1st HEFS workshop, 08/20/2014

Seminar E: Basic Theory of the HEFS Hydrologic Ensemble Post-processor (EnsPost)

James Brown

james.brown@hydrosolved.com

Contents

1. Why model hydrologic error?

Why model hydrologic error?

Recall, total flow uncertainty includes

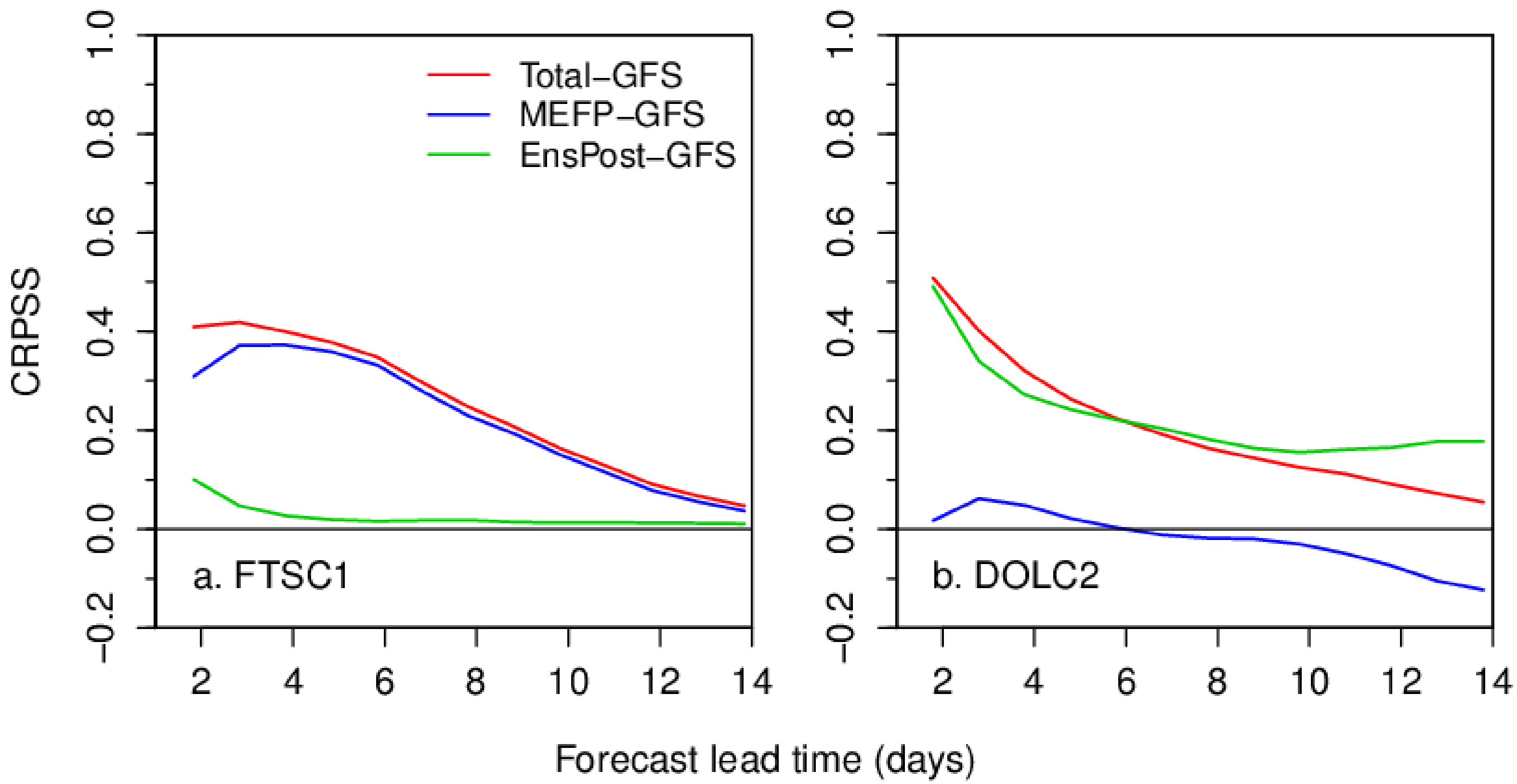
1. Meteorological forecast uncertainties/biases (MEFP)
2. Hydrologic modeling uncertainties/biases (EnsPost)

- (decision-related uncertainties not addressed here)

Hydrologic uncertainty is important!

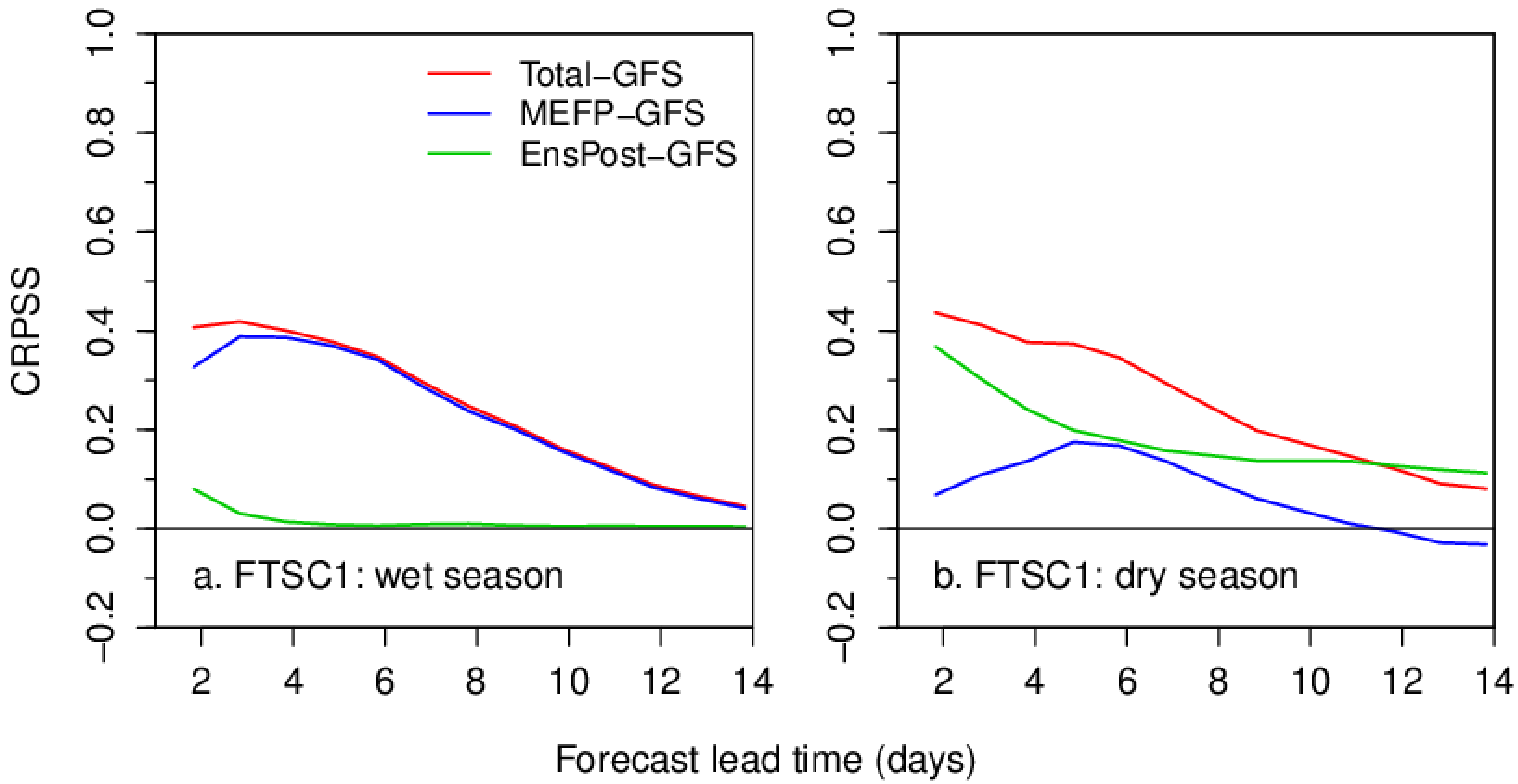
- Model structure, parameters, initial conditions, states
- Also, uncertainty from river regulations and MODs
- EnsPost aims to account for these in a lumped sense
- Without this, greater bias and less skill (e.g. MMEFS)

Importance varies between basins



- Fort Seward, CA (FTSC1) and Dolores, CO (DOLC2)
- Total skill in HEFS streamflow forecasts is similar
- Origins are completely different (FTSC1=forcing, DOLC2=flow)

But also within basins (e.g. season)



- Hydrologic uncertainty can be important under specific conditions
- In wet season (which dominates overall results), mainly MEFP skill
- In dry season, skill mainly originates from EnsPost (persistence)

2. How to model hydrologic error?

Defining an error model

Recall the definition of error

- Error = “true” (observed) value – predicted value
- Goal: model random errors statistically (uncertainty)
- Goal: remove any systematic errors (biases)

Isolating the hydrologic error

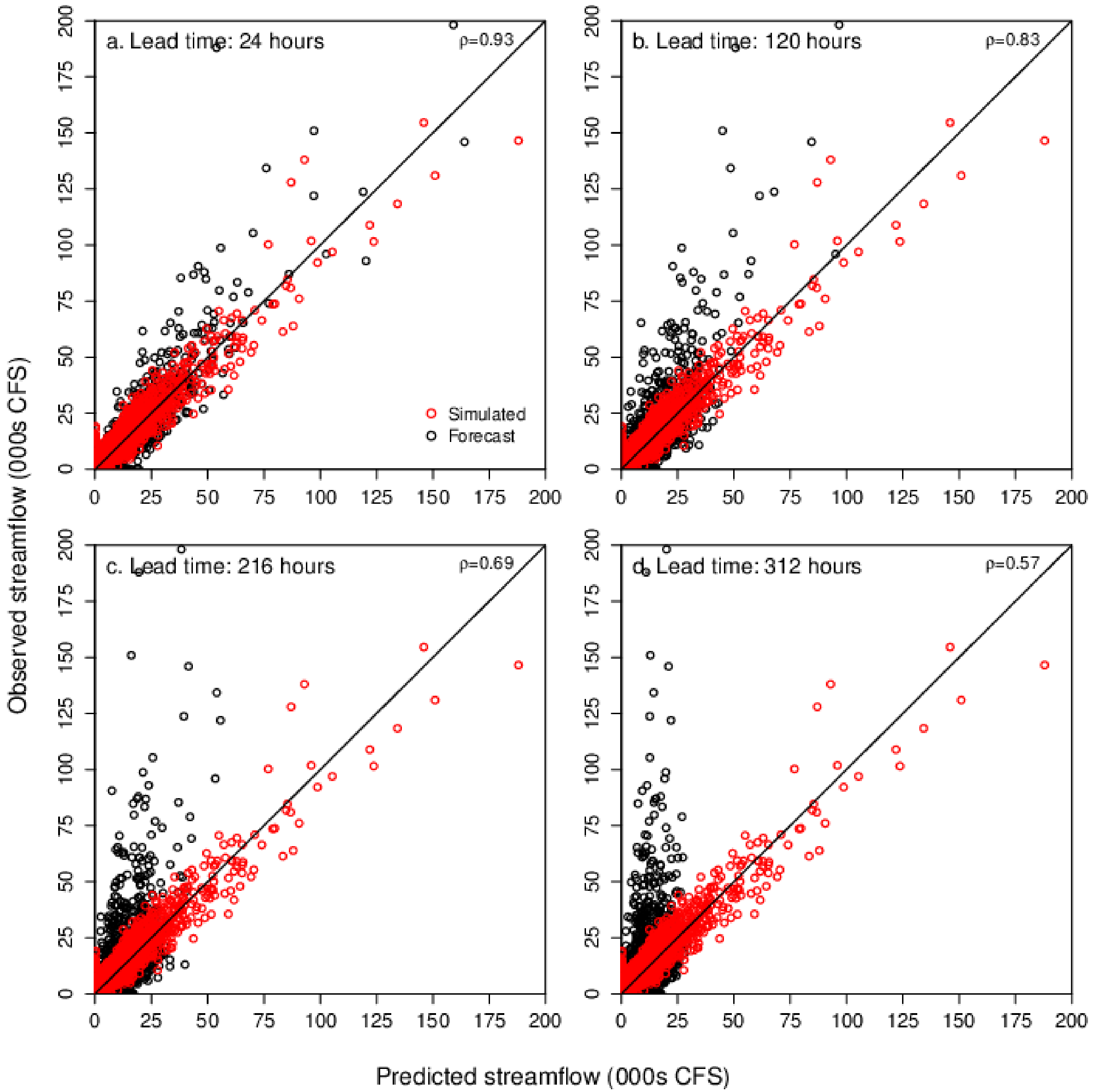
1. Forecast – observed streamflow (total error)
2. Simulated – observed streamflow (hydrologic error)

- Simulated flows are produced with observed forcing
- Add the meteorological forecast error (MEFP) later on

Why isolate hydrologic error?

Lead-time independence

- Example from Lake Oroville inflow (ORDC1) in CNRFC
- Plots show observed flows paired with forecasts (24-312 hours) and simulations
- Scatter denotes total error (forecasts) and hydrologic error (simulations)
- Total error increases with forecast lead time due to forcing error. Also, forecast biases increase!
- Hydrologic error/bias is invariant to lead time



Building a statistical model

Gather sample of historical errors

- Hydrologic error = simulated – observed streamflow

1. Collect historical pairs (ideally a large sample!)
2. Use historical errors to train a statistical model
3. Predict statistical distribution of future errors

Assumptions (there are several)

- The observed forcing is "error free" (for simulations)
- The observed streamflows are "error free"
- The MEFP adequately corrects meteorological bias

Summary

HEFS design

- Total error includes forcing and hydrology
- Forcing errors are modeled statistically by MEFP
- Hydrologic errors are modeled statistically by EnsPost

Hydrologic error

- Can be modeled with historical error sample, where…
- Hydrologic error = simulated – observed streamflow
- Hydrologic errors are invariant to forecast lead time
- Thus, one model/parameter set for all lead times

3. Structure of the EnsPost error model

Required characteristics of EnsPost

1. Model the hydrologic error only (not forcing error)
   - In HEFSv1, treat the hydrologic error as lumped
   - In future, may address error sources (e.g. DA)
2. Parsimonious, i.e. few parameters, not data hungry
3. Remove biases and add “reliable” spread
4. As a minimum, forecast climatology should be reliable
5. Capture seasonal and amount-dependent errors
6. Forecast time-series must be realistically smooth

A simple statistical model?

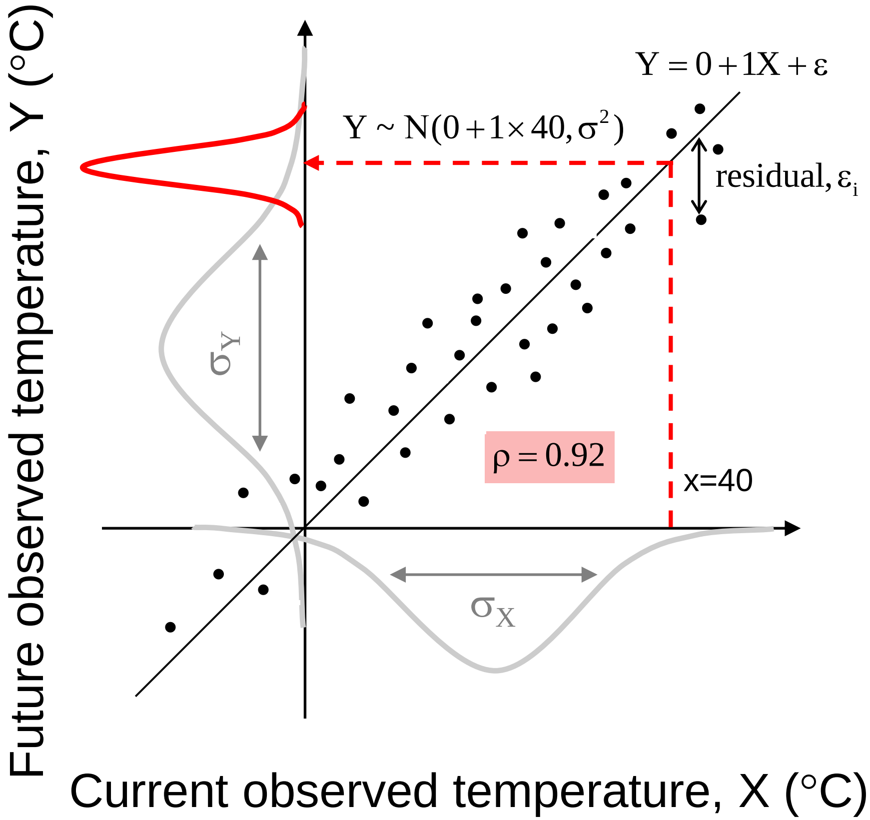
HSL

Linear regression

- Recall the following:

1. Two bivariate normal variables, (X,Y), are linearly related with correlation, ρ, that describes the strength of the relationship
2. Subject to (1), output (Y) of linear regression is normal with mean, $\alpha+\beta X$, and variance (σ^2) equal to variance of residual, ε, which is also normal

- <u>However</u>, unlike temperature, streamflow does not ordinarily follow a normal distribution

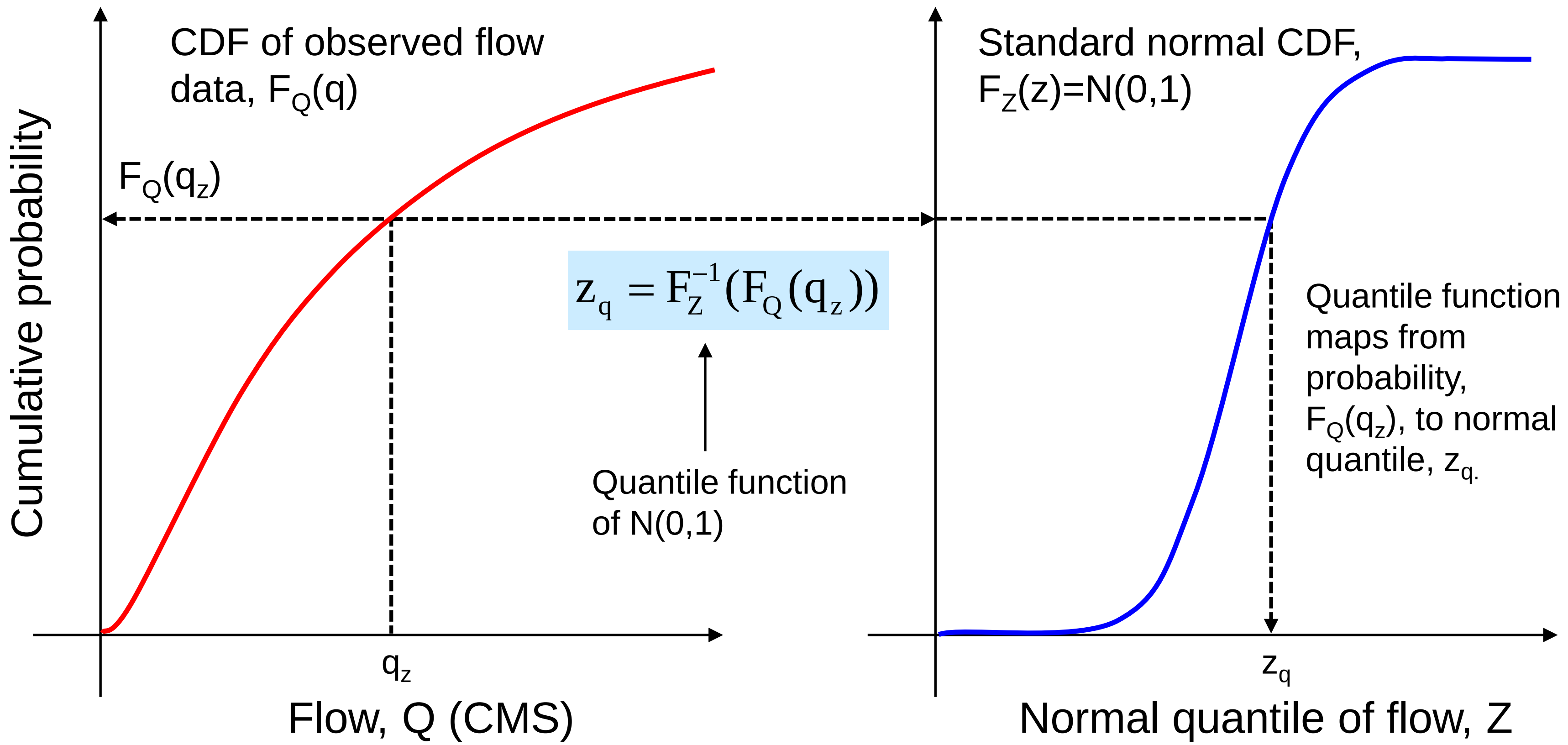


$$Y = \alpha + \beta X + \varepsilon$$

$$\beta = \rho \frac{\sigma_Y}{\sigma_X}, \varepsilon \sim \text{Normal}(0, \sigma^2)$$

Normal Quantile Transform (NQT)

Apply NQT to model variables (obs, sim)

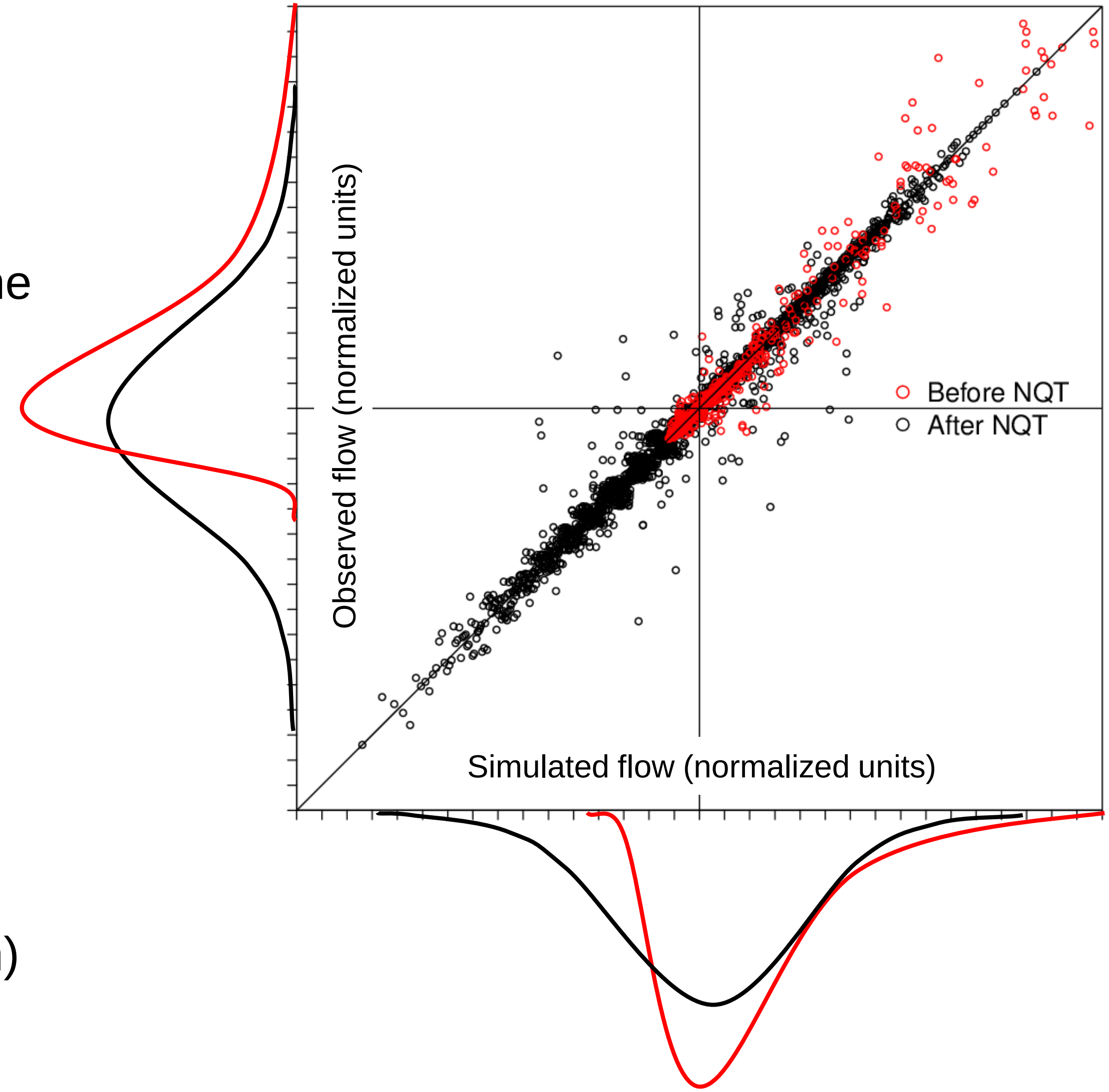


- Back-transform predictions w/ observations (reliable climatology) ✔

Effects of NQT on scatter

Scatter in normal space

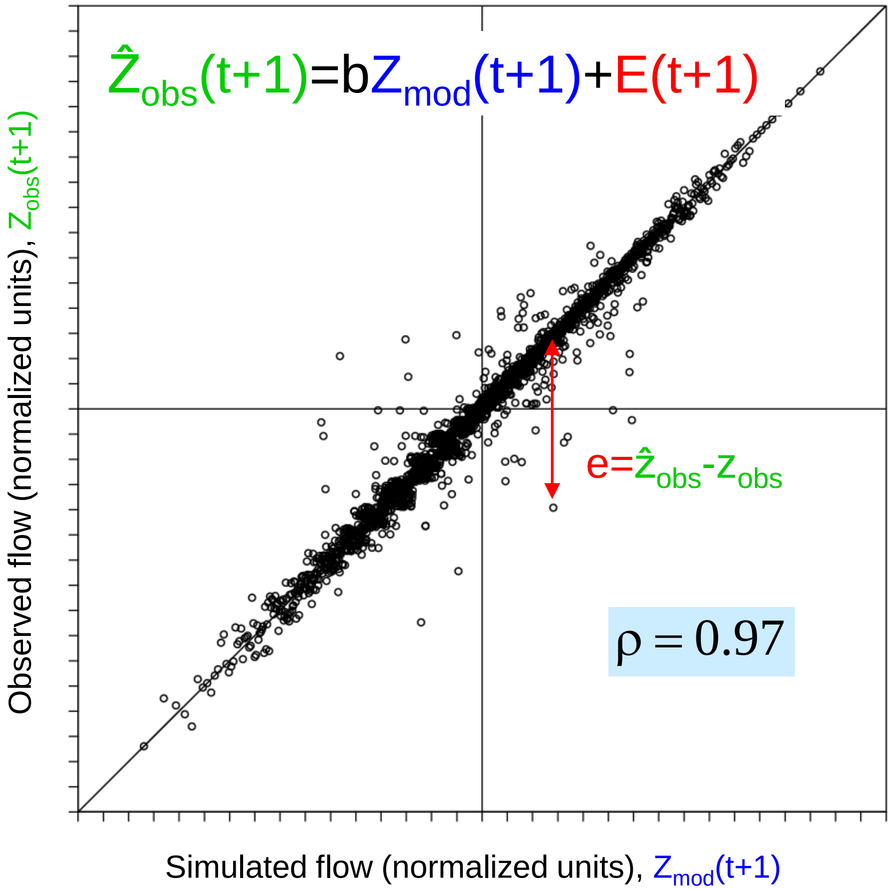
- Plot shows scatter before and after NQT
- Effects of NQT are to center the scatter at zero and stretch the scatter on each axis
- By construction, the observations and simulations are now normal on their own
- However, this does not mean that they are bivariate normal
- We assume that the variables are bivariate normal (there are ways to check this assumption)



The basic EnsPost error model

Linear regression model

- **What we want:** future observed flow at time t+1, $Z_{obs}(t+1)$, that captures hydrologic error
- **What we have:** model predicted streamflow at time t+1, $Z_{mod}(t+1)$, i.e. a hydrologic simulation
- **What we assume:** a good estimate of Z_{obs} is given by: $\hat{Z}_{obs}(t+1)=bZ_{mod}(t+1)$, where b is a regression parameter. The curve passes through the origin (0,0)
- **What we accept:** our model is imperfect (hence scatter). We have an error term, $E(t+1)$, that represents the hydrologic error



$$\hat{Z}_{obs}(t+1) = bZ_{mod}(t+1) + E(t+1)$$
$$E(t+1) \sim \mathrm{Normal}(0, \sigma^2)$$

Other useful predictors?

- Hydrologic persistence (e.g. dry conditions, snowmelt)
- Use lagged observation as predictor (cf. Adjust-Q)

Leveraging hydrologic persistence

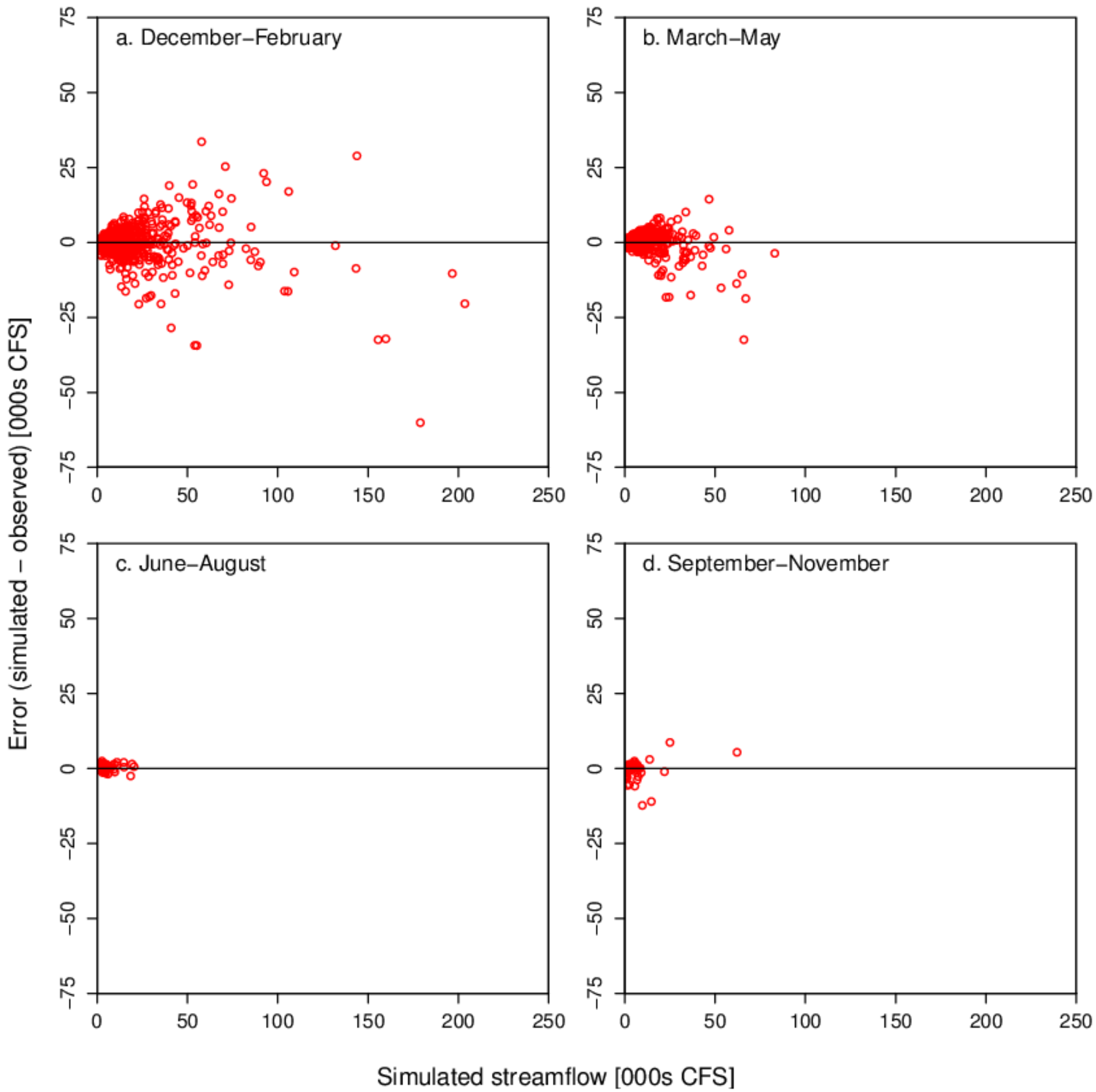
$$\hat{Z}_{obs}(t+1)=(1-b)Z_{obs}(t)+bZ_{mod}(t+1)+E(t+1)$$

- Notice (1-b)+b=1. Ensures overall unbiasedness ✓
- Autoregressive, should be smooth time-series ✓
- New information should reduce error, E(t+1)

Improving the basic model

Seasonality of error

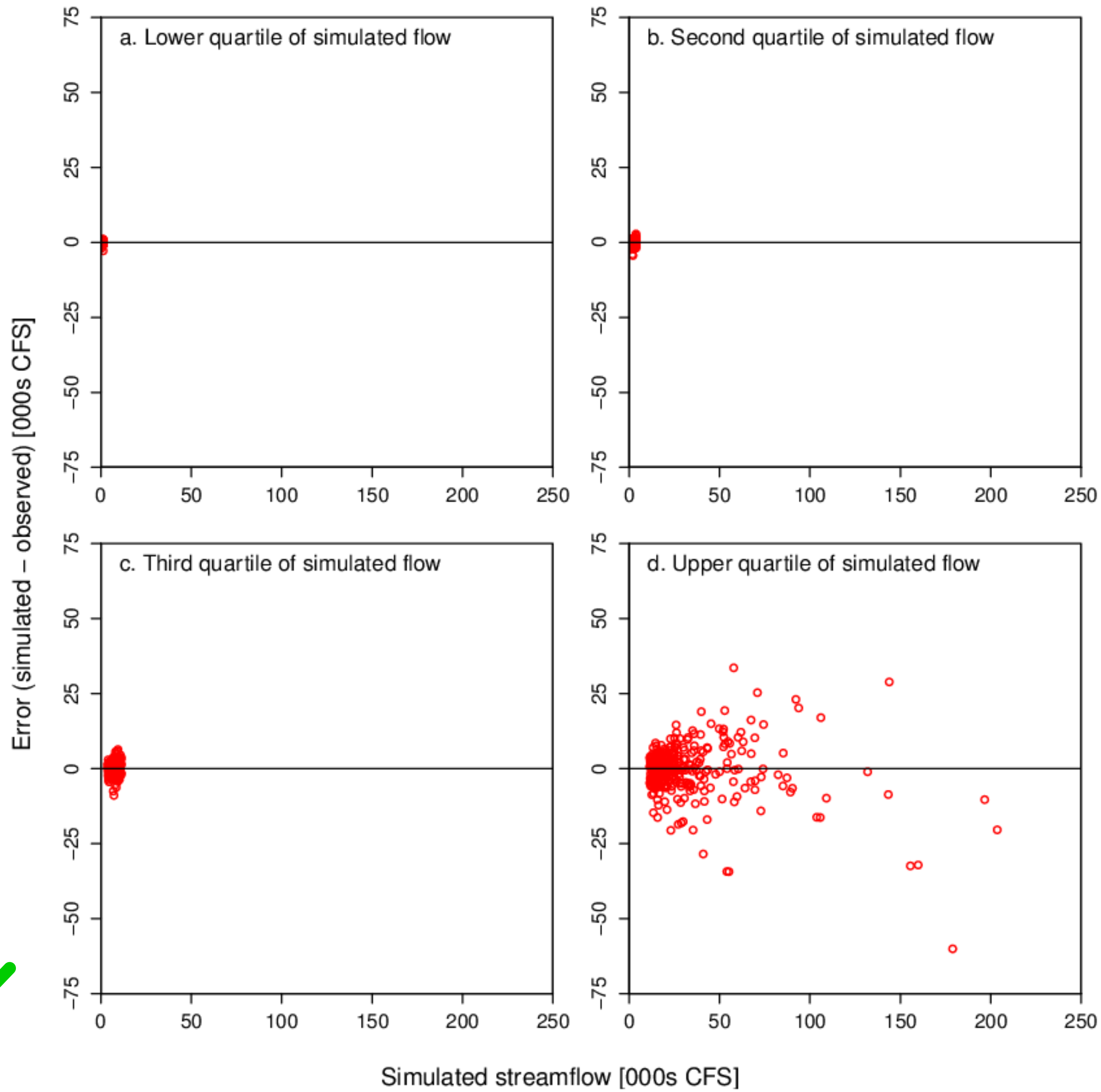
- Example of seasonality of hydrologic model error at Fort Seward (FTSC1) in CA
- Typical that hydrologic model error varies during year with seasonal climate
- Attempting to fit a single error model can be problematic
- Instead, break the paired data into seasons (e.g. two 6-month periods) and model separately, i.e. account for seasonality ✓



Improving the basic model

Amount-dependent error

- Example of amount-dependence of hydrologic model error at Fort Seward (FTSC1) in CA, Dec-May (quartiles of simulations)
- Typical that hydrologic error varies with flow amount in each season
- Fitting a single error model can be problematic
- Instead, break the paired data into categories (e.g. above/below simulation median) and model each flow category separately



HSL

Some notable characteristics

$$\hat{Z}_{obs}(t+1)=(1-b)Z_{obs}(t)+bZ_{mod}(t+1)+E(t+1)$$

1. If b=0, $\hat{Z}_{obs}(t+1)=Z_{obs}(t)+E(t+1)$: all persistence
2. If b=1, $\hat{Z}_{obs}(t+1)=Z_{mod}(t+1)+E(t+1)$: all from model

- This applies in degrees, i.e. when b is closer to 0 or 1
- Thus, b provides valuable insight about the model

3. If b=1 & $E(t+1)=0$, $\hat{Z}_{obs}(t+1)=Z_{mod}(t+1)$: clim. correction

- Can enforce this with "ER0" option, e.g. for long-range

4. Parameter estimation

Regression coefficient

Single model parameter, b

$$\hat{Z}_{obs}(t+1)=(1-b)Z_{obs}(t)+bZ_{mod}(t+1)+E(t+1)$$

- One parameter to estimate, b, a regression coefficient
- Highly parsimonious, reduces sampling noise ✓

How to estimate b using EnsPost PE?

- **Goal:** choose b so that model fits data “optimally”
- Optimal in what space? NQT or original/flow space?
- Optimal in what sense? Different error measures

Regression coefficient

Model space or flow space?

- Parameter, b, is simple to estimate in model space
- However, we care about performance in flow space
- Thus, all error measures are defined in flow space

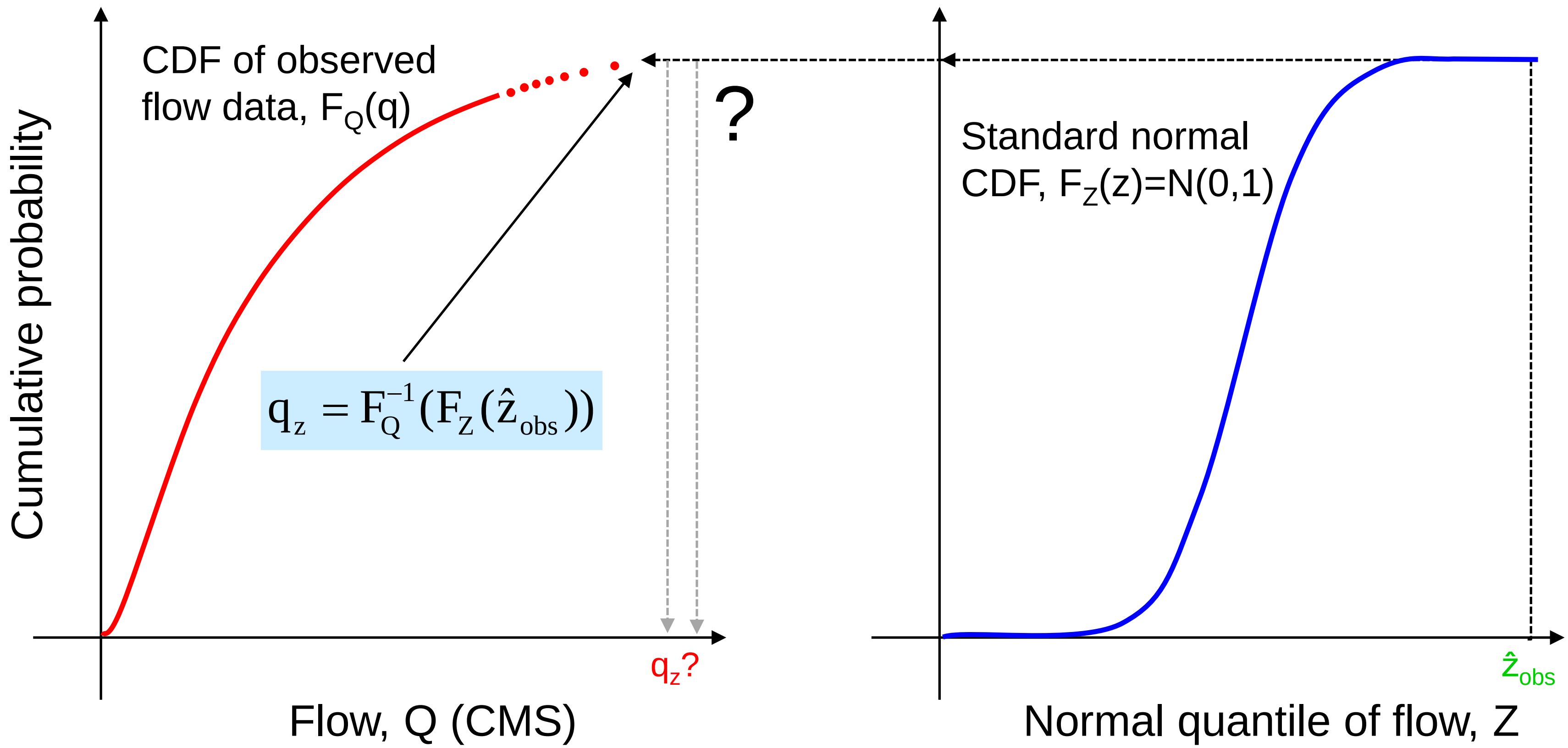
Three error measures (see extra slides)

1. Climatological error of ensemble mean
2. Conditional error of ensemble mean (MSE)
3. Conditional error of full ensemble distribution (CRPS)

- EnsPost PE allows a weighted combination of 1-3

Back-transform in upper tail

What if $\hat{Z}_{obs}$ exceeds historical data?



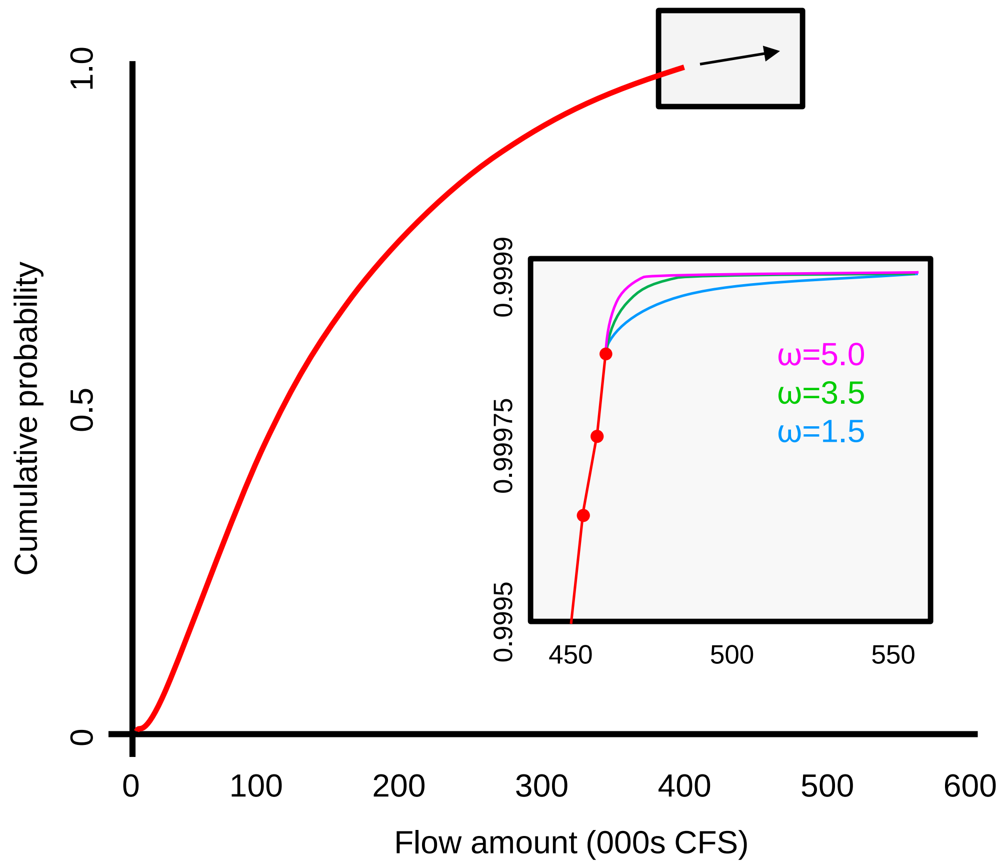
- Model predictions can exceed historical data. Need a model for tail.

Back-transform in upper tail

HSL

Upper tail is unknown

- EnsPost provides a model for the shape of the upper tail
- The "fatness" of the tail is controlled by a parameter, ω
- However, this is **guesswork**, i.e. the upper tail is unknown
- Thus, care is needed when considering flows approaching and exceeding the historical maximum (can constrain in PE)
- More generally, this is a limitation of <u>any</u> statistical technique that relies on limited historical data



$$q_z = \left[\lambda \Big/ 1 - F_z(\hat{z}_{obs}) \right]^{\frac{1}{\omega}}$$

Guidance on estimating parameters

Basic choices to make

- Seasonality (no default)
- Flow amount category (median by default)
- Error measures for optimizing regression parameter, b
- Fatness of upper tail for extreme flows, ω

Recommendations (see EnsPost manual)

- Focus efforts on choosing seasonality (no default)
- Other parameters more advanced or "trial-and-error"
- Use default settings unless time to experiment

5. Mechanics of ensemble generation in real-time

Stages in generating ensembles

Goal: add hydrologic/simulation error

- Adjust raw streamflow forecast (forcing error only)…
- …by adding simulation error to raw forecast

To generate a single ensemble trace

1. Transform observed, $z_{obs}(t)$, & forecast, $z_{mod}(t+1)$
2. Draw random sample from $E(t+1)$, namely $e(t+1)$
3. Compute $\hat{z}_{obs}(t+1)=(1-b)z_{obs}(t)+bz_{mod}(t+1)+e(t+1)$
4. Do steps 1-3 for t+2,…,t+M, substituting $\hat{z}_{obs}$ for z_{obs}
5. Back-transform $\hat{z}_{obs}(t+1)$,…, $\hat{z}_{obs}(t+M)$ to real flow units

Generating a single trace

$t = T_0$ — $t+1$ — $t+2$ — $t+3$

$z_{obs}(t)$

Raw flow forecast

$\hat{z}_{obs}(t+1) = (1-b)z_{obs}(t) + bz_{mod}(t+1) + e(t+1)$

Sample of hydrologic error at (t+1) → $+e(t+1)$

Raw flow forecast

$\hat{z}_{obs}(t+2) = (1-b)\hat{z}_{obs}(t+1) + bz_{mod}(t+2) + e(t+2)$

Sample of hydrologic error at (t+2) → $+e(t+2)$

Raw flow forecast

$\hat{z}_{obs}(t+3) = (1-b)\hat{z}_{obs}(t+2) + bz_{mod}(t+3) + e(t+3)$

Sample of hydrologic error at (t+3) → $+e(t+3)$

6. Practical considerations

Data considerations

How much calibration data is “enough”?

- Generally not an issue for EnsPost (long records)
- Record length of 20 or more years recommended
- However, it also needs to be reasonably “stationary”
- E.g., no major changes in river basin conditions

Data quality control

- Important to QC the observations and simulations
- As a statistical technique, can be sensitive to outliers
- Some error measures (for b) are sensitive (e.g. MSE)

River regulations and MODs

Can undermine EnsPost assumptions

- EnsPost treats historical data as stationary/stable
- Regulations and MODs can introduce instabilities
- Regulations difficult to isolate by season or amount
- MODs alter operational versus historical simulations

Use unregulated/unmodified flows

- If available, use natural flows in regulated basins…
- …assumes that regulations are known in real-time
- If impractical, do validation with and without EnsPost

Limitations of EnsPost

General limitations

- Lumps all hydrologic error into one residual
- In practice, different sources are highly differentiated
- Ideally, residual error would be less structured (whiter)
- For example, data assimilation will whiten residual

Specific limitations (EnsPost manual)

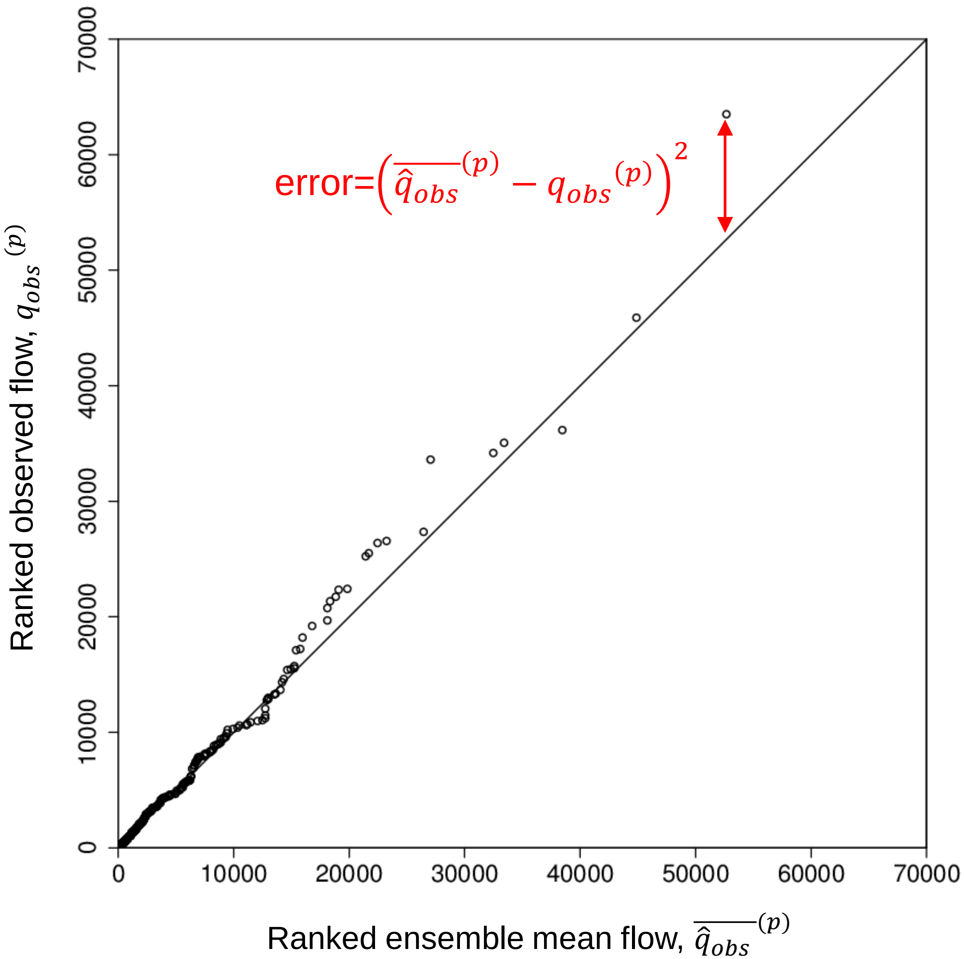
- Ephemeral streams (akin to modeling PoP in MEFP)
- Extreme events: EnsPost may add statistical noise
- Downscaling of daily flow is poor (use 6-hr observed)

Extra slides

Climatological error of mean

Climatological error

- Ensemble mean is the "best estimate" from the EnsPost
- Important that the climatology of these best estimates is similar to the climatology of the observations
- One way to show this is a quantile-quantile plot (right)
- This involves <u>separately</u> ranking the best estimates, $\overline{\hat{q}_{obs}}^{(p)}$ and the observations, $q_{obs}{}^{(p)}$ where (p) is the rank
- Compute the mean-square error of the ranked data from the diagonal (i.e. observed climatology)

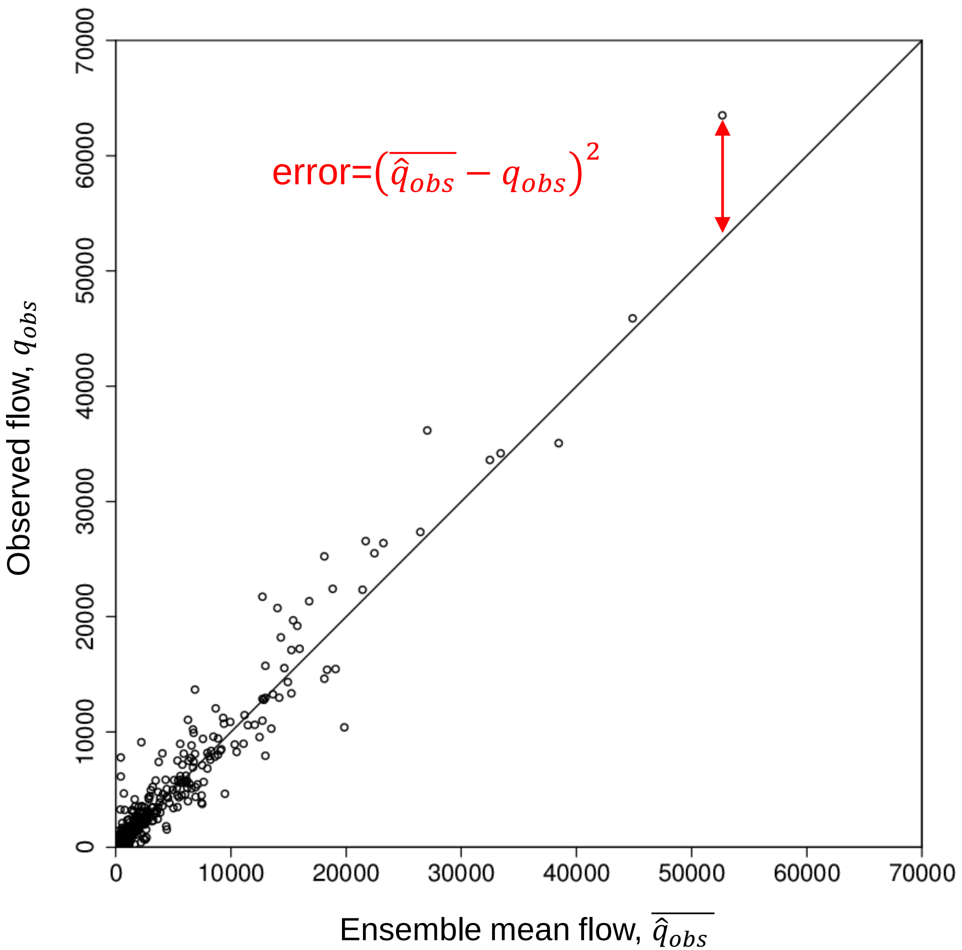


$$\text{error} = \frac{1}{n}\sum_{p=1}^{n}\left(\overline{\hat{q}_{obs}}^{(p)} - q_{obs}{}^{(p)}\right)^2$$

Conditional error of mean

Errors of paired data

- Measures defined for individual pairs, ($\overline{\hat{q}_{obs}}$, q_{obs}), are "conditional" because they preserve the relationship between the predictions and observations
- A well-known measure of conditional error is the Mean Square Error (MSE) for the pairs
- This is simply the average square deviation of the predictions from the diagonal. In this case, the prediction is the ensemble mean
- This measure is sensitive to outliers at high streamflow amounts

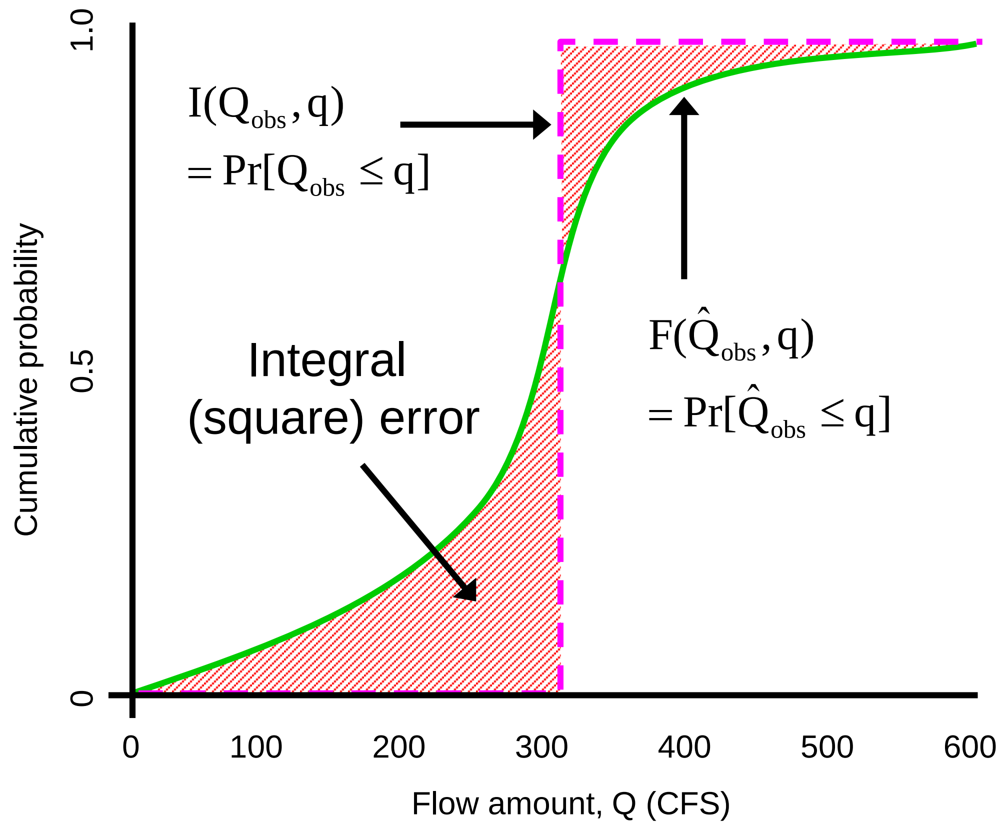


$$\text{error} = \frac{1}{n}\sum_{i=1}^{n}(\overline{\hat{q}_{obs}} - q_{obs})_i^2$$

Conditional error of ensemble

Error of full ensemble

- Also defined for individual pairs, except the full forecast probability distribution is considered
- A well-known error measure of an ensemble or probability forecast is the Continuous Ranked Probability Score (CRPS)
- Measures the integral square difference between the forecast and corresponding observation (step function). Then averaged over n pairs
- This measure is smooth and less sensitive to outliers at high streamflow amounts



$$\text{error} = \frac{1}{n}\sum_{i=1}^{n}\int [F(\hat{Q}_{obs}, q) - I(Q_{obs}, q)]_i^2 \, dq$$

Hydrologic errors are constant

Possible source of confusion

- Didn't we say the hydrologic error is constant?
- I.e. it does not vary with forecast lead time?
- Yet, $z_{obs}(t)$ reduces error at short lead times!

So what do we mean by "constant"?

- We mean the underlying error distribution
- Reached when effect of $z_{obs}(t)$ "wears off"
- Sounds like a technical detail, but it is important
- The EnsPost has one parameter, b, and it is constant