M^4: The Next-Generation AI-Based NRCS Water Supply Forecasting System

Sean W. Fleming, David C. Garen, Angus G. Goodbody, Cara S. McCarthy, Lexi C. Landers

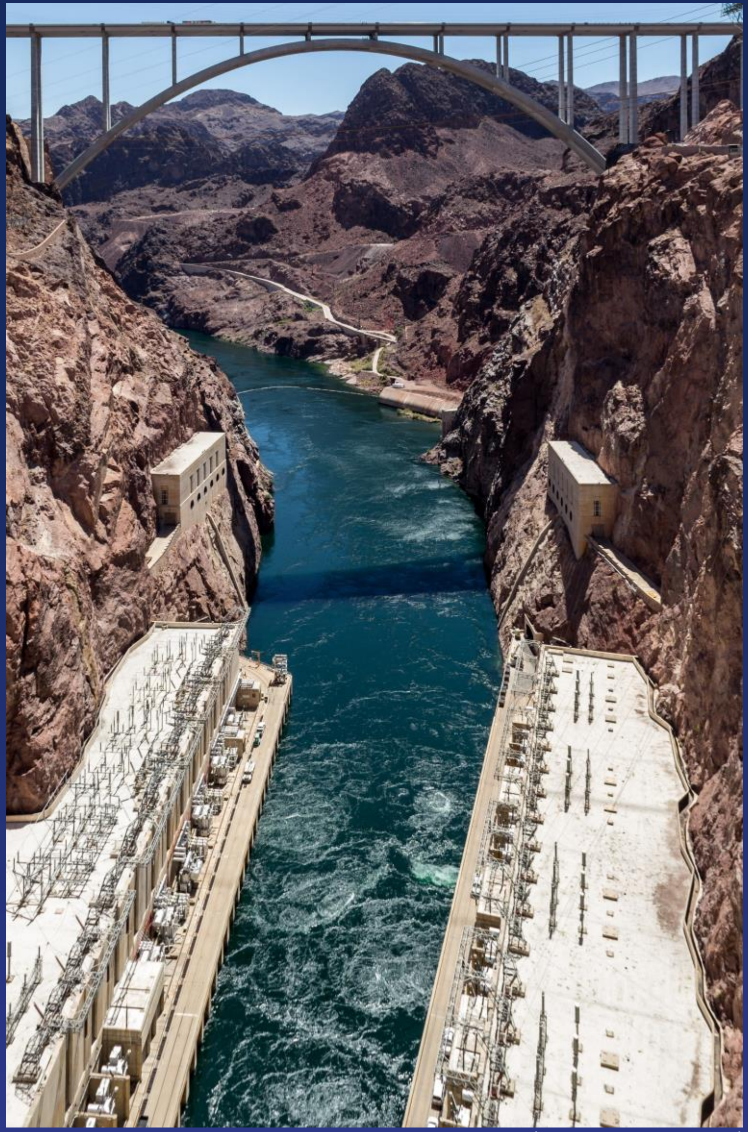
Western States Federal Agency Support Team (WestFAST) Webinar Series

20 January 2021

Punch lines

Highlights

- Developed the next generation for one of the oldest & largest operational hydrometeorological prediction systems in the American West
- Resulting novel, AI-based, hybrid metasystem is the largest transition of machine learning into a genuine operational hydrologic prediction setting to date
- Made possible by understanding needs of operational water resource science & engineering community + working hand-in-hand to build appropriate solutions
- More accurate, robust, and flexible platform will also open doors to more extensive & sophisticated forecast product integration going forward



Background

Operational water supply forecasting (WSF) in the US West

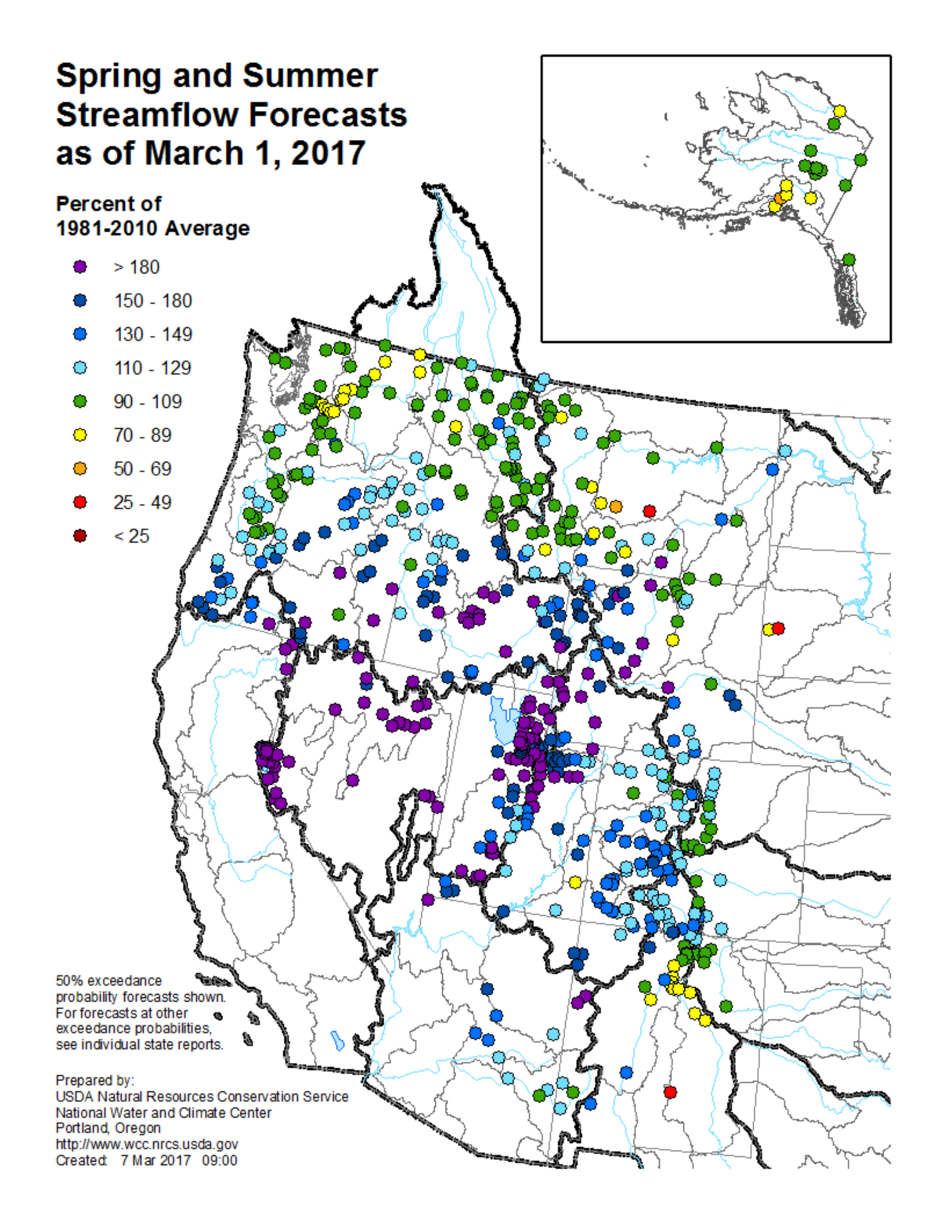
Forecasts, typically issued once or more per month, and usually beginning in winter, of upcoming spring-summer total flow volume for a given point on a given river, performed by institutions having direct accountabilities to end users around reliable generation of forecast products

- Provide the practical basis for water resource management in western North America
- Watershed hydrology models here typically emphasize snowmelt contributions to flow
- Common predictors include snowpack, precipitation, antecedent streamflow, climate indices
- Even incremental WSF model improvements can be worth > $100M/yr for just one basin

Sprague River, Upper Klamath Basin (S.W. Fleming/US Department of Agriculture NRCS NWCC)

USDA NRCS WSF

- USDA has measured snow & predicted water supply since the 1930s (e.g., SNOTEL starting in 1980s)
- National Water and Climate Center of the NRCS currently operates the largest stand-alone operational WSF system in the US West
- May be the largest statistical operational WSF system globally with > 600 forecast locations
- Each with multiple issue dates & target periods
- Several statistical and process-simulation models are currently used
- Core method: principal component regression (PCR) introduced to WSF by NRCS in early 1990s



US Department of Agriculture NRCS NWCC

Context to next-gen operational WSF system design

Preparatory steps

Thorough inventory & assessment of needs & options before undertaking system design:

- **Clearly document existing NRCS WSF system**
  First implemented about 25 years ago; operational forecasting systems are complex & organic, evolving over time
- **Assess capabilities and limitations of current system**
  Included documentation of known issues & completion of advanced statistical diagnostics
- **Comprehensively review data-driven WSF modeling progress**
  Assess for potential relevance to NRCS operations, including topics like longer lead times using climatic forcing data, more advanced statistical & machine learning methods, ensemble modeling, statistical-process simulation modeling hybrids, & other research directions
- **Assess implications of global anthropogenic climate change on WSF**
  Mainly around a need for improved seasonal prediction capabilities given an increasingly unpredictable hydroclimatic system, with lower snowpack and possibly greater variability, while water demand increases
- **Develop initial blueprint & several preliminary scoping models**
  On the basis of foregoing preliminary steps and NRCS system requirements, experiment with some concepts potentially underlying a new approach and assess their suitability for inclusion in the full prototype system

Model characteristics & selection matrix

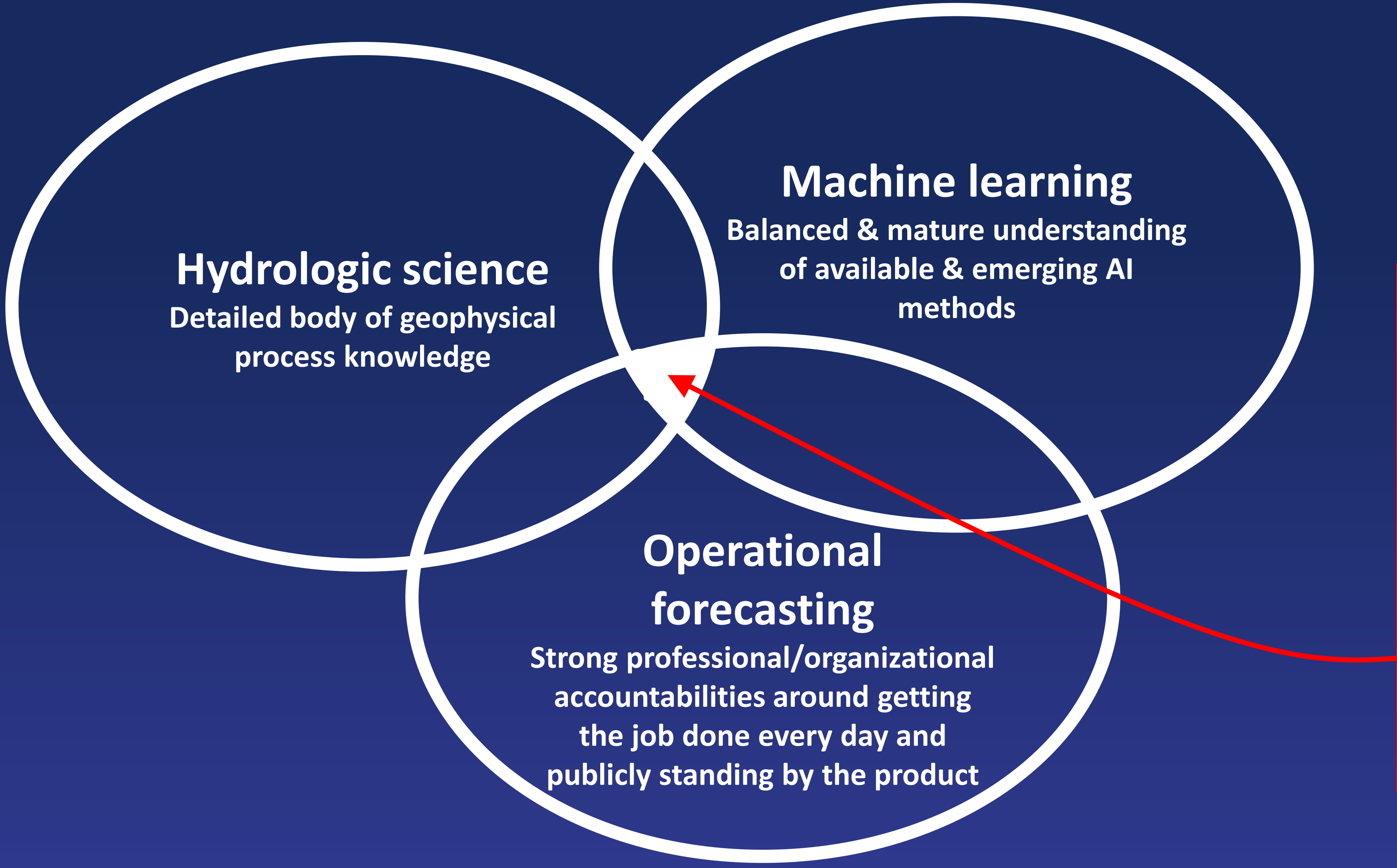
Characteristic	Process simulation	Statistical regression model	Machine learning
Cost/ease/requirements	terrible to moderate	very good to excellent	moderate to very good
Code reliability	very poor to excellent	moderate to excellent	good to excellent
Code transparency	terrible to excellent	very good to excellent	moderate to excellent
Physical interpretability	moderate to excellent	poor to very good	very poor to very good
Deterministic accuracy	terrible to very good	moderate to very good	very good to excellent
Probabilistic accuracy	terrible to moderate	moderate to excellent	moderate to excellent
Innovation/opportunities	terrible to very good	terrible to good	excellent

- **Which class of models is "best" depends on what you're aiming to do**
- **Main punch lines around machine learning as relevant to NRCS WSF:**
  - Overwhelmingly good bang for the buck as a predictive tool
  - Leverage the data science all around us in our everyday lives to produce better, cheaper WSFs
  - But can't just throw machine learning at WSF and expect it to work right: must approach this in a very application-specific way, carefully choosing – or developing – the right tools for the job

Machine learning and hydrologic prediction

- **Long and somewhat mixed history**
  - Research papers using AI for river forecasting date back to 1995
  - Yet true operationalization is mostly very recent (as far as most of us know…)
- **Some stumbling blocks**
  - Lack of probabilistic forecasts
  - Absence of a good storyline around physical hydroclimatic process for most AI models
  - Skepticism in high-stakes operational settings where existing techniques are well-established and new methods in the research literature are unproven and sometimes oversold
  - Misunderstandings and lack of technical training and professional familiarity among many physical scientists with AI, its limitations, and its strengths
- **But things have changed**
  - Just about everyone is getting used to the idea of AI
  - Tools are far more accessible, and methods are far more diversified
  - Explainability, overtraining, and prediction uncertainty estimation questions can be addressed with AI design and implementation choices… if done carefully

Convergence of knowledge, tools, and requirements



Intersection:

Need to build AI prediction systems that align with domain-expert expectations around consistency with both physical hydrologic science and organizational requirements of operational forecasting organizations

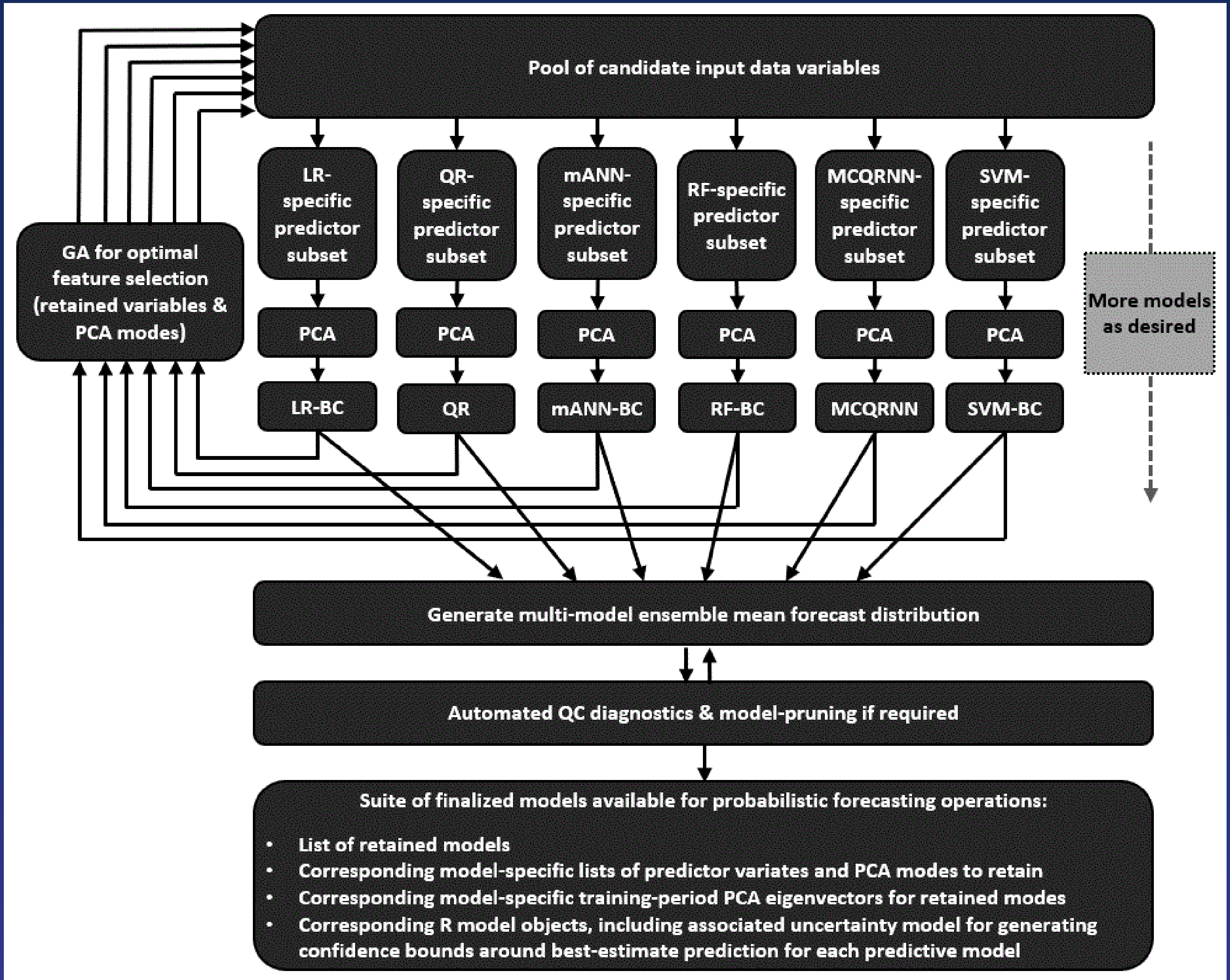
Key design criteria

1. Improved forecast accuracy
2. Improved potential for automation (largely an “over-the-loop” system)
3. Relatively low cost & good ease of development, implementation, and operation
4. Seamlessly address three known technical issues: (a) nonlinear functional forms, and (b) heteroscedastic and (c) non-Gaussian residuals
5. Modular and expandable
6. Physics-aware AI: hydrologic theory-guided & geophysically explainable forecasts
7. Balance innovation & performance gains vs. established building blocks & proven tools
8. Adopt multi-model ensemble: equifinality & model selection uncertainty over diverse geophysical environments across the US West
9. Accommodate high-dimensional predictor datasets & potential for multiple independent input signals: will only grow more important in the future with more spatially distributed inputs

Multi-method machine learning metasystem (M^4)

Solution: a hybrid prediction analytics engine

- **Several very carefully selected modeling threads: handle heteroscedastic & non-Gaussian residuals + nonlinearity, allow physicality constraints like monotonicity & non-negativity**
- **Combine with dimensionality reduction using statistical pattern recognition, genetic algorithm for feature optimization, parallelization across processor cores, flexible architecture, AutoML, physicality checks & adjustments**
- **Forms an integrated, modular, expandable, ensemble forecasting metasystem**

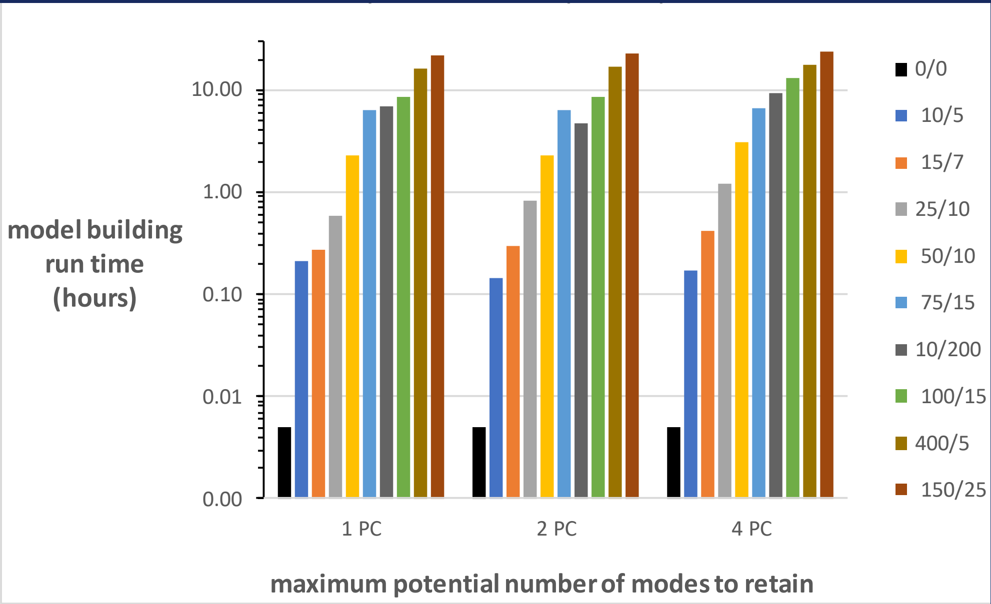


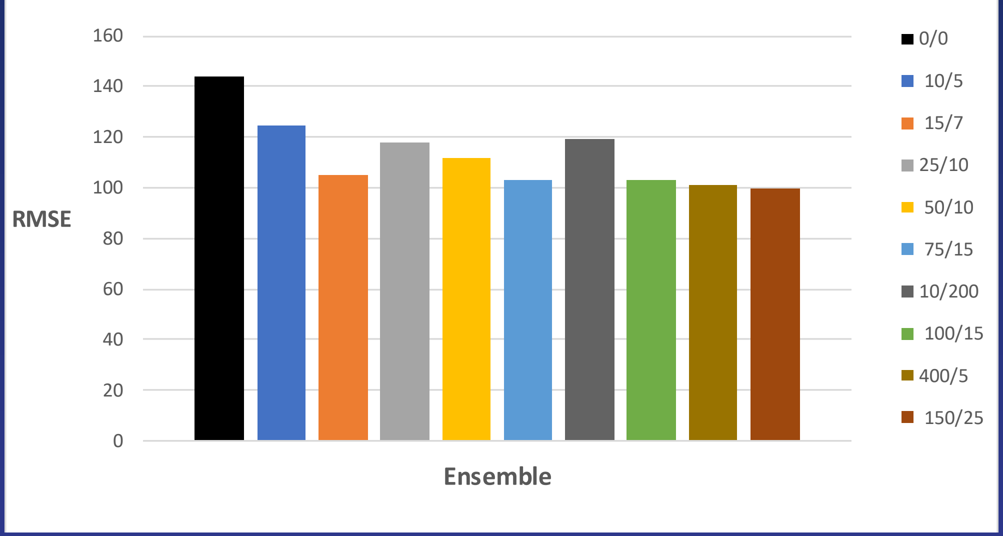
Dovetailing machine learning with process physics

- **Operational hydrologist-directed features engineering**
  - Reflects end-user knowledge around representativeness, reliability, quirks, and capabilities of potential input variables/measurement sites & geophysical interpretations of PCA modes
  - Key location in AI development process for domain experts to insert physical hydrologic knowledge
- **Reframe as a low-dimensional problem with a parsimonious solution**
  - PCA data pre-processing & compact ML architectures enable visualizing input-output relationships
- **Monotonicity constraints**
  - Some feature-target relationships are known to be significantly nonlinear but monotonic
  - Certain AI methods allow a monotonicity constraint to be enforced
- **Non-negativity constraints**
  - Negative-valued flow volume predictions are non-physical yet can happen in some prediction systems
  - Certain AI methods allow a non-negativity constraint to be enforced
  - M^4 also includes algorithmic logic to sequentially prune ensemble members if constraint is violated
- **Such theory-guided data science also encourages both regularization & explainability**

AutoML

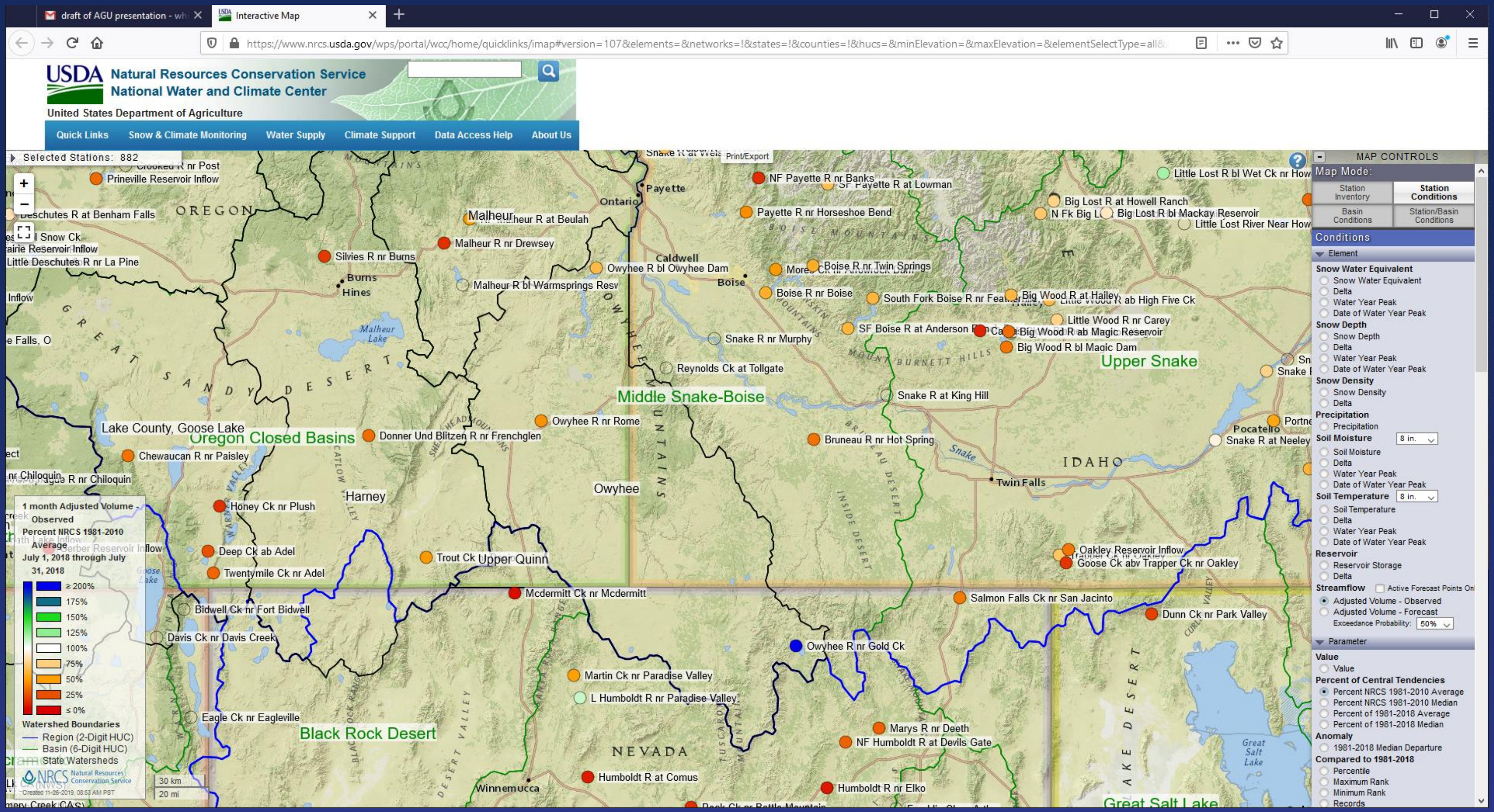
- Completely hands-off operation not desired, but streamlining/automation of many tasks is: leads to judicious use of autonomous ML
- Most optimization/decision points (including value selections for some machine learning hyperparameters) are automated with options for manual overrides
- Others have been pre-calibrated to reasonable values for this application on the basis of extensive experimentation using various NRCS WSF test cases and problem setups
- Example: 100s of metasystem runs to locate generally serviceable default values for pop size & # generations in genetic algorithm



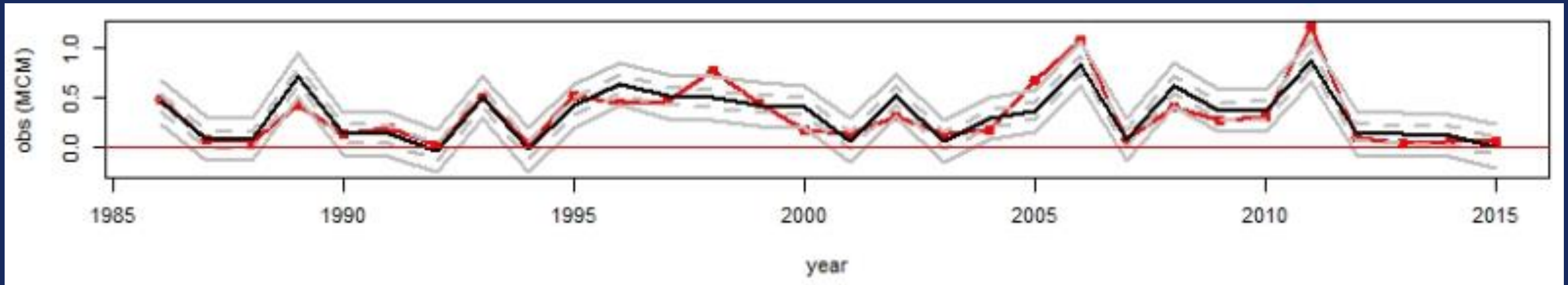


Application example

- Owyhee River near Rome, April 1 forecast of April-July volume
- Particularly troublesome forecast point
- Here, use up to four PCA modes derived from candidate input variable pool of up to 18 SNOTEL SWE & precipitation datasets

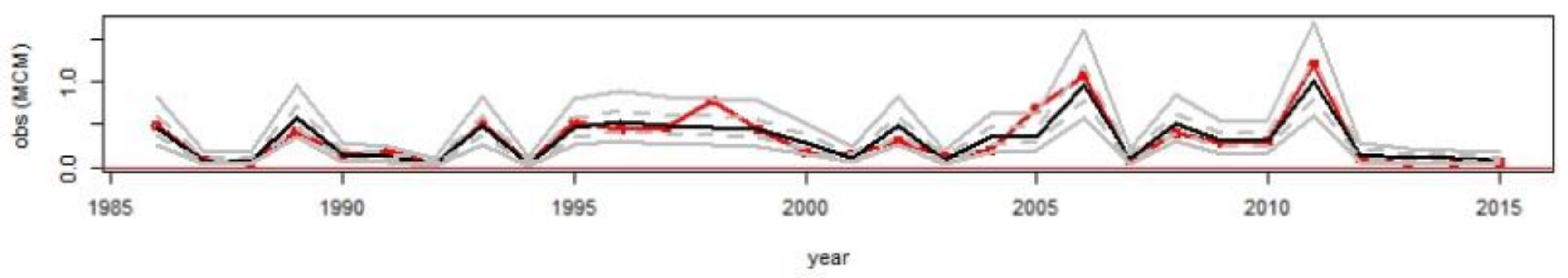


Linear PCR



- **Linear PCR solution produces some best-estimate predictions that are physically impossible: runoff volume can't be negative**
- **Homoscedastic Gaussian assumptions do not generate required time-varying prediction bound widths: too wide in low-flow years, too narrow in high-flow years**
- **This is fixable in PCR using predictand transforms – but it's a manual process, and…**
  - **Time-consuming & labor-intensive**
  - **Dependent on expert opinion & not objectively reproducible: reliability & defensibility issues**
  - **Cannot separate distributional issues from functional form issues**
  - **Basically unsatisfying: should pick models that suit the data, not change data to suit the model**

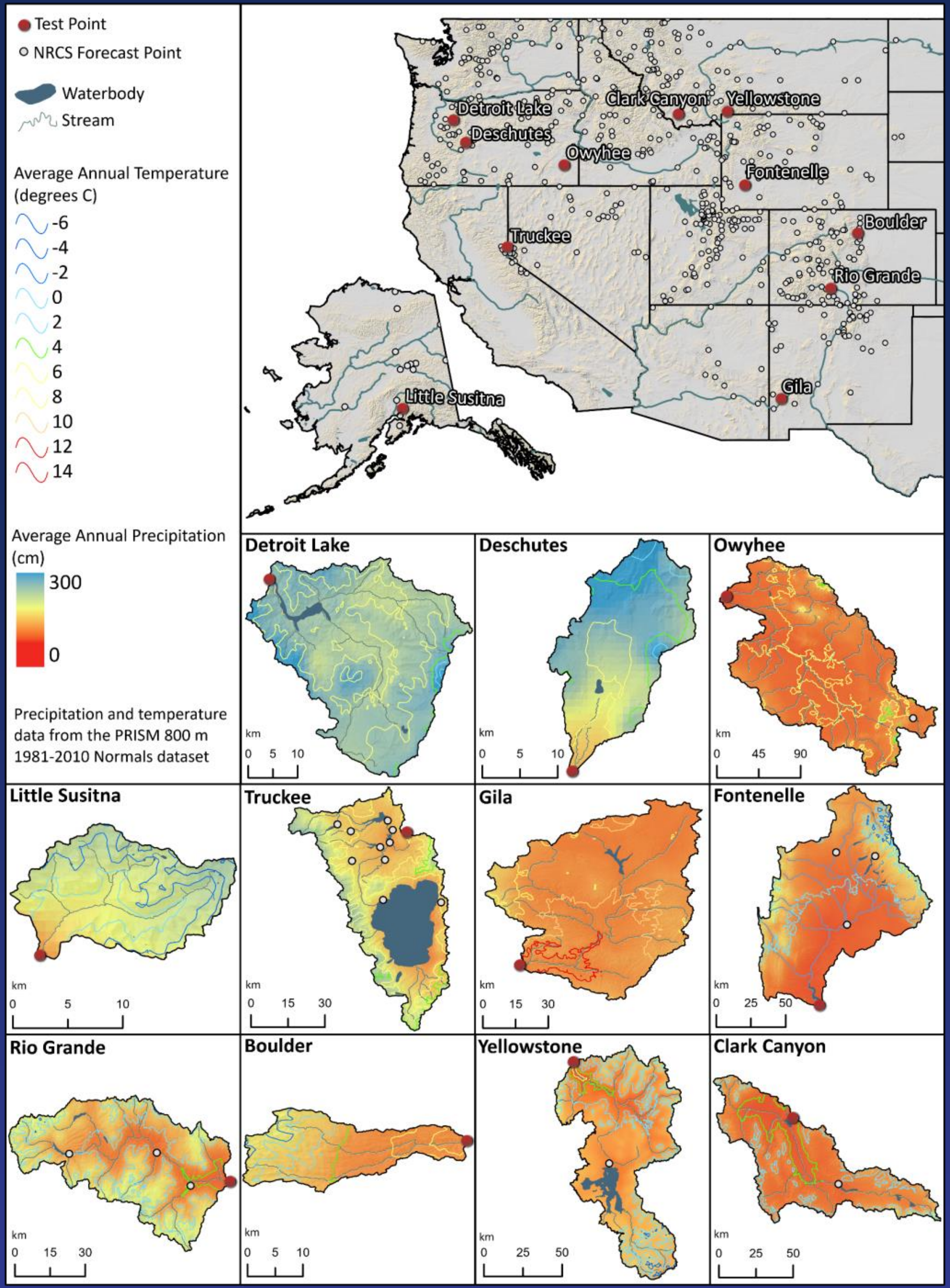
M4



- **Multi-model prediction analytics engine generates forecasts that…**
  - **are physically reasonable, in particular non-negative, from the 90% to the 10% exceedance flows**
  - **have prediction bounds that can vary in width from year to year, if needed**
  - **have prediction bounds that can be asymmetric about the best estimate, if needed**
  - **are relatively robust to outliers due to use of median as best estimate in some constituent models**
  - **are automated, don't require user intervention, & are reproducible: fast/reliable/objective**
  - **show forecast skill better than linear PCR as measured by several cross-validated performance metrics**
- **… yet is still relatively simple and cheap to build and operate**

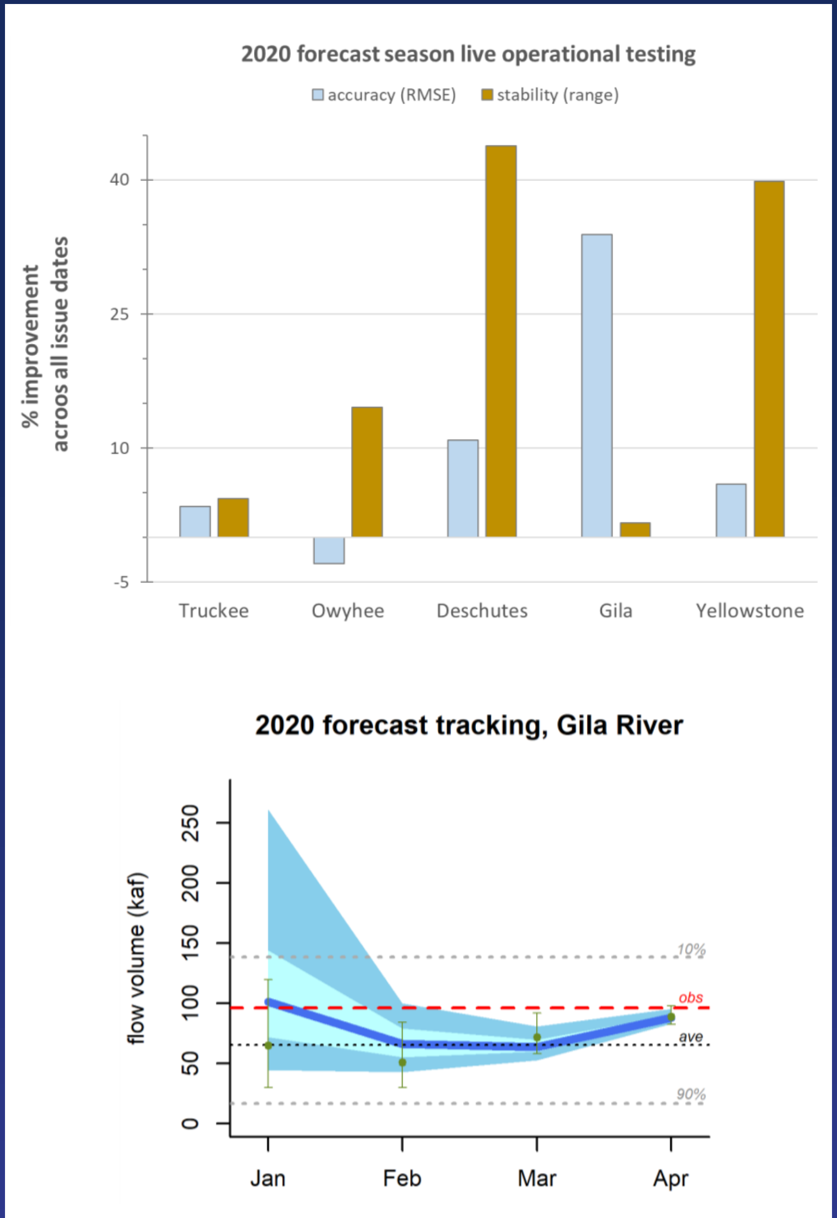
Hindcast testing

- **20 test cases spanning diverse geophysical environments & statistical characteristics across the western US and Alaska**
- **Use PCR as a challenging benchmark: well-established method about on par with ESP**
- **M^4 meets or (almost always) beats linear PCR on LOOCV deterministic & probabilistic performance metrics (improvements of 55% in R^2, 13% in RMSE, 58% in RPSS) + physicality**
- **M^4 ensemble mean more consistent than & often outperforms any of its constituent models**
- **After candidate predictor selection, completely hands-off automated use**



Operational testing

- Live testing at subset of 5 forecast locations during 2020 forecast season alongside existing operational system
- Confirmed logistical feasibility of workflows in time- & data-constrained genuine operational setting
- Easily rebuilt some models on-the-fly due to COVID-19 related snow course data absences
- Initial impressions: generally appears to generate more accurate & more stable forecasts across publication dates relative to existing system
- Confirmed physical explainability of predictions: straightforward 'storyline' for issued forecast in terms of current conditions & hydrologic processes



Paths forward

Current & next steps

- **Further software development**
  - **Streamline code & work with software development team to build production platform for large-scale operational implementation (GUI, IT/database linkages, interactive capabilities around graphics, mapping, and data pre-processing & forecast distribution post-processing)**
- **Other applied R&D directions**
  - **Continue R&D on topics not covered here: supplement PCA with NMFk; additional WSF predictors that a flexible nonlinear AI may capitalize on more effectively than traditional WSF technologies, like snow remote sensing/assimilation data (e.g., SNODAS, MODSCAG, Landsat, ASO, WWA), other remote sensing data (e.g., soil moisture initial conditions), new sources of seasonal climate prediction information (e.g., numerical climate models, data-mined predictive patterns, new ocean-atmosphere circulation indices); integrate other AIs into the modular metasystem**
- **Explore using M^4 as a more generalized forecast integration platform**
  - **Expand scope by ingesting probabilistic forecasts from external process-simulation models (e.g., NWS RFS ESPs, NRCS PRMS ESPs, NOAA National Water Model/WRF-Hydro, etc.) into the multi-technique metasystem's final ensemble – opportunity for re-establishing widespread interagency forecast coordination?**

For further information

- **Machine learning technical details of the M^4 prediction engine have been published in the peer-reviewed AI literature:**

Fleming SW, Goodbody AG. 2019. A machine learning metasystem for robust probabilistic nonlinear regression-based forecasting of seasonal water availability in the US West. *IEEE Access*, 7, 119943-119964, doi:10.1109/ACCESS.2019.2936989

- **Additional papers in progress address application/testing/performance/implications from a water resource science, engineering, and planning perspective & document other directions in machine learning R&D we are pursuing**

- **Please reach out to us with questions & comments: sean.fleming@usda.gov**