From Ensembles to Uncertainty in Hydrological Prediction

Wendy Parker
Department of Philosophy
Virginia Tech
wendyparker@vt.edu

Overview

1. A fitness-for-purpose view of model evaluation
2. Applied to Hydrological Ensemble Prediction Systems (HEPSs)
3. The users' perspective

1. A fitness-for-purpose view

Fidelity/mirror view in modeling

- Models are representations
- Model quality = representational fidelity / fit with world
- Improving a model means adding detail and/or increasing fidelity
- Long-run aim: perfect model

Some limitations of the mirror view

1. It doesn't recognize that we usually build models not just for the sake of representing the world, but in pursuit of other scientific and practical goals.
   - *predicting* when streamflow in a locale will fall above/below important thresholds
   - *understanding* why rainfall is increasing in a locale
   - *improving decisions* about flood defenses, insurance, *to protect life & property*

2. It doesn't recognize that whether goals are achieved using a model generally depends on more than how the model represents the world. It can depend also on (at least):
   - the methodology, including how the user gets initial conditions, computing power, post-processing, etc.
   - features of the model user, such as their background knowledge

3. It appears to overlook that models don't need to represent the world with perfect fidelity to be useful for various purposes.

A different view of modeling

Fidelity/mirror view

- Models are representations
- Model quality = representational fidelity / fit with world
- Improving a model means adding detail and/or increasing fidelity
- Long-run aim: perfect model

Fitness-for-purpose view

- Models are representational tools
- Model quality = fitness-for-purpose, which depends on fit with world, user, methodology, *jointly*.
- Improving a model means adjusting it to better serve purposes of interest
- Long-run aim: a parsimonious set of models that is adequate/fit for a wide range of purposes of interest

A different view of modeling

Fitness-for-purpose view

- Models are representational tools
- Model quality = fitness-for-purpose, which depends on fit with world, user, methodology, *jointly*.
- Improving a model means adjusting it to better serve purposes of interest
- Long-run aim: a parsimonious set of models that is adequate/fit for a wide range of purposes of interest

Some differences in practice

Fidelity/mirror view

- General / overall performance metrics
- Adding detail to the model and increasing its fidelity are always desirable
- Parameter uncertainty is uncertainty about ‘true’ values of parameters

Fitness-for-purpose view

- Purpose-tailored performance metrics
- Adding detail to the model or increasing its fidelity may or may not be desirable
- Parameter uncertainty is uncertainty about which parameter value (if any) would make the model best serve our purpose(s)
  - it might not be ‘true’ value of parameter
  - best values might differ for different purposes

A fitness-for-purpose view: not just for models

- A single model
- An ensemble of models
- A forecast system
- A forecast
- A dataset
- …

If this tool is used by this user, following this methodology, will purpose P be achieved?

Which purposes of interest {P} could the user achieve using this tool, given their resources (computational, background knowledge, methods)?

2. Fitness-for-purpose & HEPSs

Which purposes?

- What purposes do we hope to achieve using our HEPS?
  - accurately gauging / learning about the extent of current **uncertainty** about future hydrological conditions
  - producing calibrated and skillful **probabilistic forecasts** for important hydrological variables, like streamflow
  - providing **users** with information products that aid their decision making

- **It's not the case that fitness for these purposes must stand or fall together; a HEPS might be fit/adequate for some of them but not all.**

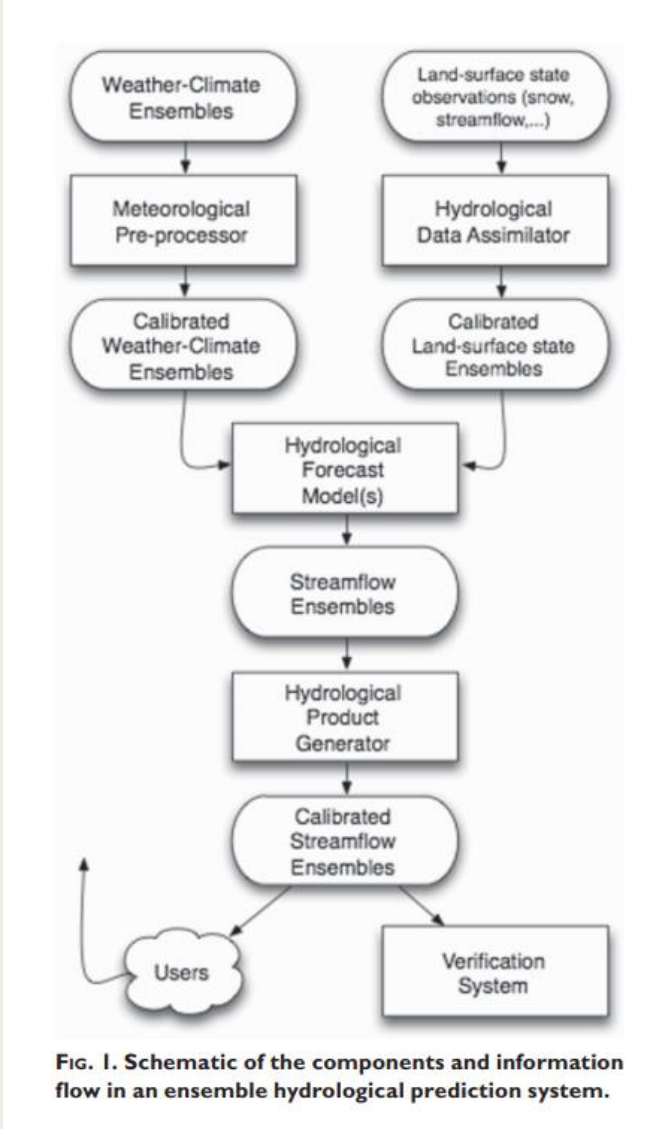


FIG. 1. Schematic of the components and information flow in an ensemble hydrological prediction system.

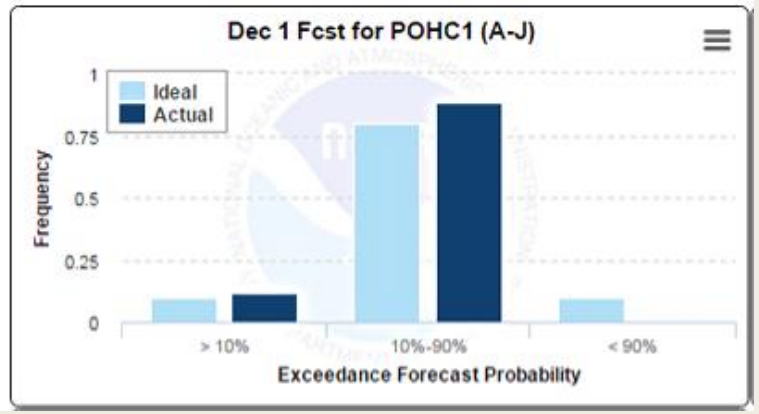
Source: Schaake et al. 2007

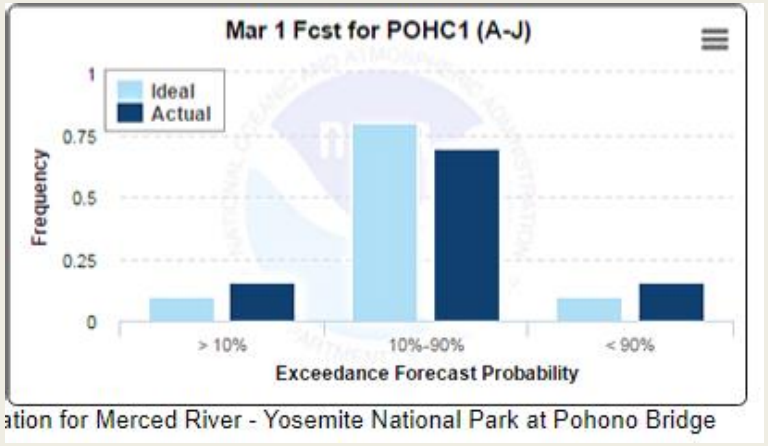
Is our HEPS fit for these purposes?

- **P**: *producing calibrated and skillful probabilistic forecasts for important hydrological variables, like streamflow*

- Our reasons for thinking YES/NO could relate to:
  - HEPS *design*: does it miss some sources of uncertainty?
  - Relevant past *performance*, e.g. verification statistics

- The answer might be YES (calibrated and skillful) for forecasts for some variables/thresholds/locations and NO (not calibrated and skillful) for others.





https://www.cnrfc.noaa.gov/WRverification_help.php

What about when it's "NO"?

- Suppose we find that our HEPS is <u>not</u> fit for ***P:*** producing calibrated & skillful probabilistic forecasts of streamflow in location L.
- Maybe we just need to adjust our "hydrological product generator" or some other element of our HEPS.
- *But, on reflection, we may have good reason to think that our current knowledge and know-how is simply <u>insufficient</u> for producing very-nearly-calibrated probabilistic forecasts for the variable in question*.
  - E.g. due to limited process understanding, structural error in non-linear models
- That is, maybe our current uncertainty about future streamflow in location L is *deeper* than any precise probability / single pdf would imply.

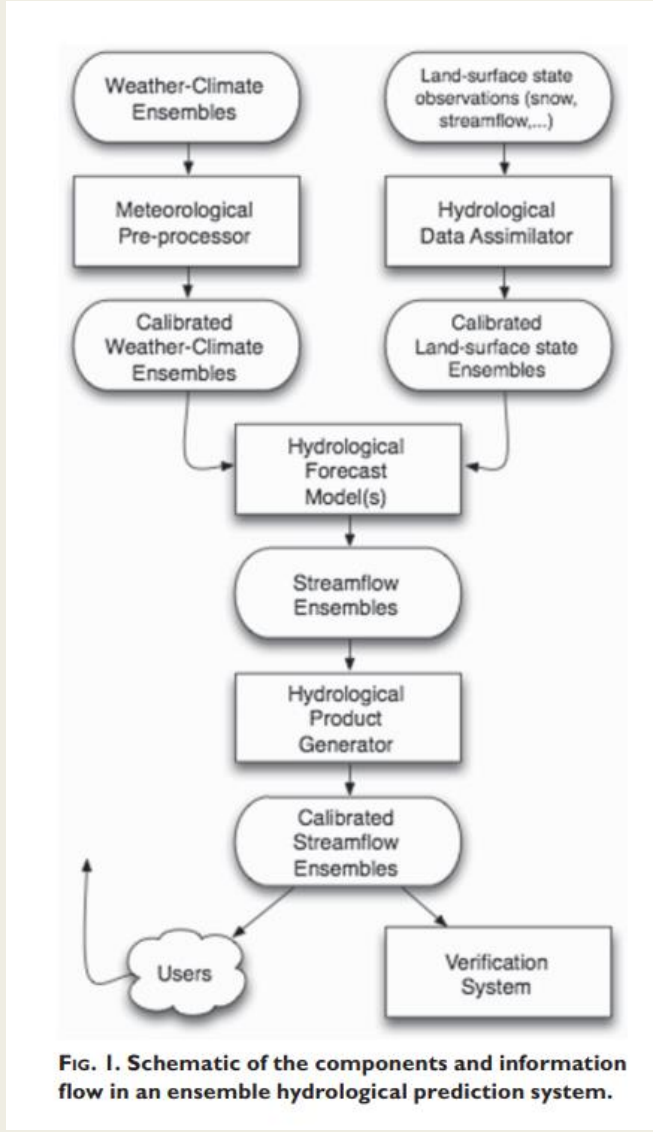


FIG. 1. Schematic of the components and information flow in an ensemble hydrological prediction system.

Source: Schaake et al. 2007

When uncertainty is deeper

- In such situations, a single probability / pdf will have a false precision; it will suggest we know more about future conditions than we actually do.

- Some *other* type of forecast product would more accurately indicate the limitations of what is known about future conditions (i.e. our uncertainty).
  - imprecise probabilistic forecasts, e.g. p(s>T) ≈ 0.33-0.66
  - forecasts of sign of streamflow change only
  - …

- That is, a HEPS that delivered one of these other types of products might have <u>greater</u> fitness for ***P***: accurately indicating the extent of current **uncertainty** about future hydrological conditions.

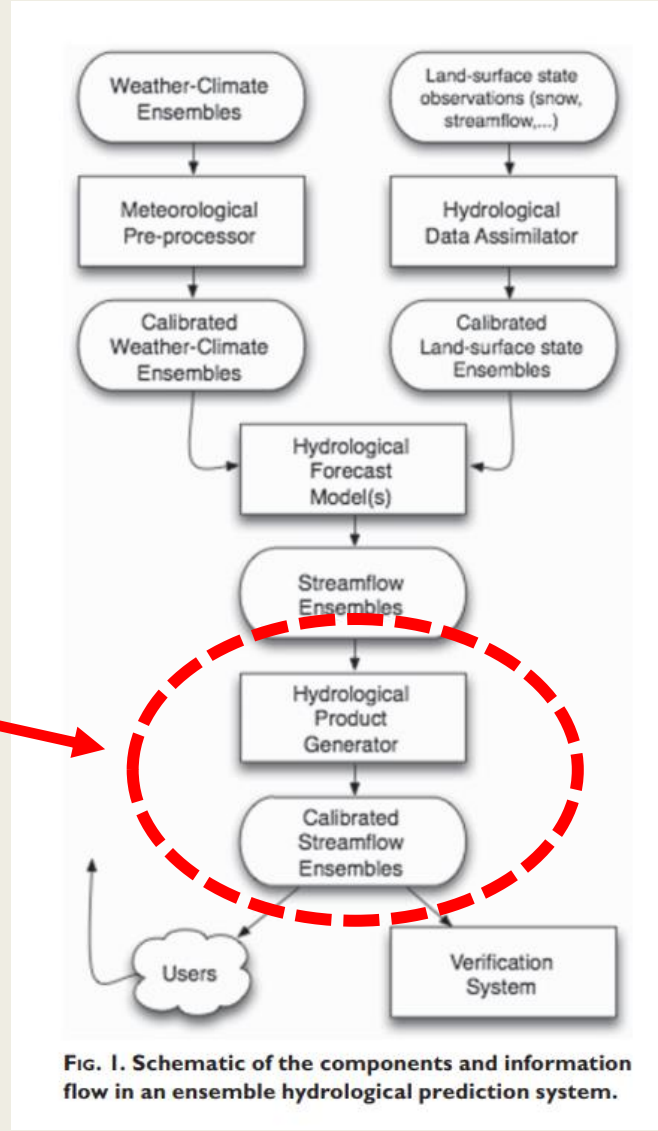


FIG. 1. Schematic of the components and information flow in an ensemble hydrological prediction system.

Source: Schaake et al. 2007

An example from climate science

Table 2 Percentage of IPCC model projections exceeding 2 °C by 2100 and expert assessed likelihood of warming exceeding 2 °C by 2100, for four different emission scenarios (see Collins et al. 2013, Sec.12.4)

Scenario	% projections with $T > 2$ °C	Assessed likelihood of $T > 2$ °C by 2100
RCP2.6	22%	Unlikely (<33%)
RCP4.5	79%	More likely than not (>50%)
RCP6.0	100%	Likely (>66%)
RCP8.5	100%	Likely (>66%)

The question should be: What we can reliably infer from our hydrological ensembles (sets of simulations)? It might **not** be precise probabilities!

Back to our HEPS purposes…

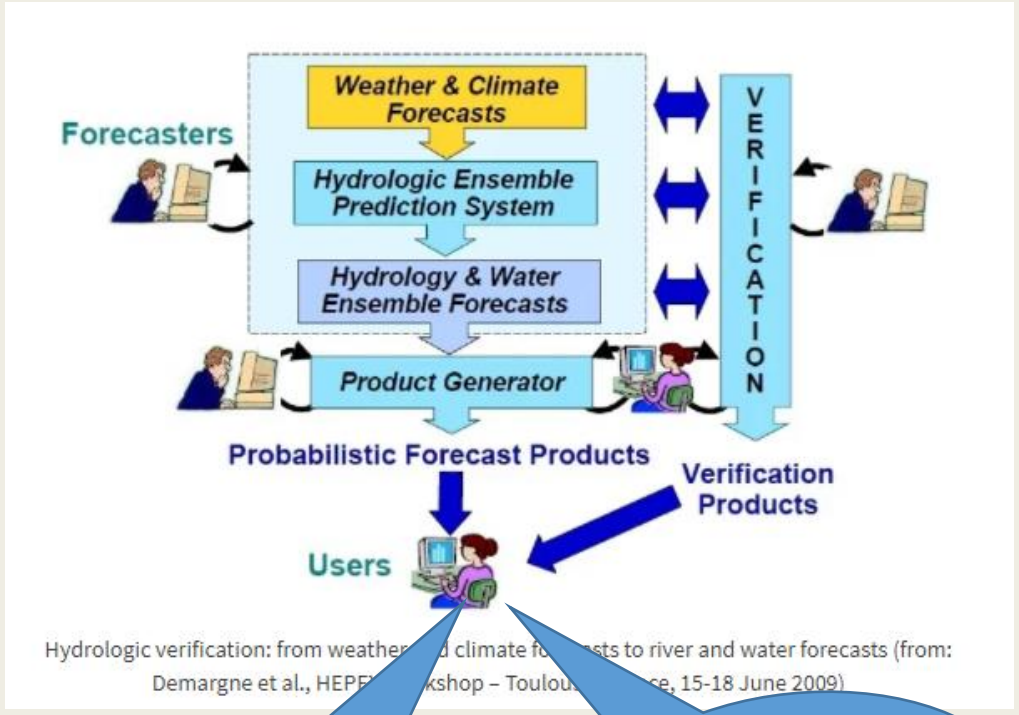
- What purposes do we hope to achieve using our HEPS?
  - gauging / learning about the extent of current **uncertainty** about future hydrological conditions
  - producing calibrated and skillful **probabilistic forecasts** for important hydrological variables, like streamflow
  - providing **users** with information products that aid their decision making

- **It's not the case that these purposes must stand or fall together.**
  - Even when we can't deliver calibrated & skillful probabilistic forecasts, we might be able to use our hydrological ensembles it to learn about our uncertainty.
  - Indeed, HEPS that (for at least some variables) deliver products in other forms might have greater fitness for the purpose of accurately indicating the extent of our **uncertainty**. (Because uncertainty IS deeper.)
  - *Would such products have greater fitness for users' decision making purposes too?*

3. The users’ perspective

Fitness for users' purposes?

- Which type of forecast products will best aid users' decision making can depend on the user, their methodology, the decision problem.
  - The savvy user vs. the inexperienced one who always takes them at face value
  - Users who employ robust decision making frameworks vs. users who try to optimize decisions
- Products that more accurately indicate the extent/depth of uncertainty will only <u>sometimes</u> have greater fitness for aiding a given user's decisions.
- For the scientist, aiming to deliver such products seems the safest option.



Take-aways…

- Try thinking in terms of **fitness-for-purpose**! It works for models, ensembles, HEPS, forecasts, datasets …. Fitness-for-purpose of a representational tool depends not just on its representational fidelity, but how it relates to user and their methodology too.

- **For which purposes are today's HEPSs fit/adequate**? If not for producing calibrated probability forecasts for X, what can we reliably infer from today's ensembles (simulations)? Imprecise probabilities for X, or trends for X, or...?

- HEPS that deliver products in a form other than precise probabilities might have **greater fitness** for the purpose of accurately indicating the extent of our **uncertainty** and for (some) **users**' practical decision making purposes too.

Some references

Bokulich, A. and Parker, W.S. 2021. “Data Models, Representation, and Adequacy-for-Purpose”. *European Journal for Philosophy of Science*.

Haasnoot, M. et al. 2014 “Fit for purpose? Building and evaluating a fast, integrated model for exploring water policy pathways”, *Environmental Modelling and Software*.

Parker, W.S. 2013. “Ensemble modeling, uncertainty and robust predictions”, *WIREs: Climate Change*.

Parker, W.S. and J.S. Risbey. 2015. “False precision, surprise and improved uncertainty assessment”, *Philosophical Transactions of the Royal Society A.*

Parker, W.S. 2020 “Model Evaluation: An Adequacy-for-Purpose View”, *Philosophy of Science*.

Mastrandrea MD, et al. 2010. *Guidance note for lead authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties*. See https://www.ipcc-wg1.unibe.ch/guidancepaper/ar5_uncertainty-guidance-note.pdf.