Super Ensemble Statistical Short-Range Precipitation Forecasts over the US and Improvements from Ocean-Area Precipitation Predictors

Thomas Smith[1], Sam Shen[2], and Ralph Ferraro[1]

1. NOAA/NESDIS/STAR and CICS/ESSIC/U. Maryland
2. San Diego State University

NOAA Satellites and Information
National Environmental Satellite, Data, and Information Service

Definitions

- Ensemble: A weighted mean of multiple realistic estimates
  - Traditionally used with dynamic GCM forecast runs with different initial conditions
  - Average used to estimate the expected value

- Statistical Ensemble: A weighted mean of different statistical estimates
  - Ensemble members may have different predictors or predictor regions or use different statistical models

- Super Ensemble: Use weights that reflect the accuracy of each ensemble member

Improvements

- Ensemble-statistical forecasting
  - Developed & tested by Shen et al. 2001 & Lau et al. 2002
  - Ensemble CCA improved seasonal U.S. T forecasts (Mo 2003)

- Method Improvements
  - Ensemble members for differences in predictor regions, predictor types, and statistical models
  - Optimal super-ensemble formed

- Data Improvements: include satellite ocean-area precipitation predictors

Statistical Super Ensemble Method

- Find predictors, p_1, p_2, …, p_n, for some property, g
- Separate models for each prediction, $f_1(p_1)$=g_1, …, $f_n(p_n)$=g_n
- Compute the n member ensemble, E[g] = $\sum_{i=1}^{n} w_n g_n$
- Optimal weights proportional to the correlation squared
- Use cross-validation to compute optimal weights

EARTH SYSTEM SCIENCE INTERDISCIPLINARY CENTER ESSIC UNIVERSITY OF MARYLAND

cics-md Cooperative Institute for Climate & Satellites UNIVERSITY OF MARYLAND

NOAA Satellites and Information

National Environmental Satellite, Data, and Information Service

Predictor & Predictand Areas: N.H. Oceans and Contiguous US

Region standard deviations, for OI SST anomalies (upper) and GPCP P anomalies (lower)

4 Ocean predictor areas:

1) Trop Pacific (23°S-23°N, 150°E-80°W)
2) Trop Atlantic (23°S-23°N, 90°W-20°E)
3) N. Pacific (20°N-60°N, 150°E-100°W)
4) N. Atlantic (20°N-60°N, 100°W-0°W)

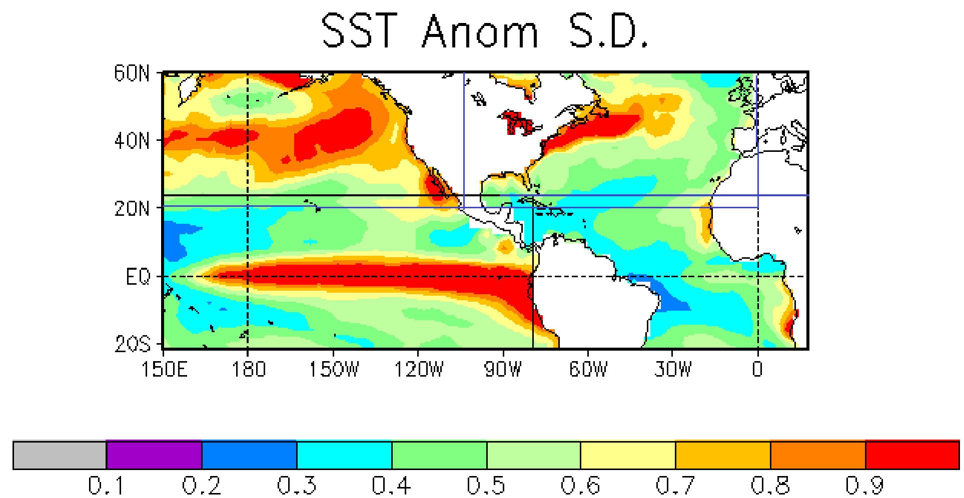
Some overlap in ocean areas

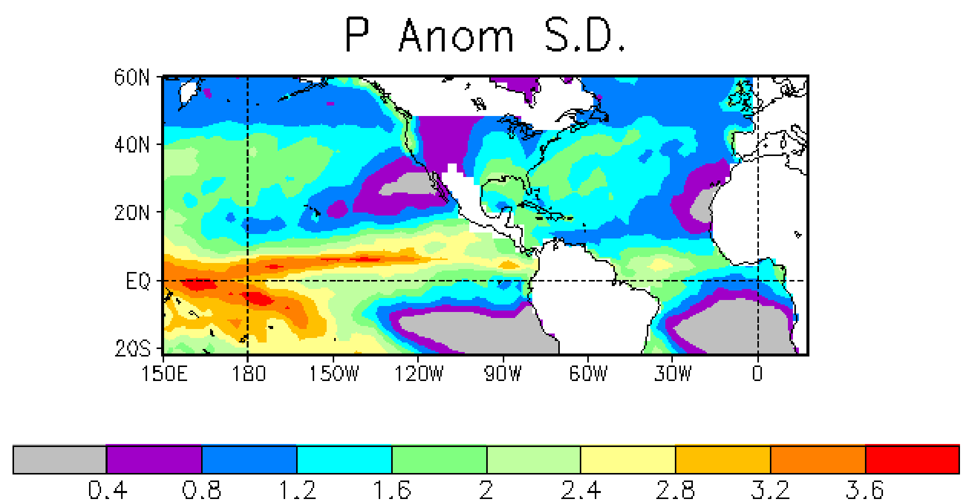
Regions likely to influence P_{US}, similar to Lau et al. (2002) areas

Predictors for ensemble:

- Ocean area $SST_k(t-1)$
- US area $P_{US}(t-1)$
- Ocean area $P_k(t-1)$

Always predict $P_{US}(t)$ anoms





One Model: Canonical Correlation Analysis (CCA)

- Used by Barnett, Banston, many others
  - Decompose predictor and predictand fields using EOFs
  - Compute CCA in spectral space
  - X-val tuning indicates that using 20 CCA modes is best

- Correlation between predictor field and the time-lagged US precipitation field used for forecast

- Separate CCA for each predictor type and region

Another: Joint Empirical Orthogonal Analysis (JEOF)

- JEOF is EOF of several fields stacked together
- Normalize predictor and time-lagged US P fields, stack together and perform EOF
- JEOF for each predictor type and region
- X-val tuning shows that 5 JEOF modes is best
- For both CCA & JEOF anomalies are forecasts, and preliminary test show separate models for different seasons are not needed

Super-Ensemble Weights

- For OI at a point, spatial correlations = 1 and weights are a function of noise/signal error variance

$$w_i = \frac{1}{1+\eta_i^2}$$

- Assume that each ensemble, x_i, is a linear function of the truth, x, with random error & maybe bias

$$x_i = \alpha_i x + \beta_i + \varepsilon_i$$

- Using definitions of variance and correlation, and we can show that weights are a function of squared correlation, $w_i = r_i^2$

- Normalize weights to avoid damping or inflation of variance, compute maps of weights

Data & Evaluations

- GPCP monthly precipitation and OI monthly SST inputs
  - 1997-2014 1dd GPCP averaged to monthly, compute anomalies
- Cross-validation testing of 0-lead monthly forecasts
  - Omit all data for the year of analysis and 3 months on either side of the year
- Data from month t-1 to predict month t
- Correlations used to evaluate skill and improvements

All-area SST CCA vs ensemble SST CCA

- CCA skill using all SST together < skill of ensemble from divided SST_i regions, i=1 to 4
- Non-ensemble SST skill similar to skill using $P_{US}(t-1)$
- All averages omit no-skill regions (correlations < 0)

Temporal correlations against GPCP computed for each month (1997-2014), averaged over the contiguous US and annually.

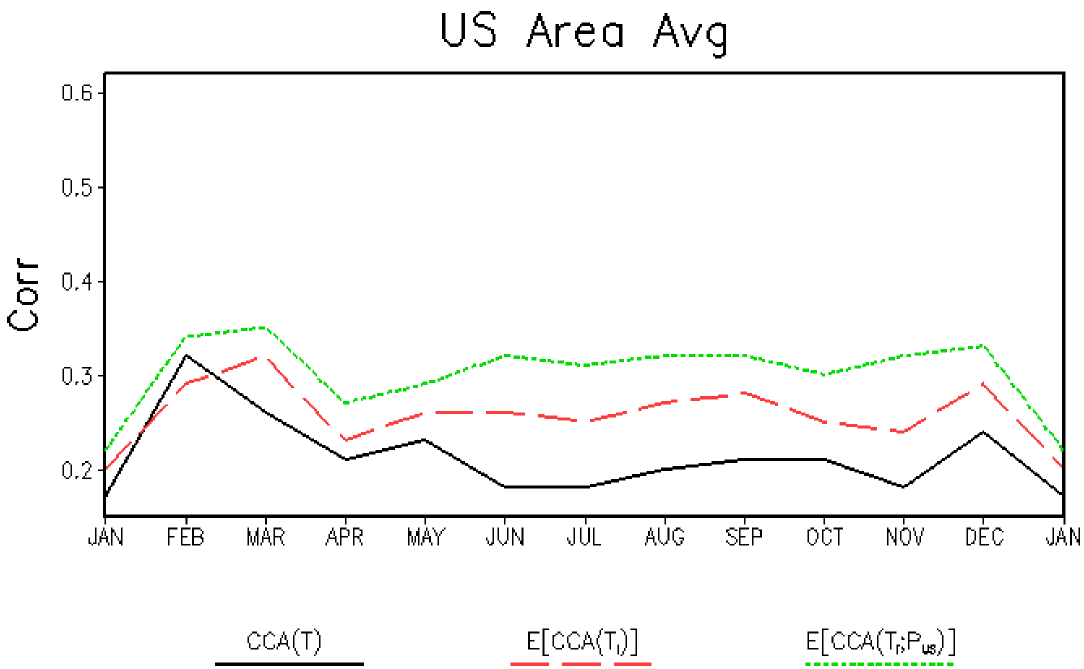
Predictors	CCA
SST(t-1)	0.22
$E[SST_i(t-1)]$	0.26
$P_{US}(t-1)$	0.22

Annual Cycle of US Average Correlation Skill

Multiple-CCA ensemble using SST(t-1) in regions almost always better than CCA using the same SST(t-1) combined

Spring-summer months most improved

Ensemble improved more when including prediction from P_{US}(t-1)

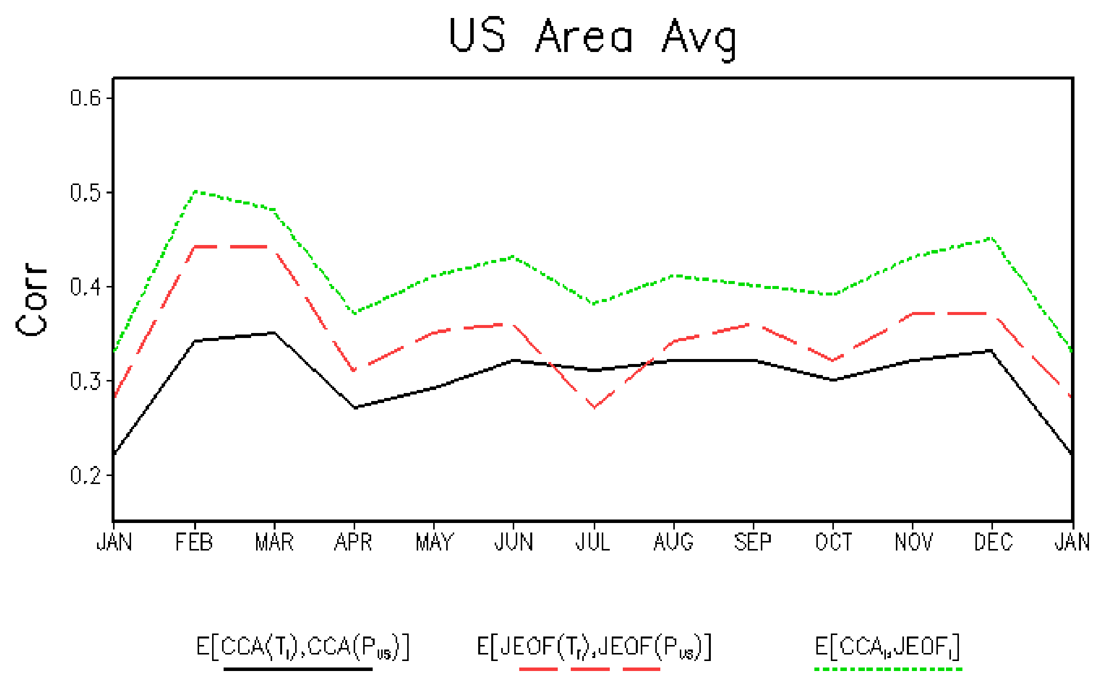


CCA vs JEOF, Annual Cycle of US Average Correlation

Ensemble using SST(t-1) and P_{US}(t-1); no oceanic P predictor members

JEOF typically better than CCA

Improved more if both JEOF and CCA members used in ensemble

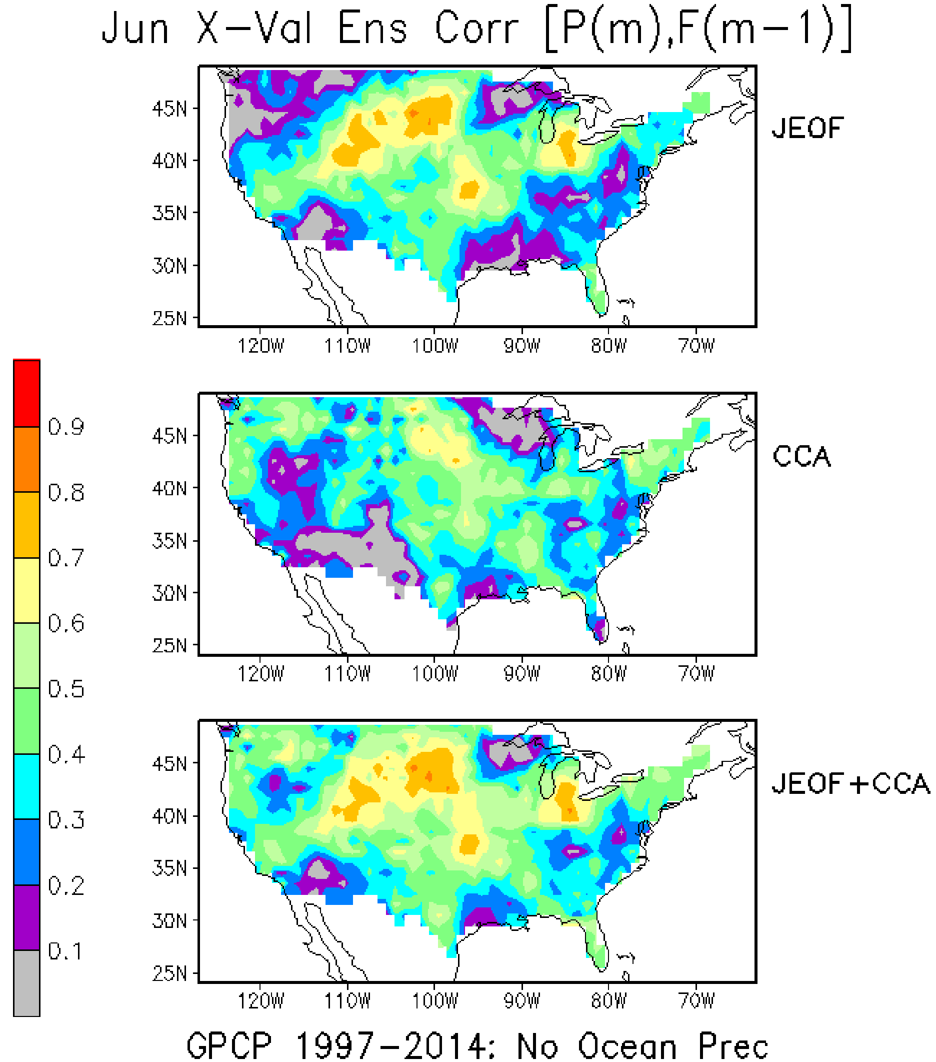


Cross-Validation Precipitation Anomaly Correlation: June, no oceanic precipitation

JEOF and CCA skill patterns similar, but not identical

Regions of high skill different in different models

Super ensemble using both takes the best of each



NOAA Satellites and Information
National Environmental Satellite, Data, and Information Service

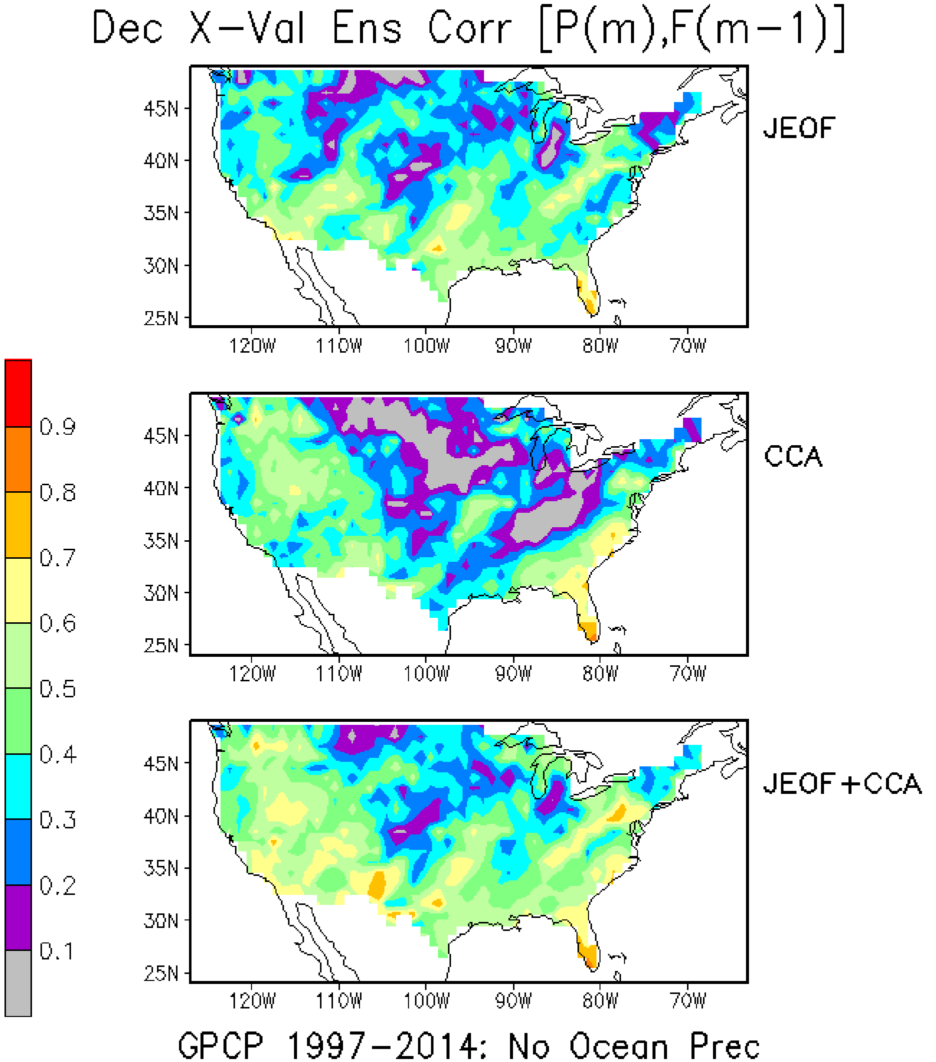
Cross-Validation Precipitation Anomaly Correlation: December, no oceanic precipitation

Both JEOF and CCA show skill gaps but in different regions

Using both expands the region of good skill

Methods Conclusions:

1) Ensembles dividing predictors into regions improves skill
2) Using ensemble members from multiple models noticeably improves skill

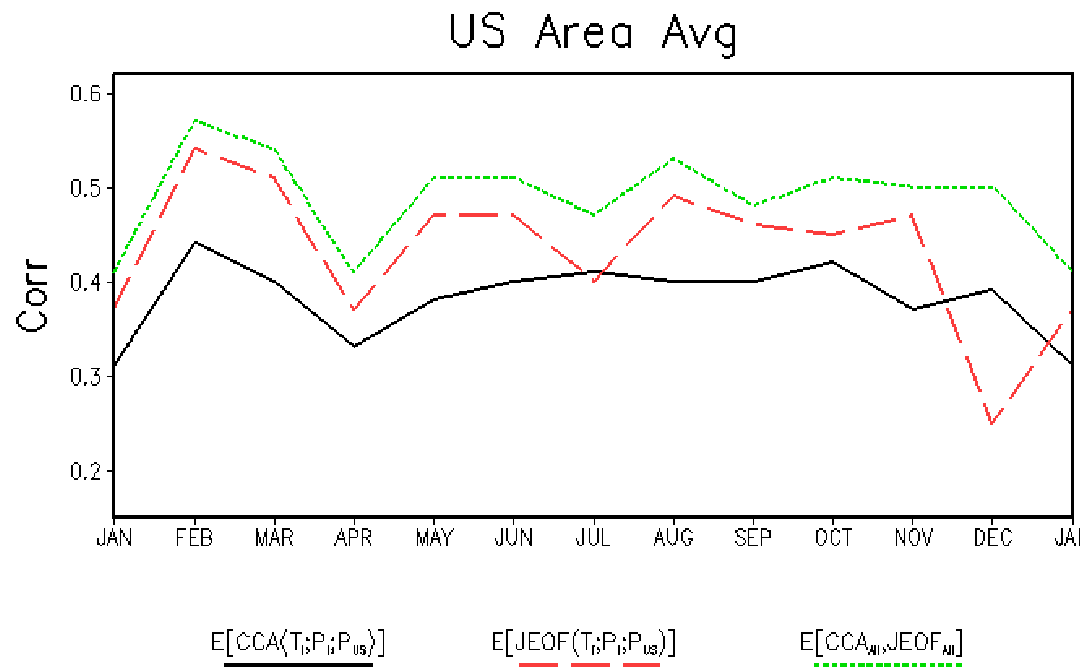


NOAA Satellites and Information
National Environmental Satellite, Data, and Information Service

Including Oceanic Precipitation in 4 Regions

Skill increases when including members with ocean area P(t-1) predictors

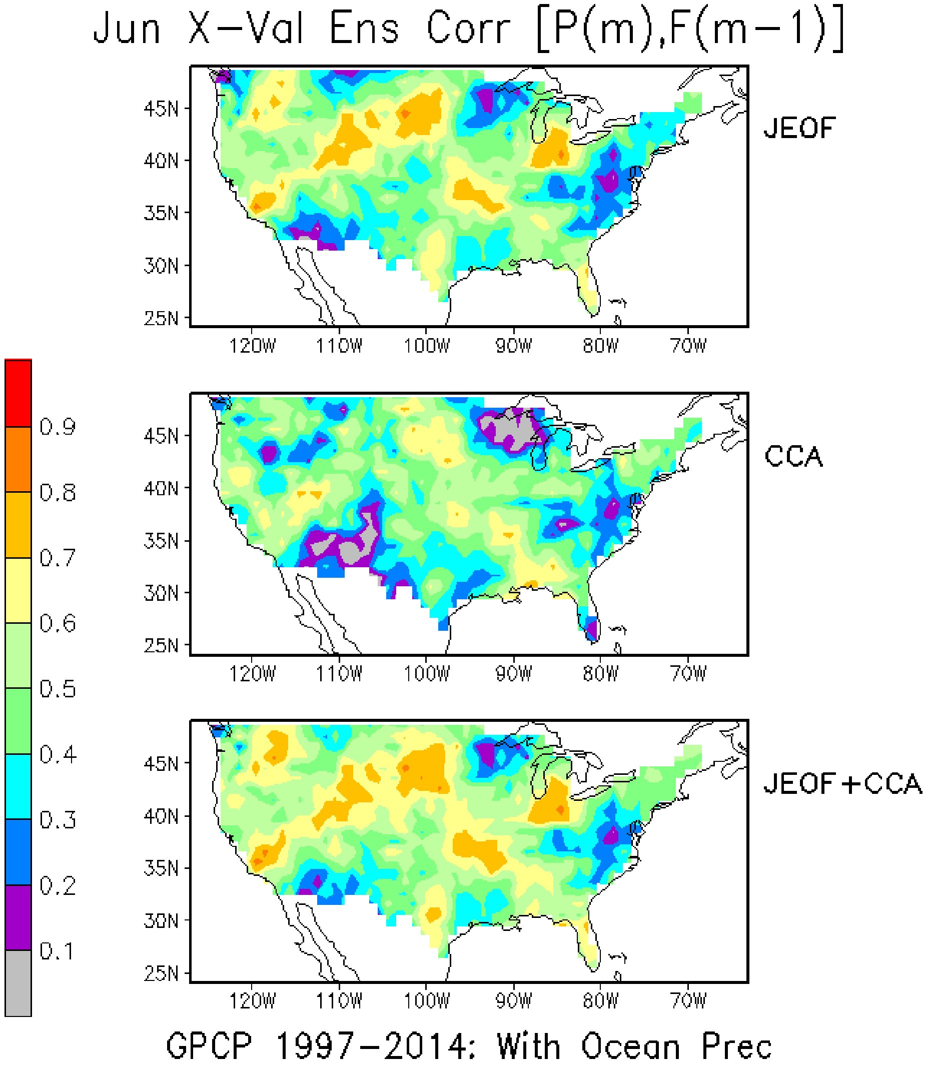
JEOF better than CCA, using both is best



Cross-Validation Precipitation Anomaly Correlation: June, with oceanic precipitation

Ocean P ensemble members improve both JEOF and CCA

JEOF still better, and combining them still gives higher skill

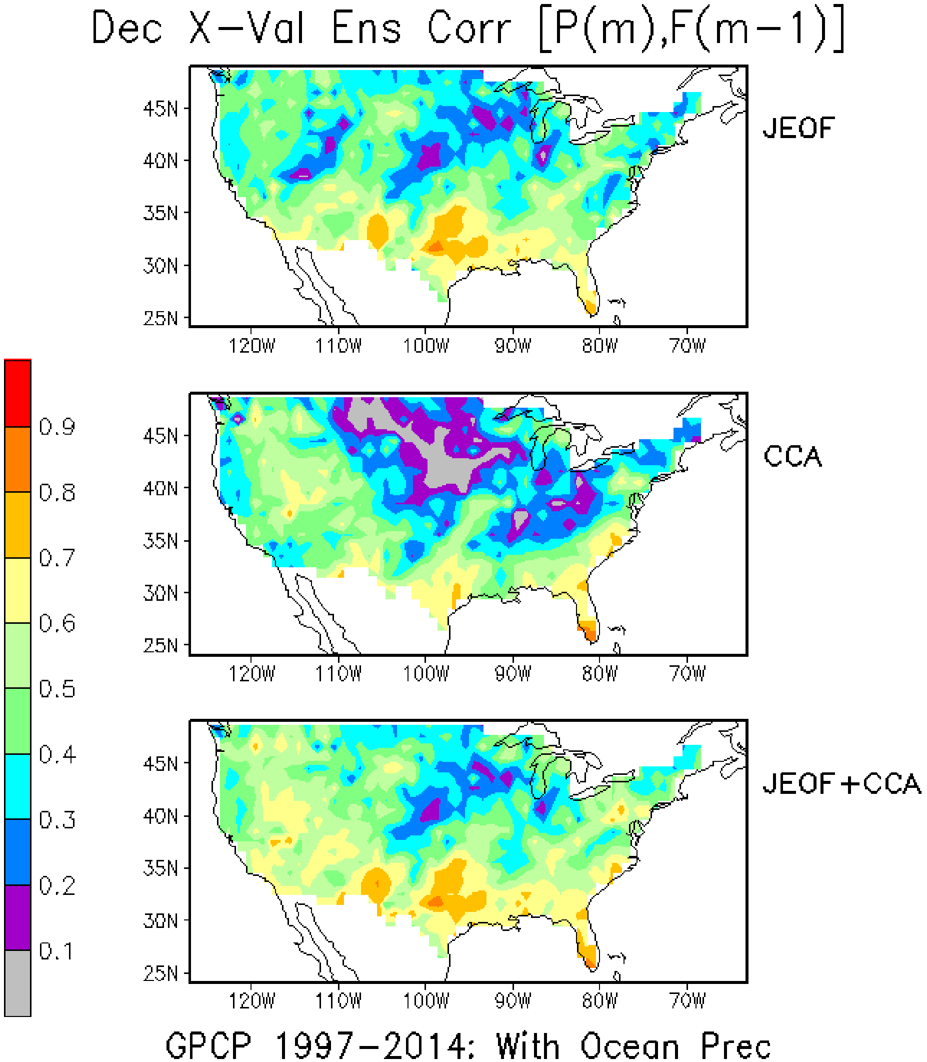


Cross-Validation Precipitation Anomaly Correlation: December, with oceanic precipitation

More regions with higher skill than the case with no oceanic precipitation: satellite-based P improves the forecast

Best skill apparently from ENSO

Low-skill regions for both JEOF and CCA not improved by combining them



NOAA Satellites and Information
National Environmental Satellite, Data, and Information Service

Skill from more than ENSO

- Skill from Tropical Pacific area SST or Precip important but not the whole story
- Combining with forecasts using SST and Precip from other regions doubles average correlation

Temporal correlations against GPCP computed for each month (1997-2014), averaged over the contiguous US and annually.

Predictors	CCA	JEOF
T_{TPac}	0.20	0.18
P_{TPac}	0.21	0.23
$E[T_i, P_{US}]$	0.31	0.35
$E[T_i, P_i, P_{US}]$	0.39	0.45

EARTH SYSTEM SCIENCE INTERDISCIPLINARY CENTER ESSIC UNIVERSITY OF MARYLAND

cics-md Cooperative Institute for Climate & Satellites UNIVERSITY OF MARYLAND

NOAA Satellites and Information
National Environmental Satellite, Data, and Information Service

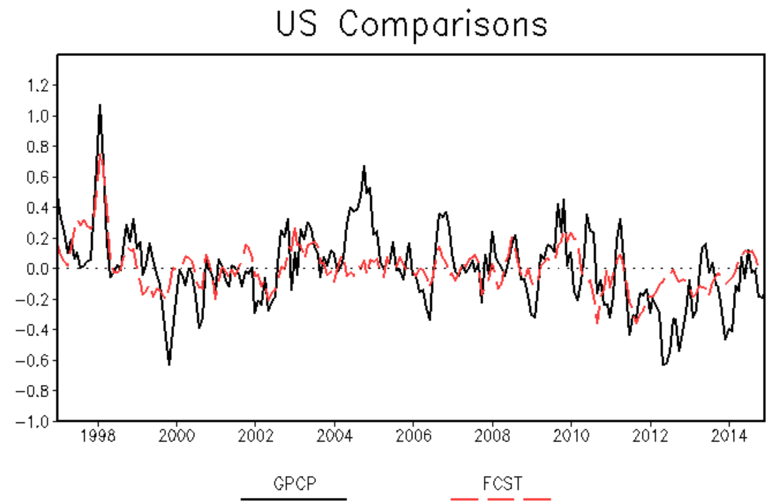
Overall Improvements from oceanic precipitation

- Adding satellite-based $P_i(t-1)$ predictors improves ensembles
- JEOF method slightly better than CCA but best-skill regions are different

Temporal correlations against GPCP computed for each month (1997-2014), averaged over the contiguous US and annually.

Predictors	CCA	JEOF	JEOF+CCA
$E[T_i, P_{US}]$	0.31	0.35	0.42
$E[T_i, P_i, P_{US}]$	0.39	0.45	0.50

US Area-Average of Forecasts vs GPCP

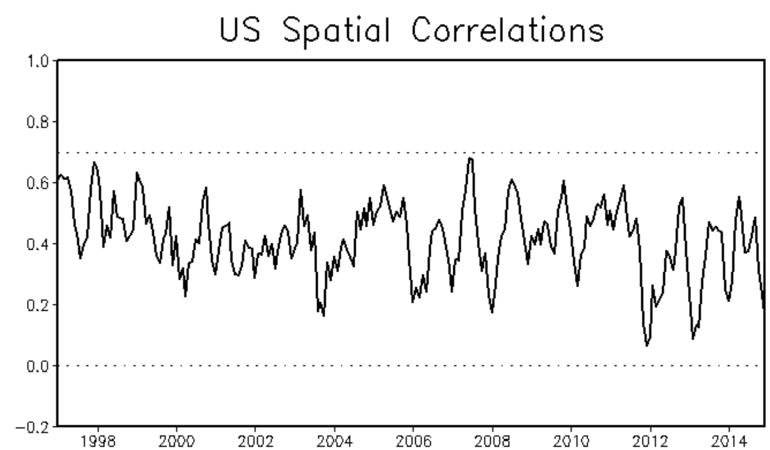


Monthly values 3-mon smoothed

Most large variations consistent, but with important misses

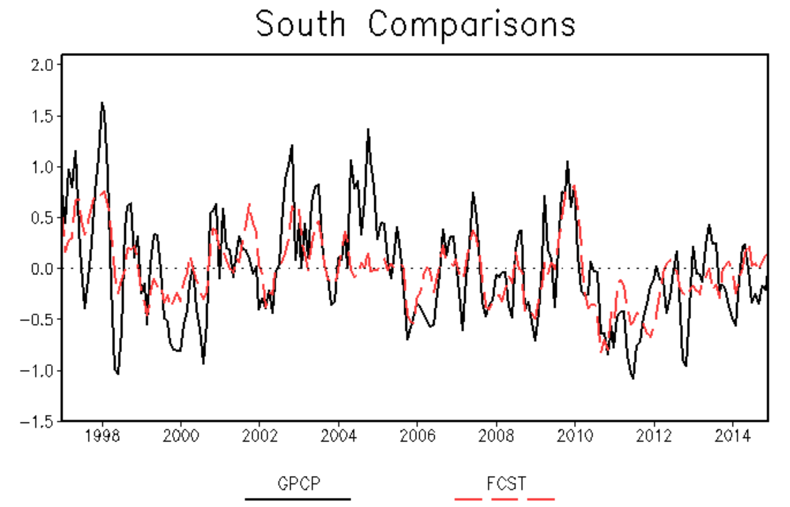
Tends to damp when it misses

Climate variations like ENSO help the correlation (avg 0.41)



NOAA Satellites and Information
National Environmental Satellite, Data, and Information Service

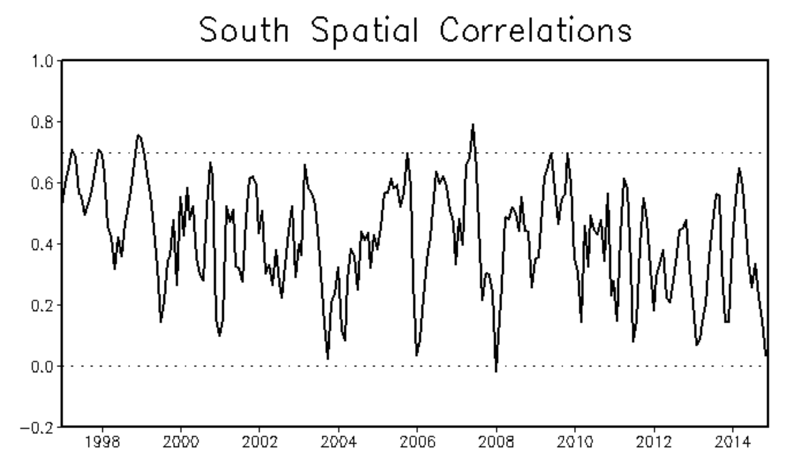
US Area South of 35°N Forecasts vs GPCP



Monthly values 3-mon smoothed

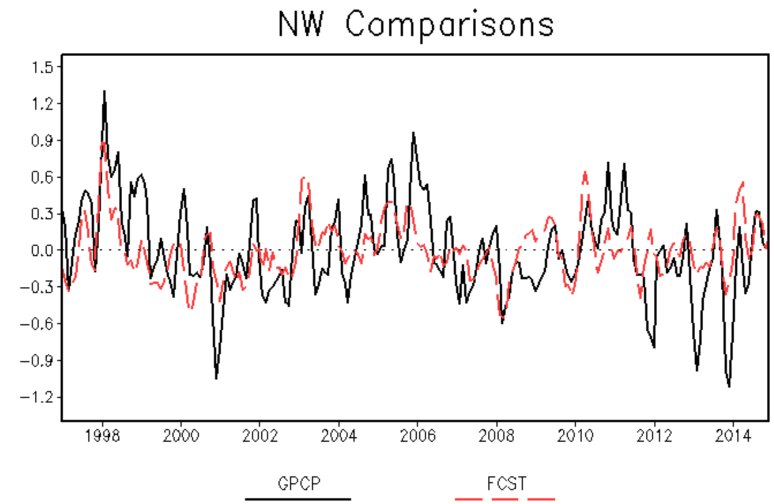
Region influenced by ENSO

Fewer misses and correlation slightly better than for entire US (avg 0.42)



NOAA Satellites and Information
National Environmental Satellite, Data, and Information Service

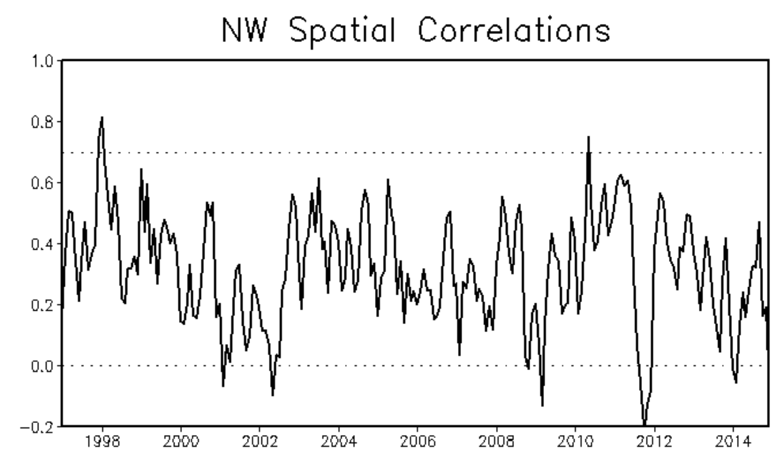
US North of 35°N, West of 100°W
Forecasts vs GPCP



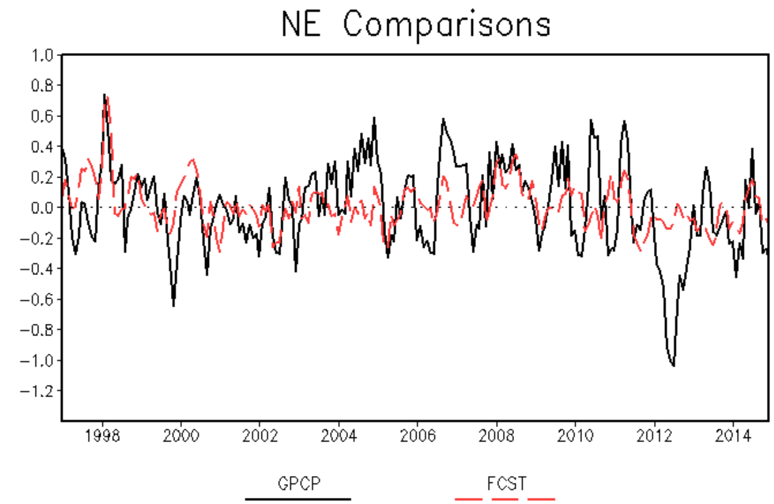
Monthly values 3-mon smoothed

Multi-decadal variations clear

More misses, correlation lower than entire US (avg 0.32)



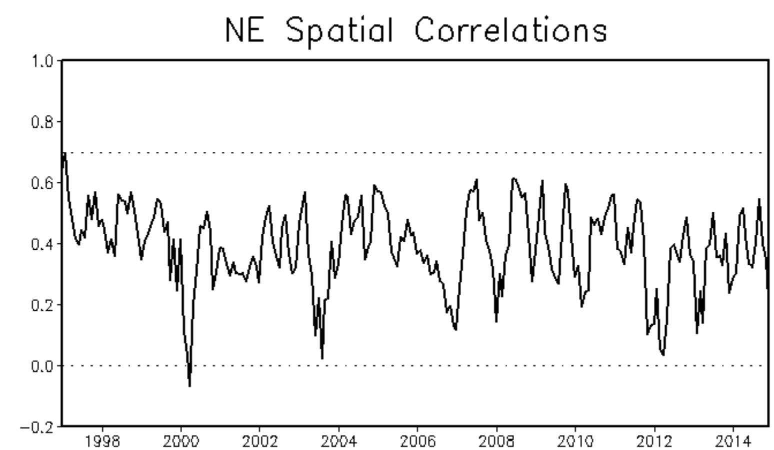
US North of 35°N, East of 100°W
Forecasts vs GPCP



Monthly values 3-mon smoothed

Forecast misses slight multi-decadal variations and some extremes

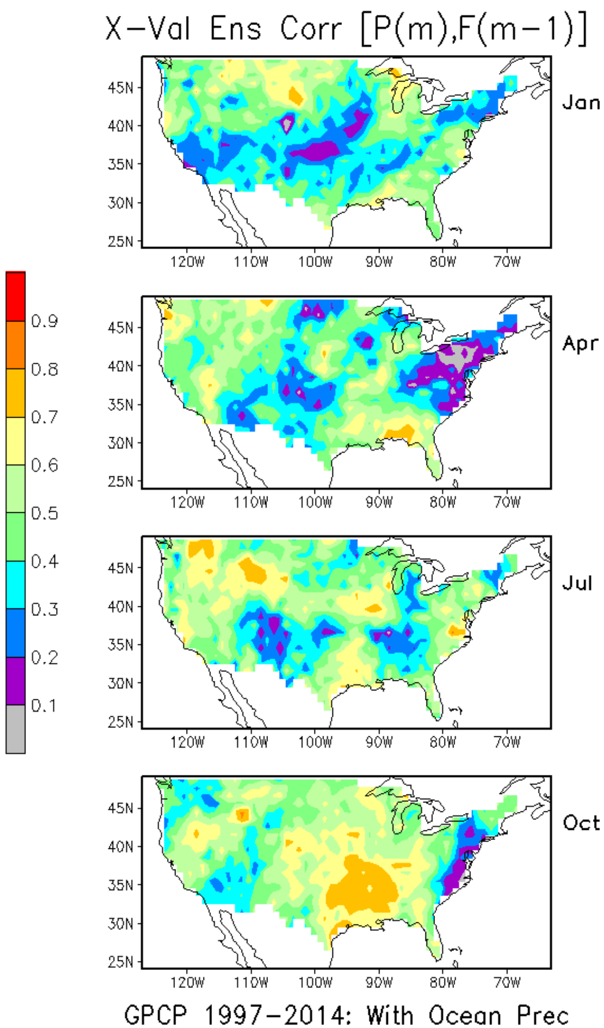
More misses, correlation slightly lower than entire US (avg 0.39)

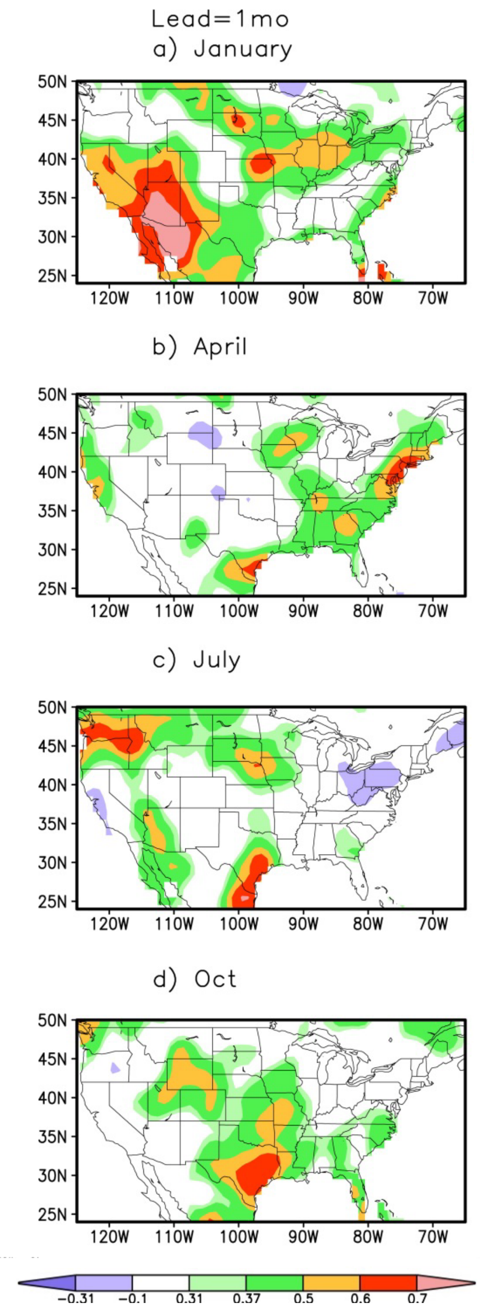


NOAA Satellites and Information
National Environmental Satellite, Data, and Information Service

Comparisons to Similar NAMME Tests

Similar Skill Levels but in Different Regions





From Mo and Lettenmaier (2014, J. Hydromet.)

EARTH SYSTEM SCIENCE INTERDISCIPLINARY CENTER
ESSIC
UNIVERSITY OF MARYLAND

cics-md
Cooperative Institute for Climate & Satellites
UNIVERSITY OF MARYLAND

NOAA Satellites and Information
National Environmental Satellite, Data, and Information Service

3-Category Validation

- For each month use the 18 years to define the lowest, middle, and highest third (below normal, normal, above normal categories)

- Find the % time forecast is in the correct third (hit)

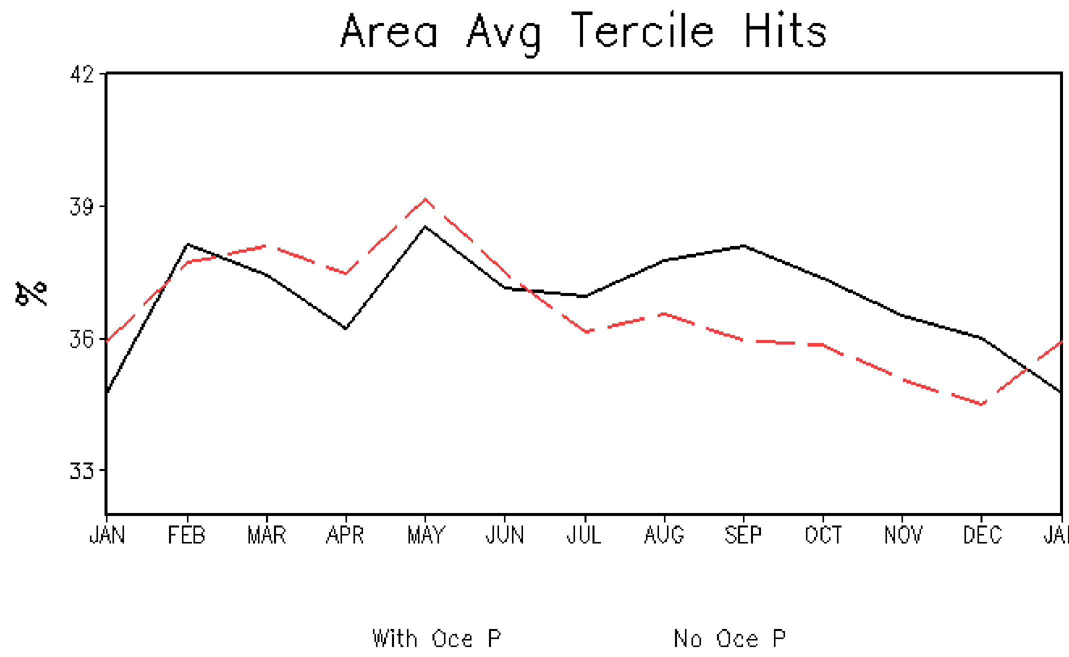
- Find the % time forecast misses by 2 categories (bad miss)

Averages for Each Month: Hits With & Without Ocean P

% Hits: forecast in correct third

Average does not change much over year

Sometimes more hits when ocean P not used, but typically better with ocean P



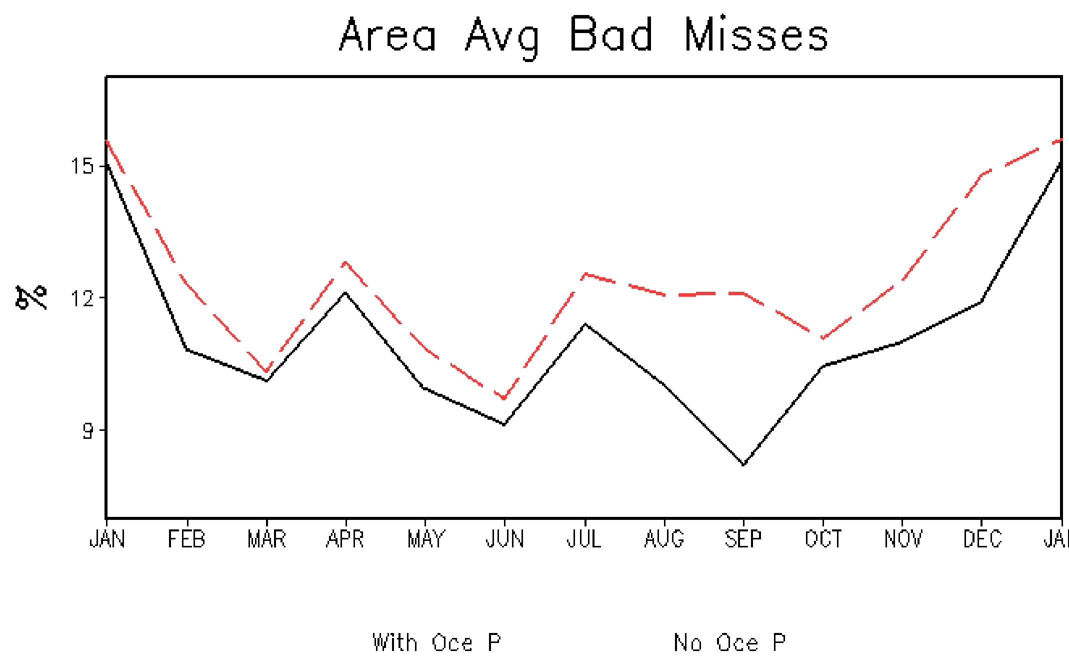
NOAA Satellites and Information

National Environmental Satellite, Data, and Information Service

Tercile Averages for Each Month: Bad Misses With & Without Ocean P

% Bad Misses: forecast upper & validate lower third or forecast lower & validate upper third

Using ocean P reduces bad misses



EARTH SYSTEM SCIENCE INTERDISCIPLINARY CENTER ESSIC UNIVERSITY OF MARYLAND

cics-md

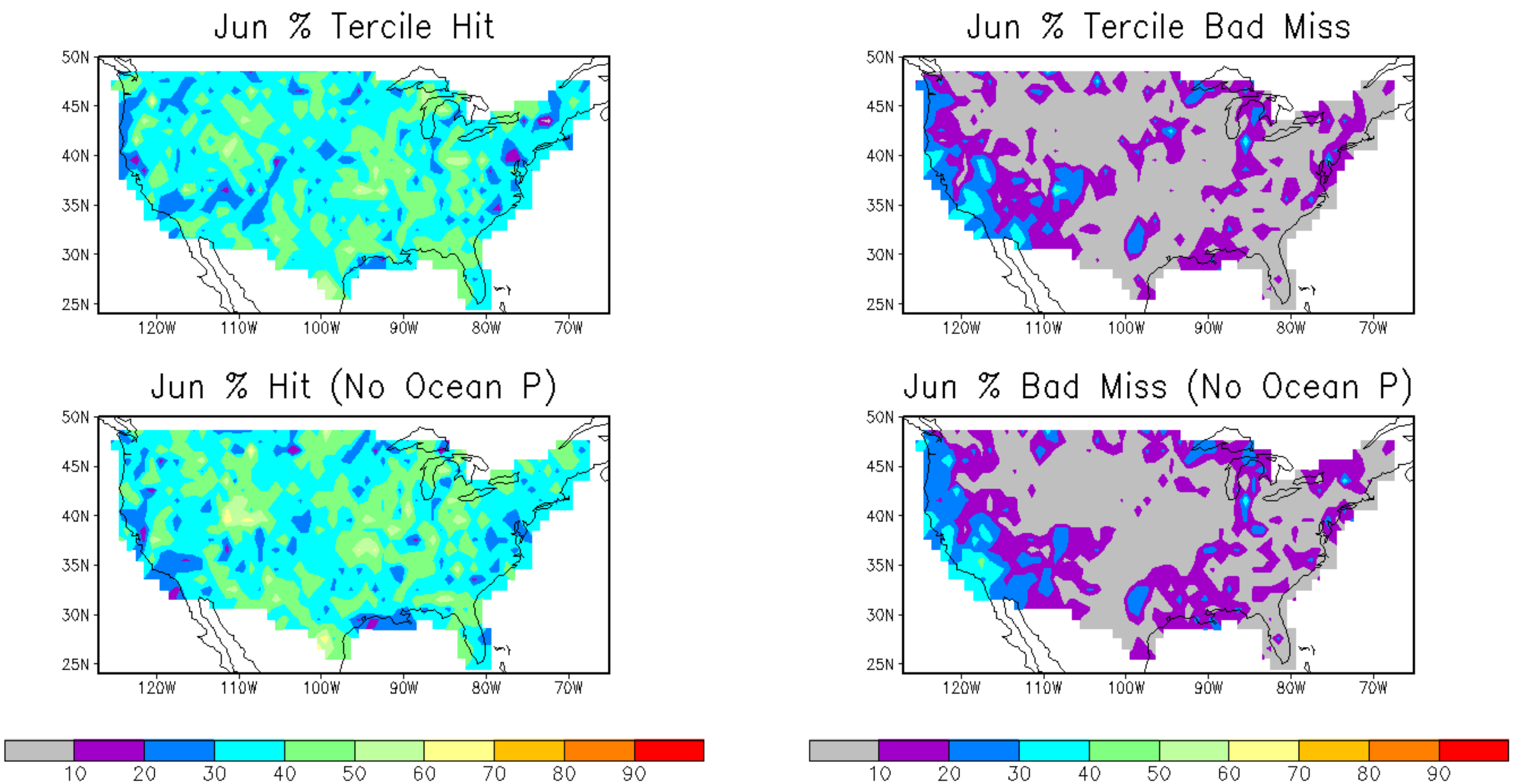
NOAA Satellites and Information

National Environmental Satellite, Data, and Information Service

Monthly June Maps: Hits & Bad Misses

Ocean P has little impact on Hits

Differences are clearer in bad misses, especially western areas



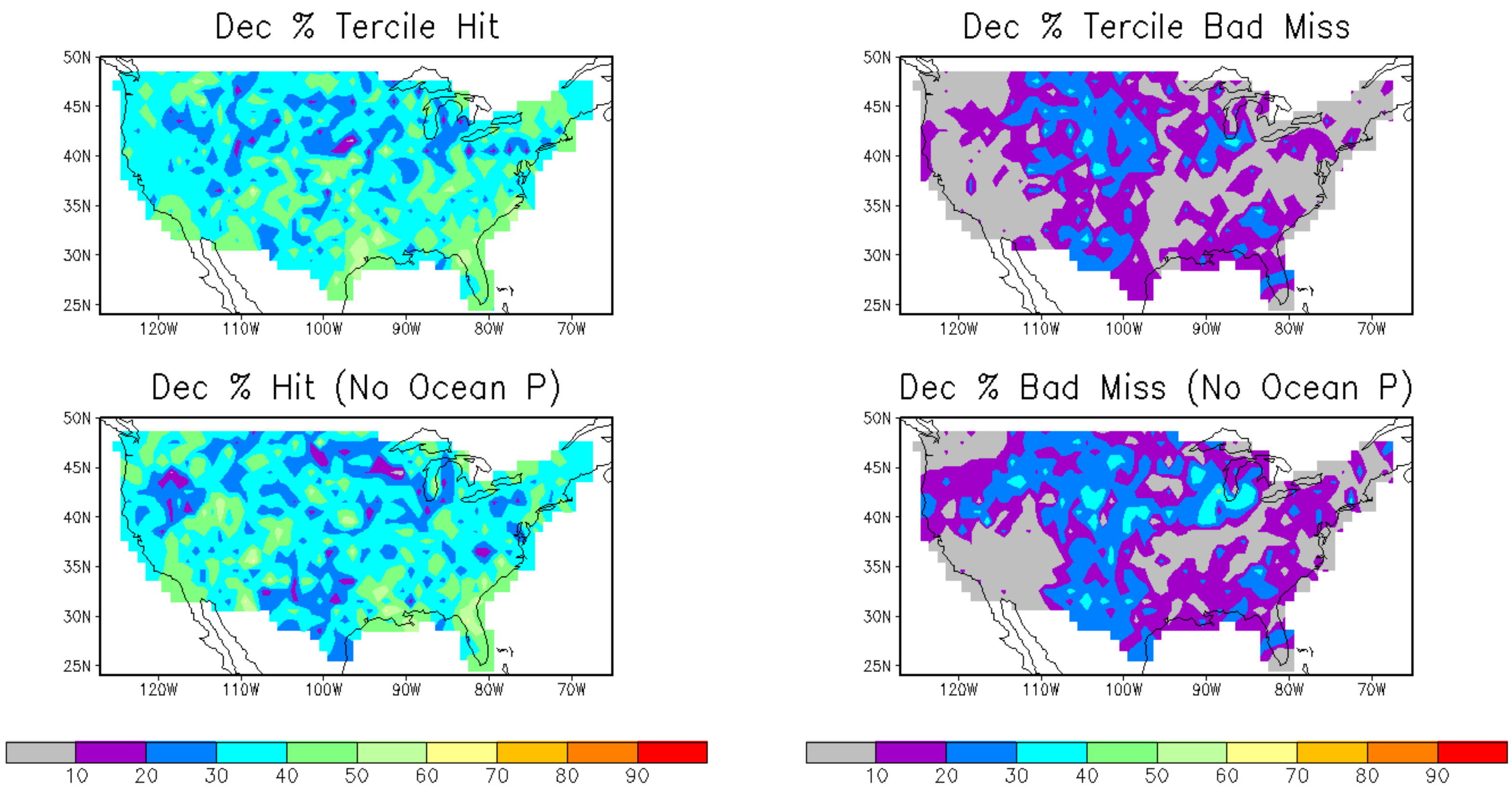
NOAA Satellites and Information

National Environmental Satellite, Data, and Information Service

Monthly December Maps: Hits & Bad Misses

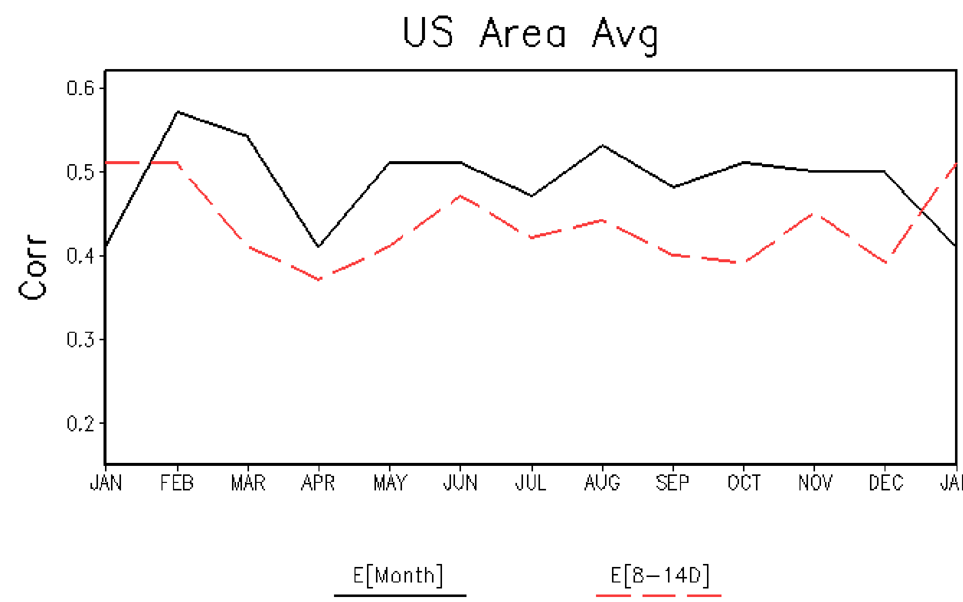
Ocean P has more impact on hits in December

Again differences are clearer in bad misses, most in mid west and southeast



Testing 8-14 Day Forecast

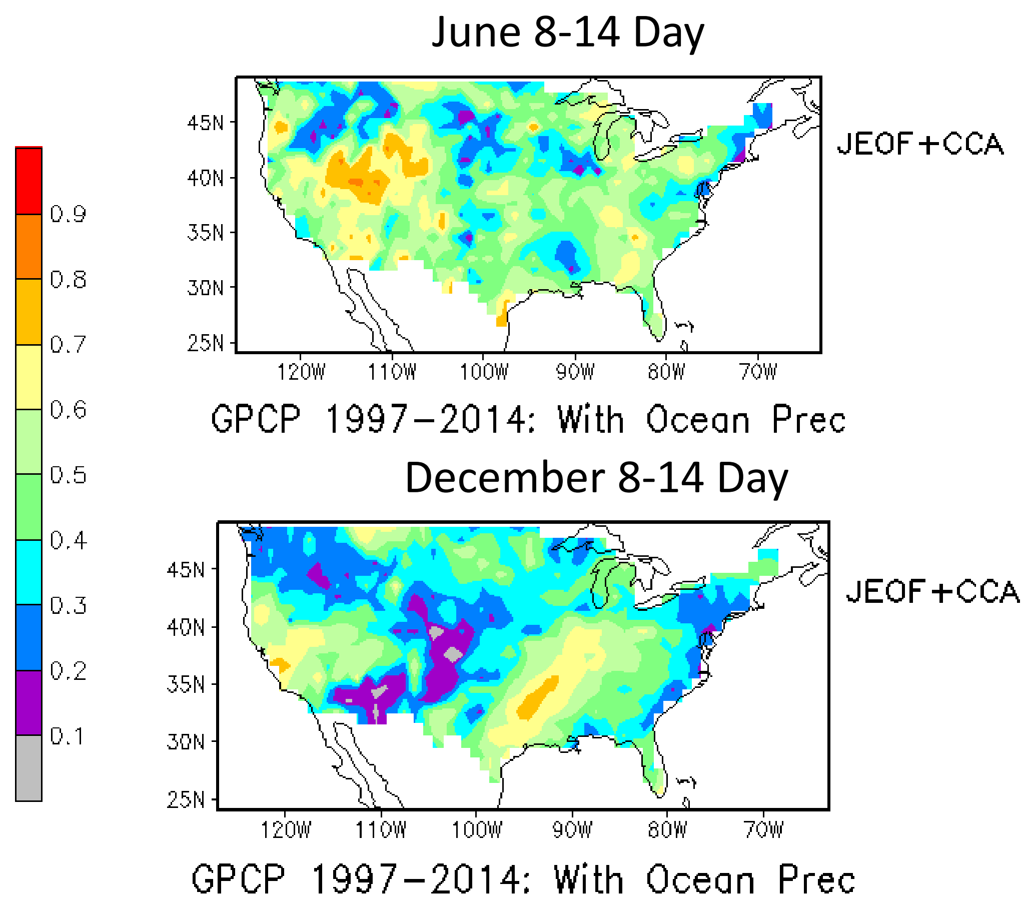
- Forecast for day 8-14 average of each month as test
- Predictors: SST for previous month, P for last week of previous month
- Ensembles of CCA+JEOF, all predictors
- Average skill similar to monthly skill but usually slightly lower



8-14 Day 1-Week Forecast Skill Patterns

Different from monthly patterns with larger areas of low skill

Need to independently test all forecasts of interest



Conclusions

- Super-ensemble-statistical forecast better than comparable non-ensemble forecasts
- JEOF better than CCA and multiple linear models gets additional information from the same predictors
- Ocean-area precipitation predictors improves US-area precipitation forecasts

Next Steps

- Need to more fully develop and test methods and new data sources
  - More lead times & more predictors
  - Other regions, test both T & P predictions
  - Funding likely needed to get more people working on the project
- Super ensembles can incorporate both statistical and dynamic predictions
- Need interested partners for improvements to become part of operational forecasts