A metrics framework to evaluate new approaches for drought monitoring and prediction

Andy Wood
NOAA / NWS
Northwest River Forecast Center

acknowledge: James McCreigh & Balaji Rajagopalan, CU

NOAA CPO Modeling, Analysis, Predictions, and Projections (MAPP) program Webinar
Virtual – February 12, 2013

Topics

- ❑ MAPP Drought Task Force Overview
  - ❑ Objectives
  - ❑ Organization
  - ❑ Accomplishments
- ❑ Framework for Drought Science Evaluation
- ❑ Application Example
  - ❑ Streamflow Prediction
  - ❑ Comment on Monitoring
- ❑ Future Directions

MAPP Drought Task Force Overview

MAPP DTF Mission

- Bring together and facilitate MAPP-funded research efforts aimed at achieving advances in capabilities to monitor and predict drought over North America
  - *assess current efforts to advance monitoring and prediction*
- Contribute to efforts to advance official national drought products including:
  - development of a DEWS by NIDIS
  - drought monitoring/prediction activities at NCEP
- Coordinate with and leverage other relevant national and international efforts (e.g. NMME, WCRP Drought Interest Group).

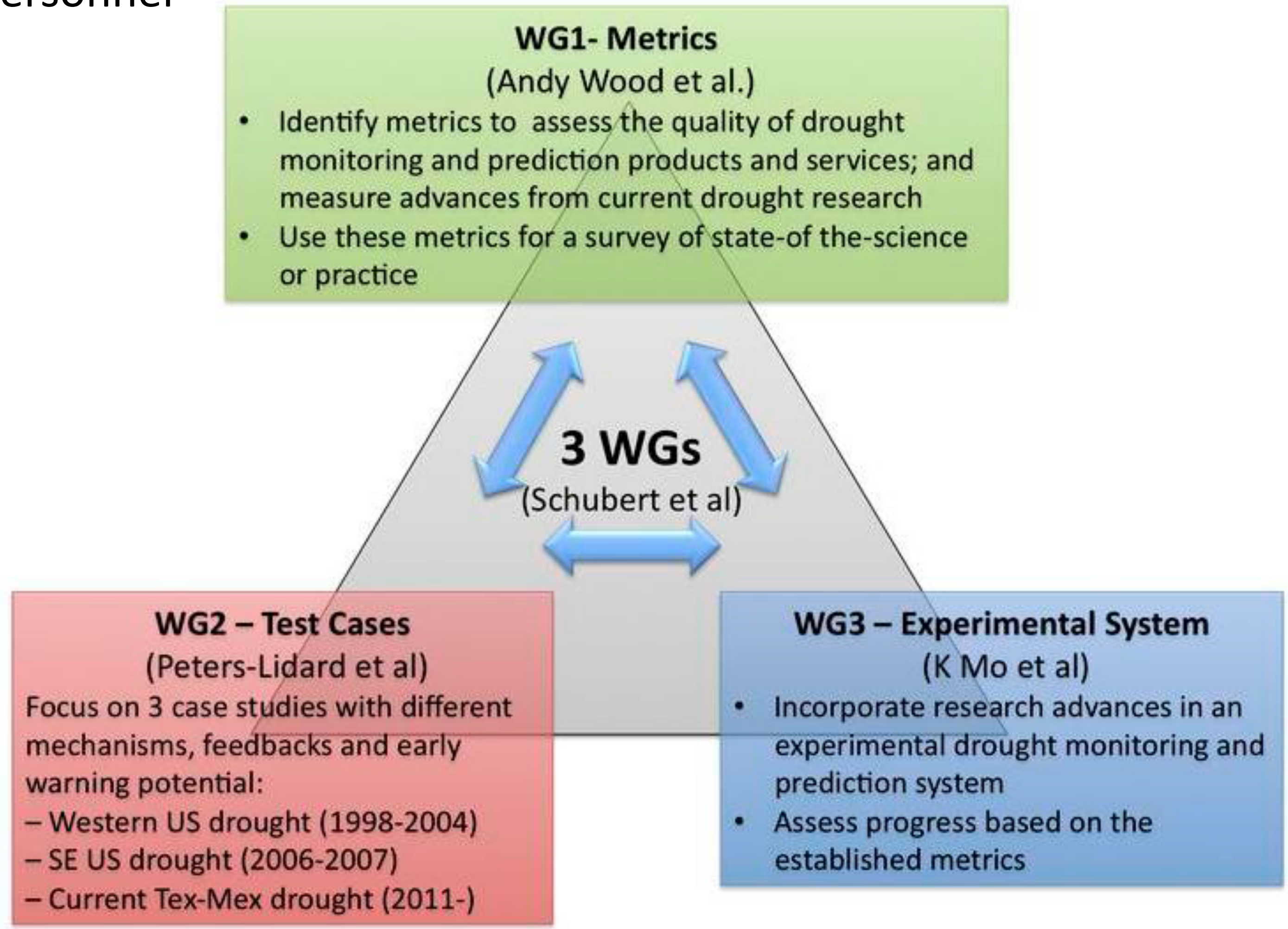
MAPP Drought Task Force Overview

Key Activities

- Communication/coordination:
  - Monthly telecons
    - *Narrative, Research-to-Capability, and Science sessions*
  - DTF Wiki page (internal) and NIDIS-DTF Web Page
    http://www.drought.gov/drought/content/resources-research/noaa-drought-task-force
  - Links to the National Multi-Model Ensemble (NMME) effort
- Organizing a special collection in JHM on "Advancing Drought Monitoring and Prediction "
  - About 20 contributed papers plus 3 overview/synthesis papers
- Developed DTF group "infrastructure" – organizes research into a de facto *testbed*

MAPP Drought Task Force Overview

- Group structure led by Annarita M. (NOAA) and Siegfried S. (NASA)
- Supported by efforts of 35+ researchers and partner organization personnel



Framework for Drought Science Evaluation

Working Group 1 – Metrics

- The Drought Task force is centrally focused on geophysical science supporting drought management
- Most variables of interest are physical, but index/category variables are included in focus

Drought Variables of Interest	Role(s)
Streamflow	Hydrological Drought
Precipitation	Meteorological Drought
Soil Moisture	Agricultural Drought
Snow Water Equivalent	Indirect / Multiple
Temperature	Indirect / Multiple
ET/PE	Indirect / Multiple
DM/SDO Category	Decision Support

Framework for Drought Science Evaluation

Central Ideas

- ❑ measure performance specific to drought phenomena & states
- ❑ explicit inclusion of benchmark capability or science

Also, where possible

- ❑ assess decision-relevant features – *eg, onset, recovery*

framework components	description
new capability	prediction or monitoring capability to be evaluated
benchmark capability	existing operational or research capability of interest. Should be stringent, including known 'easy' sources of predictability
verifying observation(s)	typically a measurement; if lacking, creativity required
analysis dataset(s)	e.g., hindcasts or observational archives that are relevant to the real-time prediction
supporting research	document literature or other resources that contain relevant, related analyses

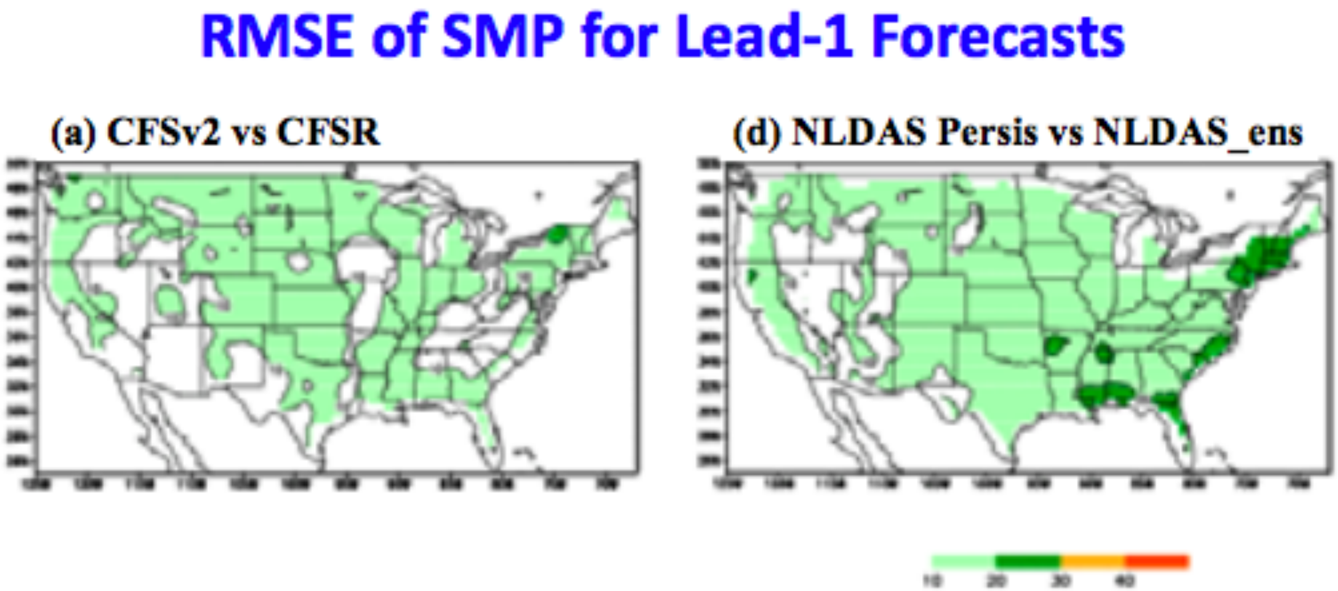
Framework for Drought Science Evaluation

- emphasize skill score framework, *eg*……….
- should indicate degree of improvement over existing science or capability
- require drought condition thresholds
- highlight features important for decisionmakers / applications

$$SS = 1 - \frac{Score(forecast)}{Score(reference)}$$

key predictand(s) – for a variable of interest:	**sample metric(s) – skill scores comparing:**
onset and recovery of drought condition duration and severity of drought condition optional: sev-dur, extent, SAD, etc.	average lead time average error (time, value) by lead
value, overall value given drought in obs or fcst period avg, max severity of drought condition	error, bias, correlation by lead (deter.) …same for ens. medians Continuous RPSS (prob.)
value indicating drought condition (probabilistic)	Brier skill score (binary) by lead Brier decompositions for reliability and resolution -- by lead
value indicating drought condition (deterministic)	CSI or ETS [POD, FAR, ROC, etc.]

Example – Overall metric

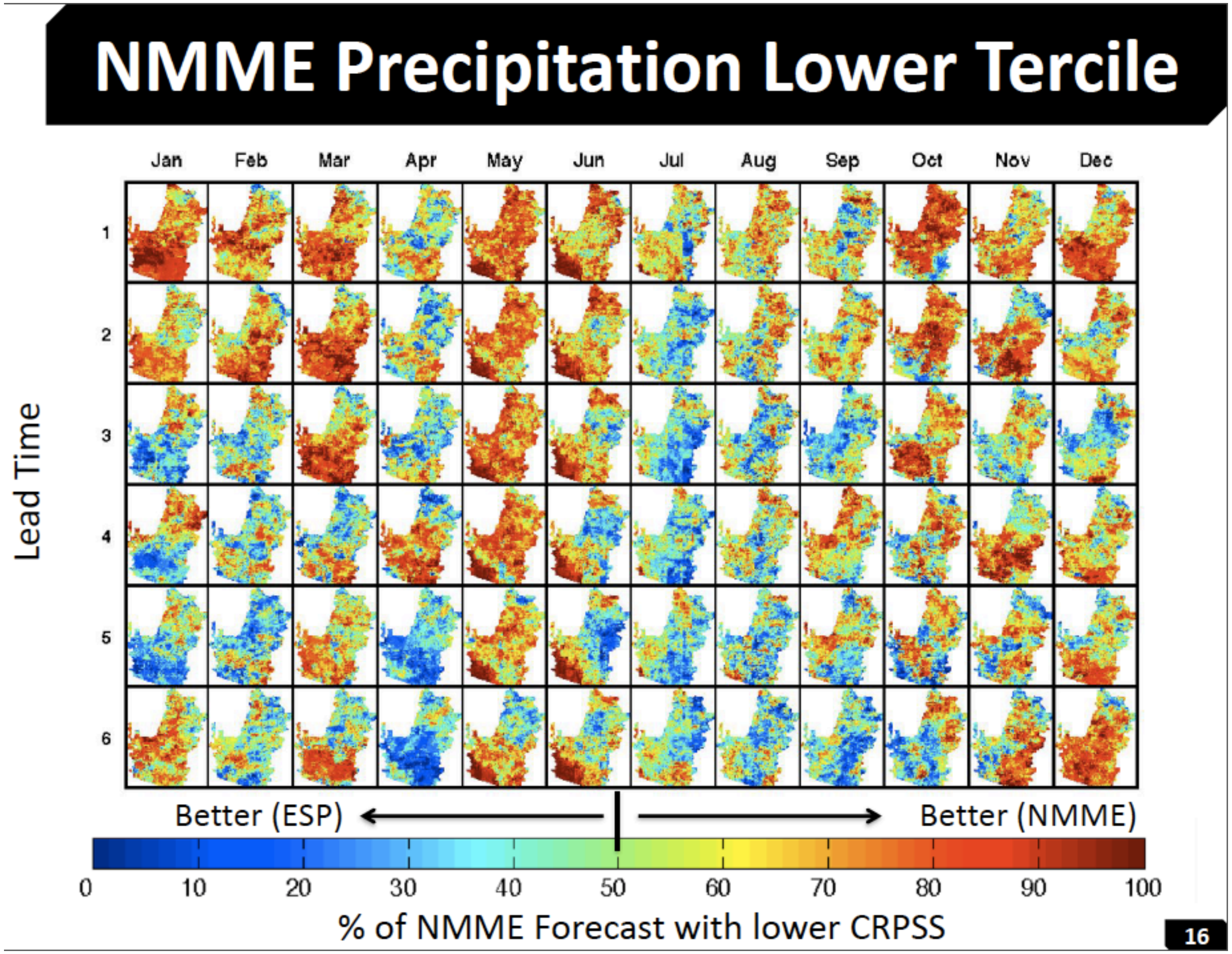


Chen & Mo (poster): **Seasonal Drought Prediction over the United States**

- a useful overall analysis – includes baselines – *but*:
- skewed/heteroskedasticvariable errors (runoff, snow, precipitation, flow) can be dominated by non-drought errors
- even for SM (perhaps beta distributed), non-drought behavior is likely the largest contribution to sample

Example – More drought-specific metric

- ❑ analyses focused on parts of the spectrum give insight
  - ❑ dry end – onset; wet end – recovery; seasonality important

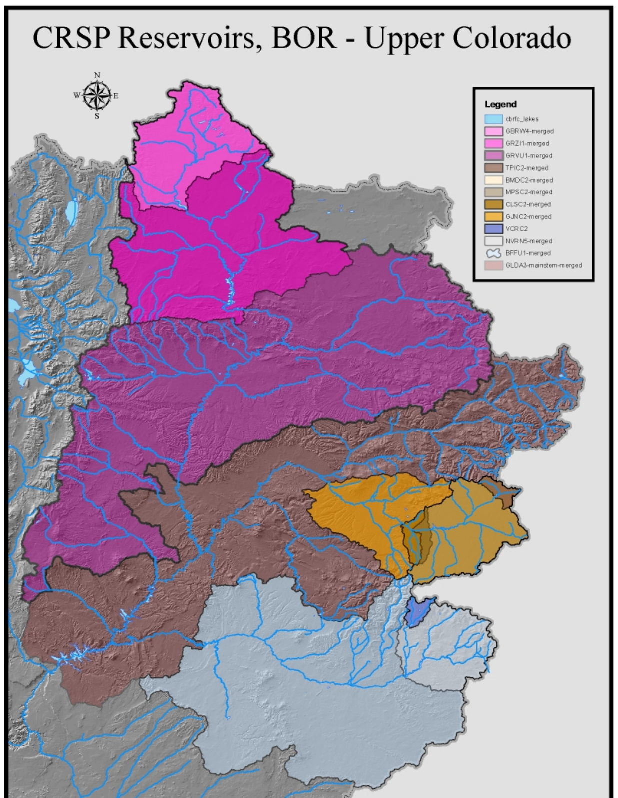


Wood (E) et al, CDPW 2012

Application Example

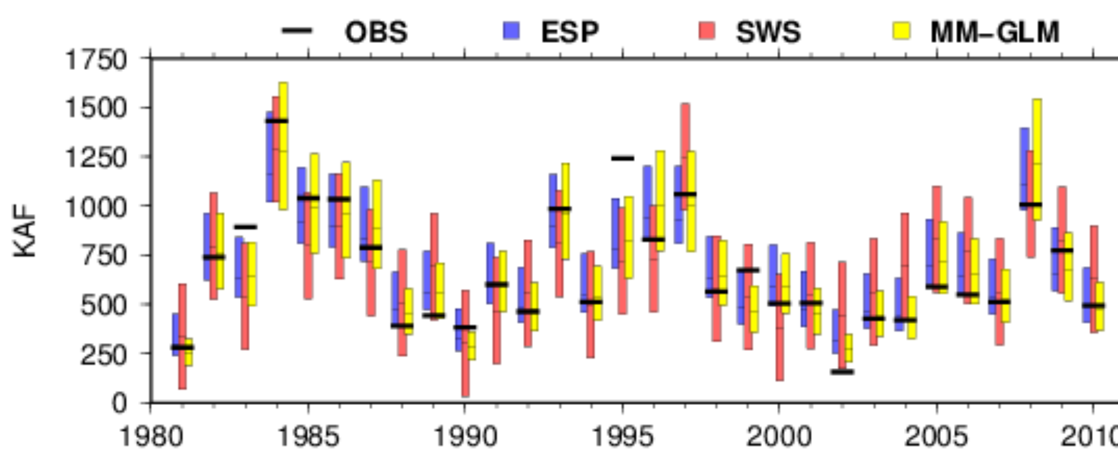
Western US Seasonal Water Supply Forecast (WSF)

- In Colorado R. basin, WSF (runoff volume predictions) can be a trigger for drought declaration, affecting 7 SWUS states
- NOAA/NWS makes 2 types of volume predictions in Jan – July period
  - ESP – model-based ensemble forecasts
  - SWS – statistical probabilistic predictions
- Subjectively merges them into a single 'NWS preferred' forecast



MAPP research question
Can these be objectively merged into a better WSF?

April 1 WSF for Gunnison R abv Blue Mesa Reservoir, CO



http://www.cbrfc.noaa.gov/testbeds/si_y2/

Application Example

Multi-model Water Supply Forecast (WSF) research

- collaboration with James McCreigh and Balaji Rajagopalan (CU)
- evaluated various approaches for merging ESP & SWS
  - *different GLMs, simple mean, Bayesian Model Avg (BMA)*
  - *focus here on 8 key locations for water supply management*

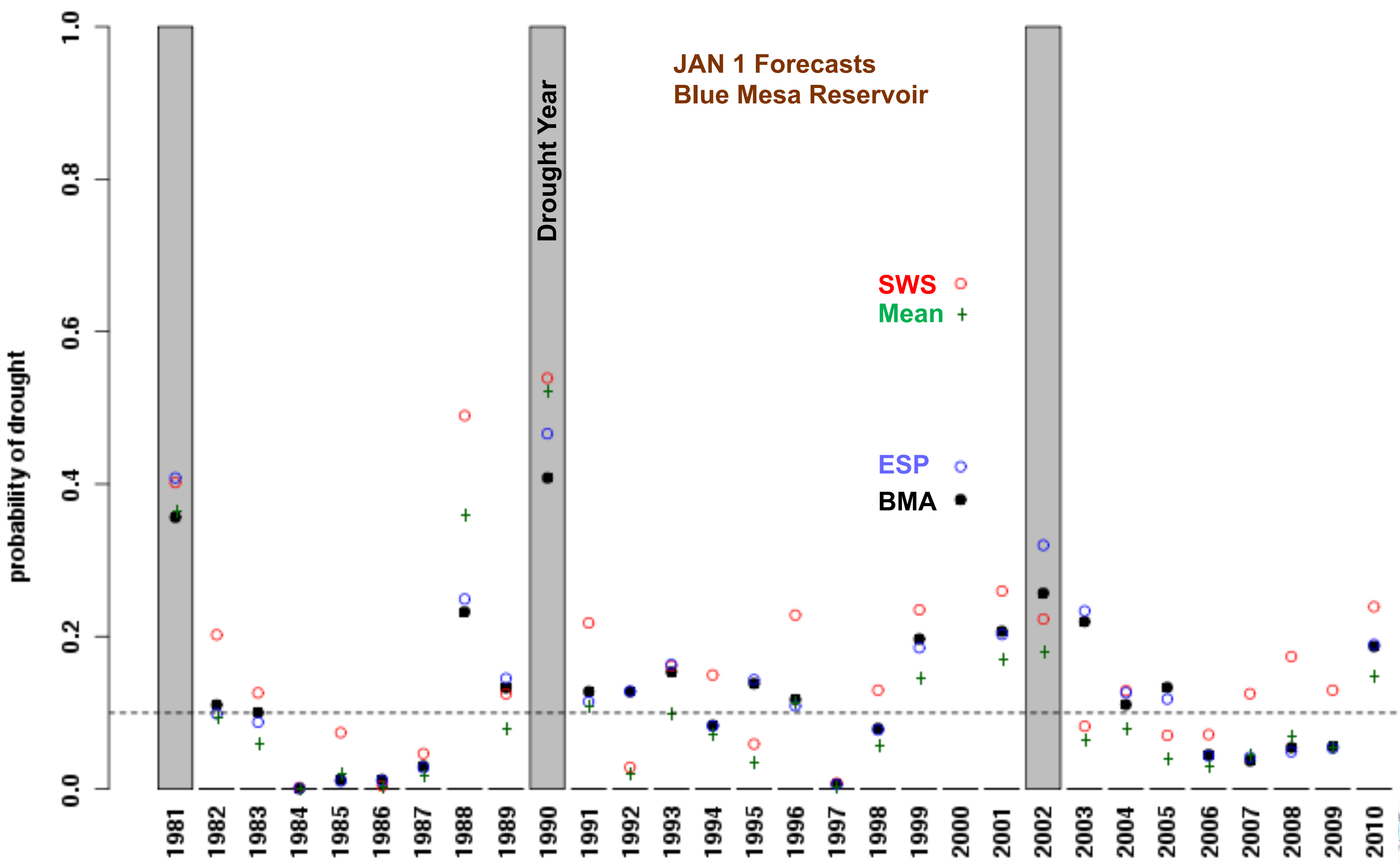
framework components	description
new capability	Objective Multi-model Water Supply Forecast
benchmark capability	**Baselines**: ESP, SWS, Simple Mean
verifying observation(s)	Observed naturalized runoff
analysis dataset(s)	Hindcasts of ESP and SWS for 30 years, 1981-2010 Cover 150-200 locations in the upper Colorado R. basin
supporting research	ample literature on WSF forecasting, less on streamflow forecast combination (eg Ajami et al)

key predictand(s) – for WSF	sample metric(s) – skill scores comparing:
value, overall	WSF median correlation by lead (deter.)
value indicating drought condition (probabilistic)	Brier skill score (binary) by lead

Application Example

define drought as 10th percentile volume runoff

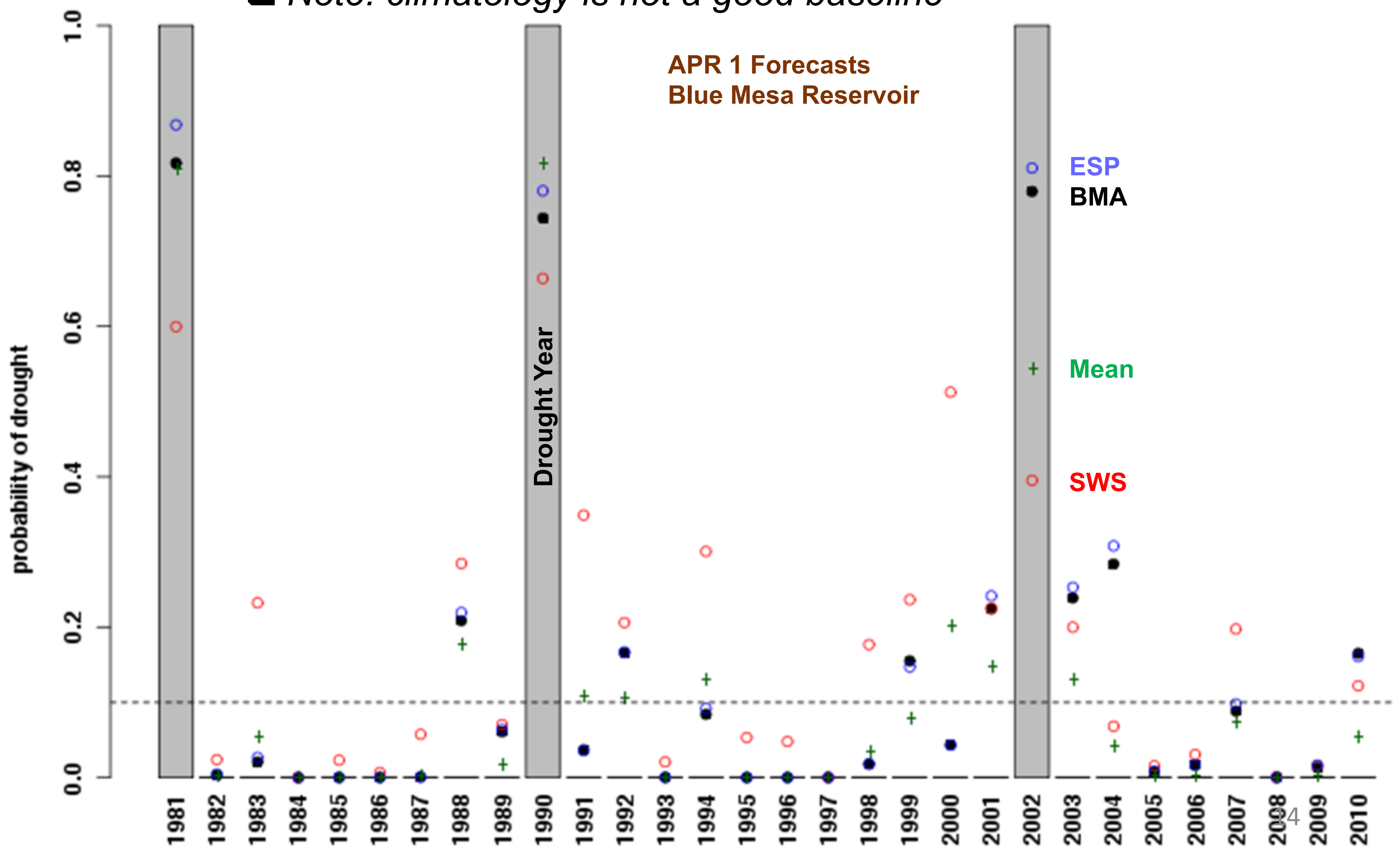
- plot shows past forecasts of probabilities of drought
- *four types shown (baselines: **ESP**, **SWS**, **Mean**; new = **BMA**)*



Application Example

define drought as 10th percentile volume runoff

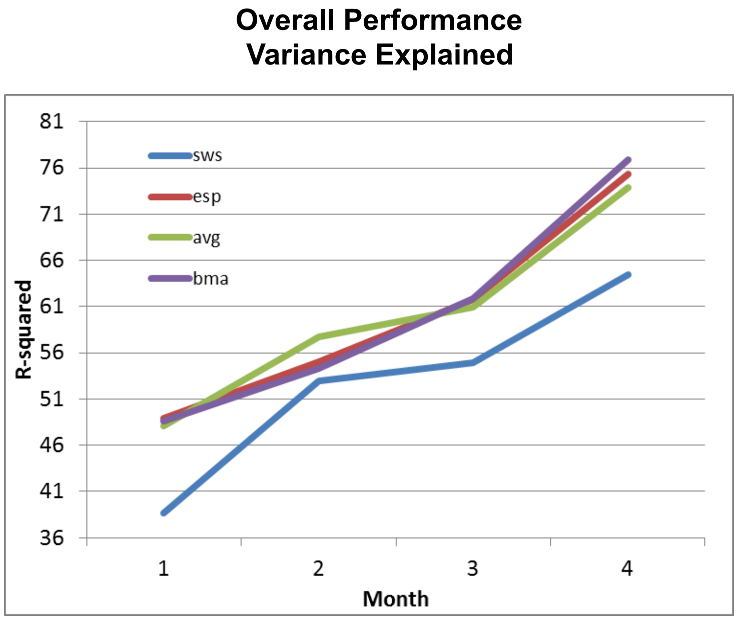
- plot shows past forecasts of probabilities of drought
- *four types shown (baselines: **ESP**, **SWS**, **Mean**; new = **BMA**)*
  - *Note: climatology is not a good baseline*

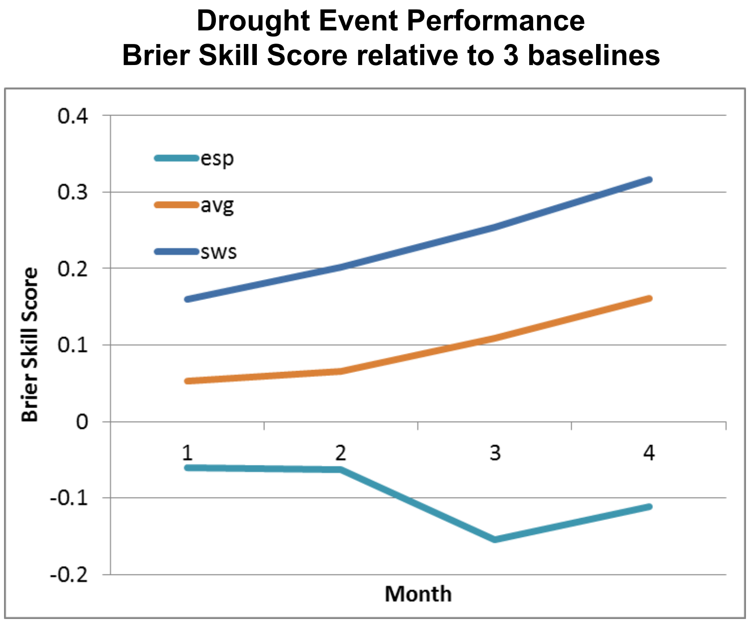


Application Example

Western US Seasonal Water Supply Forecast (WSF)

- Pooling 8 key upper Colorado locations…
- Overall, by these metrics, the BMA combination slightly improves upon the baseline forecasts in spring (March, April)
- From a *drought category* perspective, the BMA improves upon SWS and the simple mean, but *is worse than ESP in all months*





Applications – assessing our monitoring

Verification scores are straightforward for prediction… what about drought monitoring?

- There is a long standing concern about a lack of "observations" to verify 'drought monitoring'
  - e.g., can we accurately discriminate D1 from D3?

- The DTF focus is mostly on assessing our geophysical science, e.g.,
  - land surface modeling, climate prediction, climate simulation, remote sensing
  - such monitoring tools and data are verified against physical measurements in the traditional way
  - the framework goals apply

- For non-measured drought designations (eg D4), use of USDM as 'obs' has gained acceptance (note three previous talks)
  - it is a synthesis of many observational datasets, even if subjective
  - the Drought Impact Reporter (http://droughtreporter.unl.edu/) is also helpful for verification

Summary and Future Directions

- The NOAA DTF is encouraging a more standard framework for evaluating drought prediction and monitoring capabilities and science
  - *phenomenological* – focus on drought states and features in addition to overall performance
  - measured against current benchmarks or operational baselines to show progress
  - recognizing drought science applications

- Future Directions
  - progressing on assessment using major US drought case studies (eg SE, Colorado, Tex-Mex/Midwest)
  - continued interactions – e.g., Research to Capability telecon
  - complete special collection of papers
  - working toward enhanced DEWS within NIDIS

Questions?

andy.wood@noaa.gov

Example – SPI Prediction metric: correlation

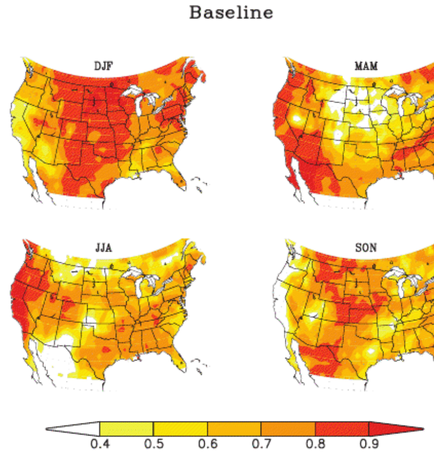


FIG. 1. The "baseline" skill of SPI6 in which random forecasts of target-season precipitation using observed historical values are combined with prior observed seasonal precipitation. Shown are the 0-lead seasonal forecasts for the target seasons of boreal (top left) winter (DJF), (top right) spring (MAM), (bottom left) summer (JJA), and (bottom right) autumn (SON). The skills are the average of 100 members of blended SPI6 calculated on the basis of randomizing the time sequence of monthly mean precipitation from GPCC for the period 1982–2008.

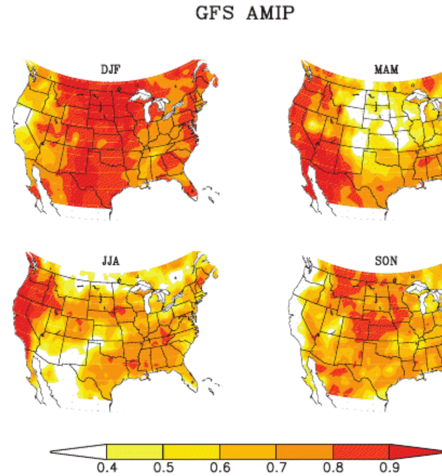


FIG. 3. Skill of the AMIP SPI6 in which AMIP simulations are used for 0-lead seasonal forecasts for the boreal (top left) winter (DJF), (top right) spring (MAM), (bottom left) summer (JJA), and (bottom right) autumn (SON). The skills are the average of a total of 10 members of blended SPI6 calculated on the basis of AMIP simulations using [illegible] 2008.

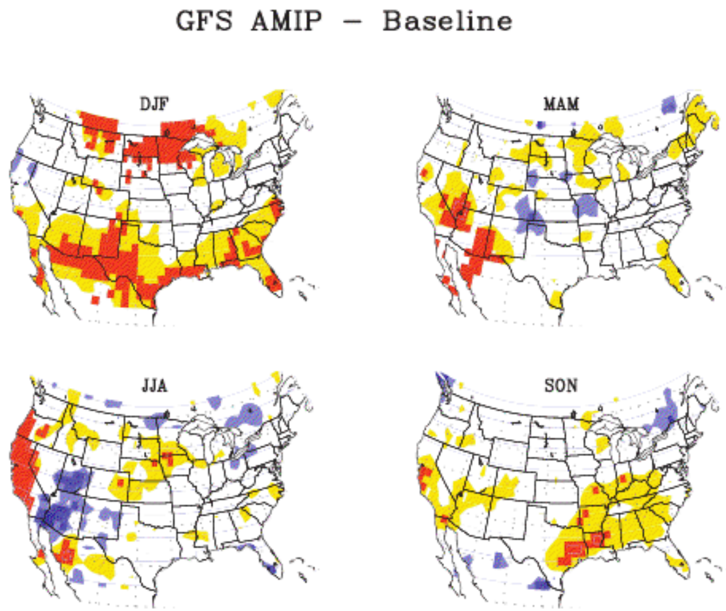
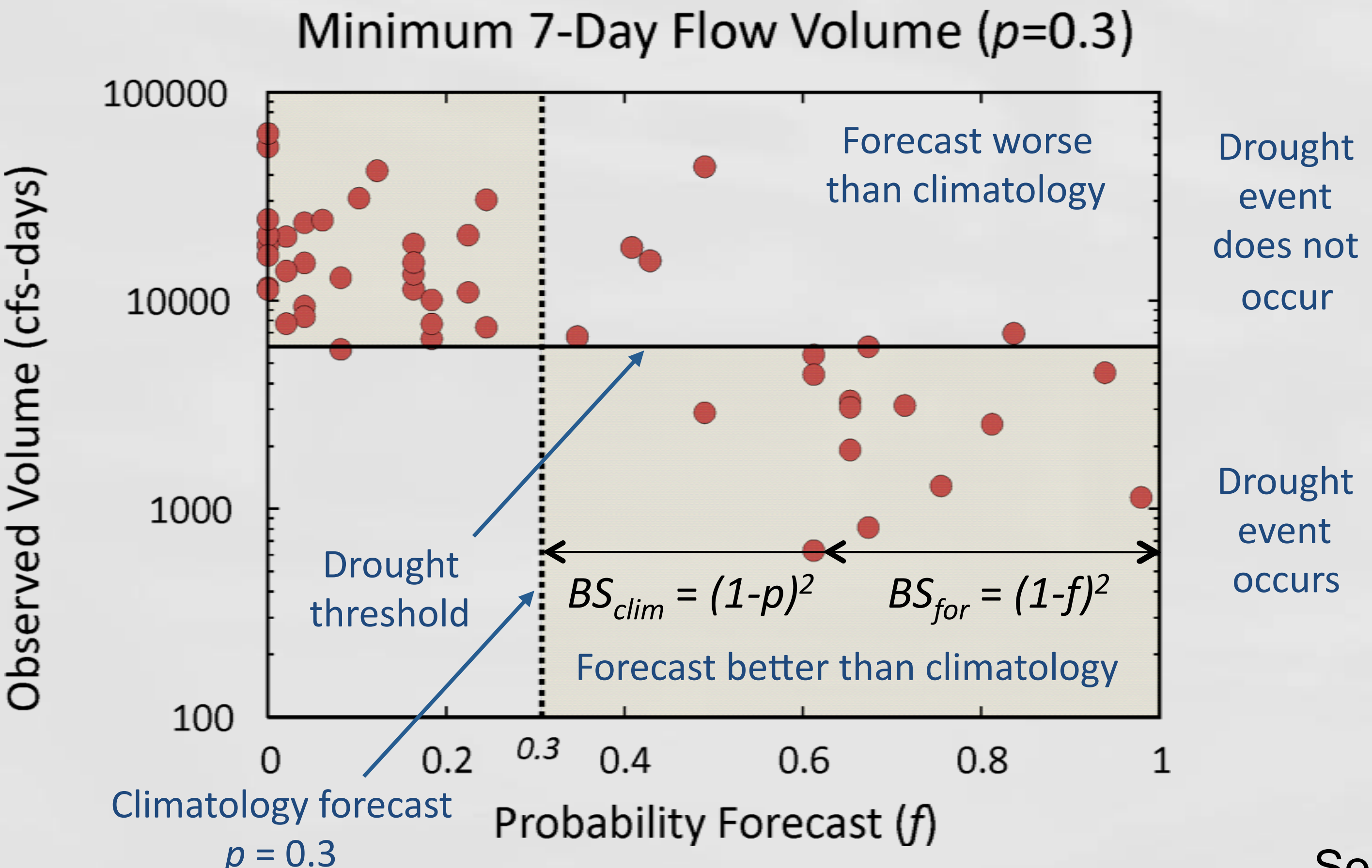


FIG. 4. Difference between the GFS AMIP SPI6 skill and the baseline SPI6 skill. Blue shading indicates values that are less than the baseline skill, yellow indicates values that are higher than the baseline skill, and red indicates values that are higher than the baseline skill at 90% significance level.

Source: Hoerling et al., JAMC 2012

Drought Event Forecast Sample

All forecasts from a verification data set



Skill Score

$$SS = 1 - \frac{\overline{BS_{for}}}{\overline{BS_{clim}}} = 0.578$$

Source: Allen Bradley