Improving Subseasonal Forecasting in the Western U.S.

Lester Mackey

Microsoft Research New England

Joint work with **Jessica Hwang & Paulo Orenstein** (Stanford),
Judah Cohen & Karl Pfeiffer (Atmospheric and Environmental Research), **Ernest Fraenkel** (MIT)

Goals

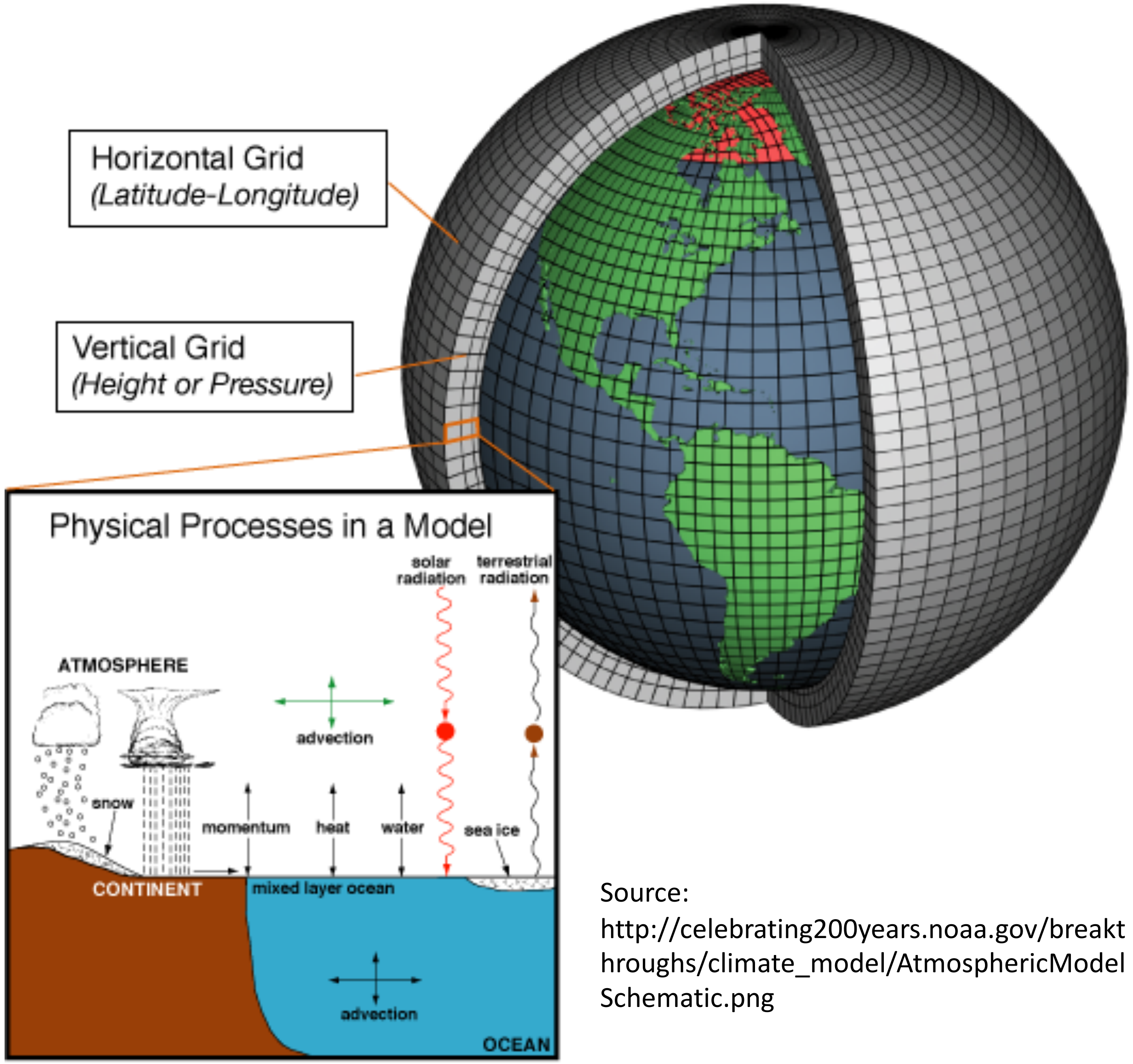
- **Awareness:** Subseasonal forecasting
- **Crowdsourced science:** The Subseasonal Climate Forecast Rodeo
- **The SubseasonalRodeo Dataset:** https://doi.org/10.7910/DVN/IHBANG
- **Machine learning:** Weighted locally linear regression, multitask model selection, multitask KNN, ensembling

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

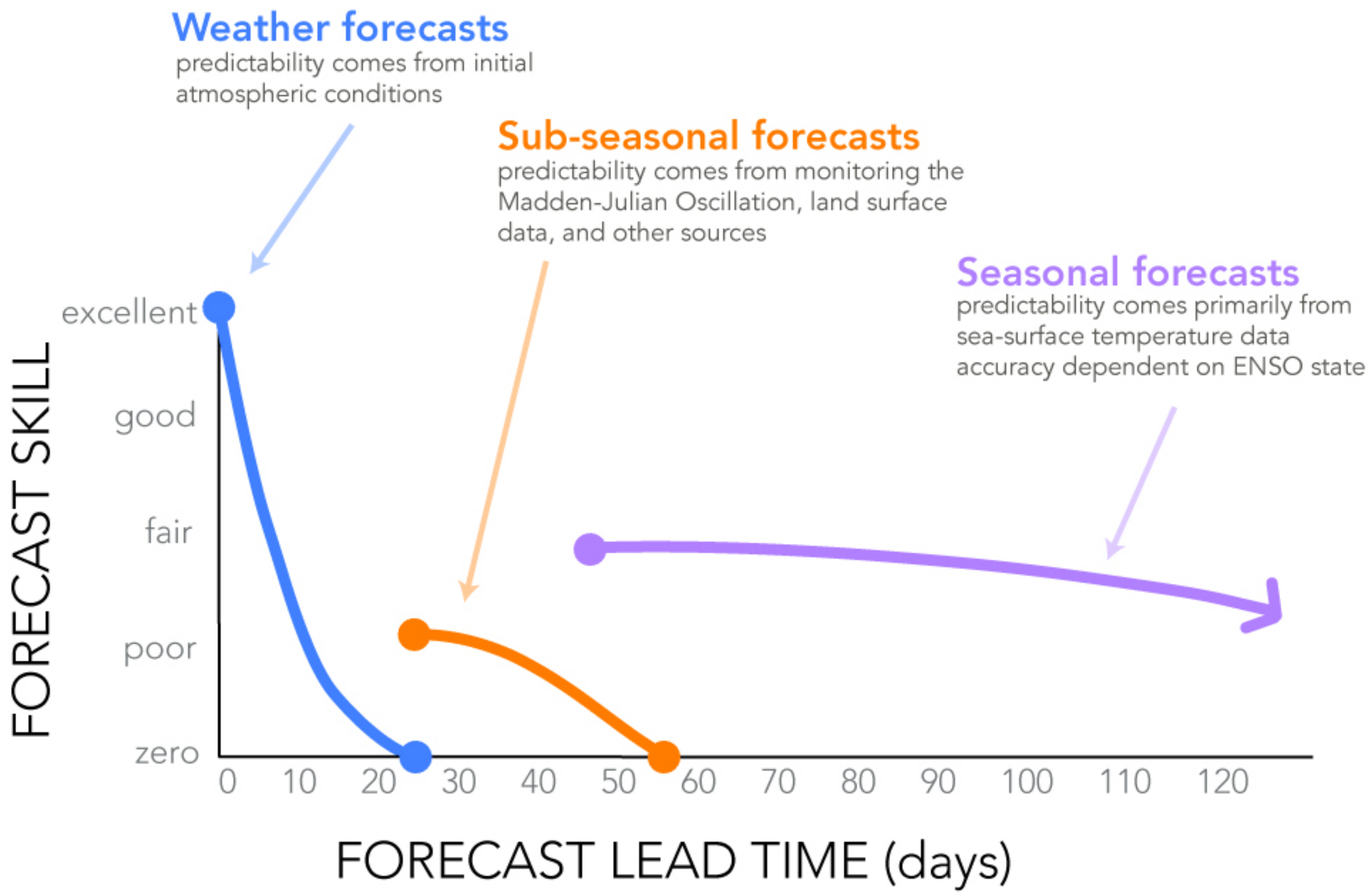
- Initialized with current weather conditions inferred from measurements
- Simulate future weather / climate by discretizing partial differential equations using supercomputers
- Accuracy limited by chaotic nature: initial error doubles every 5 days
- Ensembles with varying initial conditions / model 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 **numerical weather prediction** and **global climate 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

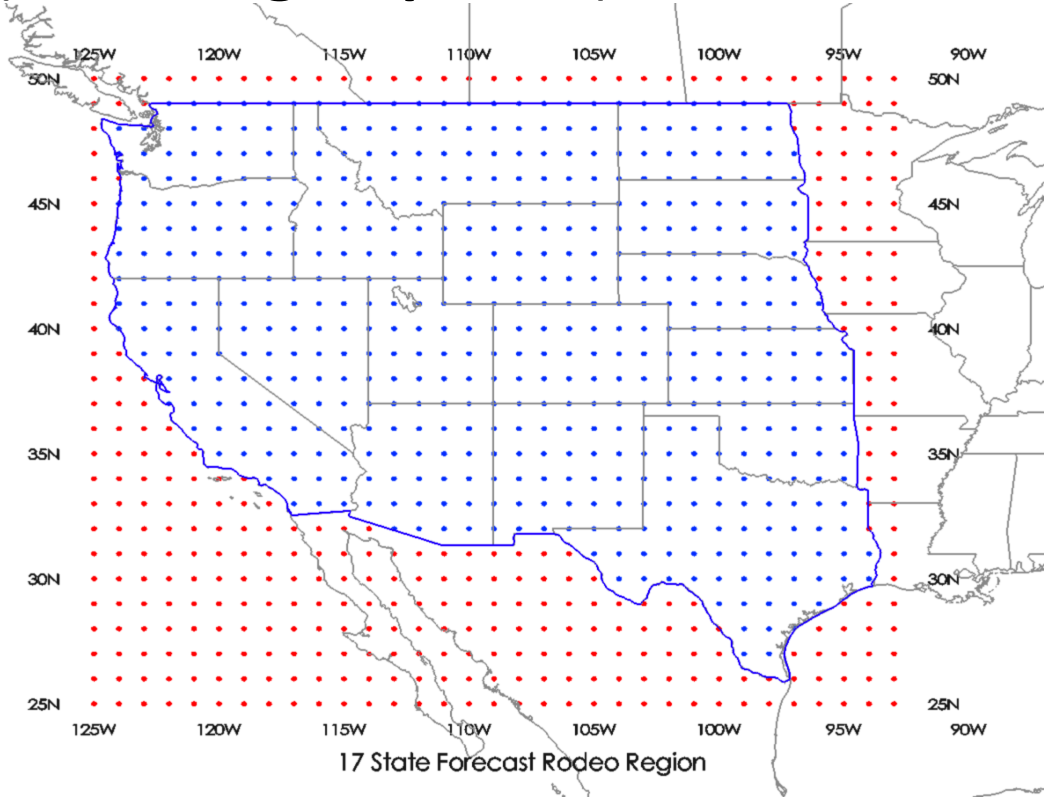
The Subseasonal Climate Forecast Rodeo

- A year long, real-time subseasonal forecasting competition
- Designed to
  - Advance science
  - Raise awareness
  - Provide an evaluation platform

Credit: David Raff, USBR

Subseasonal Rodeo Forecasts

- Four separate forecasting tasks
  - Two variables: **average temperature** (degrees C) and **total precipitation** (mm)
  - Two outlooks: **weeks 3-4** and **weeks 5-6** (forecast is over a 2-week period)
- Issued on a $1° \times 1°$ latitude-longitude grid (**G = 514 grid points**)
- Issued **every two weeks**
  - Apr 18, 2017 -- May 3, 2018, midnight GMT
  - Disqualified if two submissions missed
    - One submission was on Christmas day EST
- Uploaded to server in NetCDF format
  - Popular format for scientific array data



Acknowledgment: We would not have survived this competition without tools like CDO, wgrib2, and NCO for processing the NetCDF, GRIB2, and custom byte stream (?!) formats common in meteorological data

Subseasonal Rodeo Evaluation

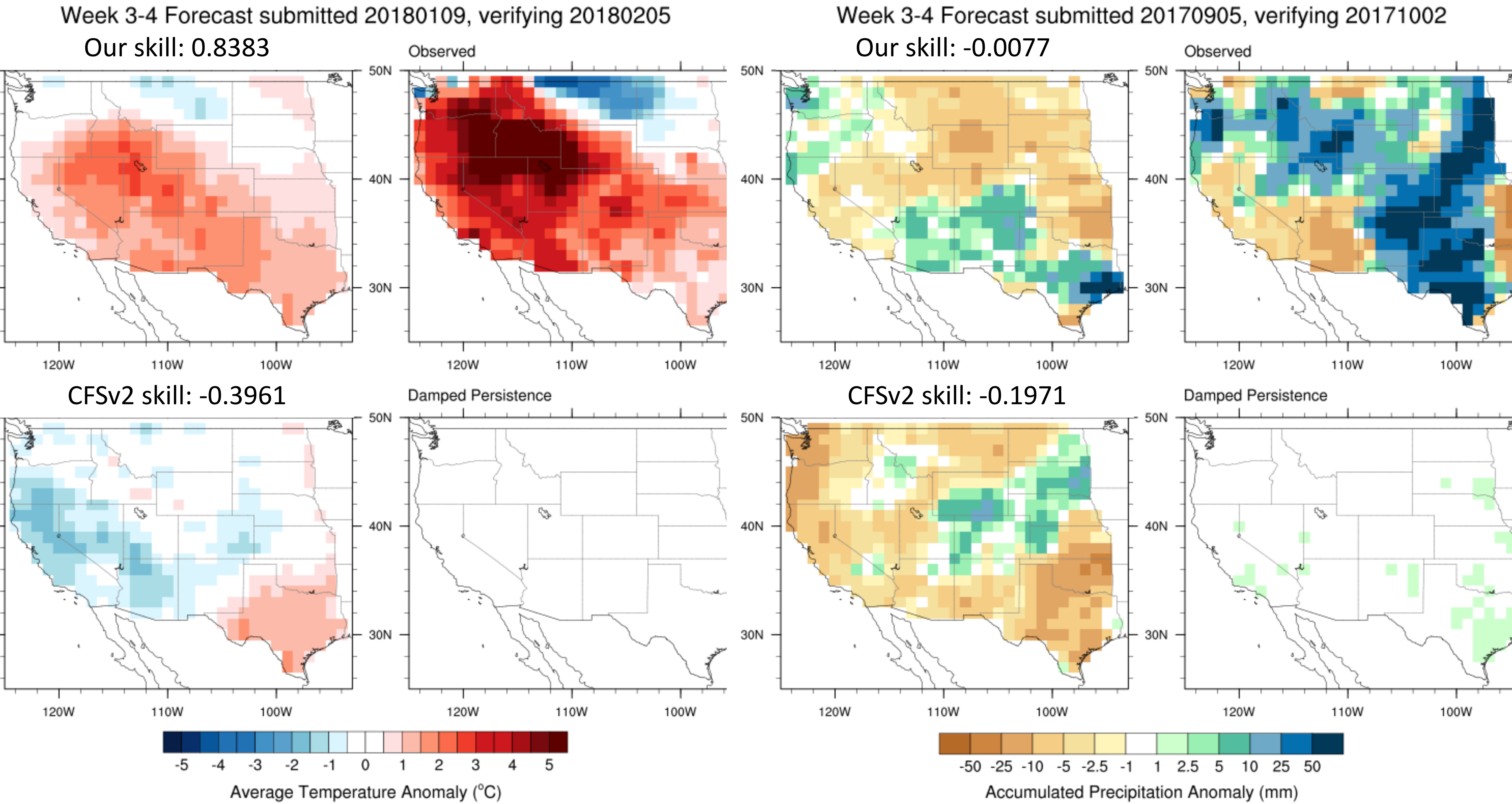
- For each 2-week period starting on date t, define
  - monthday(t), the month-day combination associated with t (e.g., January 1)
  - **observed outcomes** $\boldsymbol{y}_t \in \mathbb{R}^G$ for each grid point (temperature or precipitation)
  - **observed anomalies** $\boldsymbol{a}_t = \boldsymbol{y}_t - \boldsymbol{c}_{\text{monthday}(t)}$ where
  - **climatology** $\mathbf{c}_d \triangleq \frac{1}{30} \sum_{\substack{t\,:\,\texttt{monthday}(t)=d,\\ 1981 \leq \texttt{year}(t) \leq 2010}} \mathbf{y}_t$
    - Average outcome for month-day combo d over the climatology period, 1981-2010
- Forecasts judged on **skill** (cosine similarity) between observed anomalies and **forecast anomalies** $\widehat{\boldsymbol{a}}_t = \widehat{\boldsymbol{y}}_t - \boldsymbol{c}_{\text{monthday}(t)}$:

$$\text{skill}(\hat{\mathbf{a}}_t, \mathbf{a}_t) \triangleq \cos(\hat{\mathbf{a}}_t, \mathbf{a}_t) = \frac{\langle \hat{\mathbf{a}}_t, \mathbf{a}_t \rangle}{\|\hat{\mathbf{a}}_t\|_2 \|\mathbf{a}_t\|_2} \in [-1, 1]$$

  - Unusual objective function for machine learning
  - **Multitask** objective function: couples together the G per-grid point forecasting tasks

Subseasonal Rodeo Benchmarks

- Contestants had to outperform two benchmarks to qualify for prizes
- **Debiased Climate Forecasting System v2 (CFSv2)**
  - Operational physics-based system developed under the guidance of the U.S. National Centers for Environmental Prediction (NCEP)
  - To form debiased CFSv2 benchmark, CFSv2 forecasts for date t were
    - **Ensembled** by averaging 32 forecasts (4 model initializations and 8 lead times)
    - **Debiased** by adding the mean observed outcome for monthday(t) over 1999-2010 and subtracting the mean CFSv2 reforecast (8 lead times, 1 model initialization)
- **Damped persistence**
  - Statistical forecasting model (no exact description provided)
  - "Seasonally developed regression coefficients based on the historical climatology period of 1981-2010 that relate observations of the past two weeks to the forecast outlook periods on a grid cell by grid cell basis"


Week 3-4 Forecast submitted 20180109, verifying 20180205
Our skill: 0.8383
Observed
CFSv2 skill: -0.3961
Damped Persistence
50N
40N
30N
120W
110W
100W
-5 -4 -3 -2 -1 0 1 2 3 4 5
Average Temperature Anomaly (°C)
Week 3-4 Forecast submitted 20170905, verifying 20171002
Our skill: -0.0077
Observed
CFSv2 skill: -0.1971
Damped Persistence
-50 -25 -10 -5 -2.5 -1 1 2.5 5 10 25 50
Accumulated Precipitation Anomaly (mm)

Subseasonal Rodeo Data

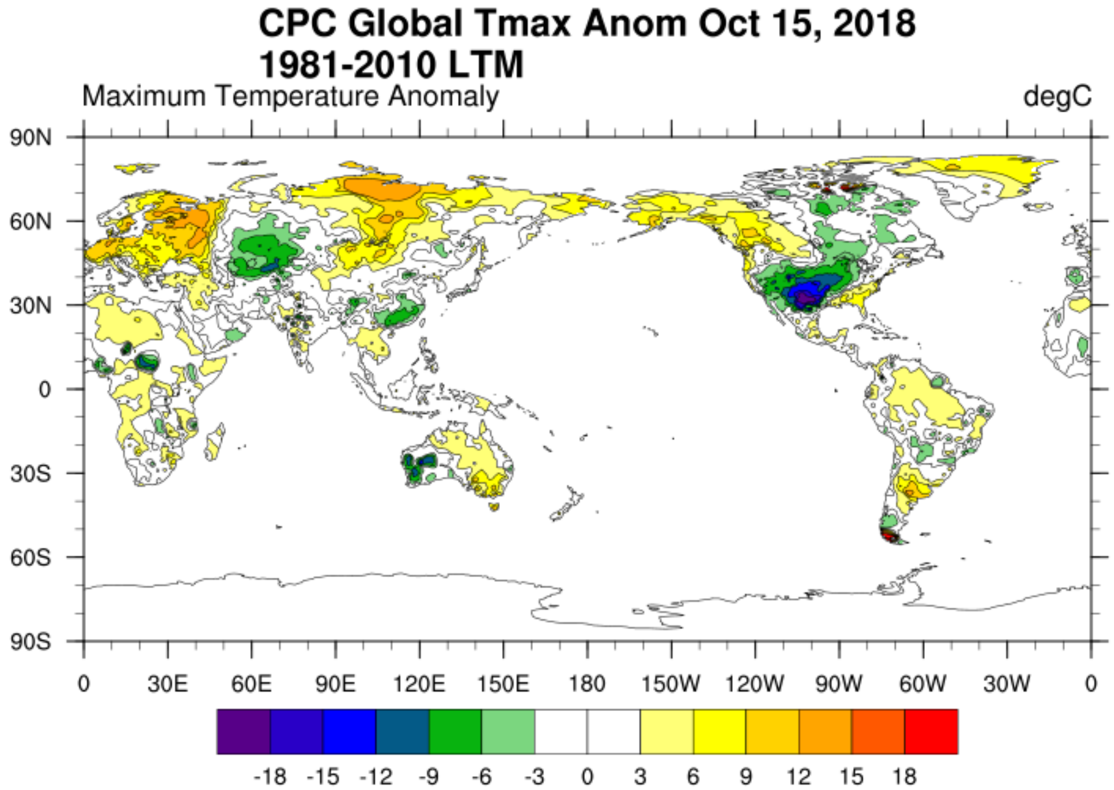
- **No data was provided!**
  - Organizers did identify which ground-truth temperature and precipitation data sources would be used for evaluation
- Contestants encouraged to use whatever data they wanted
  - If using standard physics-based model forecasts as inputs, had to demonstrate significant improvement over those models

Our SubseasonalRodeo Dataset

- To train and evaluate our predictive models, we constructed a **SubseasonalRodeo dataset** from diverse data sources with varying file formats, spatial layouts, and measurement frequencies
- Organized as a collection of Python Pandas objects in HDF5 format
  - Spatial variables (vary with the target grid point but not the target date)
  - Temporal variables (vary with the target date but not the target grid point)
  - Spatiotemporal variables (vary with both the target grid point and the target date)
- Gridded data interpolated to $1°\times 1°$ grid (using distance-weighted average interpolation) and restricted to contest grid points
- Daily measurements replaced with averages (or, for precipitation, sums) over ensuing 2-week period
- Released via the Harvard Dataverse https://doi.org/10.7910/DVN/IHBANG

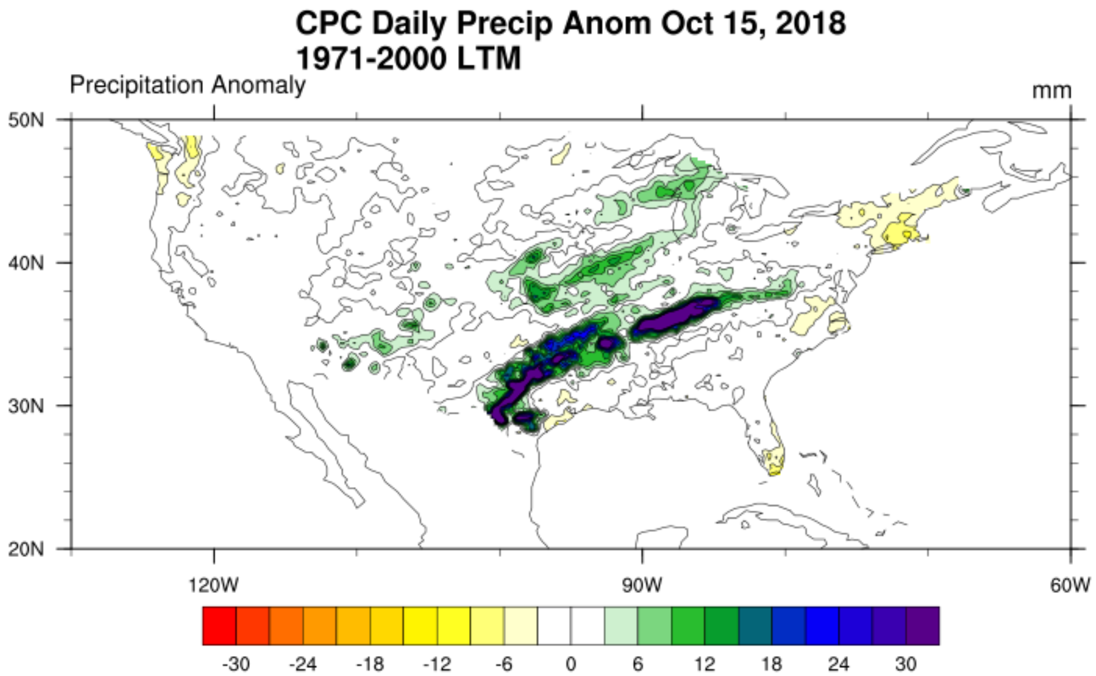
Our SubseasonalRodeo Dataset

- **Temperature**
  - Source: NOAA's Climate Prediction Center (CPC) Global Gridded Temperature dataset
  - Daily max and min temperature at 2 meters (tmax and tmin) from 1979 onwards
  - Official contest target temperature variable: $\mathrm{tmp2m} = \frac{\mathrm{tmax}+\mathrm{tmin}}{2}$



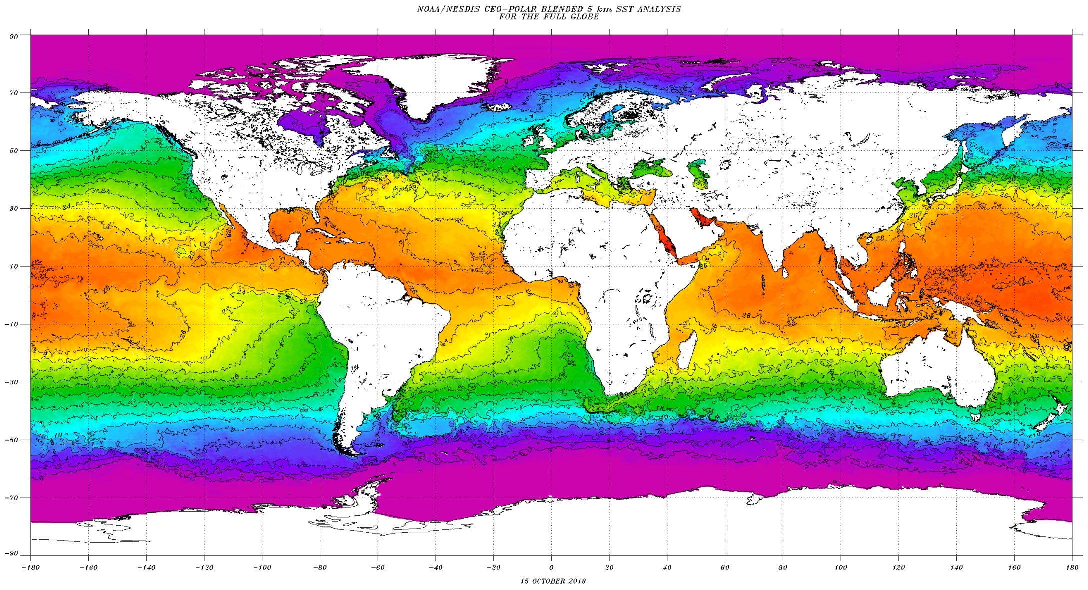
Our SubseasonalRodeo Dataset

- **Precipitation**
  - Source: NOAA's CPC Gauge-Based Analysis of Global Daily Precipitation (Xie, Chen, and Shi 2010)
  - Daily precipitation (precip) data from 1979 onward
  - Augmented with daily U.S. precipitation data from 1948-1979 from the CPC Unified Gauge-Based Analysis of Daily Precipitation over CONUS



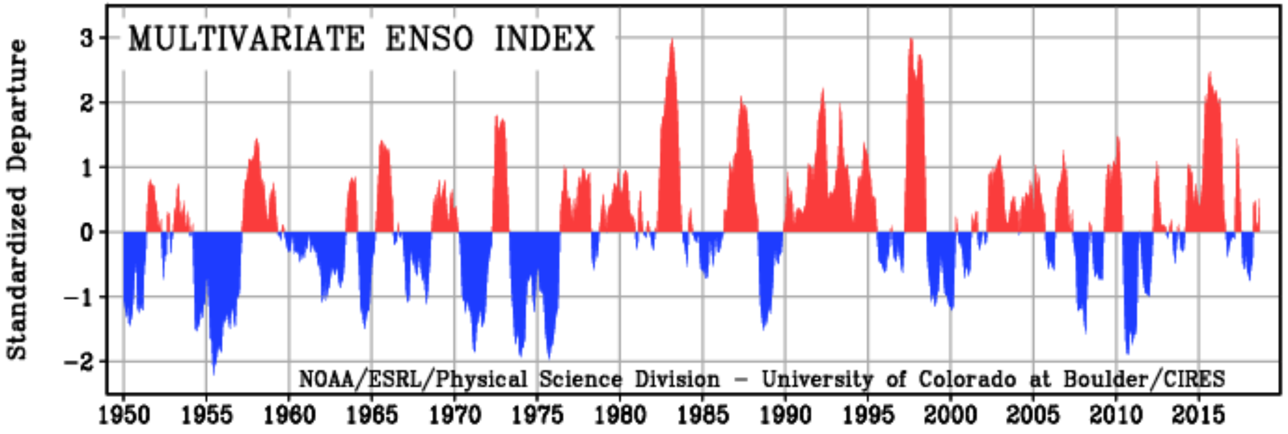
Our SubseasonalRodeo Dataset

- **Sea surface temperature and sea ice concentration**
  - Source: NOAA’s Optimum Interpolation Sea Surface Temperature (SST) dataset (Reynolds et al. 2007)
  - Daily SST and sea ice concentration data, from 1981 to the present.
  - After interpolation, we extracted the top three principal components (PCs) across grid points in the Pacific basin region (20S to 65N, 150E to 90W), $(\mathrm{sst}_i)_{i=1}^{3}$ and $(\mathrm{icec}_\mathrm{i})_{i=1}^{3}$



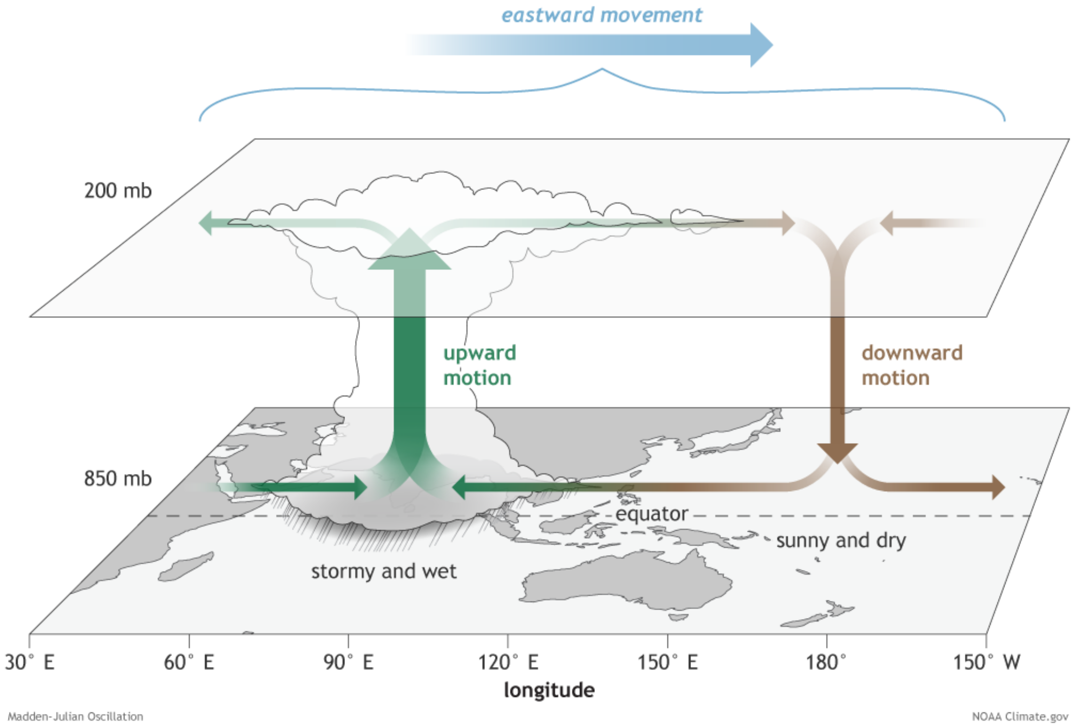
Our SubseasonalRodeo Dataset

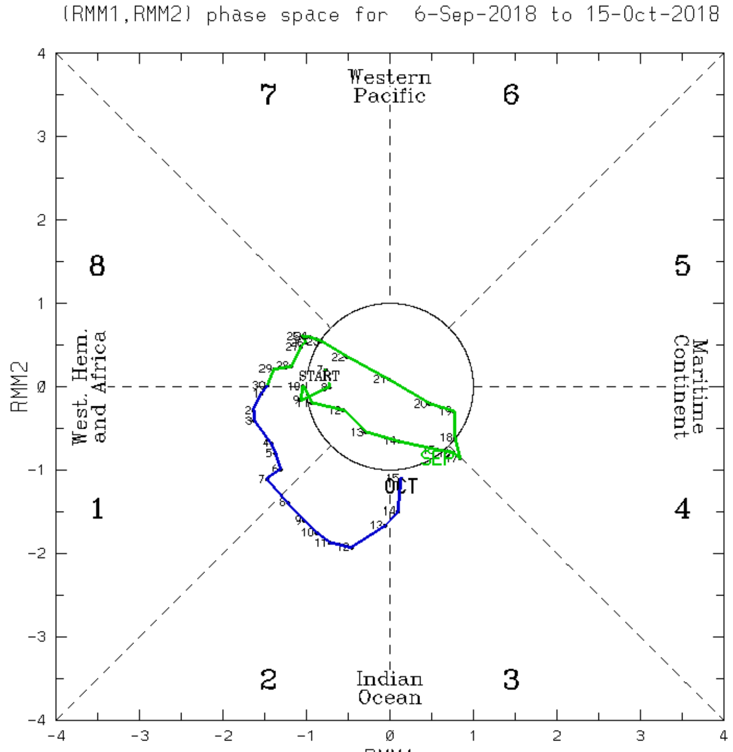
- **Multivariate ENSO index (MEI)**
  - Source: NOAA/Earth System Research Laboratory (Wolter 1993; Wolter and Timlin 1998)
  - Bimonthly MEI values (mei) from 1949 to the present
  - El Niño/Southern Oscillation (ENSO) is an irregularly periodic variation in winds and SSTs over the tropical eastern Pacific Ocean that affecting global climate variability on interannual timescales
  - MEI is a scalar summary of six variables (sea-level pressure, zonal and meridional surface wind components, SST, surface air temperature, and sky cloudiness) associated with ENSO



Our SubseasonalRodeo Dataset

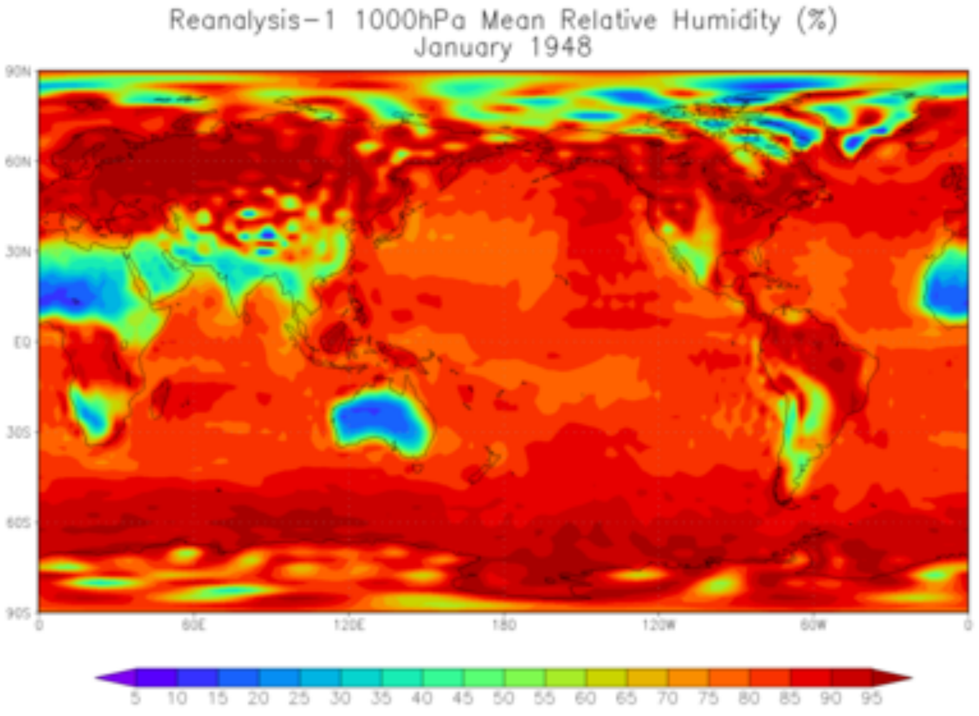
- **Madden-Julian oscillation (MJO)**
  - Source: Australian Government Bureau of Meteorology (Wheeler and Hendon 2004)
  - Daily MJO amplitude and phase values since 1974 (we do not aggregate)
  - MJO is a metric of tropical convection on daily to weekly timescales and can have significant impact on the western United States' subseasonal climate.

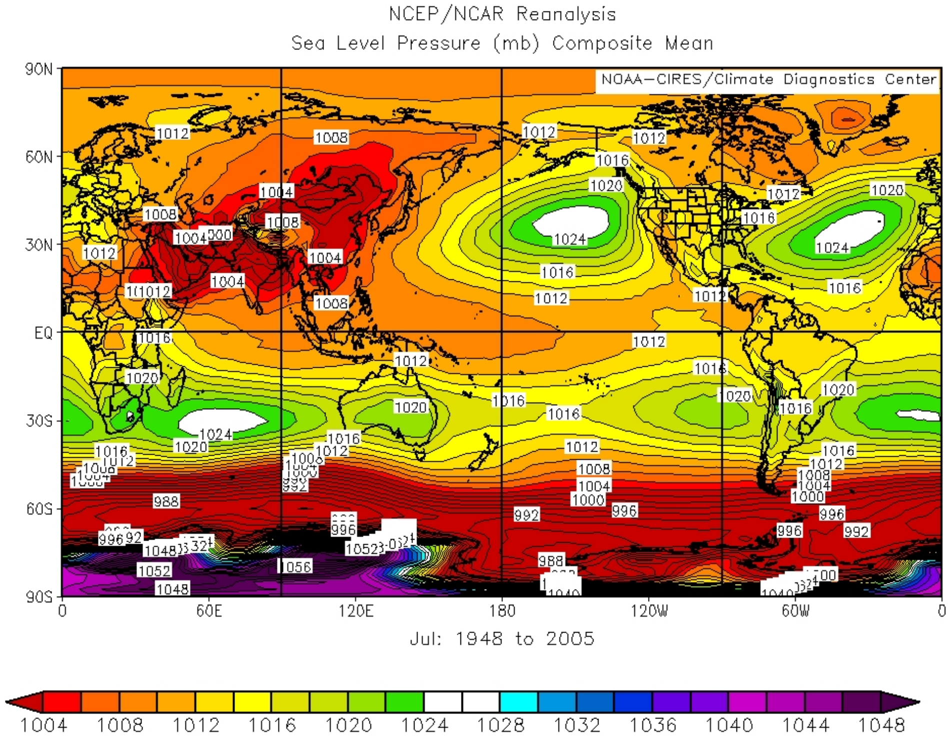




Our SubseasonalRodeo Dataset

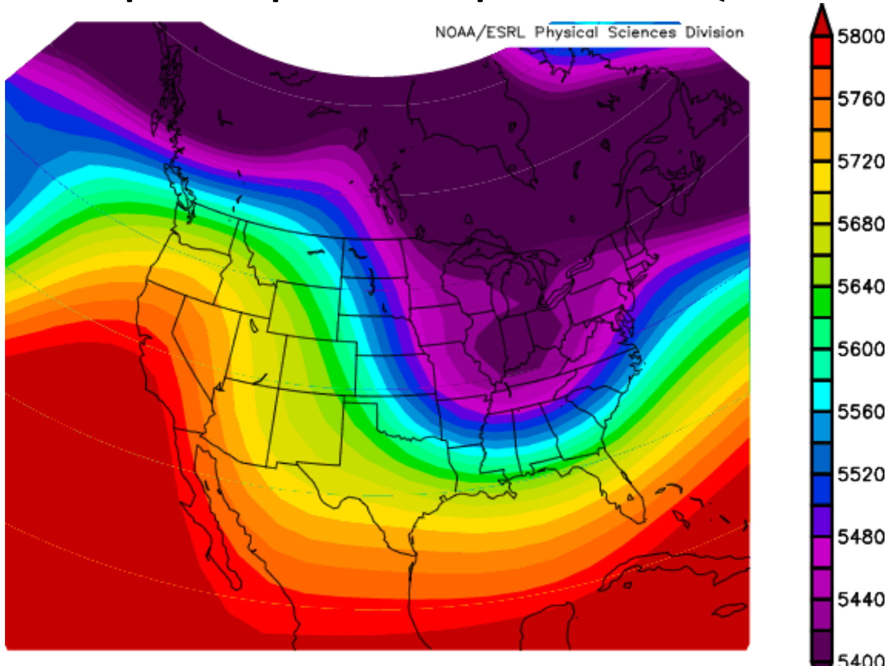
- **Relative humidity and pressure**
  - Source: NOAA's National Center for Environmental Prediction (NCEP)/National Center for Atmospheric Research Reanalysis dataset (Kalnay et al. 1996)
  - Daily relative humidity (rhum) near the surface from 1948 to the present
  - Daily pressure at the surface (pres) from 1979 to the present.



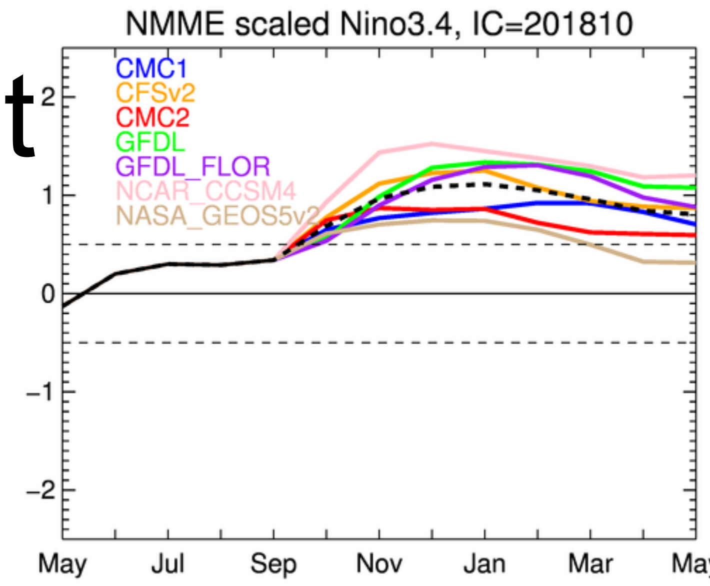


Our SubseasonalRodeo Dataset

- **Geopotential height**
  - Source: NCEP Reanalysis dataset (Kalnay et al. 1996)
  - Daily mean height at which 10mb of pressure occurs since 1948
  - Captures variability in the Arctic polar vortex, a large-scale low-pressure area lying near the North Pole
  - Extracted the top three principal components $(\text{wind\_hgt\_10}_{\text{i}})_{i=1}^{3}$



Our SubseasonalRodeo Dataset



- **North American Multi-Model Ensemble (NMME)**
  - Source: IRI/LDEO Climate Data Library (Kirtman et al. 2014)
  - Monthly forecasts of physics-based models from North America modeling centers
    - Cansips, CanCM3, CanCM4, CCSM3, CCSM4, GFDL-CM2.1-aer04, GFDL-CM2.5 FLOR-A06 and FLOR-B01, NASA-GMAO062012, and NCEP-CFSv2.
  - Each forecast contains monthly mean predictions from 0.5 to 8.5 months ahead.
  - Derived 2-week forecasts from a weighted average of the monthly predictions with weights proportional to the number of target period days that fell into each month.
    - Formed an equally-weighted average (nmme_wo_ccsm3_nasa) of all models save CCSM3 and NASA (which were not reliably updated during the contest).
    - Also created by averaging the most recent monthly forecast of each model save CCSM3 and NASA (nmme0_wo_ccsm3_nasa).

Our Forecasting Models

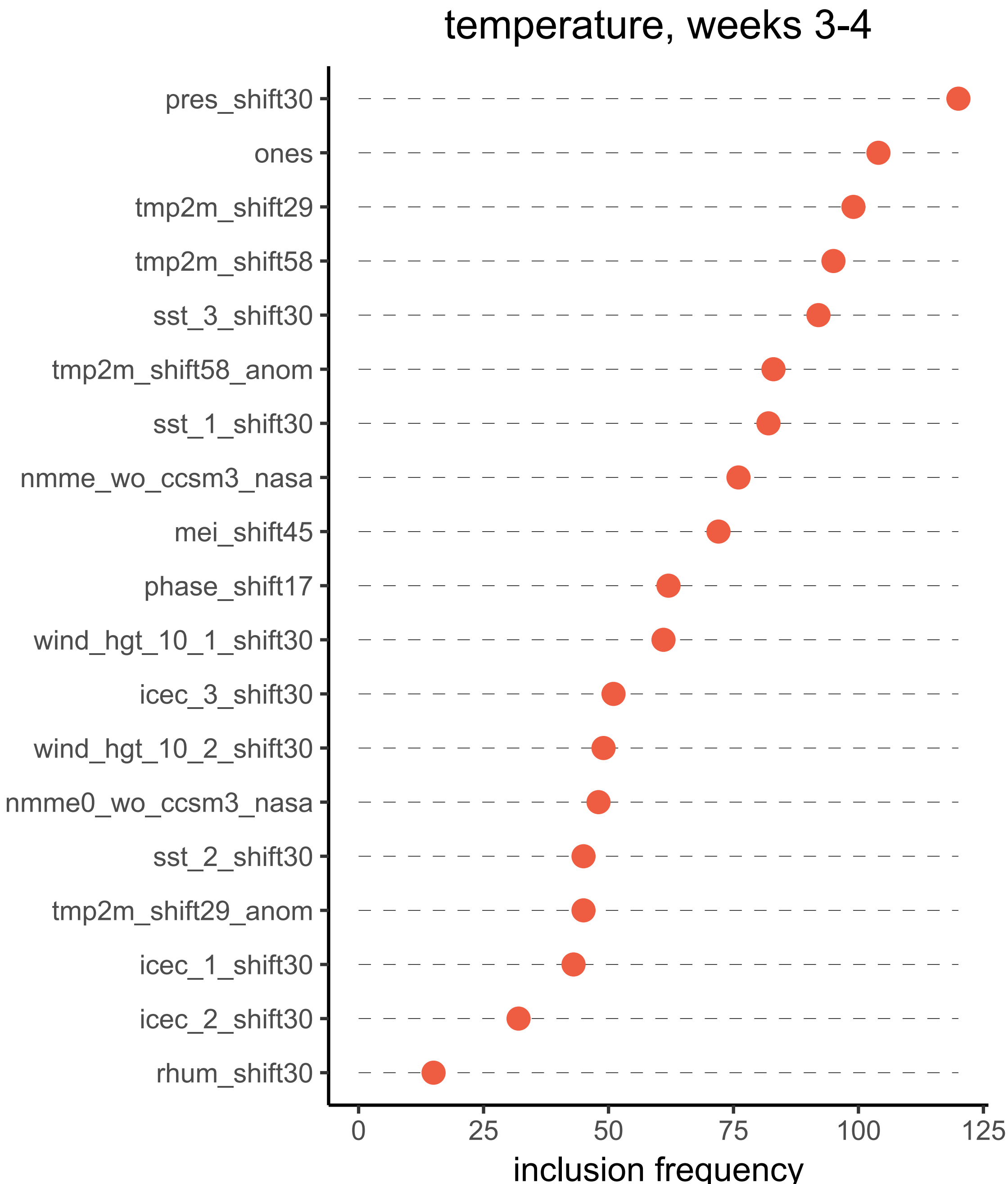
- **MultiLLR: Local Linear Regression with Multitask Feature Selection**
  - Incorporates lagged measurements from all data sources
  - Prunes irrelevant regressors using skill-based multitask feature selection
- **AutoKNN: Multitask Nearest Neighbor Autoregression**
  - Identifies dates most similar to target using skill-based similarity measure
  - Regresses onto observed temperature or precipitation of similar dates and fixed lags
- **Ensemble:** Averages the normalized predicted anomalies

$$\hat{\mathbf{a}}_{\mathrm{ensemble}} \triangleq \frac{1}{2}\frac{\hat{\mathbf{a}}_{\mathrm{multillr}}}{\left\|\hat{\mathbf{a}}_{\mathrm{multillr}}\right\|_2} + \frac{1}{2}\frac{\hat{\mathbf{a}}_{\mathrm{autoknn}}}{\left\|\hat{\mathbf{a}}_{\mathrm{autoknn}}\right\|_2}.$$

  - **Proposition** *If the average of the individual model skills is positive, then the ensemble skill is strictly greater than the average of the individual skills.*

Local Regression with Multitask Feature Selection (MultiLLR)

1. **Extract lagged measurements** of SubseasonalRodeo variables as regression features $\boldsymbol{x}_{t,g}$
   - Lag based on measurement availability
   - "shift$l$" indicates that measurements were from l days prior
   - "anom" indicates that anomalies are used instead of raw values
   - "ones" is a constant feature always = 1 (used in lieu of an intercept)



Local Regression with Multitask Feature Selection (MultiLLR)

2. Combine features using **local linear regression**
   - Locality determined by the day of the year
   - Uniform weights ($w_{t,g} = 1$), no offsets ($b_{t,g} = 0$)

Algorithm 1 Weighted Local Linear Regression (LLR)

input test day of year d^*; span s; training outcomes, features, offsets, weights $(y_{t,g}, \mathbf{x}_{t,g}, b_{t,g}, w_{t,g})_{t \in \mathcal{T}, g \in \{1,\dots,G\}}$

$\mathcal{D} \triangleq \{t \in \mathcal{T} : \frac{365}{2} - ||\mathsf{doy}(t) - d^*| - \frac{365}{2}| \leq s\}$

for grid points $g = 1$ **to** G **do**

$\hat{\boldsymbol{\beta}}_g \in \operatorname{argmin}_{\boldsymbol{\beta}} \sum_{t \in \mathcal{D}} w_{t,g}(y_{t,g} - b_{t,g} - \boldsymbol{\beta}^\top \mathbf{x}_{t,g})^2$

output coefficients $(\hat{\boldsymbol{\beta}}_g)_{g=1}^G$

Local Regression with Multitask Feature Selection (MultiLLR)

3. **Feature selection:** select subset of relevant features for each target date
    - Motivation: Not all features are relevant at all times of year
    - Use a customized backward stepwise procedure to prune features
        - Start with all features included in the model
        - Until termination
            - Fit LLR model with each remaining candidate feature removed and evaluate predictive performance
            - If, in each case, performance is reduced substantially, keep all remaining features and terminate
            - Otherwise, remove the feature that reduces skill the least
    - Selection is multitask: variables selected jointly for all grid points, while coefficients are fit separately for each grid point
    - Performance measured via leave-one-year-out cross-validated average skill
        - For each year in the training set, we find the date t in that year with the same day-of-year as the target date, withhold one year of data surrounding t from the training set, and measure the skill of the trained model at predicting the withheld anomalies $\boldsymbol{a}_t$
        - The average of these skills across all years is our predictive performance measure

Multitask k–Nearest Neighbor Autoregression (AutoKNN)

1. **Extract features from historical measurements of the outcome variable**
   - Constant "ones" feature and 3 lagged anomaly measurements
     - Lags = 29, 58, and 365 days for weeks 3-4 or 43, 56, and 365 days for weeks 5-6
   - Observed anomaly patterns of the outcome variable on similar dates in the past
     - Similarity measure based on skill objective and measured jointly for all grid points
     - Compute mean skill over a history of H = 60 days, starting 1 year prior to target date ($\ell = 365$)
     - Extract anomalies of k = 20 most similar dates for temperature and k = 1 for precipitation

---

Algorithm 3 Multitask k-Nearest Neighbor Similarities

---

input test date t^*; training anomalies $(\mathbf{a}_t)_t$; lag ℓ; history H

  for all training dates t **do**

    $\mathtt{sim}_t = \frac{1}{H} \sum_{h=0}^{H-1} \text{skill}(\mathbf{a}_{t-\ell-h}, \mathbf{a}_{t^*-\ell-h})$

output similarities $(\mathtt{sim}_t)_t$

Multitask k–Nearest Neighbor Autoregression (AutoKNN)

2. Combine features using **weighted local linear regression**
   - Locality determined by the day of the year
   - Climatology offsets $b_{t,g} = c_{\text{monthday}(t),g}$, so target variable is anomaly $a_{t,g}$ rather than raw measurement $y_{t,g}$
   - Weights $w_{t,g} = 1/\sum_g \left(a_{t,g} - \frac{1}{G}\sum_f a_{t,f}\right)^2$ to mimic cosine similarity objective

---

Algorithm 1 Weighted Local Linear Regression (LLR)

---

input test day of year d^*; span s; training outcomes, features, offsets, weights $(y_{t,g}, \mathbf{x}_{t,g}, b_{t,g}, w_{t,g})_{t\in\mathcal{T}, g\in\{1,\ldots,G\}}$

$\mathcal{D} \triangleq \{t \in \mathcal{T} : \frac{365}{2} - ||\mathtt{doy}(t) - d^*| - \frac{365}{2}| \leq s\}$

for grid points $g = 1$ **to** G **do**

$\hat{\boldsymbol{\beta}}_g \in \operatorname{argmin}_{\boldsymbol{\beta}} \sum_{t\in\mathcal{D}} w_{t,g}(y_{t,g} - b_{t,g} - \boldsymbol{\beta}^\top \mathbf{x}_{t,g})^2$

output coefficients $(\hat{\boldsymbol{\beta}}_g)_{g=1}^{G}$

Ensembling

- Our final forecast is an ensemble of the MultiLLR and AutoKNN predictions
  - We average of the normalized predicted anomalies of the two models:

$$\hat{\mathbf{a}}_{\text{ensemble}} \triangleq \frac{1}{2}\frac{\hat{\mathbf{a}}_{\text{multillr}}}{\|\hat{\mathbf{a}}_{\text{multillr}}\|_2} + \frac{1}{2}\frac{\hat{\mathbf{a}}_{\text{autoknn}}}{\|\hat{\mathbf{a}}_{\text{autoknn}}\|_2}.$$

- **Proposition** *If the average of the individual model skills is positive, then the ensemble skill is strictly greater than the average of the individual skills.*

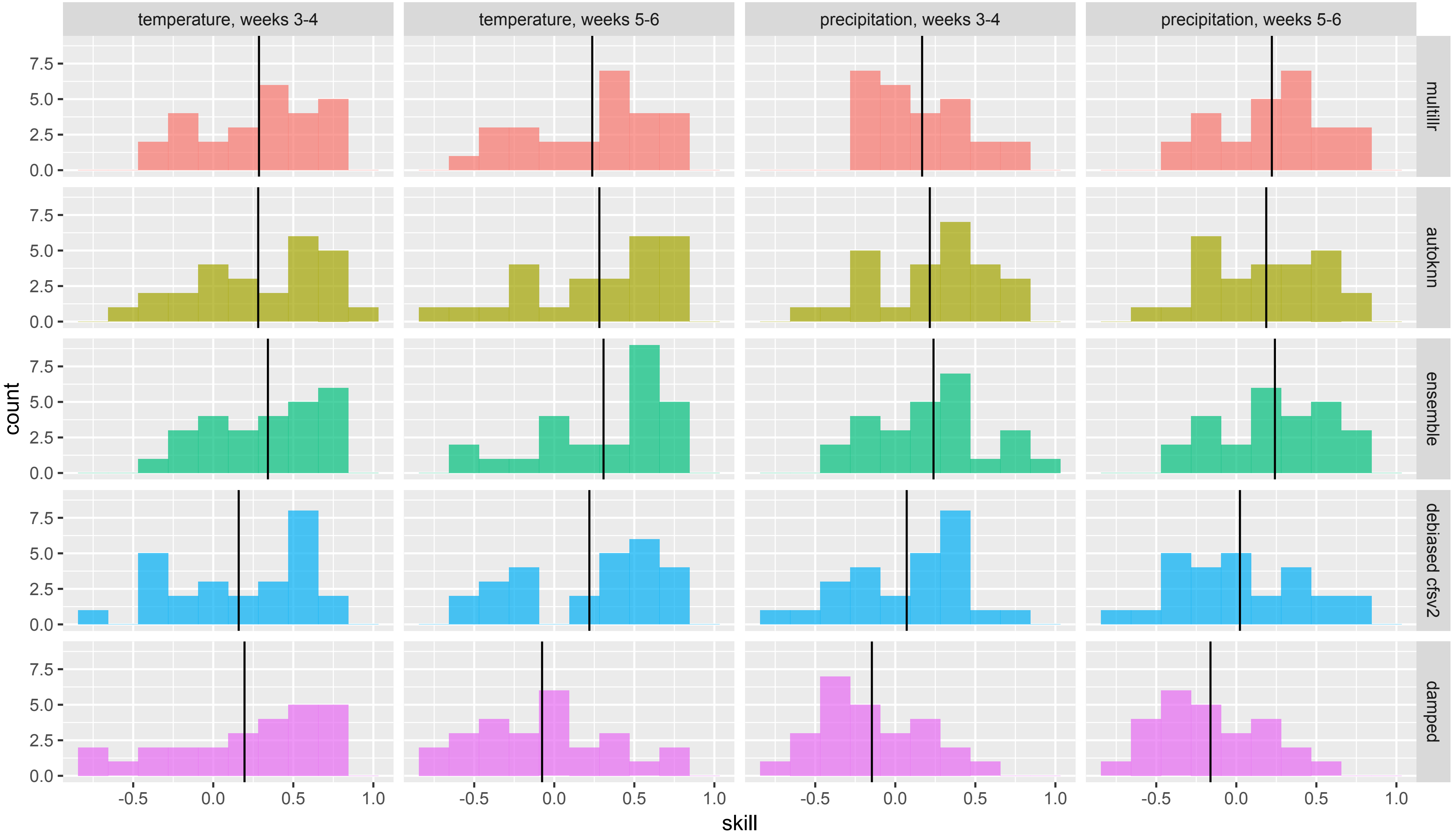
Contest Period Evaluation

Table 1 Average contest-period skill of MultiLLR, AutoKNN, the proposed ensemble of MultiLLR and AutoKNN, the official contest debiased-CFSv2 baseline, the official contest damped-persistence baseline, and the top-performing competitor in the Forecast Rodeo contest.

task	multillr	autoknn	ensemble	debiased cfsv2	damped	top competitor
temperature, weeks 3-4	0.2856	0.2807	**0.3414**	0.1589	0.1952	0.2855
temperature, weeks 5-6	0.2371	0.2817	**0.3077**	0.2192	-0.0762	0.2357
precipitation, weeks 3-4	0.1675	0.2156	**0.2388**	0.0713	-0.1463	0.2144
precipitation, weeks 5-6	0.2219	0.1870	**0.2412**	0.0227	-0.1613	0.2162

- All three proposed methods outperform both contest baselines in all four tasks
- The ensemble outperforms the top Rodeo competitor in all four tasks
- Note: the competitor skills represent the real-time evaluations of forecasting systems that may have evolved over the course of the competition

Contest Period Evaluation



Distribution of contest-period skills: baselines tend to have more extreme negative skills

Historical Forecast Evaluation

- We next evaluate the performance of our methods over 2011 – 2017 (all years following the climatology period)
- We reconstruct the debiased CFSv2 baseline (**rec-deb-cfs**) following contest guidelines and using the CFSv2 Operational Forecast dataset
  - The forecast is an ensemble of 8 lead times but only 1 model initialization (the other model initialization forecasts were released in real time but deleted after 1 week)
- We also evaluate a three-way ensemble of MultiLLR, AutoKNN, and rec-deb-cfs (**ens-cfs**):

$$\hat{\mathbf{a}}_{\text{ens-cfs}} \triangleq \frac{1}{3}\frac{\hat{\mathbf{a}}_{\text{multillr}}}{\|\hat{\mathbf{a}}_{\text{multillr}}\|_2} + \frac{1}{3}\frac{\hat{\mathbf{a}}_{\text{autoknn}}}{\|\hat{\mathbf{a}}_{\text{autoknn}}\|_2} + \frac{1}{3}\frac{\hat{\mathbf{a}}_{\text{rec-deb-cfs}}}{\|\hat{\mathbf{a}}_{\text{rec-deb-cfs}}\|_2}$$

Historical Forecast Evaluation

	temperature, weeks 3-4					temperature, weeks 5-6				
year	multillr	autoknn	ensemble	rec-deb-cfs	ens-cfs	multillr	autoknn	ensemble	rec-deb-cfs	ens-cfs
2011	0.2479	0.3664	0.3433	**0.4598**	0.4563	0.2685	0.3240	0.3646	0.3879	**0.4405**
2012	0.0879	**0.3135**	0.2173	0.1397	0.2181	0.2765	0.3205	**0.3529**	0.1030	0.3316
2013	0.0944	0.2011	0.1688	**0.2861**	0.2711	**0.2397**	0.0531	0.1895	0.1211	0.1858
2014	0.1682	0.2775	0.2803	0.3018	**0.3591**	0.1448	0.3056	0.2596	0.1936	**0.3311**
2015	0.3673	0.3885	0.4339	0.2857	**0.4383**	0.1487	0.3939	0.2970	0.4234	**0.4311**
2016	0.3098	0.3502	0.3663	0.2490	**0.3887**	0.2277	0.2882	**0.3023**	0.0983	0.2799
2017	0.2856	0.2807	**0.3414**	0.0676	0.3239	0.2371	0.2817	**0.3077**	0.1708	0.2993
all	0.2230	0.3111	0.3073	0.2557	**0.3508**	0.2204	0.2810	0.2962	0.2142	**0.3279**

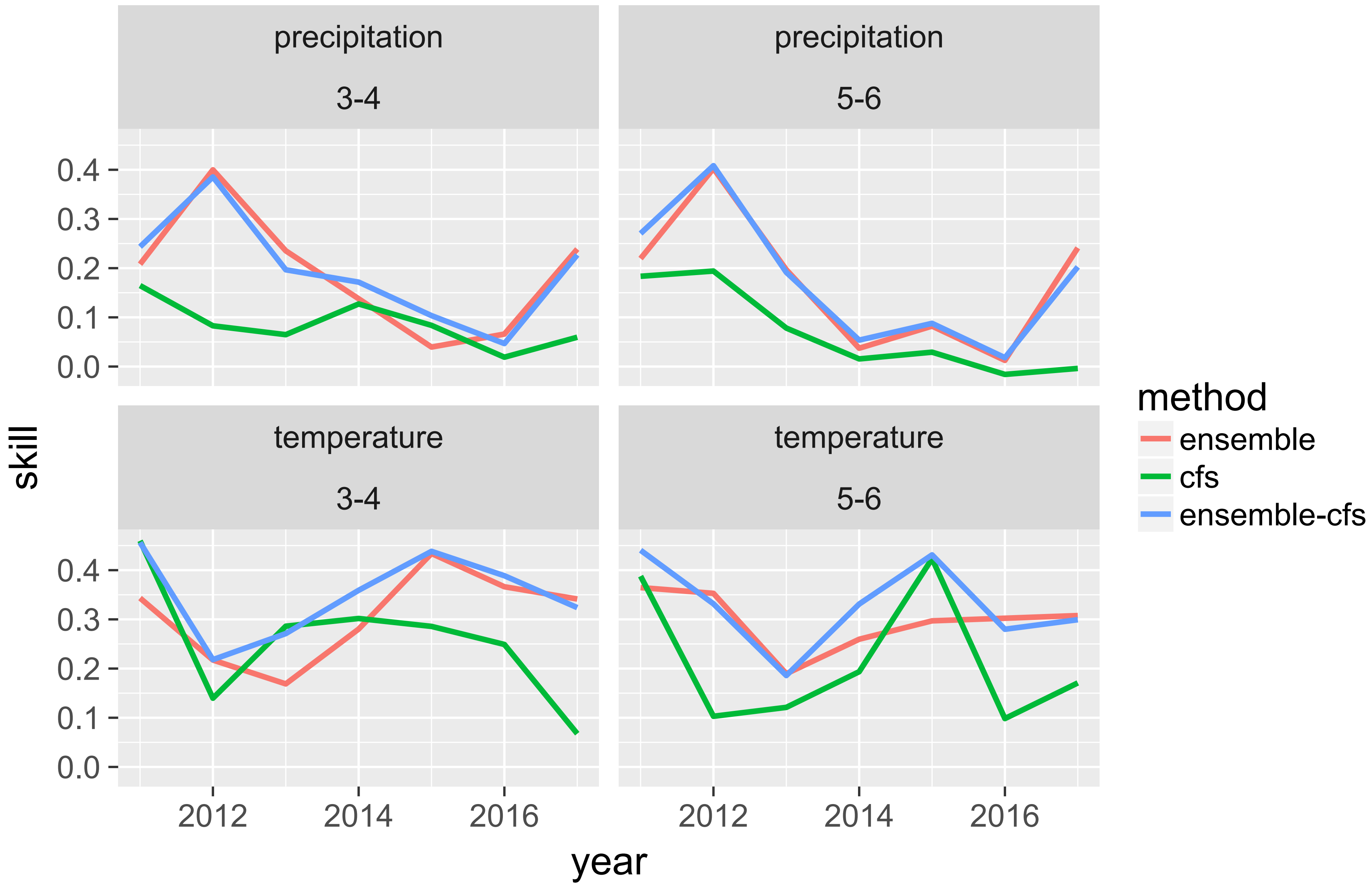
- AutoKNN and ensemble have higher mean skill than debiased CFSv2 on both tasks; MultiLLR has higher mean skill on weeks 5-6
- Ensemble improves over debiased CFSv2 by 20% for weeks 3-4 and 38% for weeks 5-6
- Ens-cfs improves over debiased CFSv2 by 37% for weeks 3-4 and 53% for weeks 5-6

Historical Forecast Evaluation

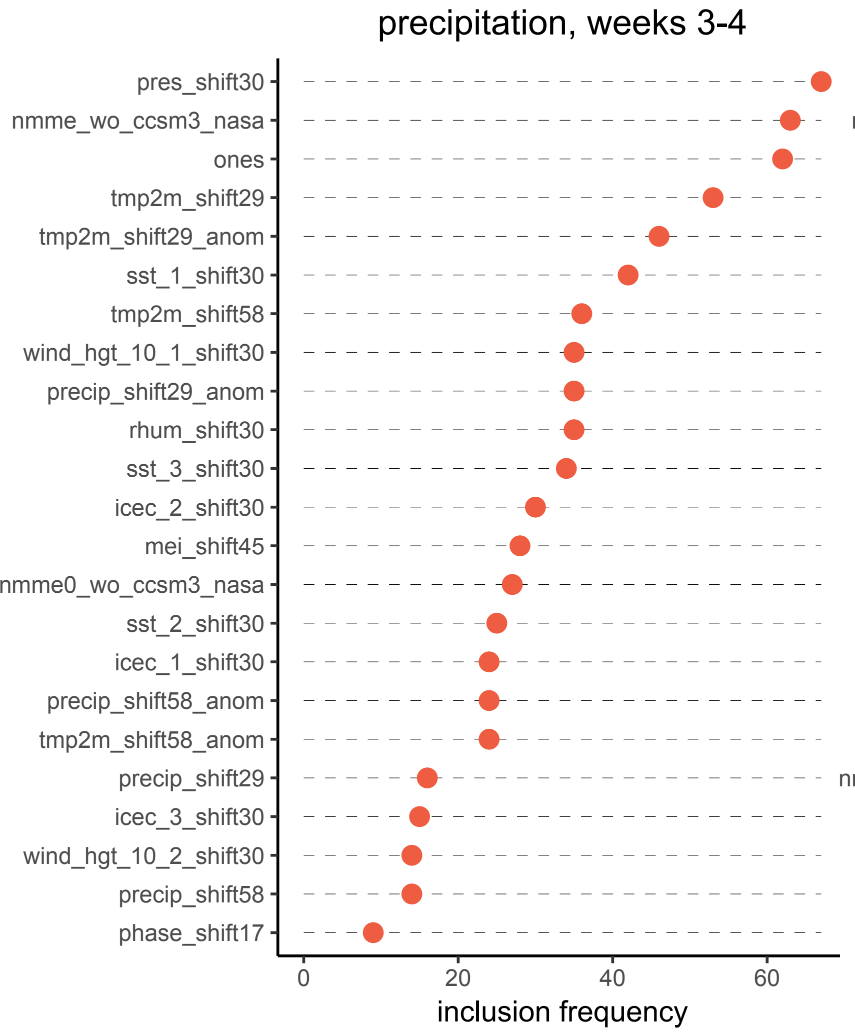
	precipitation, weeks 3-4					precipitation, weeks 5-6				
year	multillr	autoknn	ensemble	rec-deb-cfs	ens-cfs	multillr	autoknn	ensemble	rec-deb-cfs	ens-cfs
2011	0.1332	0.2173	0.2081	0.1646	**0.2435**	0.1371	0.2132	0.2195	0.1835	**0.2704**
2012	0.3219	0.3648	**0.3999**	0.0828	0.3854	0.2879	0.3943	0.4026	0.1941	**0.4083**
2013	0.1922	0.2026	**0.2353**	0.0648	0.1967	0.1394	0.1784	**0.1969**	0.0782	0.1915
2014	0.0799	0.1208	0.1378	0.1272	**0.1716**	-0.0404	**0.0818**	0.0372	0.0155	0.0537
2015	0.0631	-0.0053	0.0396	0.0837	**0.1035**	0.0701	0.0204	0.0822	0.0292	**0.0878**
2016	**0.1436**	-0.0568	0.0660	0.0190	0.0467	**0.1022**	-0.0930	0.0125	-0.0160	0.0180
2017	0.1675	0.2156	**0.2388**	0.0596	0.2270	0.2219	0.1870	**0.2412**	-0.0038	0.2026
all	0.1573	0.1513	0.1893	0.0860	**0.1964**	0.1312	0.1403	0.1703	0.0691	**0.1755**

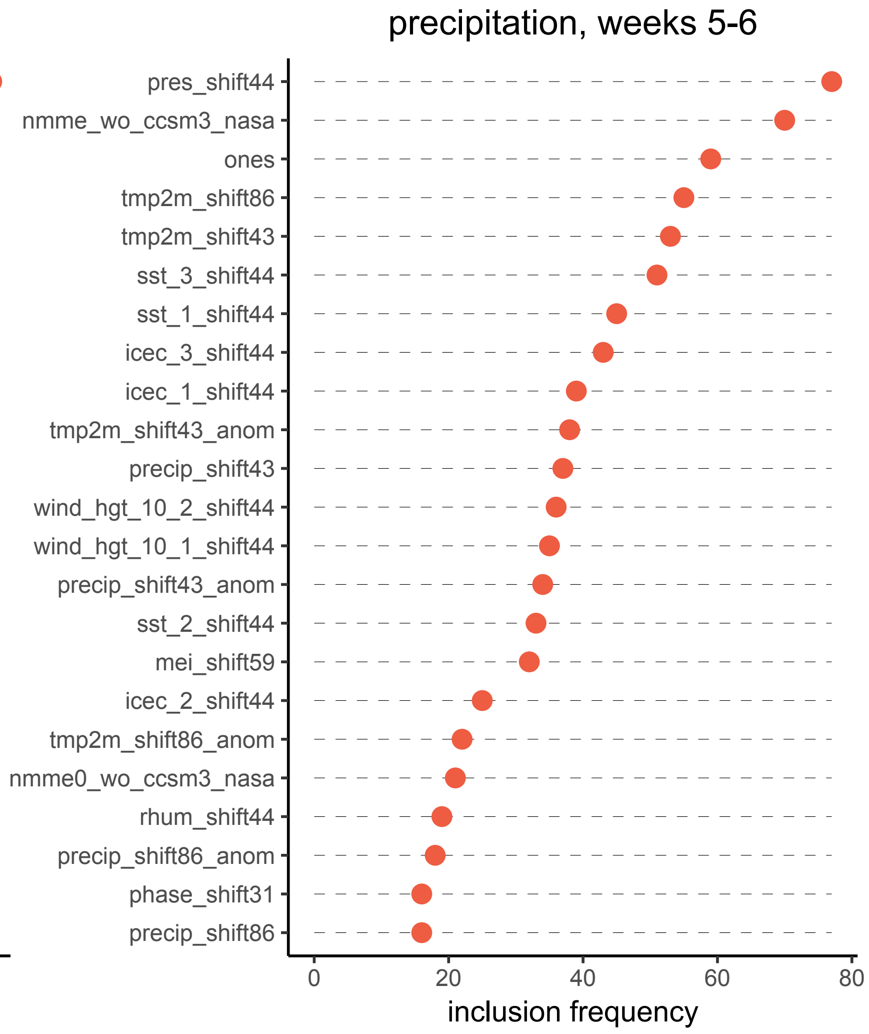
- MultiLLR, AutoKNN, and ensemble have higher mean skill than debiased CFSv2 on both tasks
- Ensemble improves over debiased CFSv2 by 120% for weeks 3-4 and 146% for weeks 5-6
- Ens-cfs improves over debiased CFSv2 by 128% for weeks 3-4 and 154% for weeks 5-6

Historical Forecast Evaluation



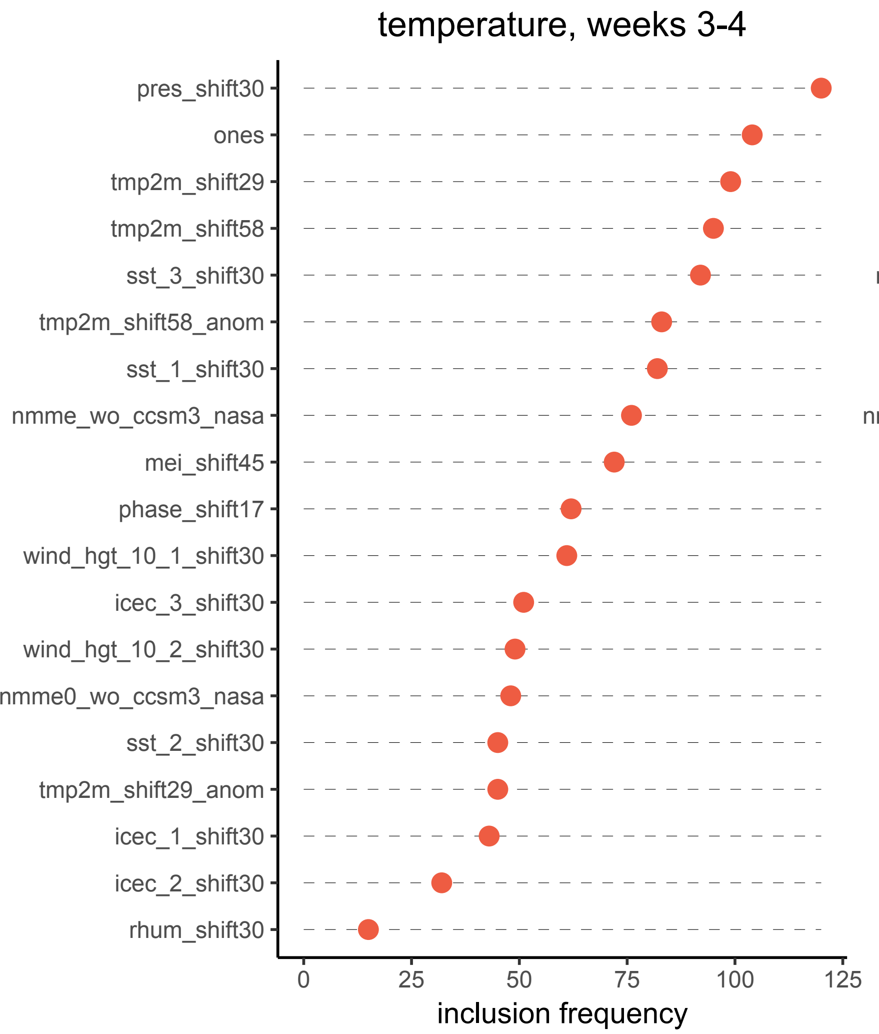
Exploring MultiLLR Feature Selection

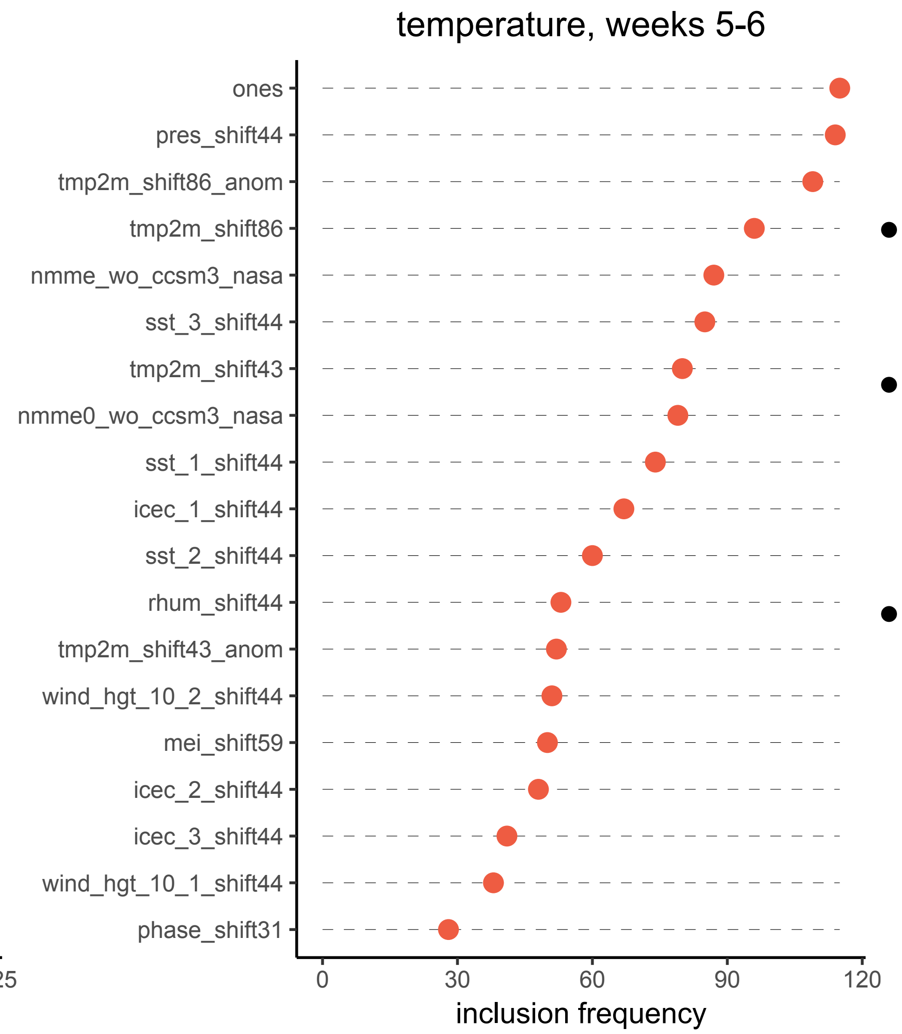




- Median number of selected features: 4 of 23 for weeks 3-4, 5 of 23 for weeks 5-6
- **Pressure**, ones, and **lagged temperature** are in top 4 features for all 4 tasks
- NMME frequently selected for precipitation tasks

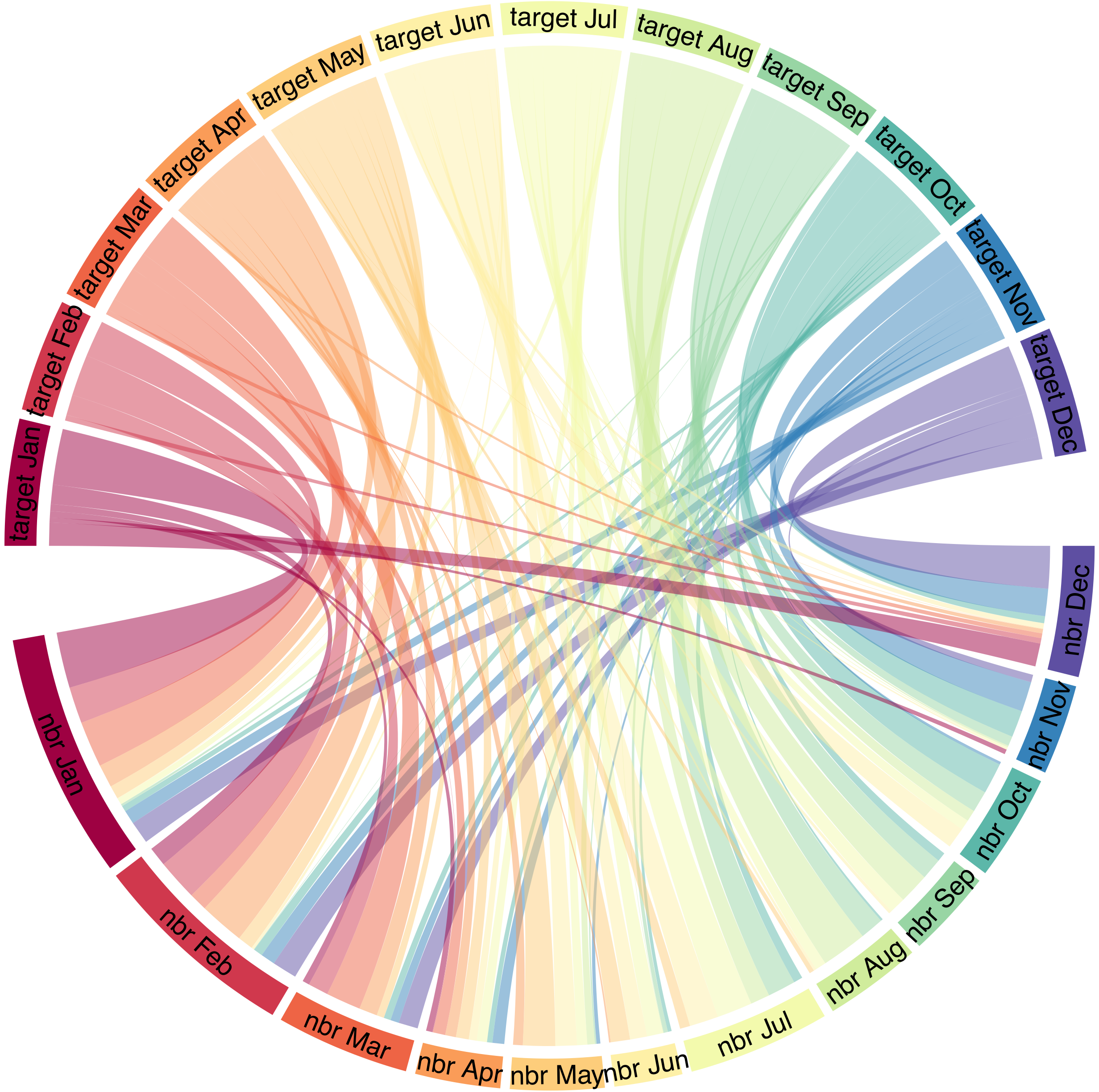
Exploring MultiLLR Feature Selection





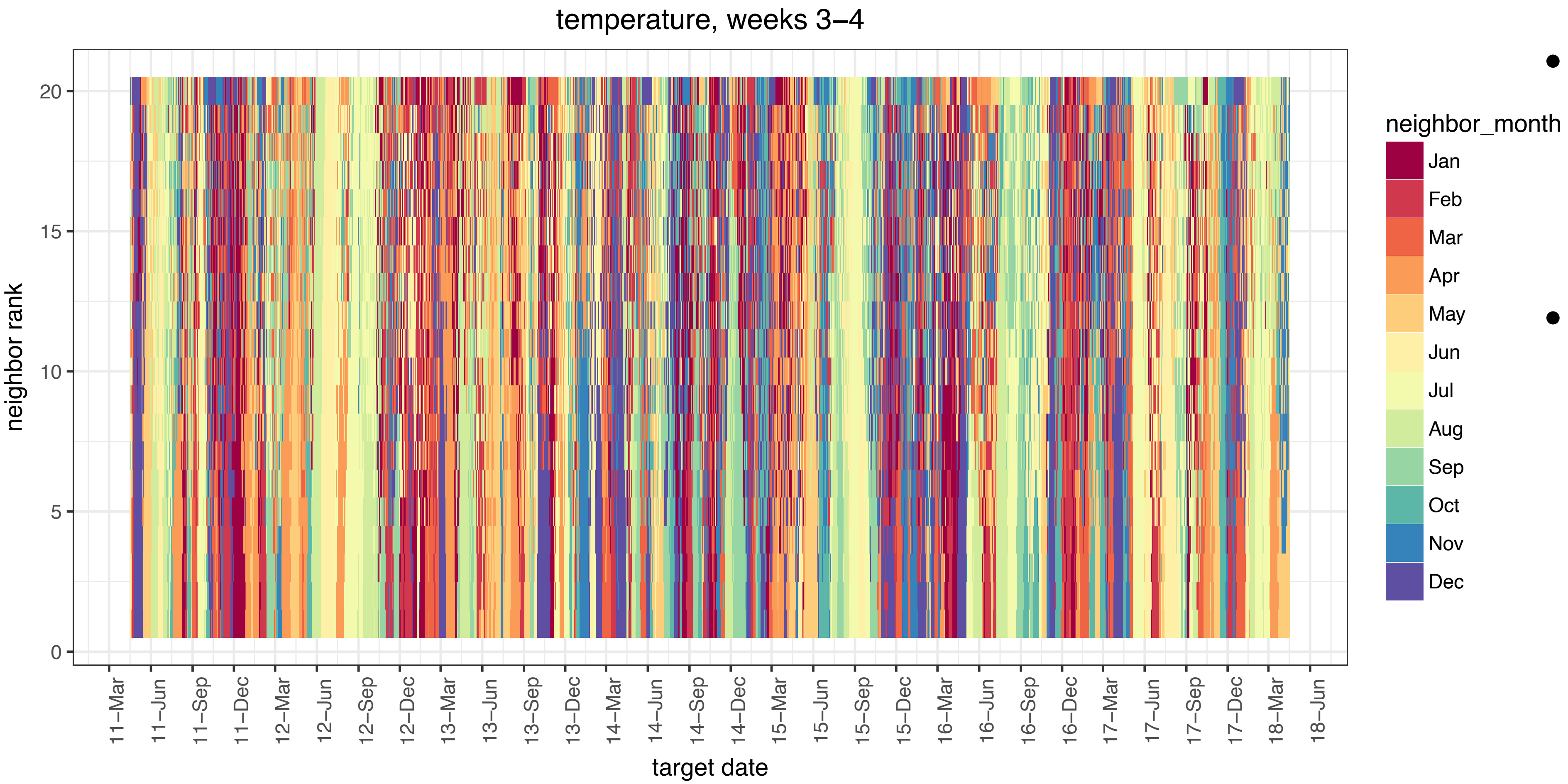
- Median number of selected features: 7 of 20
- **Pressure**, ones, and **lagged temperature** are in top 4 features for all 4 tasks
- NMME selected less frequently than for precipitation

Exploring AutoKNN Neighbor Selection



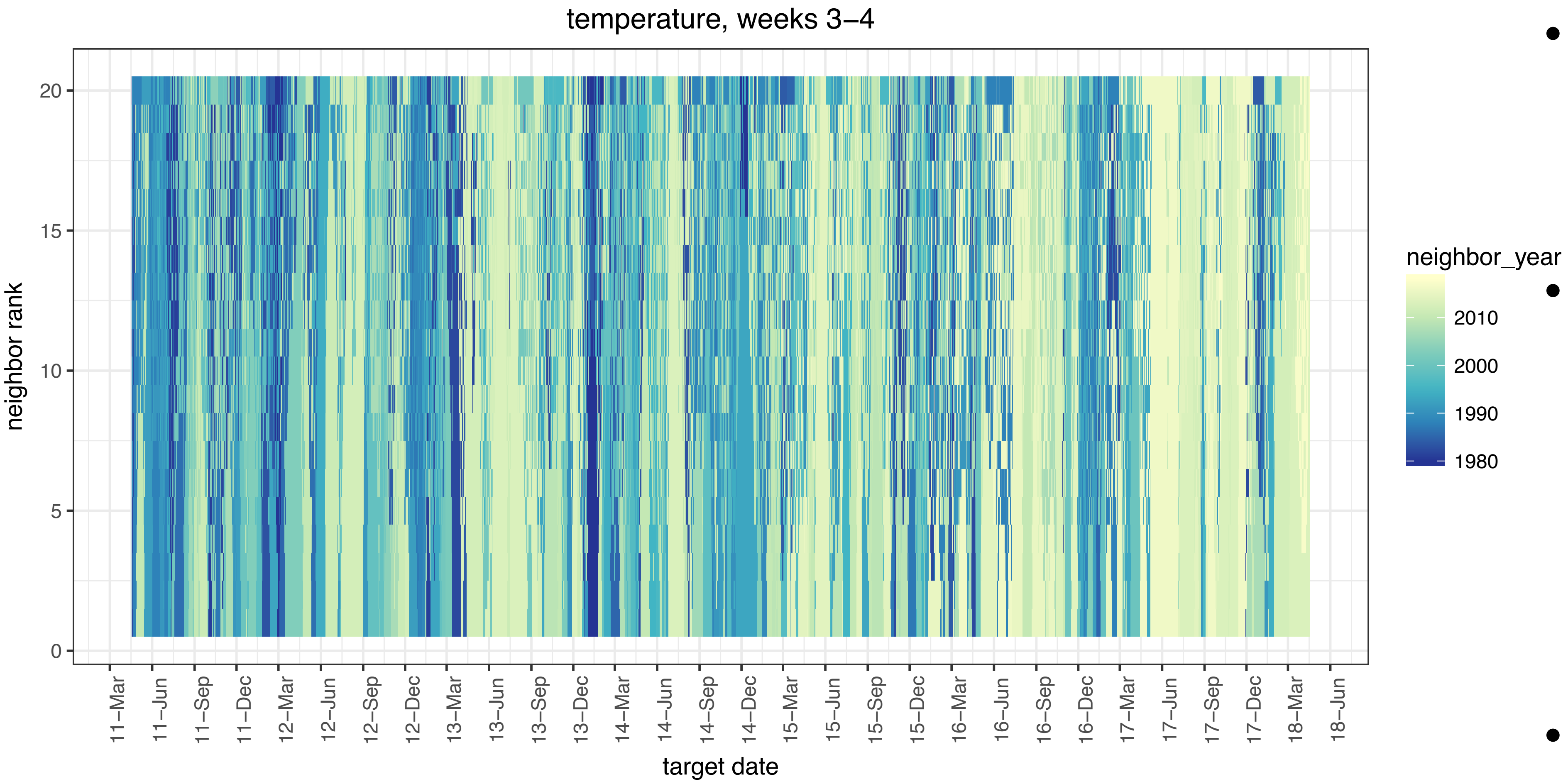
- Month distribution of most similar neighbor learned by AutoKNN for **precipitation weeks 3-4**, as a function of month of the target date
- **Top neighbor typically from same time of year**: summer targets typically have summer neighbors and winter targets typically have winter neighbors
- The same trend does not hold for temperature: neighbors are scattered about the year for all target month

Exploring AutoKNN Neighbor Selection



- Month of top 20 neighbors for temp weeks 3-4 by target date
- Vertical striations suggest that **neighbor months tend to be homogenous**: neighbors tend to come from similar times of year

Exploring AutoKNN Neighbor Selection



- Year of top 20 neighbors for temp weeks 3-4 by target date
- Vertical striations suggest that **neighbor years tend to be homogenous**: neighbors tend to come from similar years
- For post-2015 targets, most neighbors from post-2010 (consistent with record high temp)

The End

For more details, see

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

https://arxiv.org/abs/1809.07394

Photo Credit: BLM Photo