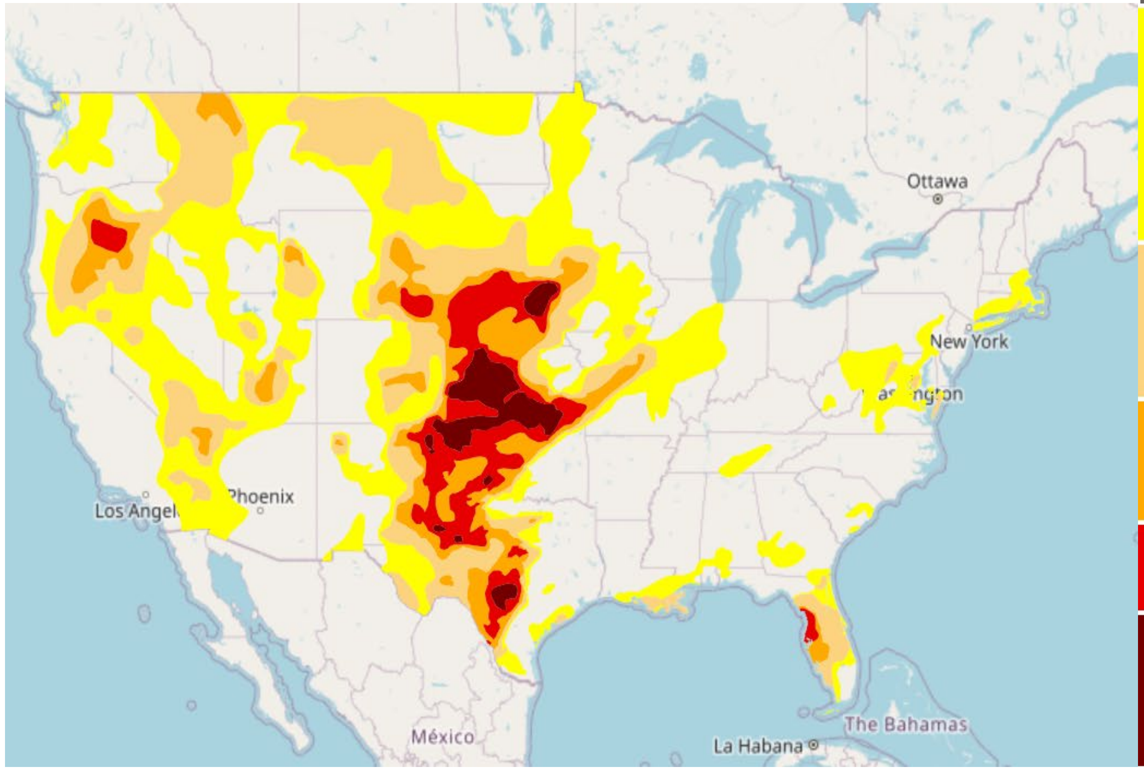
Use of Machine Learned Mutual Information Between USDM and Correlated Drought Factors

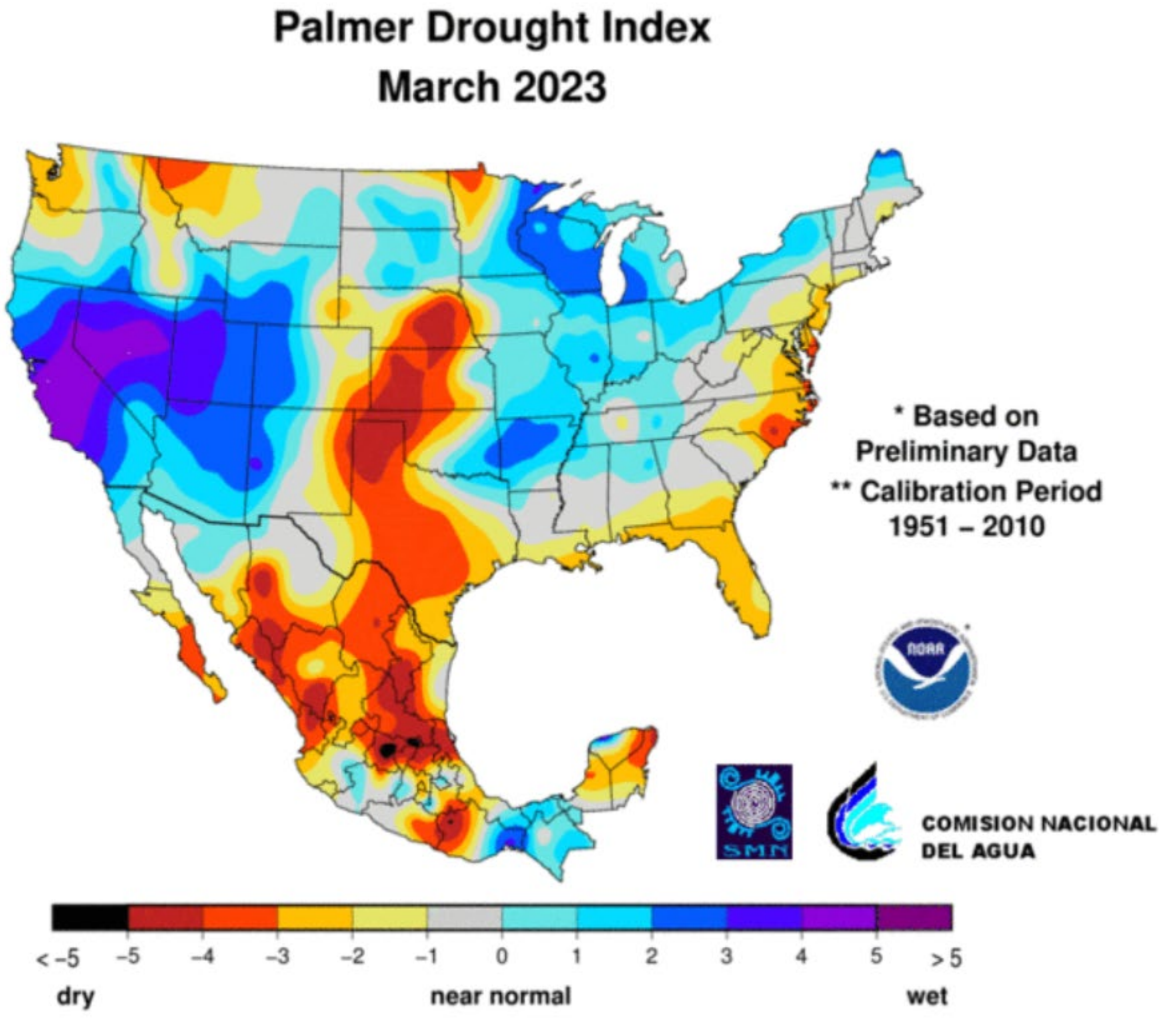
Michael Shaw, NIDIS, iSciences, michael.shaw@noaa.gov
Steve Ansari, NIDIS, steve.Ansari@noaa.gov
Soni Yatheendradas, NASA, U. Md., soni.yatheendradas-1@nasa.gov
David Mocko, NASA, SAIC, david.mocko@nasa.gov
Justin Fain, NIDIS, iSciences, justin.fain@noaa.gov

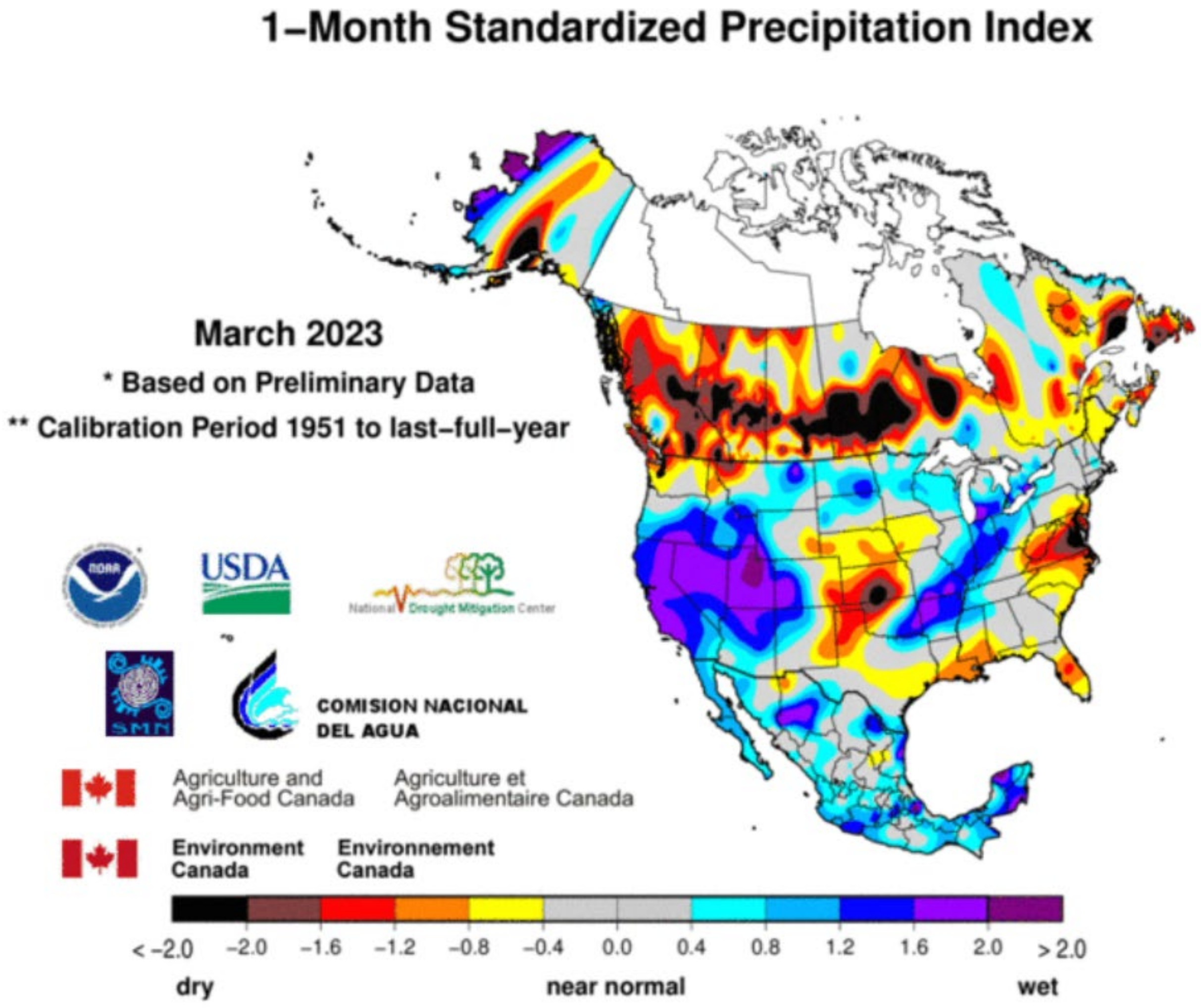
USDM … an authoritative depiction of drought conditions



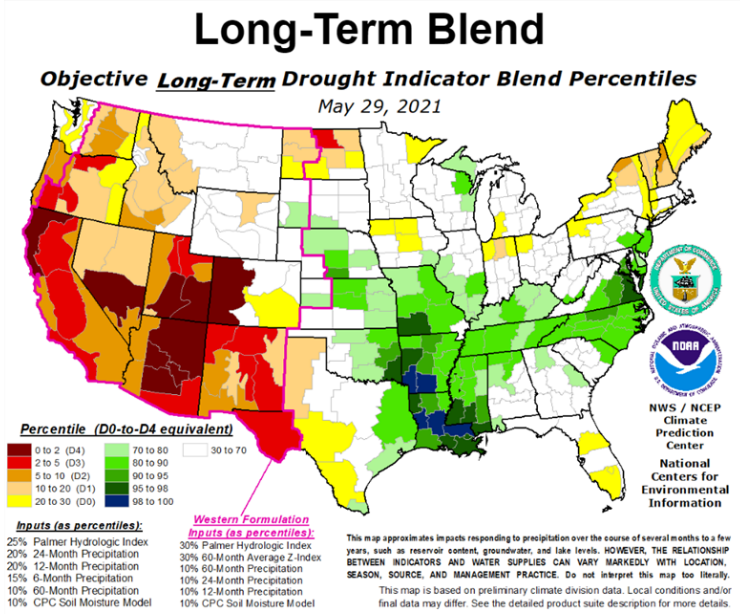
Category	Description	Possible Impacts	Ranges: Palmer Drought Severity Index (PDSI)	Ranges: CPC Soil Moisture Model (Percentiles)	Ranges: USGS Weekly Streamflow (Percentiles)	Ranges: Standardized Precipitation Index (SPI)	Ranges: Objective Drought Indicator Blends (Percentiles)
D0	Abnormally Dry	Going into drought: • short-term dryness slowing planting, growth of crops or pastures Coming out of drought: • some lingering water deficits • pastures or crops not fully recovered	-1.0 to -1.9	21 to 30	21 to 30	-0.5 to -0.7	21 to 30
D1	Moderate Drought	• Some damage to crops, pastures • Streams, reservoirs, or wells low, some water shortages developing or imminent • Voluntary water-use restrictions requested	-2.0 to -2.9	11 to 20	11 to 20	-0.8 to -1.2	11 to 20
D2	Severe Drought	• Crop or pasture losses likely • Water shortages common • Water restrictions imposed	-3.0 to -3.9	6 to 10	6 to 10	-1.3 to -1.5	6 to 10
D3	Extreme Drought	• Major crop/pasture losses • Widespread water shortages or restrictions	-4.0 to -4.9	3 to 5	3 to 5	-1.6 to -1.9	3 to 5
D4	Exceptional Drought	• Exceptional and widespread crop/pasture losses • Shortages of water in reservoirs, streams, and wells creating water emergencies	-5.0 or less	0 to 2	0 to 2	-2.0 or less	0 to 2

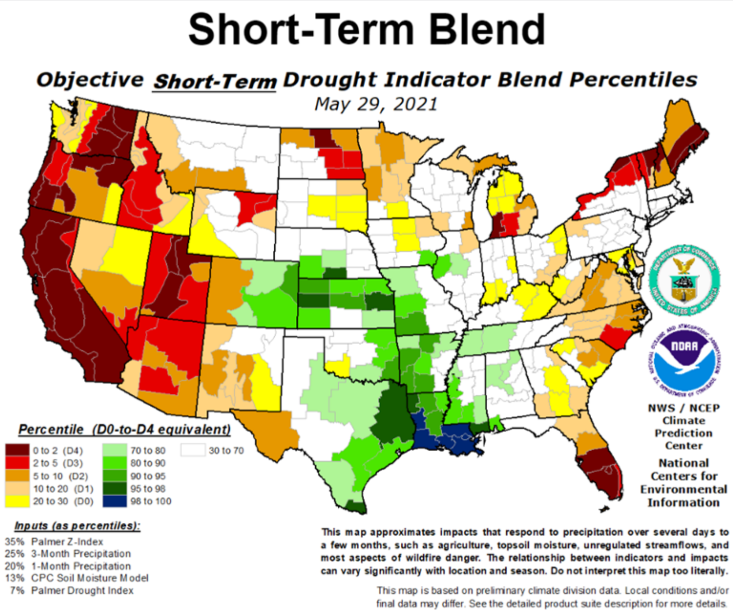
Drought indices to do with factors shaping USDM reported conditions





CPC experimental blends of drought indicators…





The ***Short-Term Blend*** approximates drought-related impacts that respond to precipitation (and secondarily other factors) on time scales ranging from a few days to a few months, such as **wildfire danger, non-irrigated agriculture, topsoil moisture, range and pasture conditions, and unregulated streamflows.**

The ***Long-Term Blend*** approximates drought-related impacts that respond to precipitation on time scales ranging from several months to a few years, such as **reservoir stores, irrigated agriculture, groundwater levels, and well water depth.**

It should be noted that the relationship between indicators and impacts varies, sometimes markedly, with location and season. This is particularly true of water supplies, which are additionally dependent on the source (or sources) tapped, management practices, and legal mandates. Exercise caution when attempting to relate these maps to specific impact implications for a particular location and time of year. The blend-to-impact correlation is not always direct, and will vary spatially and temporally.

The following bullets describe the composition of these experimental blends:

- These products are generated using the Climate Prediction Center's real-time daily & weekly climate division data, and the National Climatic Data Center's monthly climate division data archive, back to 1932.
- The indices used in the blends and their weights are as follows:
  - ***SHORT-TERM:*** 35% Palmer Z-Index; 25% 3-Month Precipitation; 20% 1-Month Precipitation; 13% Climate Prediction Center Soil Moisture Model; and 7% Palmer (Modified) Drought Index.
  - ***LONG-TERM:*** 25% Palmer Hydrologic Drought Index; 20% 12-Month Precipitation; 20% 24-Month Precipitation; 15% 6-Month Precipitation; 10% 60-Month Precipitation; 10% Climate Prediction Center Soil Moisture Model.
- All parameters are first rendered as percentiles with respect to 1932-2000 data using a percent rank method. Most parameters are ranked relative to the National Climatic Data Center's historic climate division data for the current month, except for the Z-Index which is rendered relative to all months on record (this introduces evaporative seasonality into the short-term blend).

So…

- **What**? **Mutual Information (MI) and related Fractional Information (FI) statistics to learn importance of dozens of drought indicators wrt 20 years of the U.S. Drought Monitor to multiple indicators** (NASA contract) with Climate Engine climatologies (Climate Engine [CE] contract using Google Earth Engine) on Google Cloud Platform ([GCP] Google contract).
- **Why**?
  - Provide guidance on dozens of indicators commonly used for drought monitoring
  - Provide lists of top indicators in state, county, watershed, and more
  - Produce new gridded blends – Multi Indicator Drought Indices (MIDI) with normalized FI weighting vs CPC weights

https://www.drought.gov/drought-research/quantifying-relative-importance-multiple-drought-indicators-us-drought-monitor

Soni Yatheendradas, David M. Mocko, Christa Peters-Lidard and Sujay Kumar, *Quantifying the Importance of Selected Drought Indicators for the United States Drought Monitor*, Submitted to JHM.

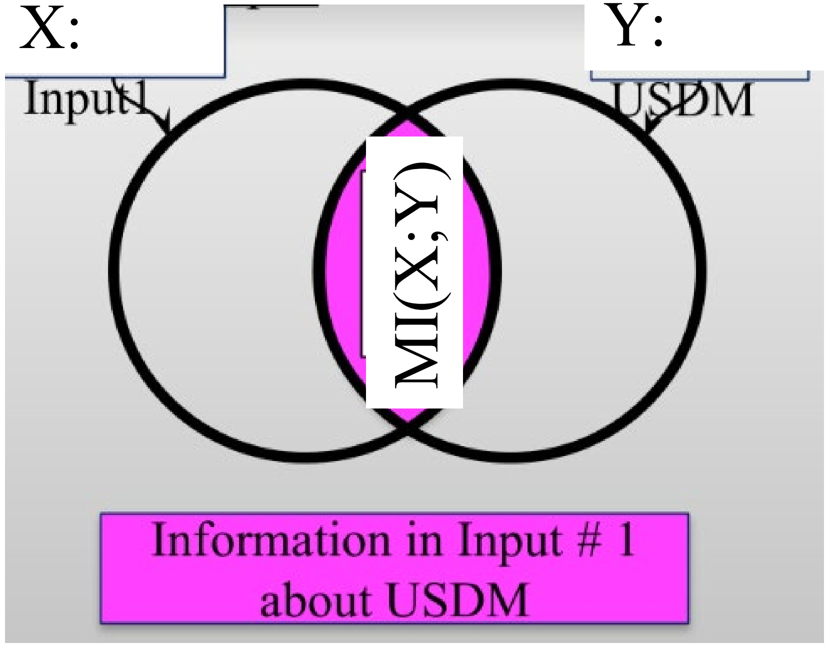
MI and FI

- Based only in data, avoiding model assumptions – “use only what assumptions known” a la maxent; mutual information, MI, divided by USDM info entropy, H ⍰ fractional information, FI
- Applicable across regions, scales, and associated properties (LULC, soils, etc)
- Provides for indicator and variable importances with respect to a particular quantity of interest (USDM category)

$$MI(X;Y) = H(X) - H(X|Y) = -\sum_n \sum_z P(x_n, y_z) \log(P(x_n)/P(x_n|y_z))$$

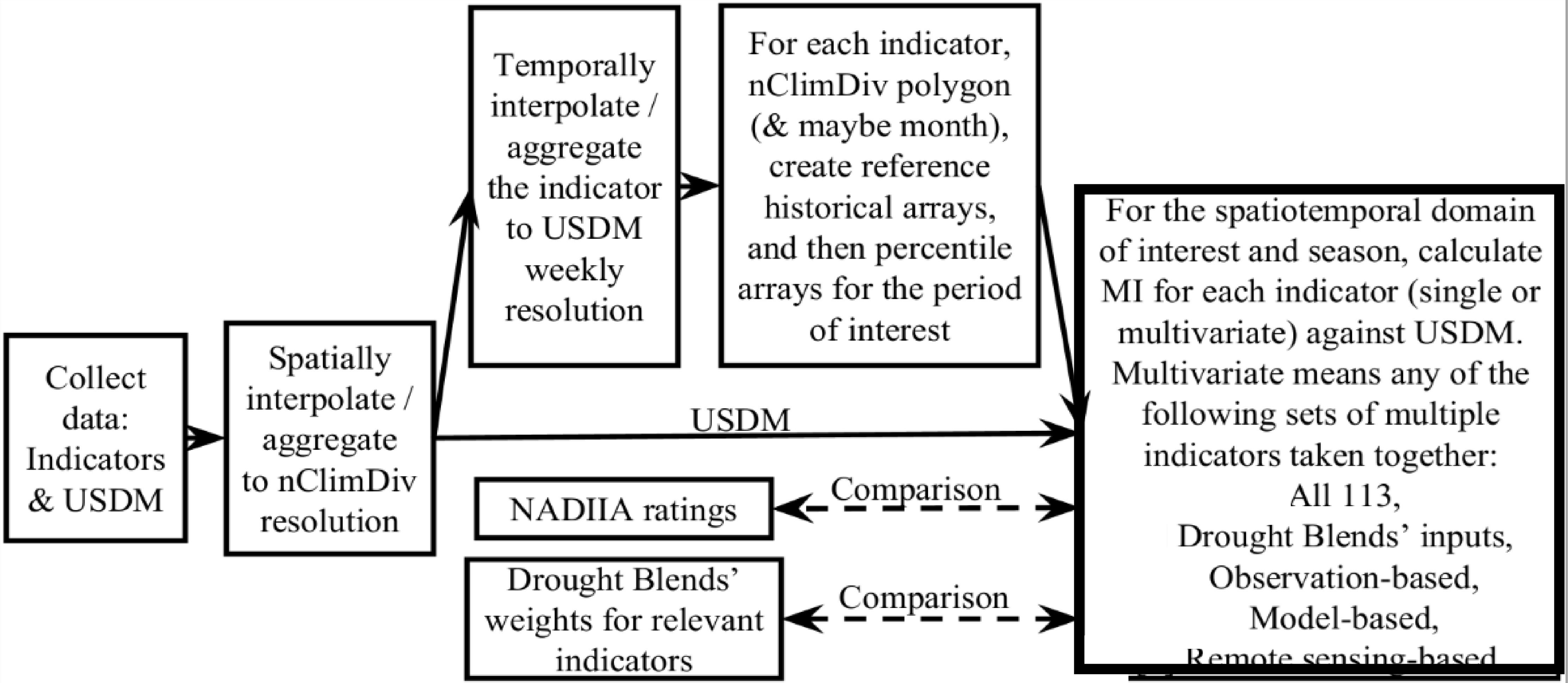
$$H(X) = -\sum_{n=1}^{N} P(x_n) \log P(x_n)$$

$$H(X|Y) = -\sum_n \sum_z P(x_n, y_z) \log P(x_n|y_z)$$



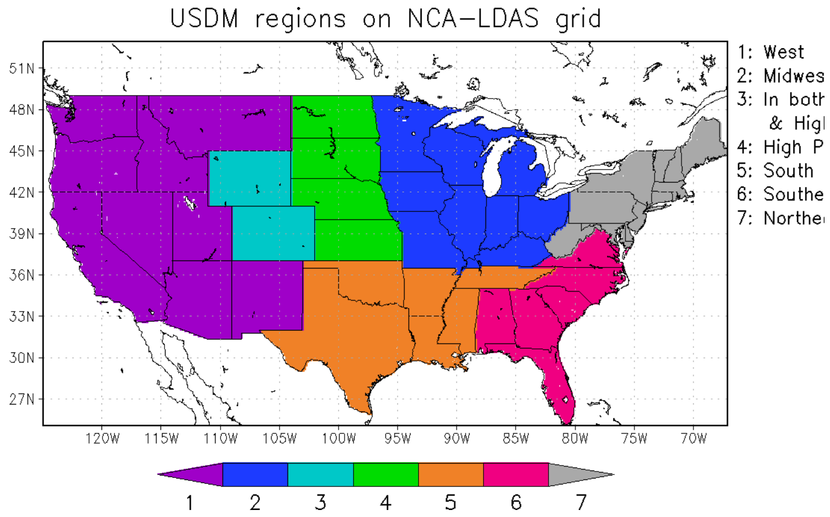
- But, computationally expensive

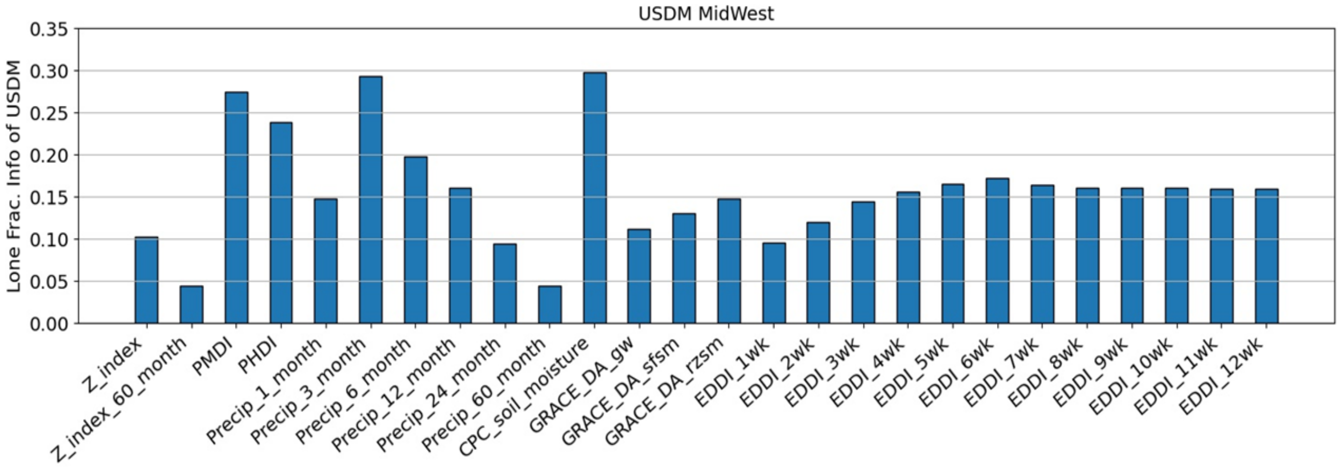
FI data flow



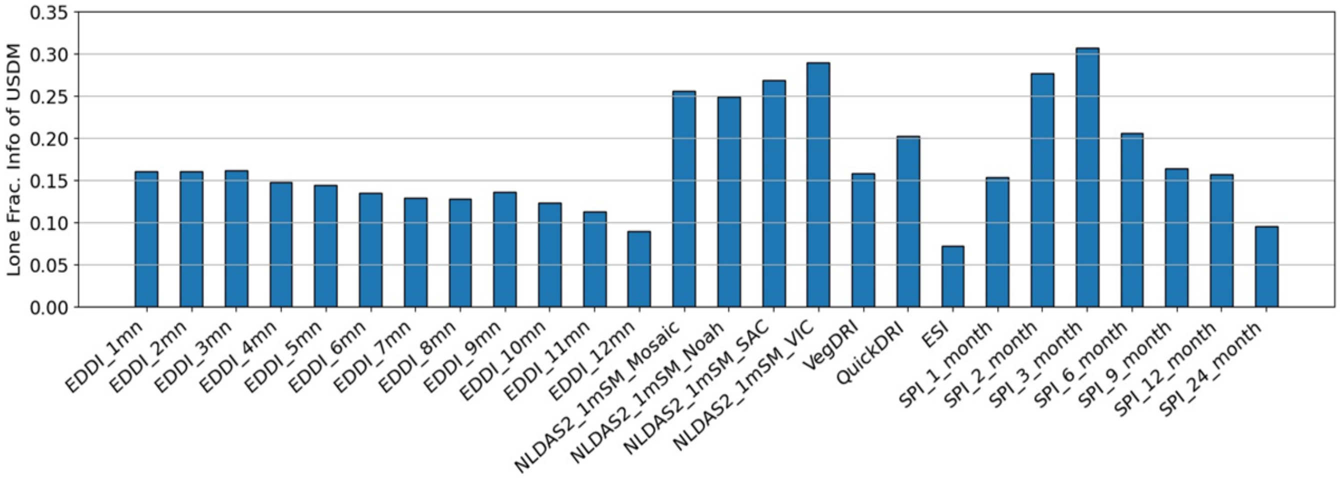
Collect data: Indicators & USDM
Spatially interpolate / aggregate to nClimDiv resolution
Temporally interpolate / aggregate the indicator to USDM weekly resolution
For each indicator, nClimDiv polygon (& maybe month), create reference historical arrays, and then percentile arrays for the period of interest
USDM
For the spatiotemporal domain of interest and season, calculate MI for each indicator (single or multivariate) against USDM. Multivariate means any of the following sets of multiple indicators taken together:
All 113,
Drought Blends' inputs,
Observation-based,
Model-based,
NADIIA ratings
Comparison
Drought Blends' weights for relevant indicators
Comparison

Example of indices' FI relative to USDM category
USDA Midwest Region





- Indices not used in CPC or NDMC blends emerge as important
- Different importance or "weight" emerges
- Precip products FI highest 2/3-month
- EDDI also highest 1/3-month
- High FI for top 1-m soil moisture
- High FI for Palmers
- VegDRI/QuickDRI moderately high FI

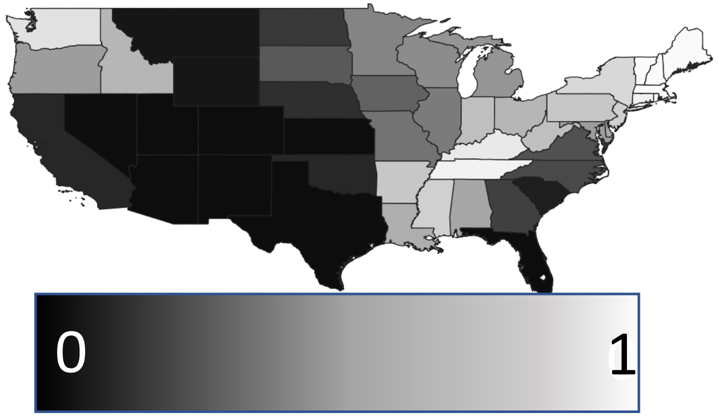


Blend components

For each region, need these **<u>FIs</u>** and **<u>percentiles</u>** (a la CPC short and long term blends' "linear model")

- Full (all season annual)
- Seasonal
- Flash

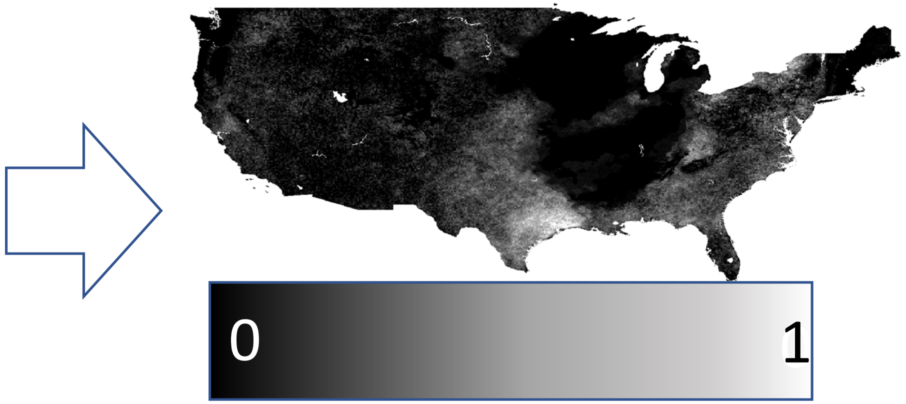
Aggegated blend component



1 week EDDI percentile

Normalized FI

Weighted blend component

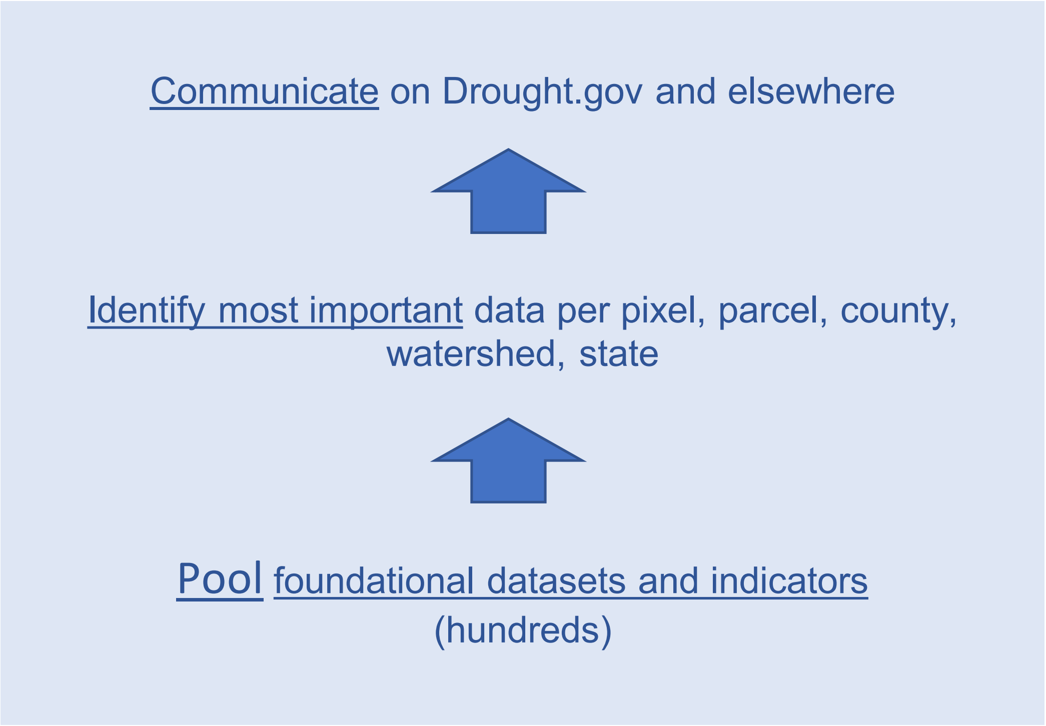


Status

- Recently upped tally to 58 FI files (eventually 113 indicators, some seasonal) from NASA, with more coming – time/resource intensive; logistics wrt NASA supercomputing cluster upgrades to be completed by June (future "in house" GCP processing may help wrt additional indices)
- Recently received and reviewing/adapting code to compute FI in house (GCP)
- Developed code to fetch/compute index percentiles (CE) & generate linear model blends, zonally aggregate (state, USDM region, watershed, e.g.)
- Adding more indices/variables in CE for percentile calculations with collaborators

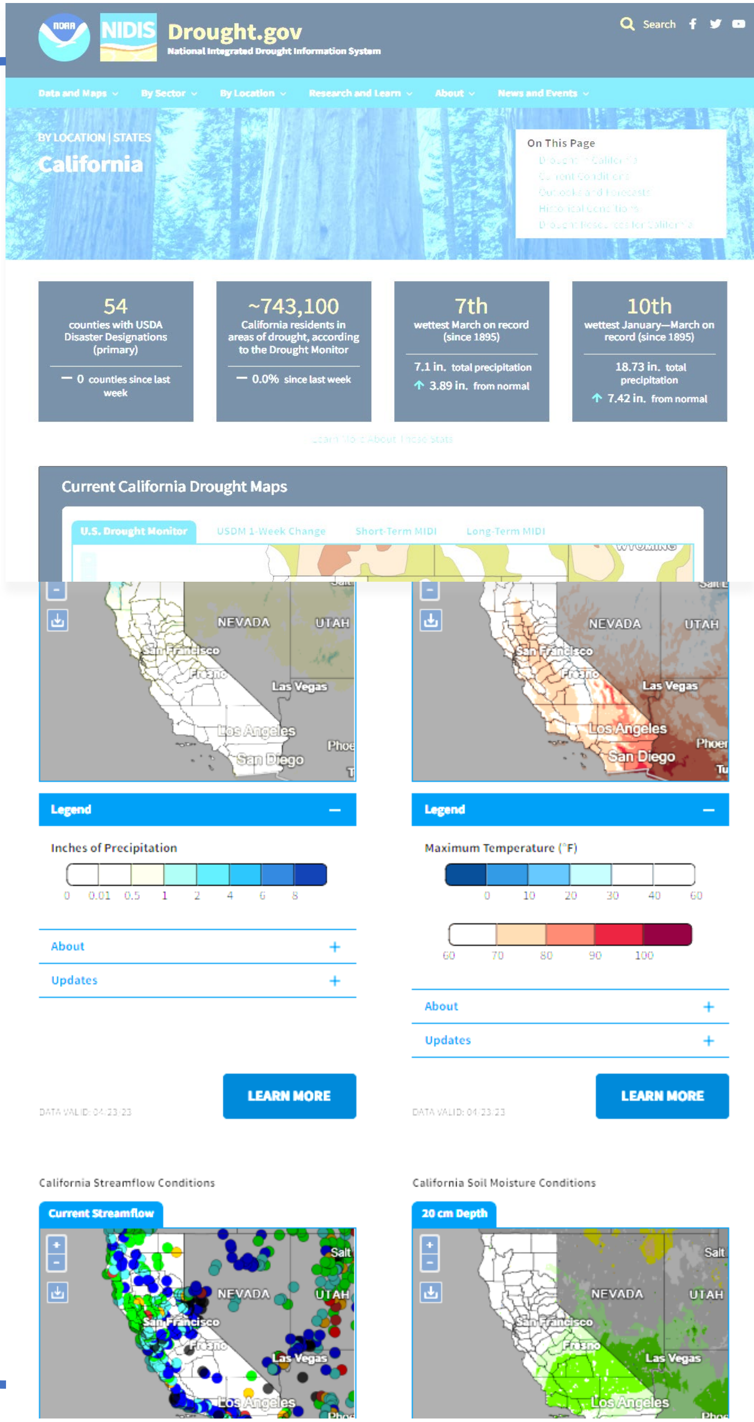
Next steps

- **Complete migration to cloud**.
- Begin conducting **in house calculations, complete FI generation**
- Build out **more climate engine index percentiles** – operational production; 360+ already
- Survey flash drought trends; consistent and inconsistent index FI to gauge utility of more generic **flash drought FI generation and blends**
- **Engage with stakeholders** and page adaptation (automation of seasonal index presentation)
- **Further complementary stochastics**. Random forest? Examination of related entropic considerations relative to hydrologic topology and associated "flow partitioning" (runoff, streamflow, energy fluxes e.g. ET) and expectations concerning what maximal mutual information of which optimal indices/variables might be expected to perform at what scales/locations ?



MI, CE, and GCP being invoked for:

- Mechanistic **understanding**
- **MIDIs** (Multi Indicator Drought Indices)
  - Forecasting ("linear model" from forecast variable –based indices)
  - Monitoring ("linear model" from remote sensing assets'…)
- **drought.gov indices to present when, where** (state, county, watershed pages: short term versus long term – **regional stakes matter**…)
- …**Stakeholder engagement** and index priority reflected to some extent implicitly vis a vis DM author localized priorities' use in MI/FI calculation



Thank you!

Michael Shaw, NIDIS, iSciences, michael.shaw@noaa.gov
Steve Ansari, NIDIS, steve.ansari@noaa.gov
Soni Yatheendradas, NASA, soni.yatheendradas-1@nasa.gov
David Mocko, NASA, david.mocko@nasa.gov
Justin Fain, NIDIS, iSciences, justin.fain@noaa.gov

Notables

- Appears insensitive to SPI and SPEI PDF
- Unclear what interpolation and aggregation sensitivity is
- As compared to NADIIA, some similarities, some divergences

Existing Blends

CPC Short-Term	NDMC Short-Term	CPC Long-Term Western	NDMC Long-Term Western	CPC Long-Term non-Western	NDMC Long-Term non-Western
Palmer Z-index (35%)	1-month SPEI (35%)	PHDI (30%)	9-month SPEI (30%)	PHDI (25%)	9-month SPEI (25%)
3-month nClimDiv precipitation (25%)	3-month SPI (25%)	60-month Z-index (30%)	60-month SPEI (30%)	24-month nClimDiv precipitation (20%)	24-month SPI (20%)
1-month nClimDiv precipitation (20%)	1-month SPI (20%)	60-month nClimDiv precipitation (10%)	60-month SPI (10%)	12-month nClimDiv precipitation (20%)	12-month SPI (20%)
CPC soil moisture (13%)	Noah 0-100cm soil moisture (13%)	24-month nClimDiv precipitation (10%)	24-month SPI (10%)	6-month nClimDiv precipitation (15%)	6-month SPI (15%)
PMDI (7%)	9-month SPI (7%)	12-month nClimDiv precipitation (10%)	12-month SPI (10%)	60-month nClimDiv precipitation (10%)	60-month SPI (10%)
		CPC soil moisture (10%)	Noah 0-200cm soil moisture (10%)	CPC soil moisture (10%)	Noah 0-200cm soil moisture (10%)

Sample drought event top FI indicators