CW3E Hydrologic Modeling/Forecasting Efforts: From Process-Based to Data-Driven

Ming Pan, CW3E
(CW3E Hydrology/S2S/Weather Teams and Partners)

Session: Improving Inflow Forecasts - Progress on Hydrologic Forecasting

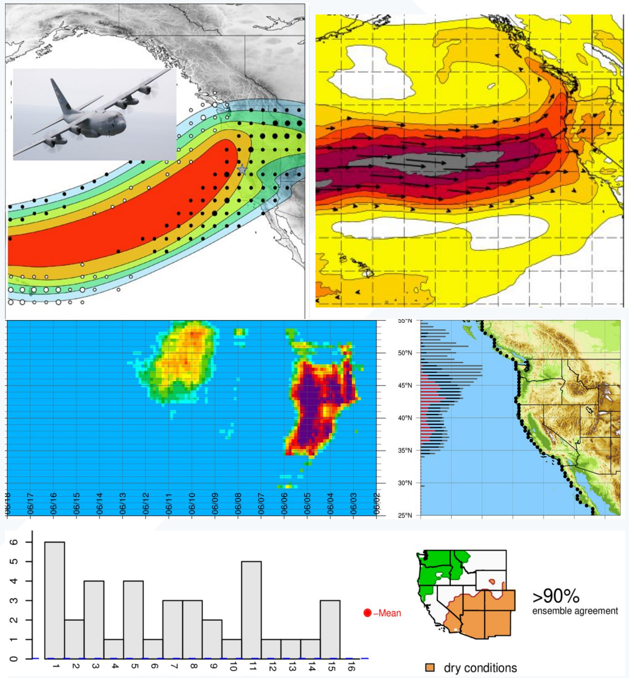
Wednesday
3 August 2022

2022 FIRO Workshop
2-4 August 2022
Hosted by the Center for Western Weather and Water Extremes

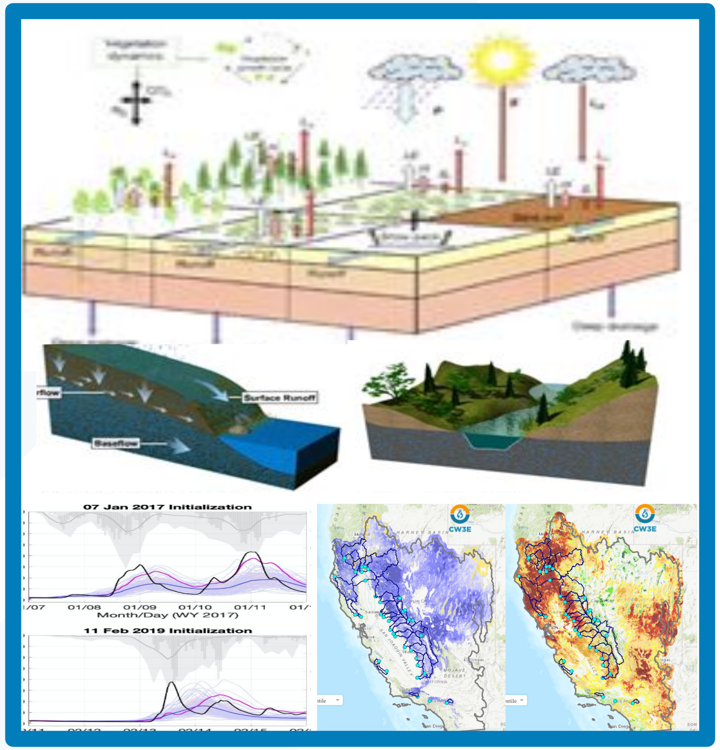
Center for Western Weather and Water Extremes

Background: role of hydrologic modeling in FIRO

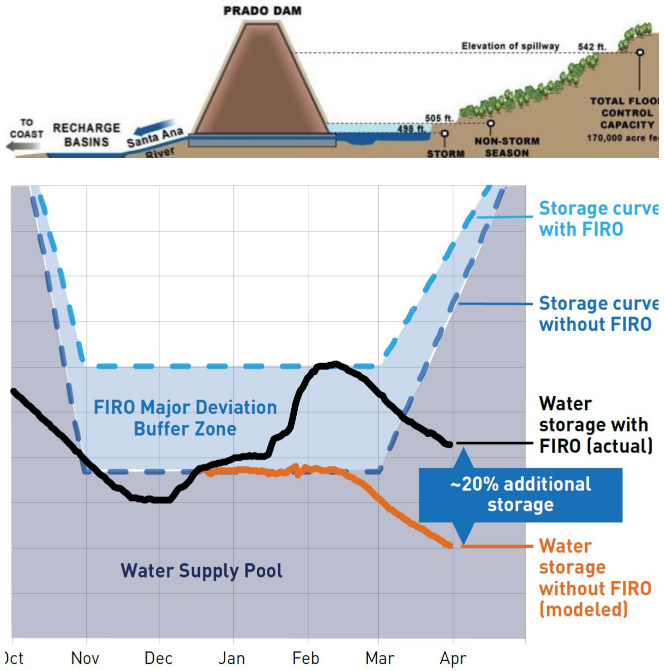
Weather/Climate



Hydrology



Engineering/App

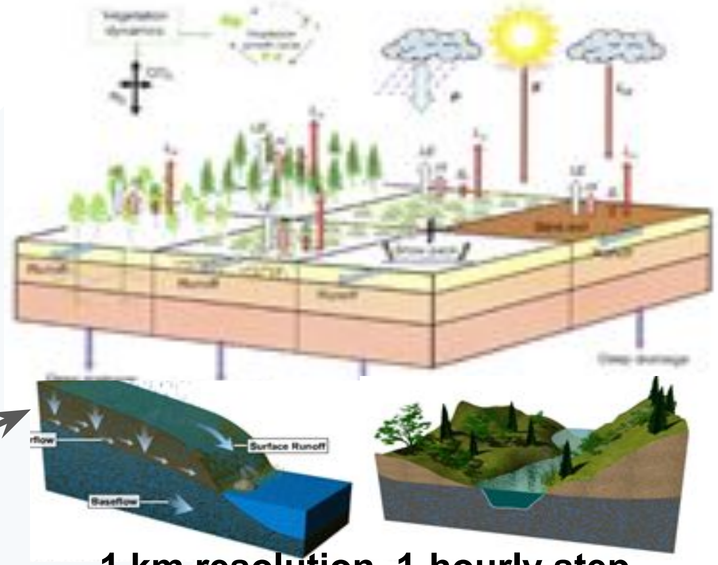


CW3E Hydrologic Monitoring/Forecasting: Basic Framework

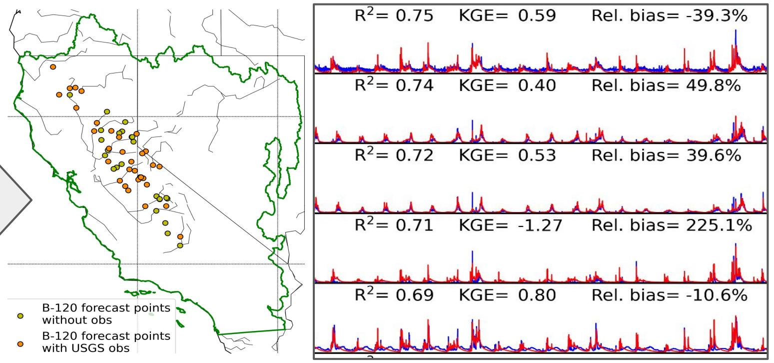
Input: Prec, Temp, Shortwave, Longwave, Pressure, Humidity, Wind (u/v)

AORC Calibrated Parameters

WRF-Hydro
(National Water Model)



1 km resolution, 1-hourly step



Output: Q, SM, SWE, ET, ...

Post-processing for Bias Correction & Skill Enhancement

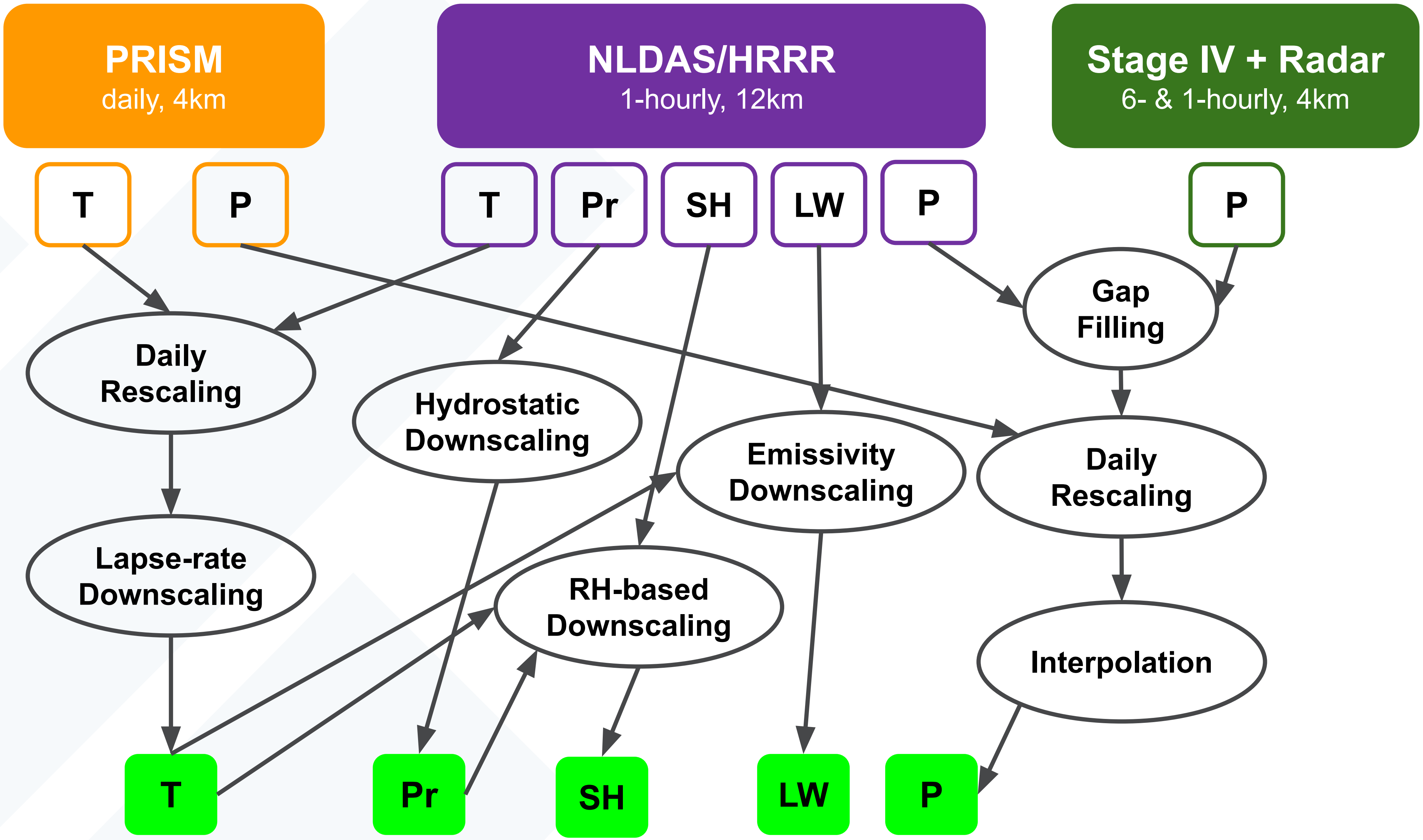
Reanalysis + Reforecast (Hindcast)	NRT Monitor	Short Fcst	Seasonal Fcst

1979 2002 T-6m T-4d T+0 T+7d T+15d T+6m

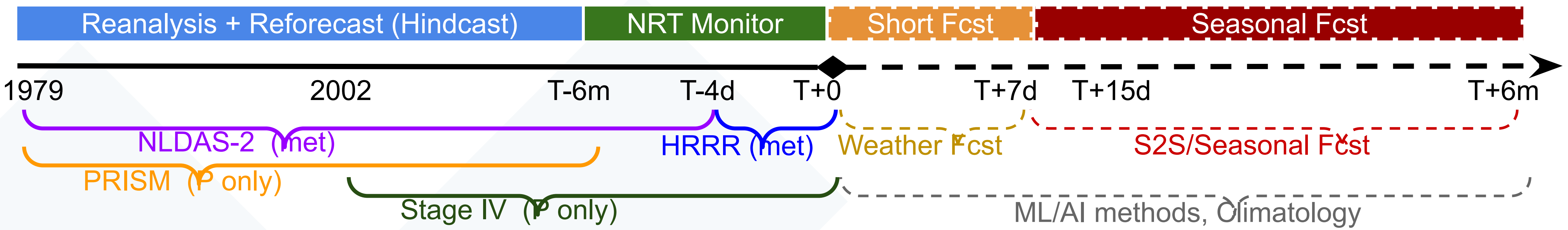
NLDAS-2 (met)
HRRR (met)
Weather Fcst
S2S/Seasonal Fcst
PRISM (P only)
Stage IV (P only)
ML/AI methods, Climatology

CW3E Hydrologic Monitoring/Forecasting: Forcing Data Engine

1-km, 1-hourly, Near Real Time & Retrospective Surface Meteorological Forcing Data



CW3E Hydrologic Monitoring/Forecasting: Forcing Data Engine

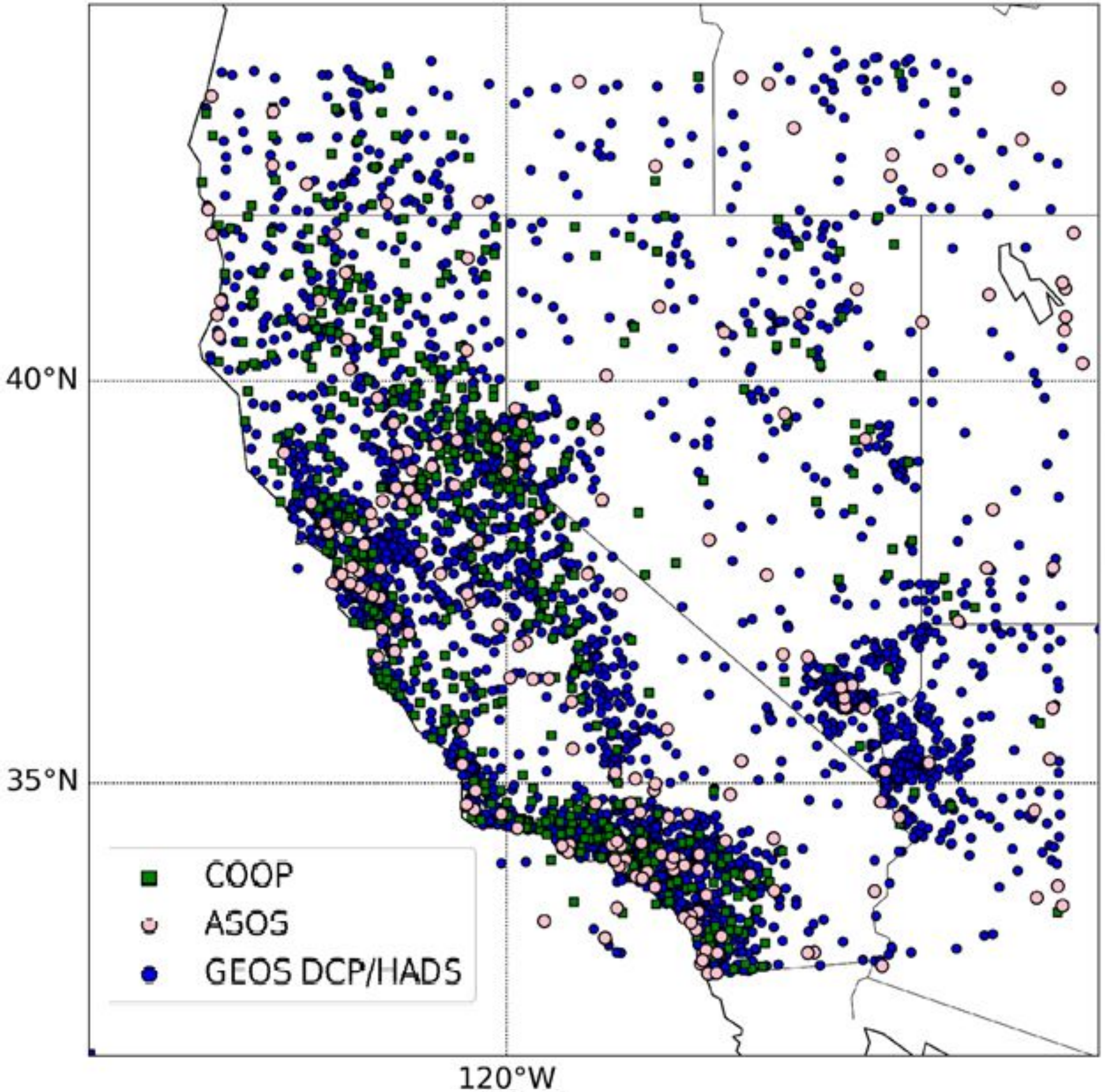


Dataset	Res	Freq	Domain	Period	Latency	Format	Sub-folder
NLDAS-2	0.125°	hourly	CONUS	1979-	3-4 days	Grib-1	nldas2
HRRR	3 km	hourly	US	2021-	hours	Grib-2	hrr
Stage-IV	4 km	hourly	US	2002-	realtime: hours archive: 10 days	Grib-2	stage4
PRISM	4 km	daily	US	1979-	provisional: 1 mon recent: 6 mon	NetCDF	prism
AORC	1 km	hourly	CONUS	2011-2019	retrospective	NetCDF	aorc
WestWRF	9 km	hourly	WUS	10 days lead	hours	NetCDF	westwrf
CCA	4 km	monthly	CNRFC	6 mon lead	5 days	NetCDF	cca

CW3E Hydrologic Monitoring/Forecasting: Forcing Data Engine

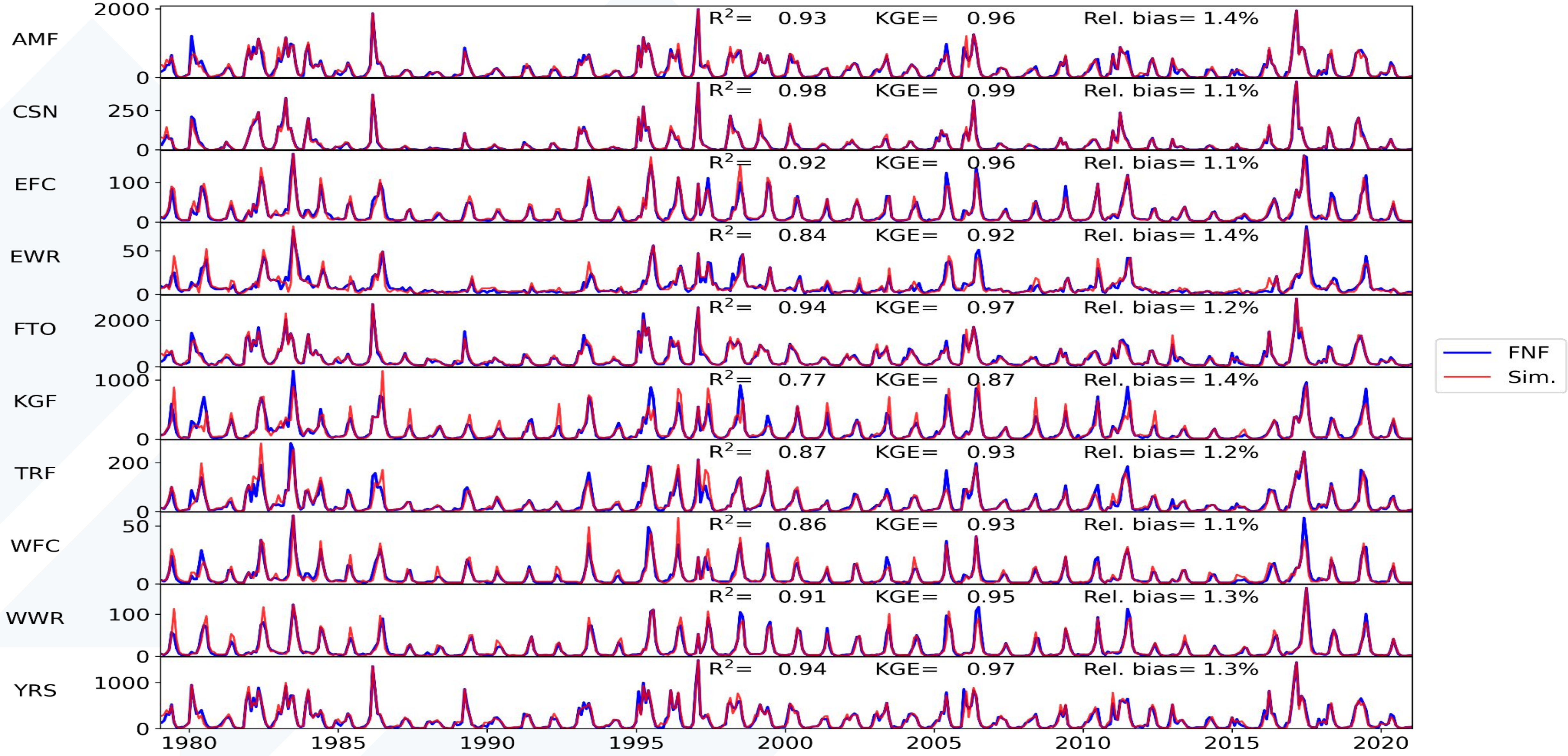
QPE improvement (work in progress):

- Hourly gauge-based estimates with site QC
- Sampling bias
- Orographic
- Snow undercatch
- Freezing level (precip phase)
- …

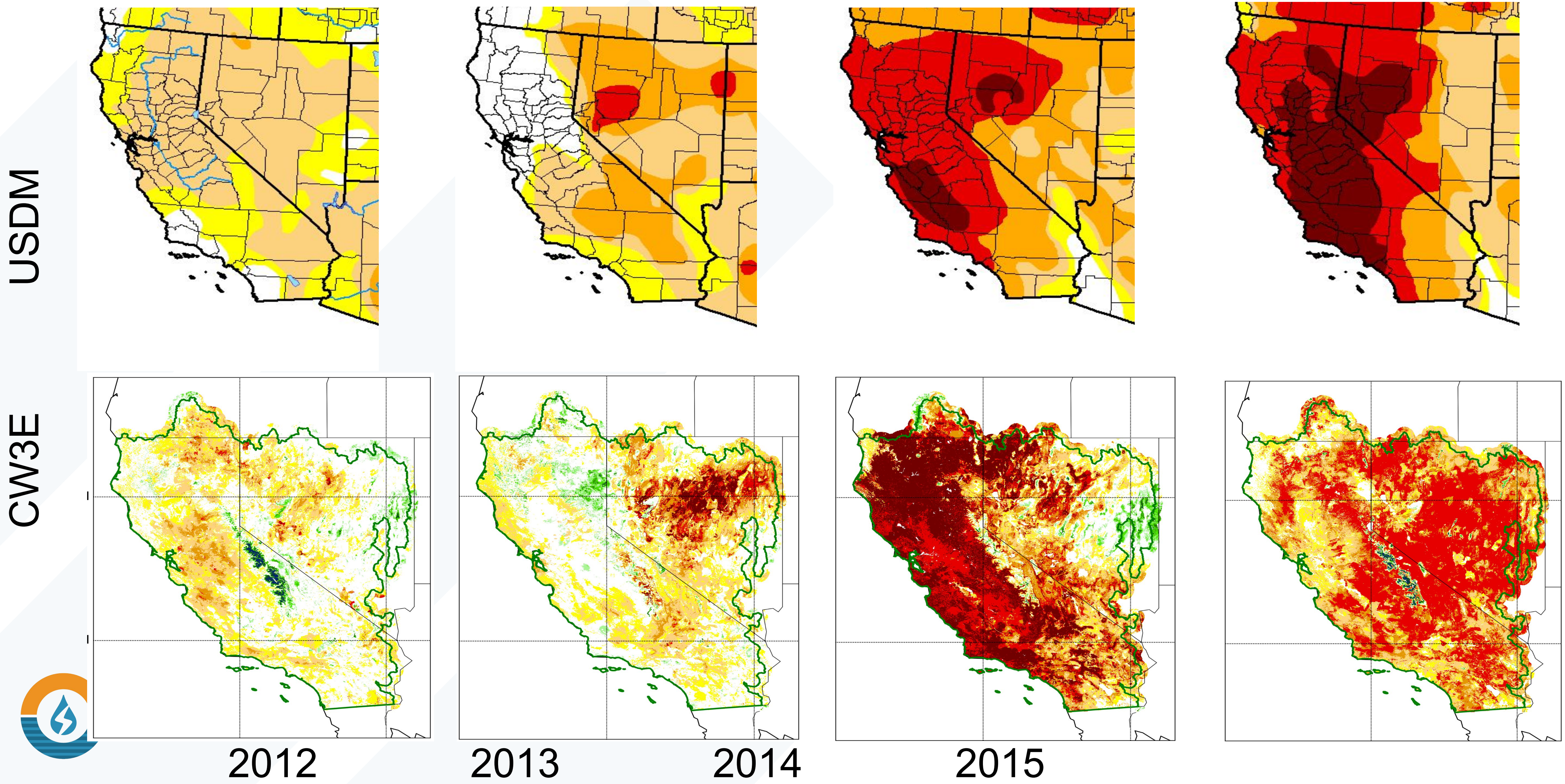


CW3E Hydrologic Monitoring/Forecasting: Reanalysis

Comparison with monthly unimpaired flow 1979/01-2020/12

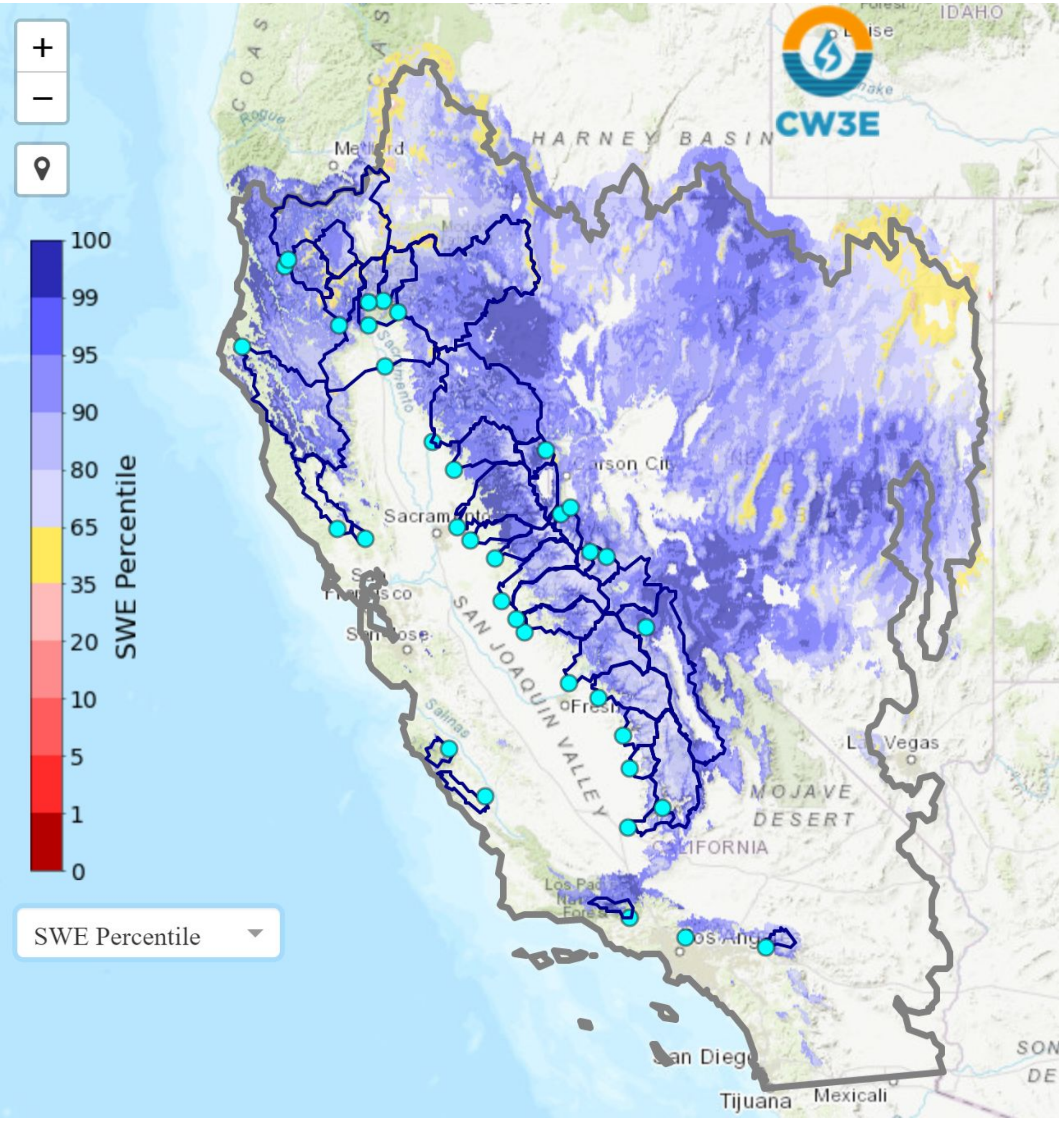


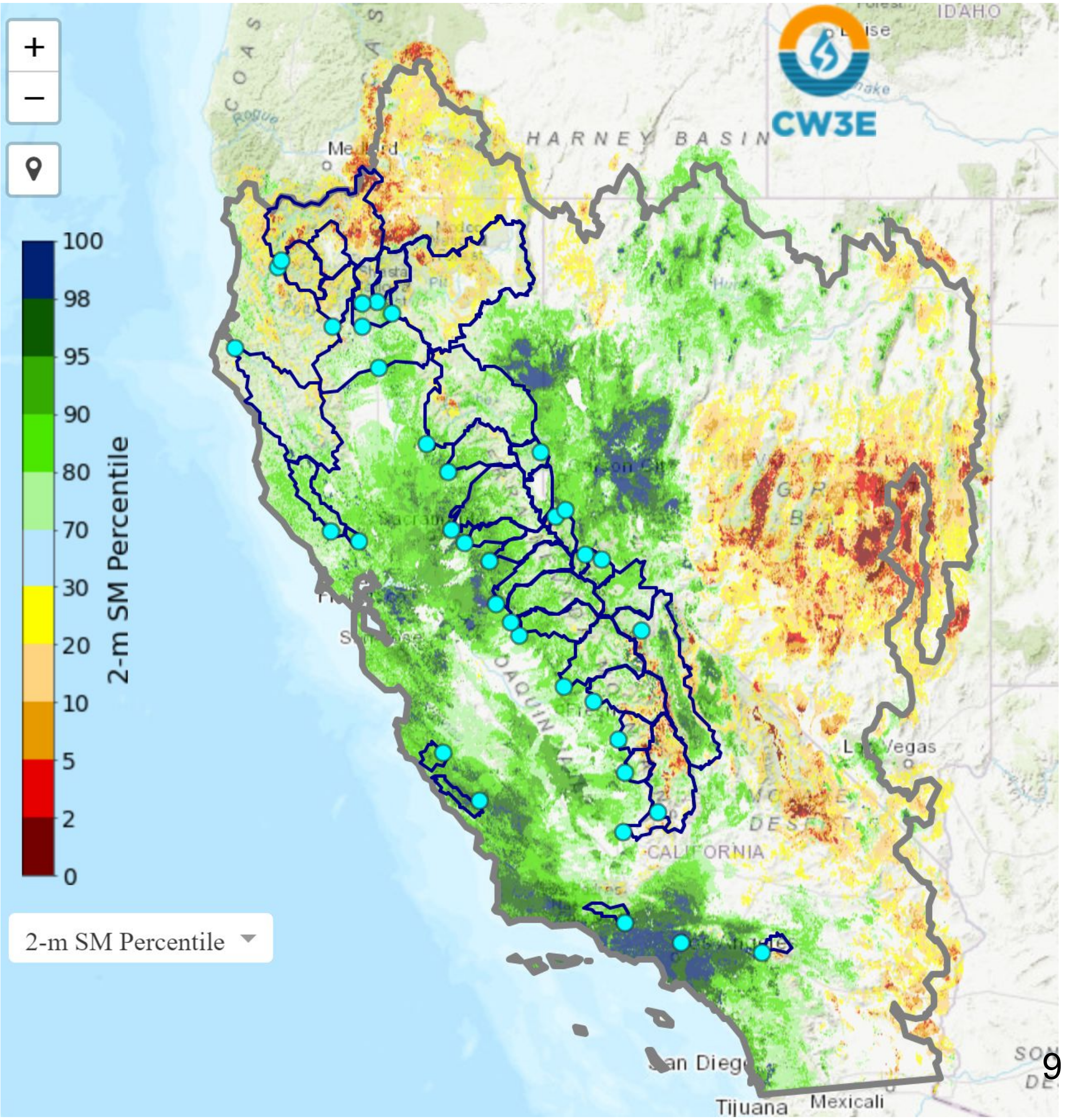
CW3E Hydrologic Monitoring/Forecasting: Reanalysis



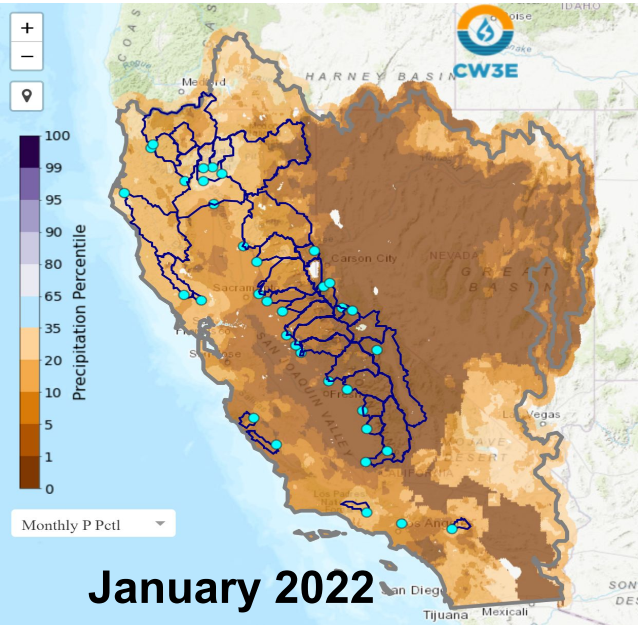
CW3E Hydrologic Monitoring/Forecasting: NRT Monitoring

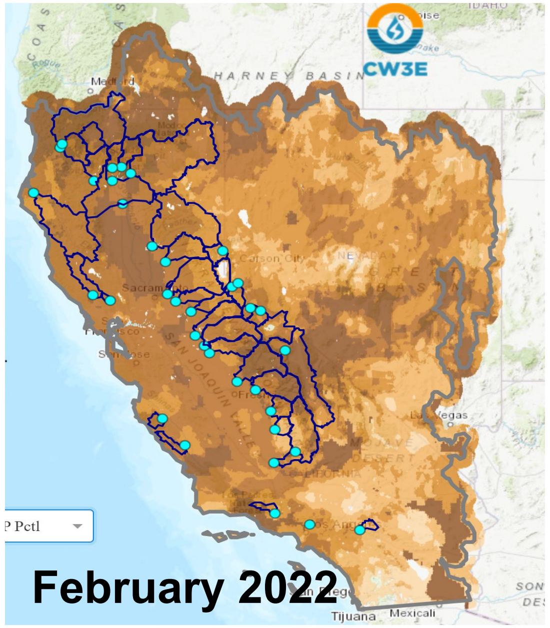
SWE Percentile **2021-12-31** **Soil Moisture Percentile**

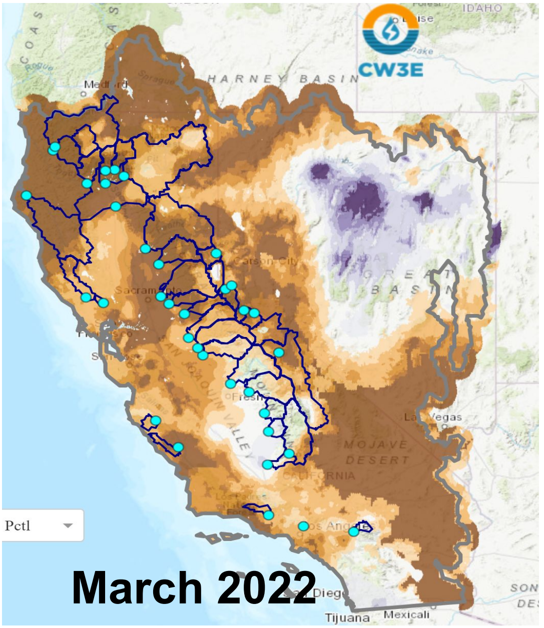


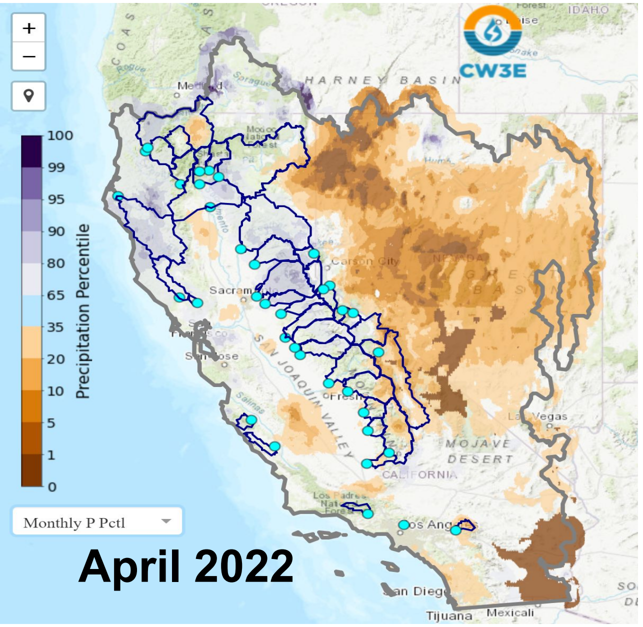


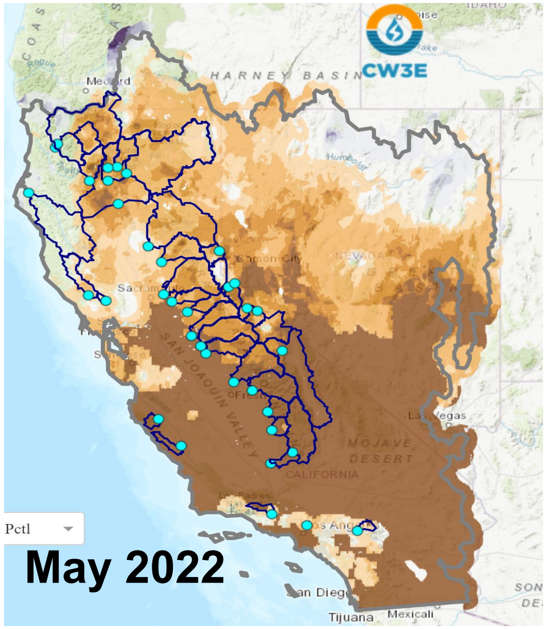
CW3E Hydrologic Monitoring/Forecasting: NRT Monitoring





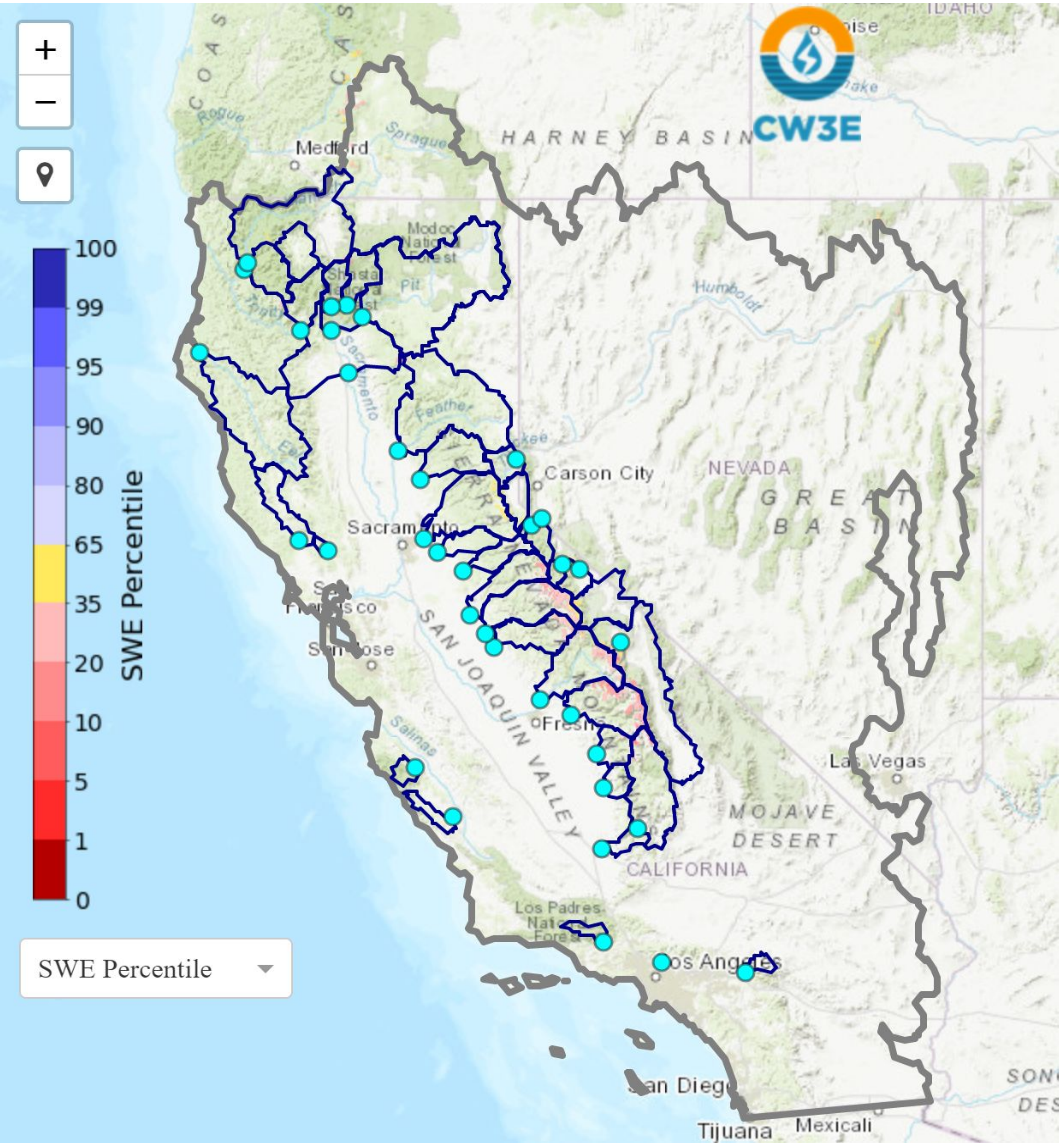


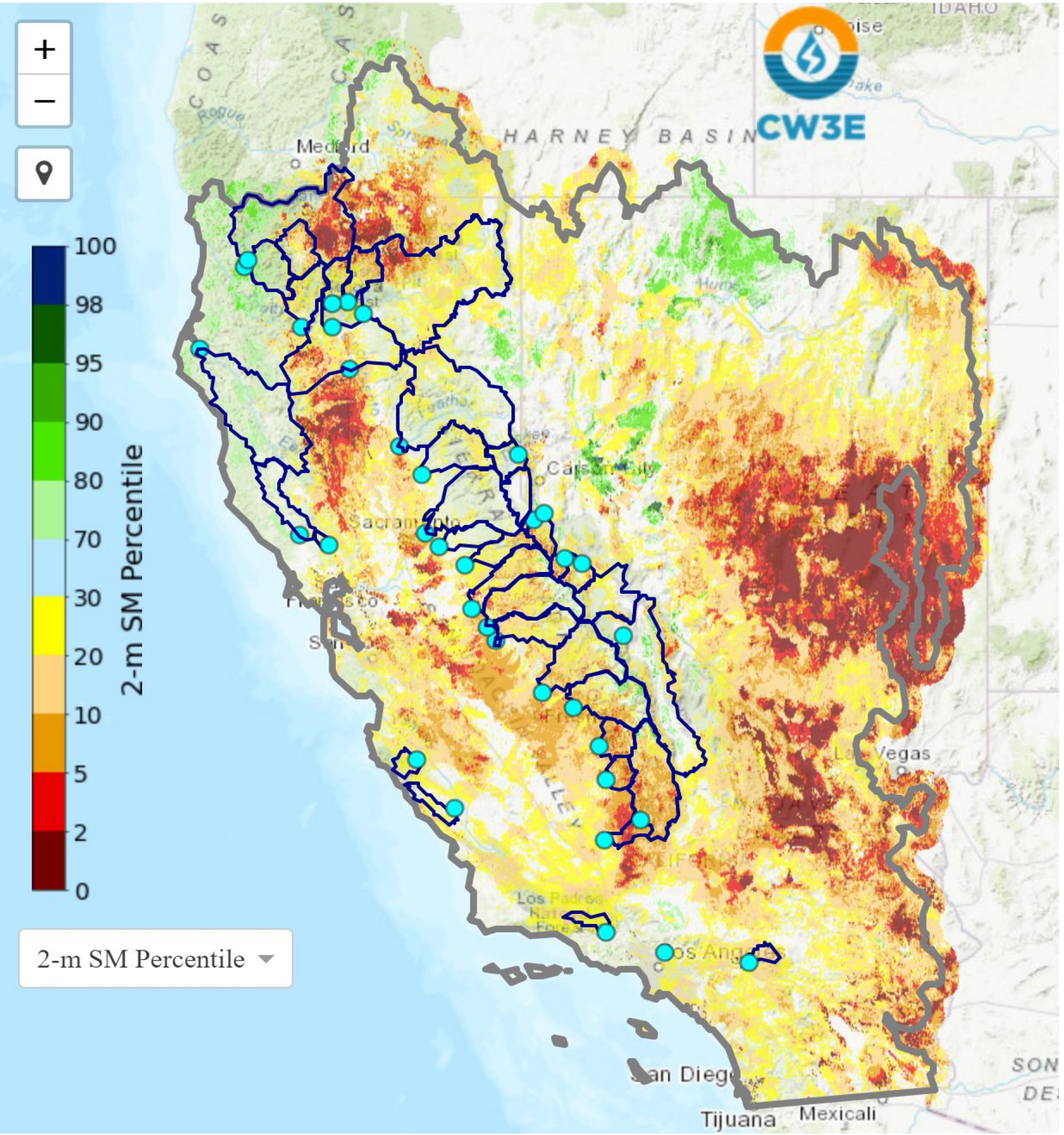




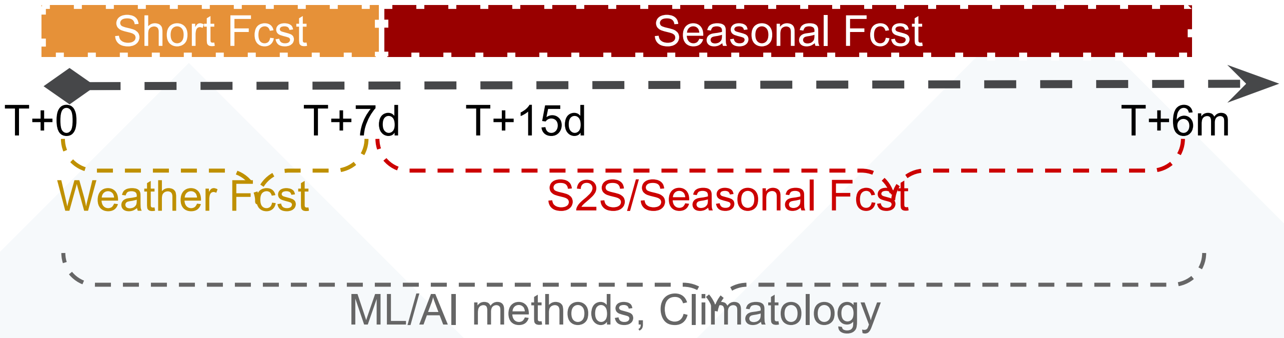
CW3E Hydrologic Monitoring/Forecasting: NRT Monitoring

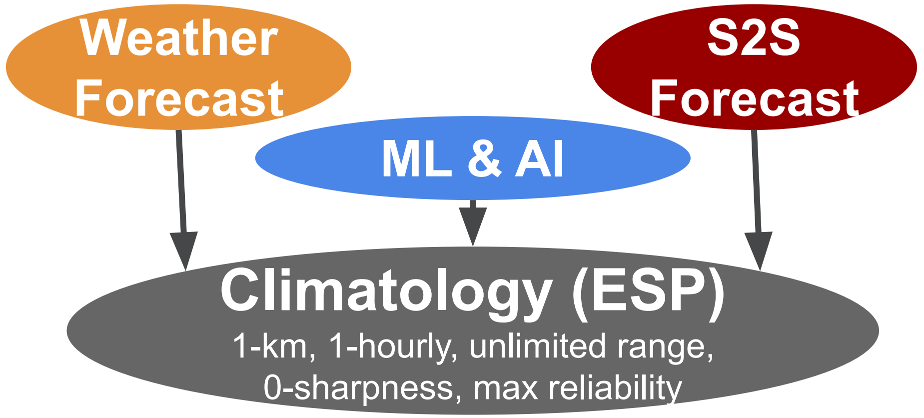
SWE Percentile **2022-05-31** **Soil Moisture Percentile**

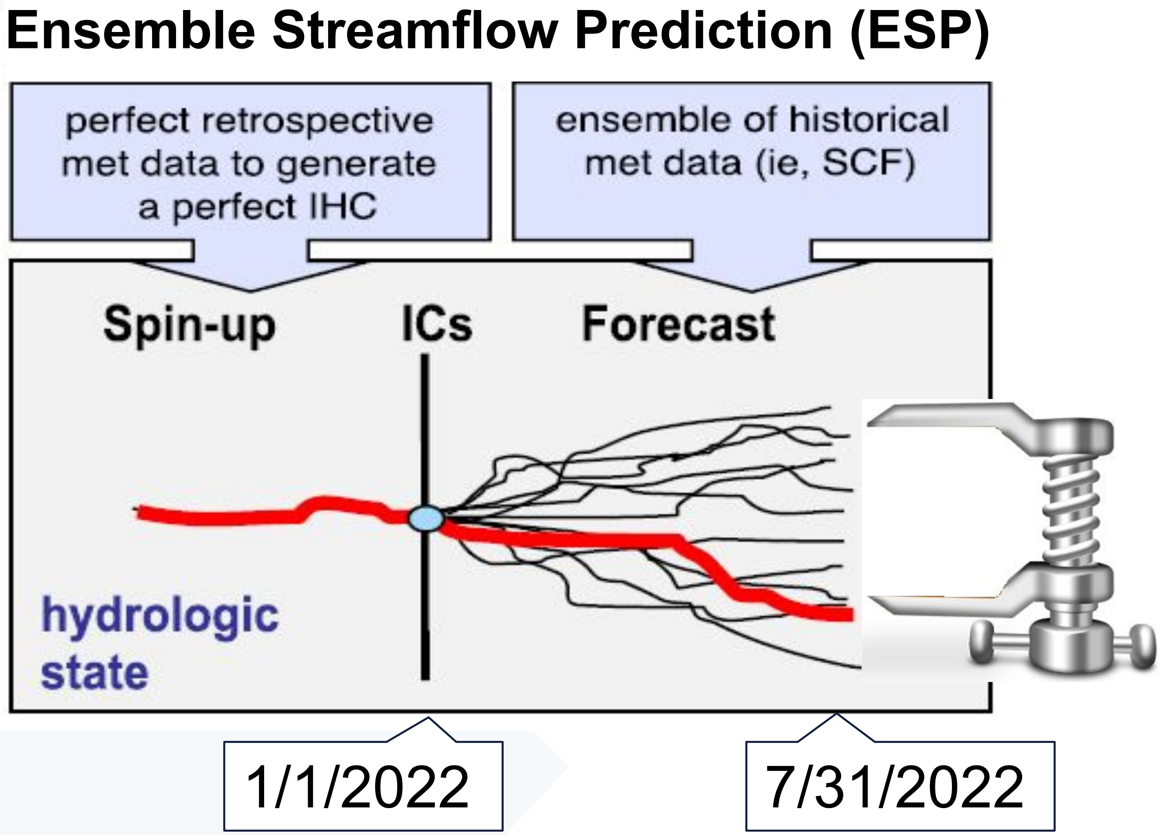




CW3E Hydrologic Monitoring/Forecasting: Forecasting

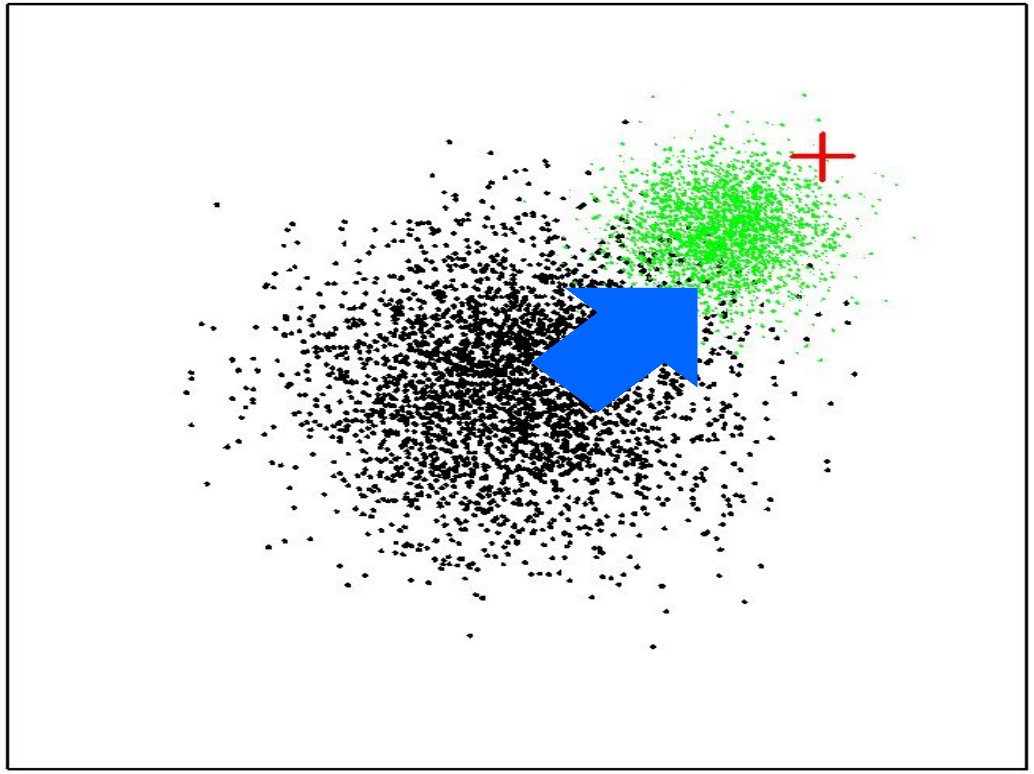






CW3E Hydrologic Monitoring/Forecasting: Forecasting

EnKF



EnKF state update is a weighted sum of inputs:

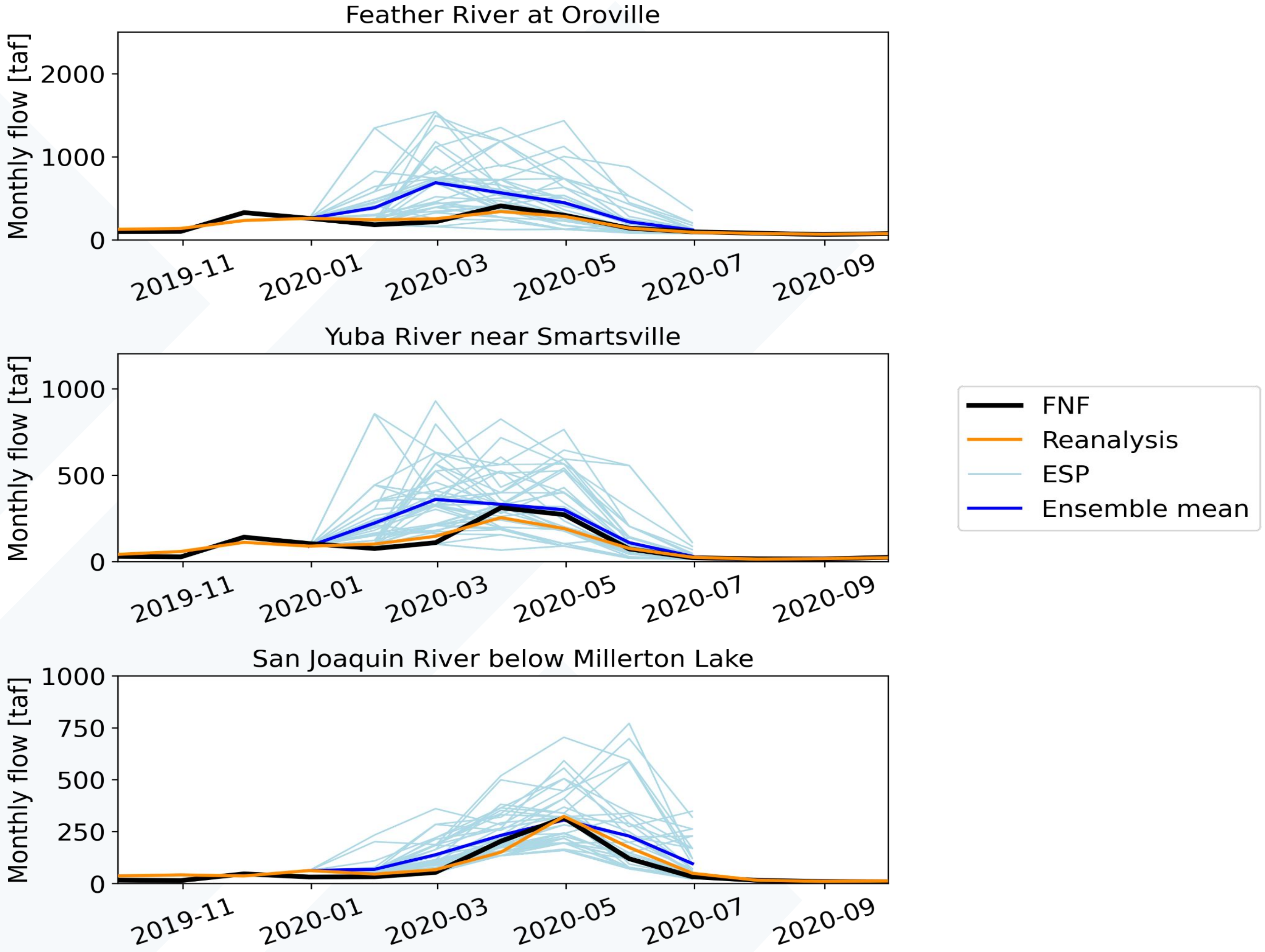
$$\mathbf{x}''^{(i)}_t = \mathbf{x}'^{(i)}_t + \mathbf{K}_t\left(\mathbf{z}_t - \mathbf{H}_t\mathbf{x}'^{(i)}_t\right)$$
$$= w_x\mathbf{x}'^{(i)}_t + w_z\mathbf{z}_t$$

$\mathbf{x}'^{(i)}_t$: prior estimate (ESP)
$\mathbf{x}''^{(i)}_t$: posterior estimate (merged)
$\mathbf{z}_t$: additional forecast info

How to determine weights: if CCA forecast explains r^2 of the variability, then the combined (fused) forecast ensemble should have its spread reduced by r^2, therefore:

$$w_x = 1 - r^2 \quad \text{and} \quad w_z = r^2$$

CW3E Hydrologic Monitoring/Forecasting: Forecasting



ESP+CCA Hindcasts initialized on 2020-02-01

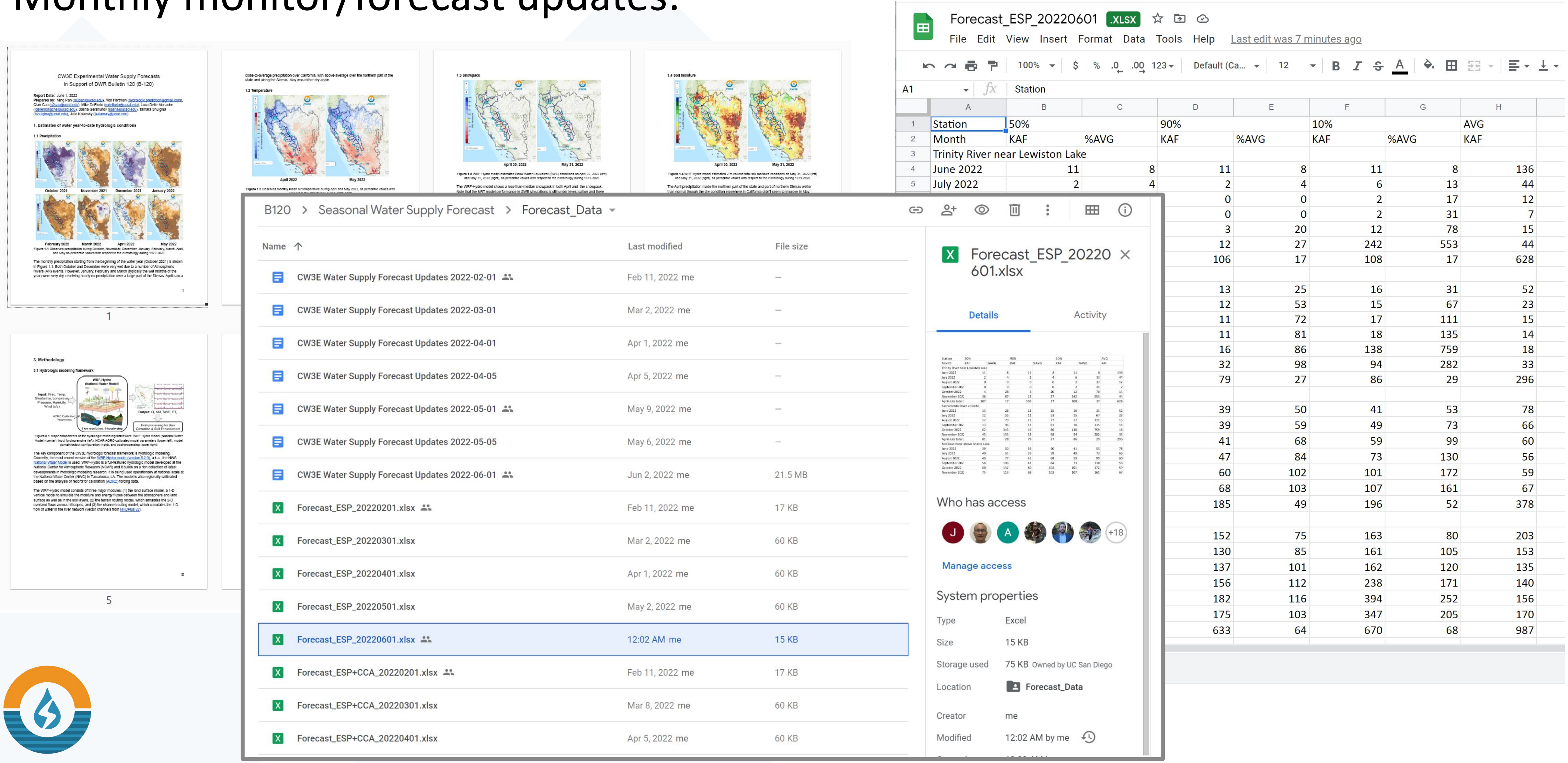
Zoom-in plots

CW3E Hydrologic Monitoring/Forecasting: Forecasting

Station	50%		90%		10%		AVG
Month	KAF	%AVG	KAF	%AVG	KAF	%AVG	KAF
Feather River at Oroville							
June 2022	95	32	94	32	107	36	298
July 2022	80	55	80	55	89	62	144
August 2022	70	72	68	70	92	95	97
September 2022	65	76	62	72	92	107	86
October 2022	99	101	74	74	245	248	99
November 2022	338	184	123	67	739	403	183
April-July total 2022	897	**54**	895	54	913	55	1666
Yuba River near Smartsville							
June 2022	44	24	44	24	64	35	185
July 2022	19	34	16	28	28	50	56
August 2022	13	62	12	58	23	110	21
September 2022	20	113	16	92	33	185	18
October 2022	121	378	29	91	243	761	32
November 2022	215	250	147	170	477	554	86
April-July total 2022	532	**55**	531	55	573	59	973
American River below Folsom Lake							
June 2022	87	35	86	35	113	46	245
July 2022	22	32	20	29	23	34	69
August 2022	5	31	5	31	12	72	16
September 2022	10	81	8	65	12	98	12
October 2022	20	74	19	69	177	660	27
November 2022	201	236	80	93	836	982	85
April-July total 2022	741	**61**	735	60	781	64	1223

CW3E Hydrologic Monitoring/Forecasting: Forecasting

Monthly monitor/forecast updates:



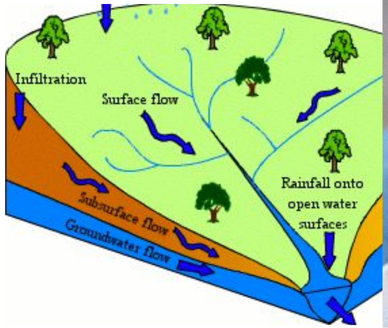
Data-Driven Alternative Modeling: Machine Learning (ML)

Machine Learning: statistics on steroids

- Any inputs, any predictive connections, any functional relationships, exhaustive extraction, …
- Requires data for training, over-fitting, curse of dimensionality, …

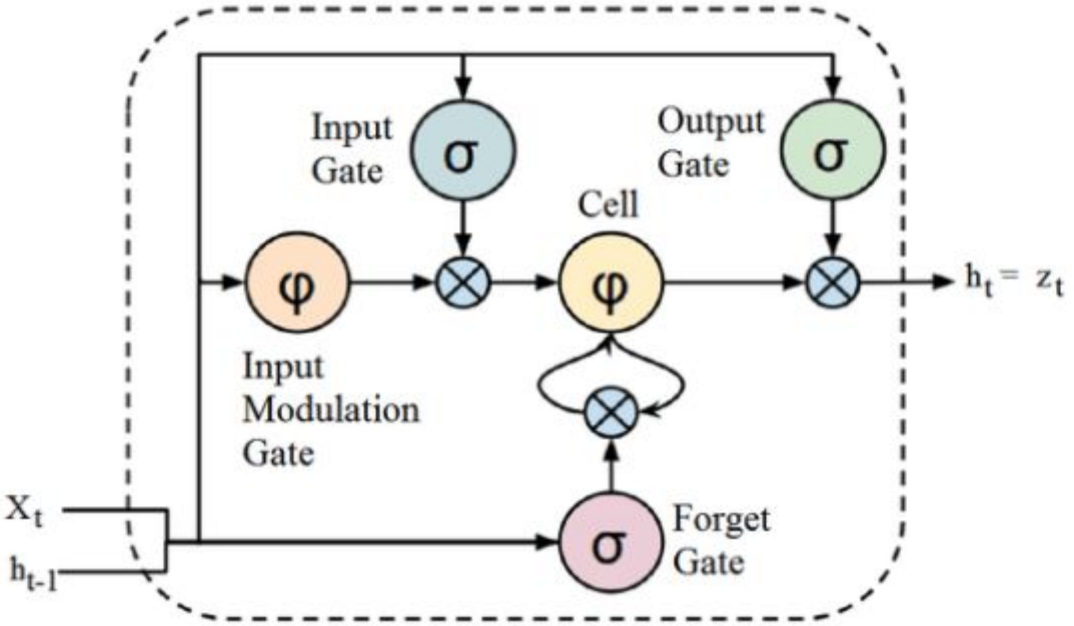
Runoff Hydrology:

- Many processes at many spatial and temporal scales – what matters at what time/location/scale/circumstance?
- Low dimensionality: watershed runoff is the primary target)
- *Lack measurements to verify assumed physical processes*
- *Still need tuning against data*



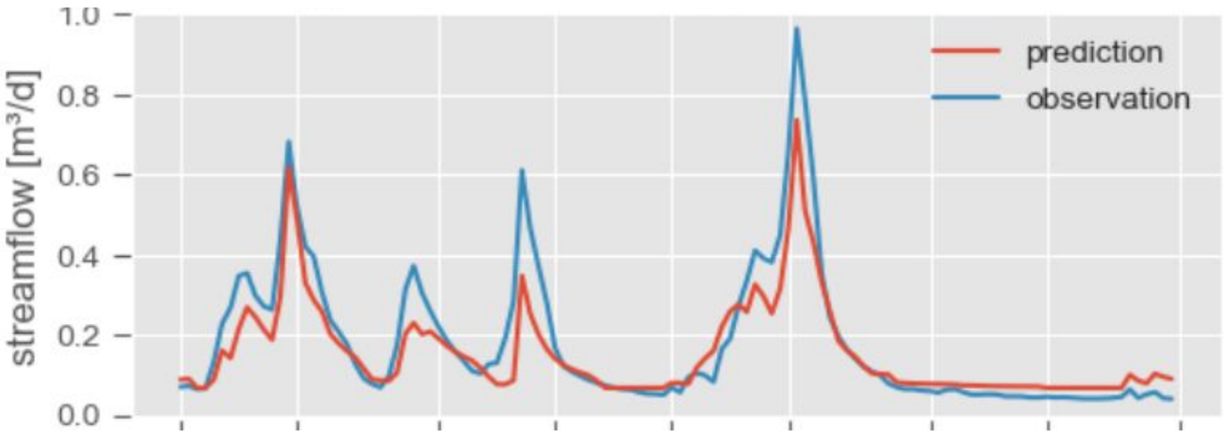
Are you sure? Can you verify?
Why so dependent on training?

Data-Driven Alternative Modeling: Machine Learning (ML)

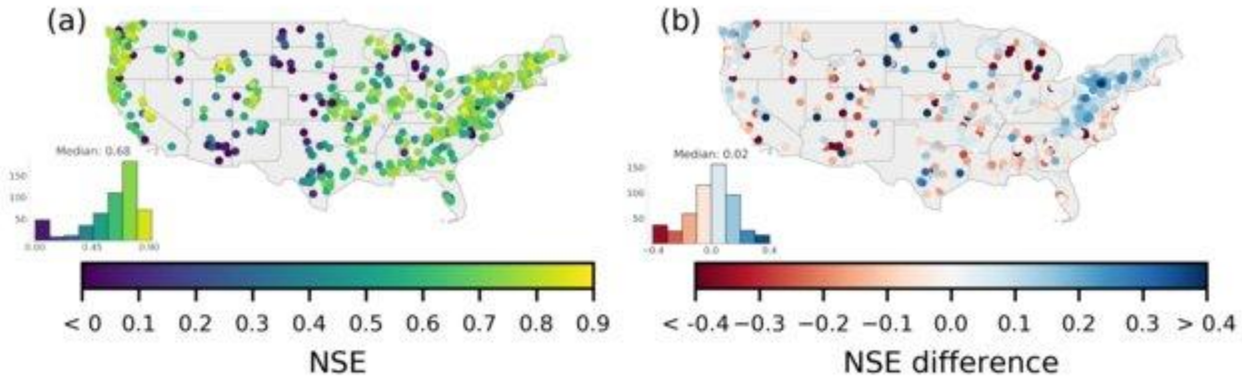


Long Short Term Memory (LSTM)

State-space (Markovian) type of dynamic processes



(Drost, 2022)



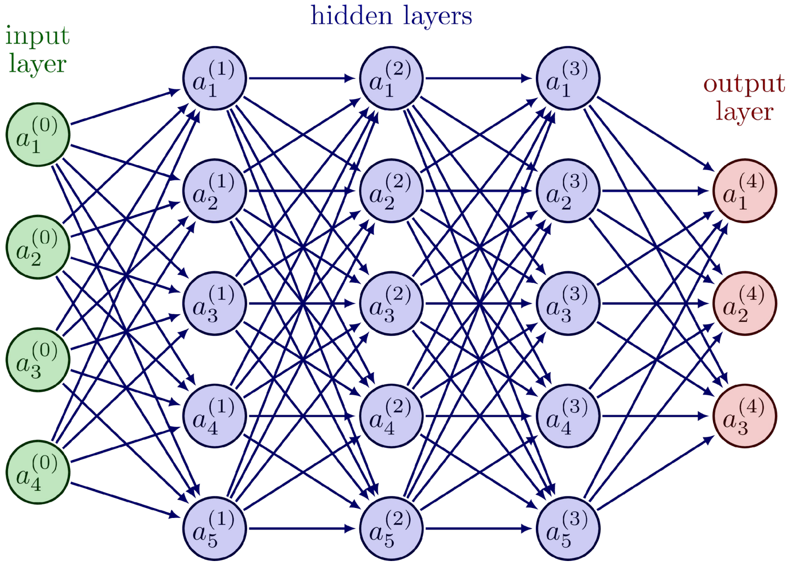
(Kratzart et al., 2018)

Discussions

How we're trying to improve inflow modeling:

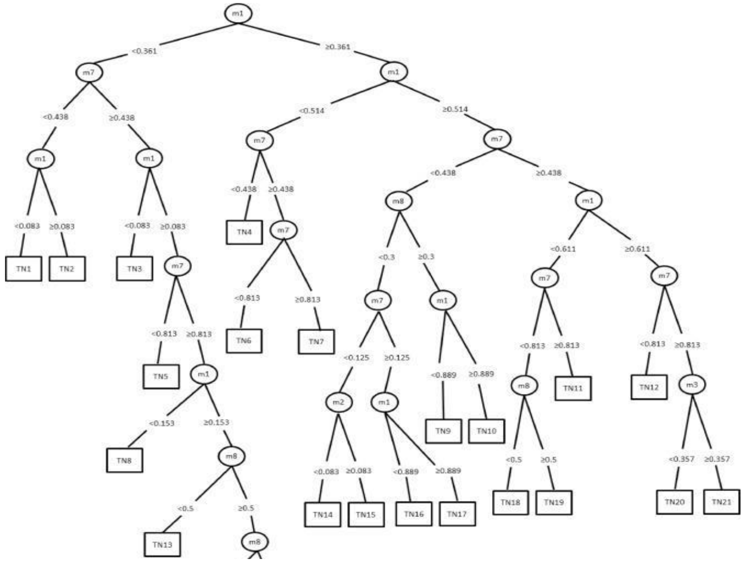
- Data
  - Better QPE/QPF (orographic, undercatch, sampling bias, phase, etc.)
  - Long-term consistency (bias correction, data-driven approaches)
- Modeling
  - Physical process based modeling (at fine scales)
  - Data driven modeling like ML (at watershed and daily scales)
- Assimilation/fusion (snow, soil moisture, groundwater, etc.)
  - Traditional DA
  - Differentiable ML

Data-Driven Alternative Modeling: Machine Learning (ML)



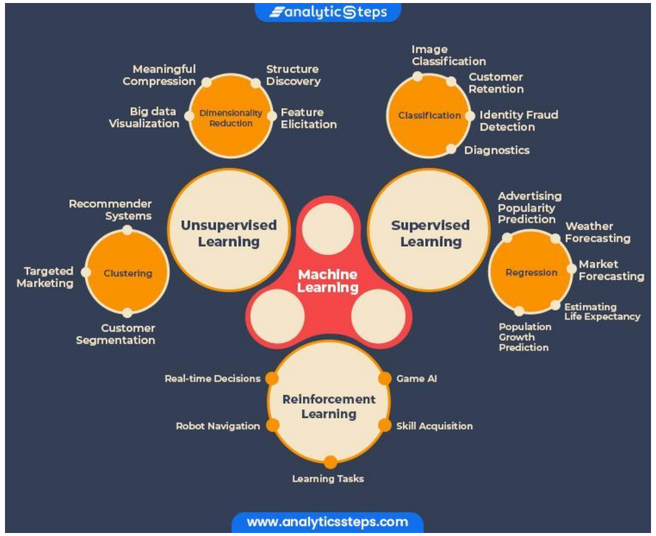
Neural Network (NN)

Universal-fit regressions



Regression Trees

Conditional processes



Many many more …

Ming Pan:

- Lead hydrologic modeling/forecasting research at CW3E since 2021
- Background in hydrologic modeling, remote sensing, data assimilation
- Research scholar at Princeton University before joining CW3E
- PhD in hydrology from Princeton University (2006)