the NOAA's dynamical model; CFSv2 (under the context of Subseasonal forecast Rodeo competition).

How?

- Two linear regression models: a local linear regression model with multitask feature selection (MultiLLR) and a weighted local autoregression enhanced with multitask k-nearest neighbor features (AutoKNN).

Why?

- Immense societal value, having an impact in a wide variety of domains.
- Forecasts based on dynamical models with changes on land/sea processes. There is a big room for improvement on skill, particularly on weeks 2 to 4 onwards.
- Computationally efficient methods that exploited the multitask, i.e., multiple grid point, nature of our problem and incorporated the unusual forecasting skill objective function.

Data

	Type	Freq	Range	Resolution
Temperature (tmax and tmin)	Reanalysis	daily	1979 - 2019	1ºx1º
Precipitation	Gauge-Based Analysis	daily	1979 - 2019	1ºx1º
SST & ICEC	Reanalysis PCs	-	1981 - 2010	1ºx1º ?
ENSO index (MEI)	6 variables aggregation	Bimonthly	1949 - 2019	scalar
MJO	Phase and amplitude	Daily	1974-2019	scalar
RH & Sfc. Pressure	Reanalysis	daily	1948-2019	1ºx1º
Geopotential Height	Reanalysis PCs	daily	1948-2010	1ºx1º ?
NMME	Dynamical Forecast	monthly (weight avg)	-	1ºx1º

Methods

The **MultiLLR** model introduces candidate regressors from each data source in the SubseasonalRodeo dataset and then prunes irrelevant predictors using a multitask backward stepwise criterion designed for the forecasting skill objective.

- Variables are selected for a target date jointly for all grid points, while the coefficients associated with those variables are fit independently for each grid point using local linear regression.

- Specifically, the training data for a given target date is restricted to a 56-day (8-week) span around the target date’s day of the year (s = 56).
    - If the target date is May 2, 2017, the training data consists of days within 56 days of May 2 in any year.

- The skill for a target date t is the cosine similarity achieved by holding out a year’s worth of data around t, fitting the model on the remaining data, and predicting the outcome for t.

Methods

Autoknn is a weighted local linear regression with features derived exclusively from historical measurements of the target variable (temperature or precipitation).

- When predicting weeks 3-4, we include lagged temperature or precipitation anomalies from 29 days, 58 days, and 1 year prior to the target date; when predicting weeks 5-6, we use 43 days, 86 days, and 1 year.

- In addition to fixed lags, we include the constant intercept ones and the observed anomaly patterns of the target variable on similar (cosine similarity) dates in the past.
  - 20 Knns for temperature 1 for precipitation.

- To predict a given target date, we regress onto the three fixed lags, the constant intercept feature ones, and either knn1 through knn20 (for temperature) or knn1 only (for precipitation), treating each grid point as a separate prediction task.

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Results

task	multillr	autoknn	ensemble	contest debiased cfsv2	damped	top competitor
temperature, weeks 3-4	0.3079	0.2807	**0.3451**	0.1589	0.1952	0.2855
temperature, weeks 5-6	0.2562	0.2817	**0.3025**	0.2192	-0.0762	0.2357
precipitation, weeks 3-4	0.1597	0.2156	**0.2364**	0.0713	-0.1463	0.2144
precipitation, weeks 5-6	0.1876	0.1870	**0.2315**	0.0227	-0.1613	0.2162

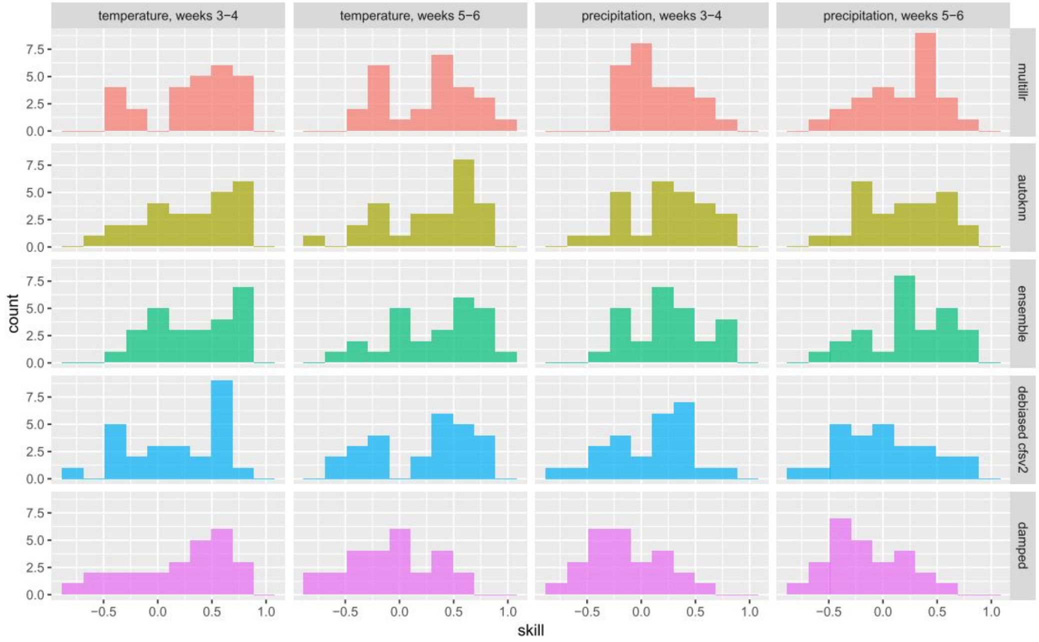


Figure 1: Distribution of contest-period skills of the proposed models MultiLLR and AutoKNN, the proposed ensemble of MultiLLR and AutoKNN (*ensemble*), the official contest debiased-CFSv2 baseline, and the official contest damped-persistence baseline (*damped*). See Section 5.2 for more details.

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Results

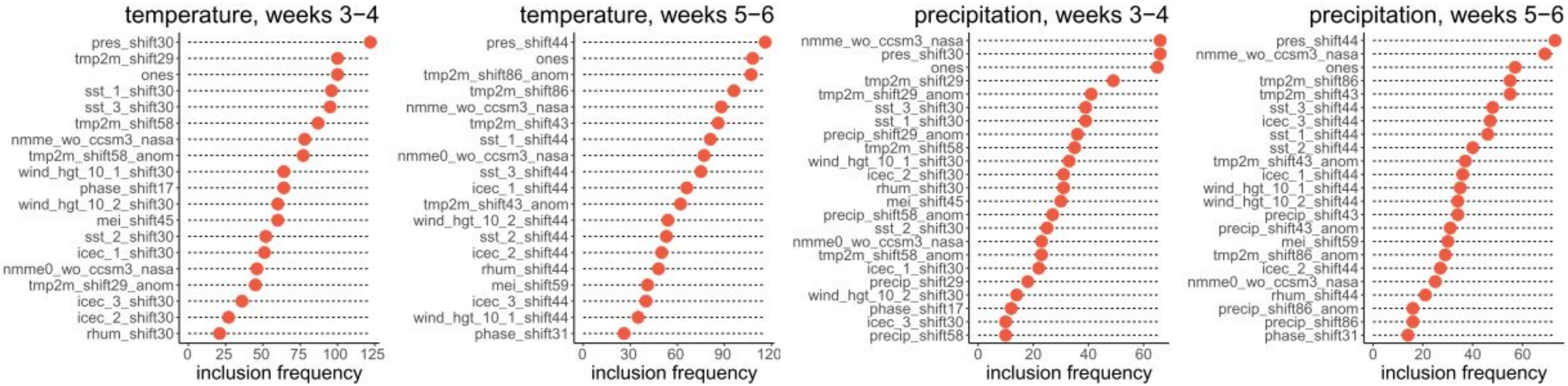
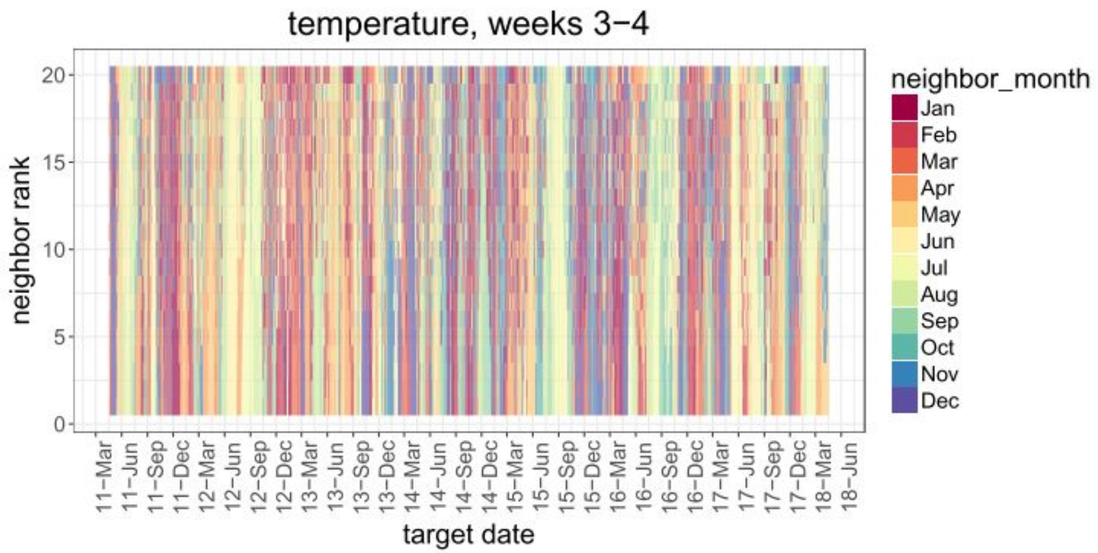
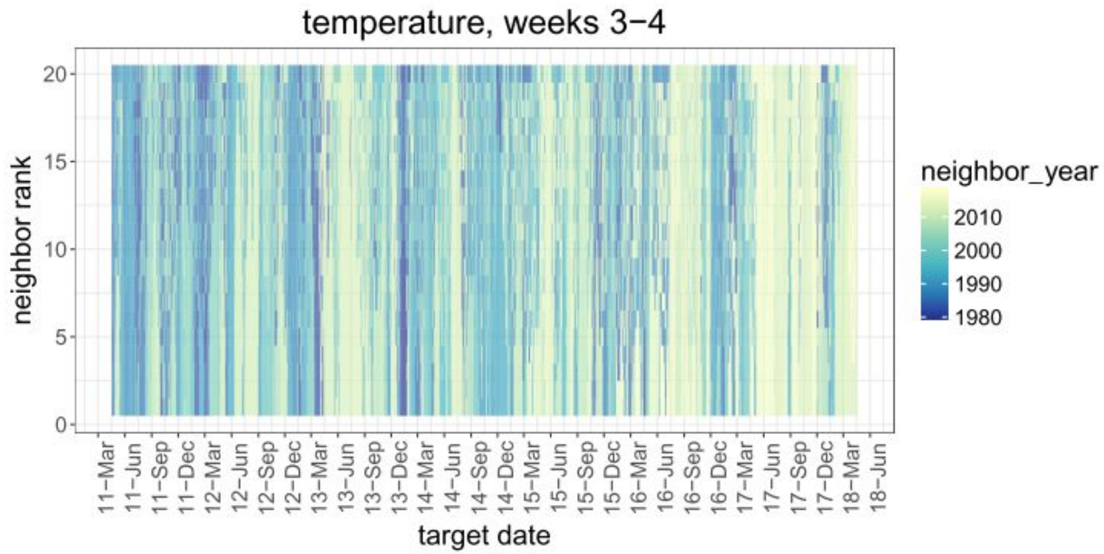
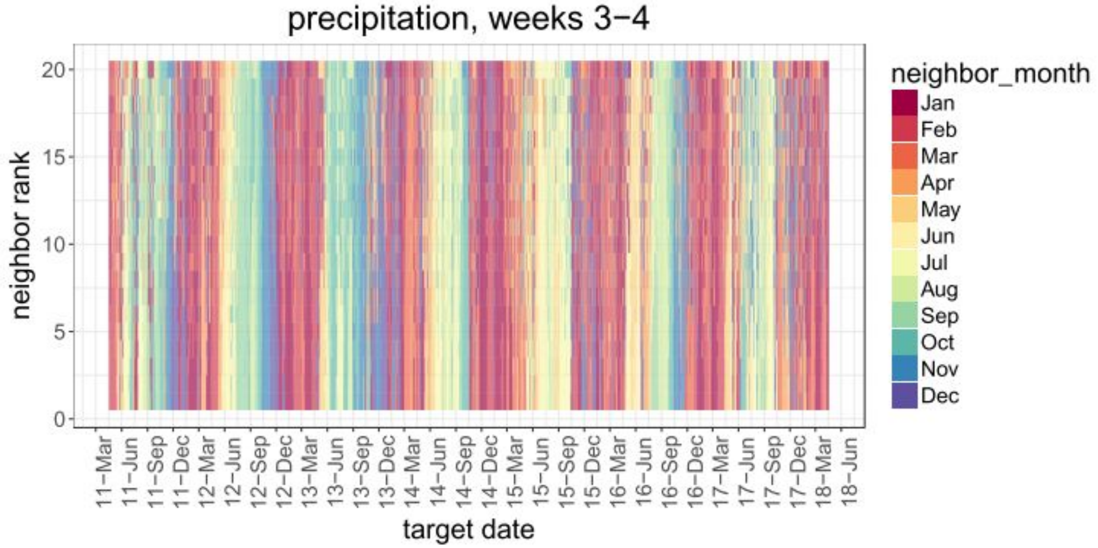


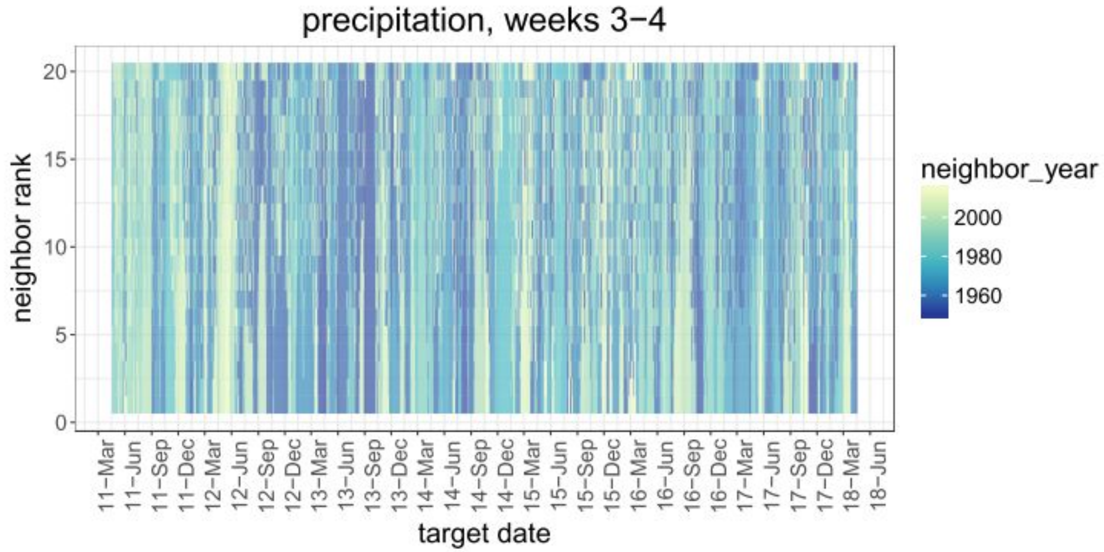
Figure 2: Feature inclusion frequencies of all candidate variables for local linear regression with multitask feature selection (`MultiLLR`) across all target dates in the historical forecast evaluation period (see Section 5.4).

Results


temperature, weeks 3−4
neighbor rank
0
5
10
15
20
target date
11−Mar 11−Jun 11−Sep 11−Dec 12−Mar 12−Jun 12−Sep 12−Dec 13−Mar 13−Jun 13−Sep 13−Dec 14−Mar 14−Jun 14−Sep 14−Dec 15−Mar 15−Jun 15−Sep 15−Dec 16−Mar 16−Jun 16−Sep 16−Dec 17−Mar 17−Jun 17−Sep 17−Dec 18−Mar 18−Jun
neighbor_month
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec


temperature, weeks 3−4
neighbor rank
0
5
10
15
20
target date
11−Mar 11−Jun 11−Sep 11−Dec 12−Mar 12−Jun 12−Sep 12−Dec 13−Mar 13−Jun 13−Sep 13−Dec 14−Mar 14−Jun 14−Sep 14−Dec 15−Mar 15−Jun 15−Sep 15−Dec 16−Mar 16−Jun 16−Sep 16−Dec 17−Mar 17−Jun 17−Sep 17−Dec 18−Mar 18−Jun
neighbor_year
2010
2000
1990
1980


precipitation, weeks 3−4
neighbor rank
0
5
10
15
20
target date
11−Mar 11−Jun 11−Sep 11−Dec 12−Mar 12−Jun 12−Sep 12−Dec 13−Mar 13−Jun 13−Sep 13−Dec 14−Mar 14−Jun 14−Sep 14−Dec 15−Mar 15−Jun 15−Sep 15−Dec 16−Mar 16−Jun 16−Sep 16−Dec 17−Mar 17−Jun 17−Sep 17−Dec 18−Mar 18−Jun
neighbor_month
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec


precipitation, weeks 3−4
neighbor rank
0
5
10
15
20
target date
11−Mar 11−Jun 11−Sep 11−Dec 12−Mar 12−Jun 12−Sep 12−Dec 13−Mar 13−Jun 13−Sep 13−Dec 14−Mar 14−Jun 14−Sep 14−Dec 15−Mar 15−Jun 15−Sep 15−Dec 16−Mar 16−Jun 16−Sep 16−Dec 17−Mar 17−Jun 17−Sep 17−Dec 18−Mar 18−Jun
neighbor_year
2000
1980
1960

Conclusions

- We release a new SubseasonalRodeo dataset suitable for training and benchmarking subseasonal forecasts.
- We introduce two subseasonal regression approaches tailored to the forecast skill objective, one of which uses only features of the target variable.
- We introduce a simple ensembling procedure that probably improves average skill whenever average skill is positive.
- We show that each regression method alone outperforms the Rodeo benchmarks, including a debiased version of the operational U.S. Climate Forecasting System (CFSv2), and that our ensemble outperforms the top Rodeo competitor.
- We show that, over 2011-2018, an ensemble of our models and debiased CFSv2 improves debiased CFSv2 skill by 40-50% for temperature and 129-169% for precipitation.

Critique

- Very heterogeneous dataset -> wide variety of preprocessing techniques applied (weighthing, PCs, regridding...)
  - Wide variety of regressors (anomalies, time-lagged)
  - Complexity and trivial decissions
- What about more complex models (DL, XGboost...) ?
  - Non-linear spatial information
  - Temporal sequence information
- Is the skill score used the best solution for each variable?
- For the multiLLR; the top regressors list makes sense?

Context

Aim of the article:

- Explore a wide variety of ML for subseasonal forecasting, extending the current state of the art implementations

How?

- Implementation of multiple ML models (XGboost, Lasso, LSTM-FNN)

Why?

- Immense societal value, having an impact in a wide variety of domains.
- Forecasts based on dynamical models with changes on land/sea processes. There is a big room for improvement on skill, particularly on weeks 2 to 4 onwards.

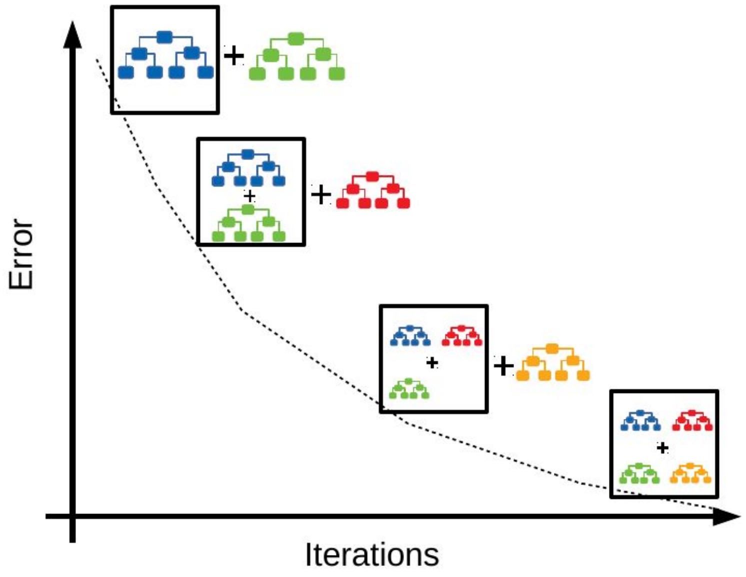
Data

Table 1: Description of climate variables and their data sources.

Type	Climate variable	Description	Unit	Spatial coverage	Data Source
Spatial-temporal	tmp2m	Daily average temperature at 2 meters	C°	US mainland	CPC Global Daily Temperature [13]
	sm	Monthly Soil moisture	mm		CPC Soil Moisture [25, 53, 12]
	sst	Daily sea surface temperature	C°	North Pacific & Atlantic Ocean	Optimum Interpolation SST (OISST) [41]
	rhum	Daily relative humidity near the surface (sigma level 0.995)	%	US mainland and North Pacific & Atlantic Ocean	Atmospheric Research Reanalysis Dataset [28]
	slp	Daily pressure at sea level	Pa		
	hgt10 & hgt500	Daily geopotential height at 10mb and 500mb	m		
Temporal	MEI	Bimonthly multivariate ENSO index	NA	NA	NOAA ESRL MEI.v2 [62]
	Niño 1+2, 3, 3.4, 4	Weekly Oceanic Niño Index (ONI)			NOAA National Weather Service, CPC [41]
	NAO	Daily North Atlantic Oscillation index			NOAA National Weather Service, CPC [3, 51]
	MJO phase & amplitude	Madden-Julian Oscillation index			Australian Government BoM [60]

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Methods



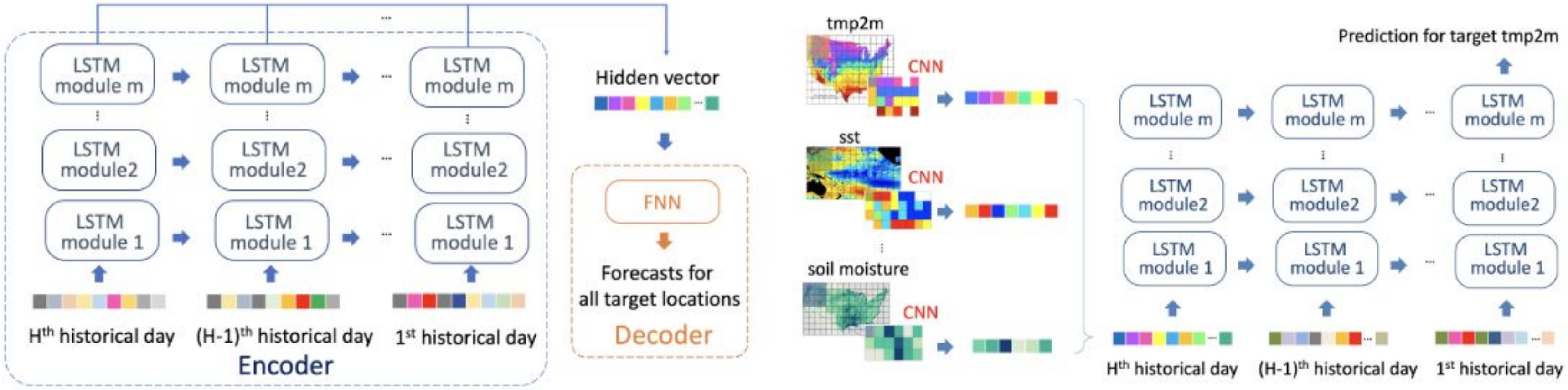
- **Autoknn and MultiLLR**
- **Multitask Lasso**
  - “Multilinear regression with added penalty coefficient”
- **Gradient boosting trees**
- **State-of-the-art climate baseline:**
  - Both are Least Square (LS) linear regression models [59]. The first model has predictors as climate indices, such as NAO index and Niño indices, which are used to monitor ocean conditions. The predictor of the second model is the most recent anomaly of the target variable, i.e., anomaly temperature of week -2 & -1, with which the model, also known as damped persistence [52] in climate science, is essentially a first-order autoregressive model.

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Methods

- **Encoder (LSTM)-Decoder (FNN)**
  - **Input of the model is features extracted spatially from covariates (PCA).**
  - **The temporal components of covariates are handled by feeding features of each historical date into an LSTM Encoder recurrently.**
  - **The output of each date from LSTM is sent jointly to a two-layer FNN network using ReLU as an activation function. The output of the FNN Decoder is the predicted average temperature of week 3 & 4 over all target locations.**
- **CNN-LSTM**
  The proposed CNN-LSTM model directly learns the representations from the spatial-temporal data using CNN components [31]. CNN extracts features for each climate variable at all historical dates separately. Then, the extracted features from the same date are collected and fed into an LSTM model recurrently. The temperature prediction for all target locations is done by an FNN layer taking the output of the LSTM's last layer from the latest input.

Methods



(a) Encoder (LSTM)-Decoder (FNN)

(b) CNN-LSTM

Figure 2: Architectures of the designed DL models. (a) Encoder (LSTM)-Decoder (FNN) includes a few LSTM layers as the Encoder, and two fully connected layers as the Decoder. (b) CNN-LSTM consists of a few convolutional layers followed by an LSTM.

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Results

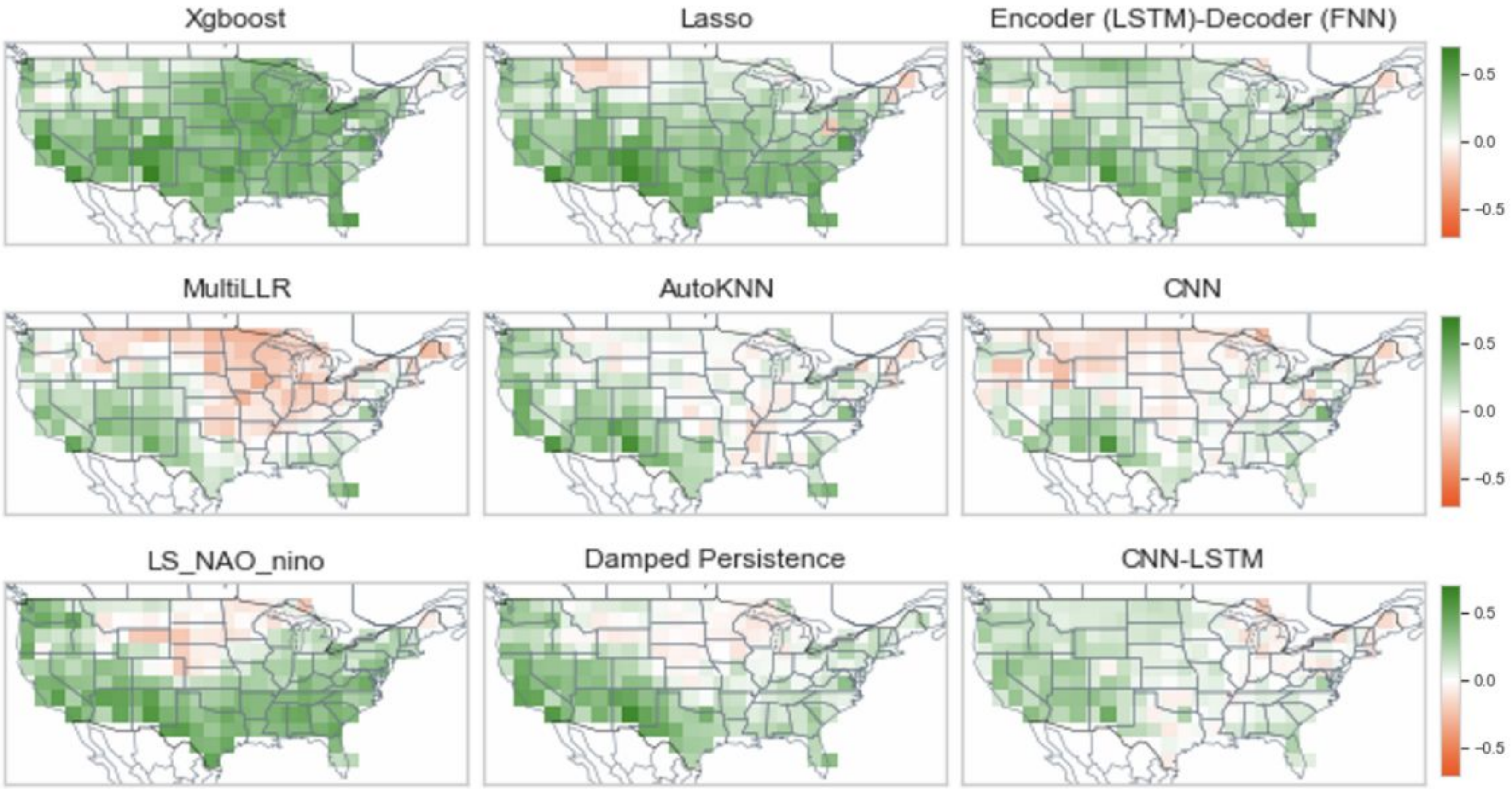
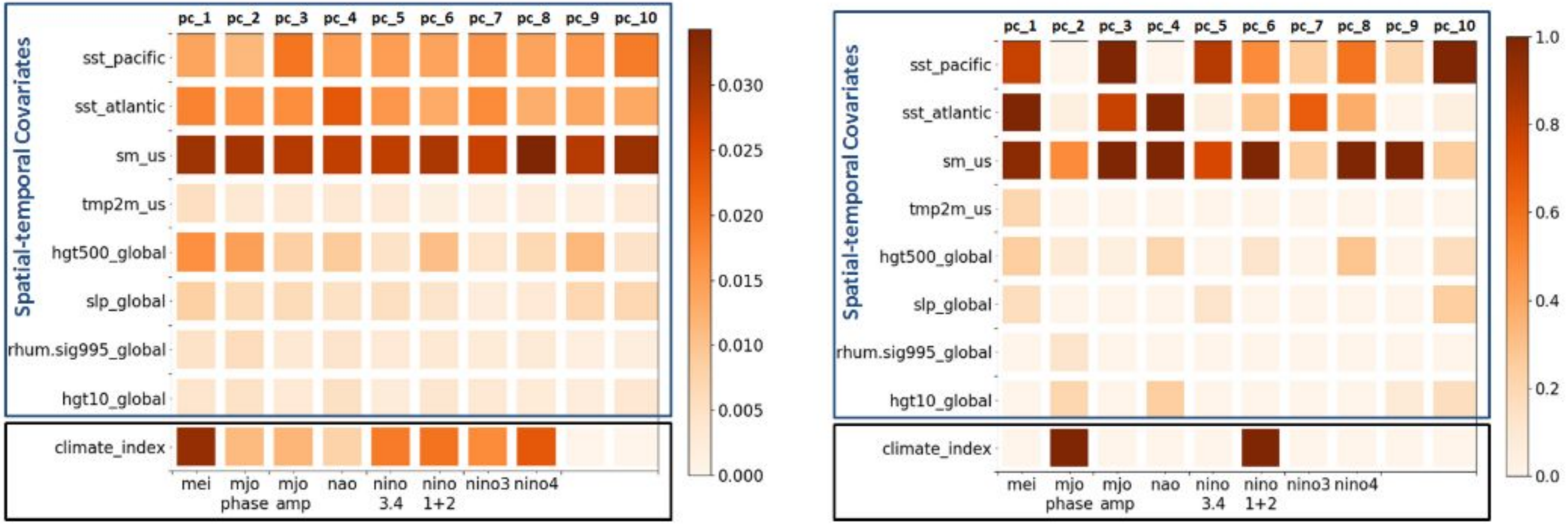


Figure 3: Temporal cosine similarity over the US mainland of ML models discussed in Section 4 for temperature prediction over 2017-2018. Large positive values (green) closer to 1 indicates better predictive skills. Overall, XGBoost and Encoder (LSTM)-Decoder (FNN) perform the best. Qualitatively, coastal and south regions are easier to predict than inland regions (e.g., Midwest).

Results



(a) XGBoost (b) Lasso

Figure 4: Feature importance scores computed from (a) XGBoost and (b) Lasso. Darker color means a covariate is of the higher importance. The first 8 rows contains the top 10 principal components

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Conclusions

- We illustrate the difficulty of SSF due to the complex physical couplings as well as the unique nature of climate data, i.e., strong spatial-temporal correlation and high-dimensionality.

- We show that suitable ML models, e.g., XGBoost, to some extent, capture predictability for sub-seasonal time scales from climate data, and persistently outperform existing approaches in climate science, such as climatology and the damped persistence model.

- We demonstrate that even though DL models are not the obvious winner, they still show promising results with demonstrated improvements from careful architectural choices. With further improvements, DL models present a great potential topic for future research.

- We find that ML models tend to select covariates from the land and ocean, such as soil moisture and El Niño indices, and rarely select atmospheric covariates, such as 500mb geopotential height

Critique

- What about more complex models (DL, XGboost...) ?
  - Non-linear spatial information
  - Temporal sequence information
- Is the skill score used the best solution for each variable?
- For the multiLLR; the top regressors list makes sense?
- Failure on CNN, PCA preferred method for extracting spatio-temporal information from fields.
  - Different approaches?