Testing modeling frameworks or… how can we improve predictive skill from our data?

Mary C. Hill, University of Kansas; Emeritus, U.S. Geological Survey

Laura Foglia, Technical University of Darmstadt (soon at UC Davis)

Ming Ye, Florida state University

Lu Dan, Oak Ridge National Laboratory

Frameworks for Model Analysis

- Look for surprises
- Surprises can indicate

  New understanding about reality

  or

  Model limitations or problems
- Fast methods are convenient -- enable routine evaluation. Computationally frugal, parallelizable methods

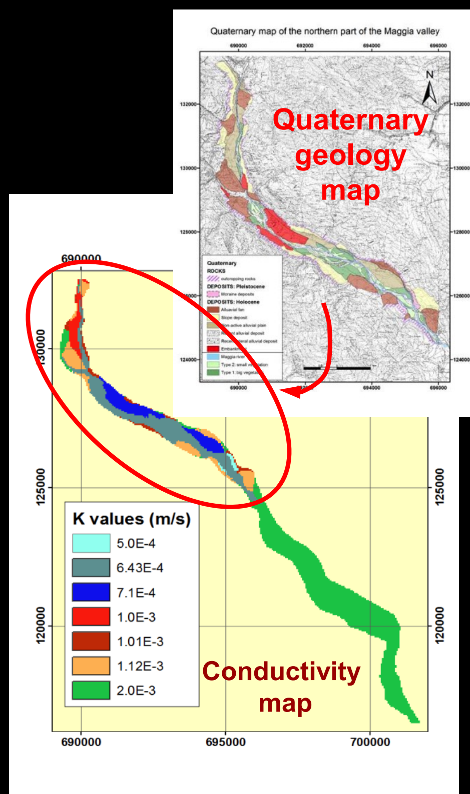
What do I mean by a modeling framework?

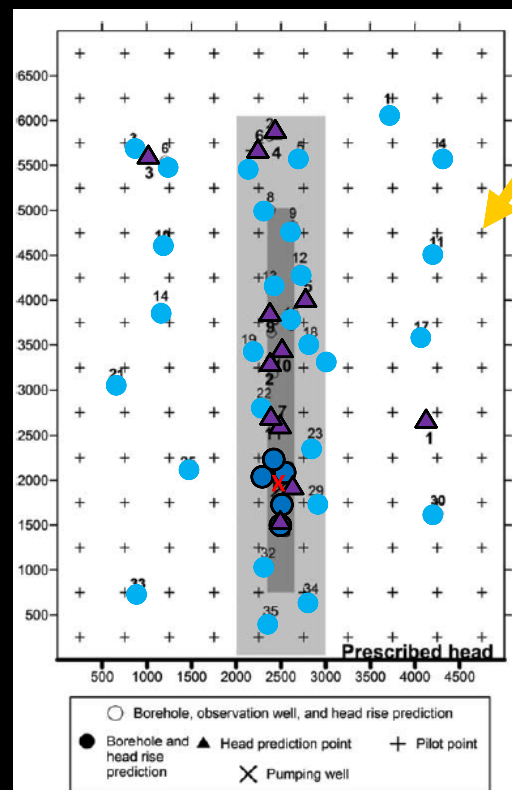
- Methods used to
  - Parameterize system characteristics
  - Compare to data
  - Adjust for data (calibration, tuning)
  - Sensitivity analysis
  - Uncertainty quantification

What do I mean by a modeling framework?

- Methods used to
  - Parameterize system characteristics

Define zones or simple interpolations and estimate relatively few parameter values





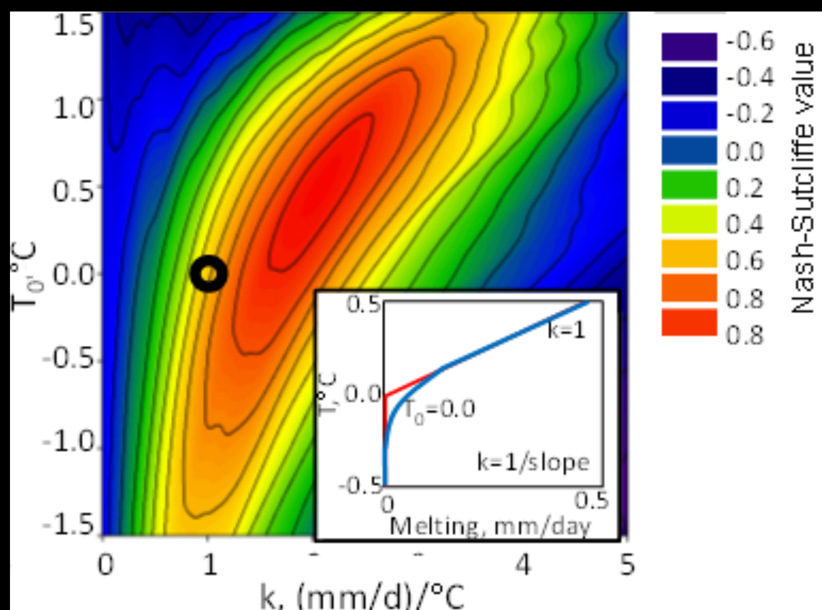
Define many interpolation points. Estimate a parameter value at each point.

What do I mean by a modeling framework?

- Methods used to
  - Parameterize system characteristics
  - Compare to data
    - Graphical
    - Statistics like sum of squared residuals

What do I mean by a modeling framework?

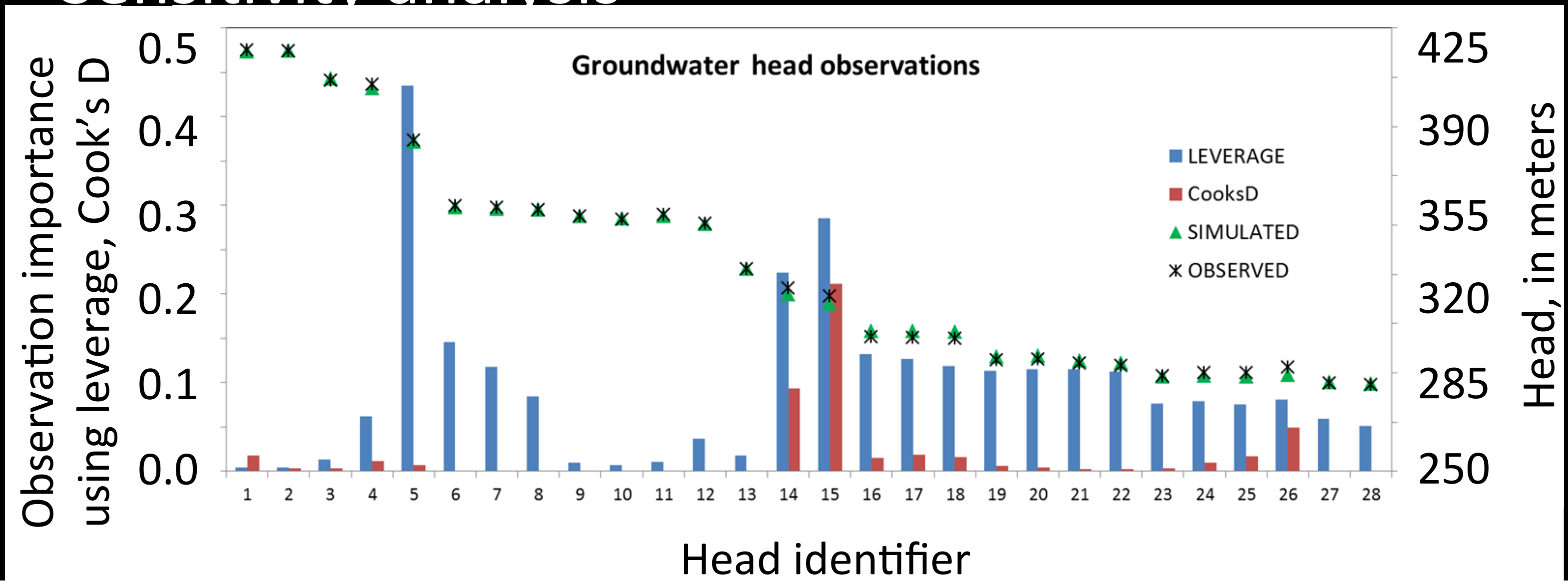
- Methods used to
  - Parameterize system characteristics
  - Compare to data
  - Adjust for data (calibration, tuning)



What do I mean by a modeling framework?

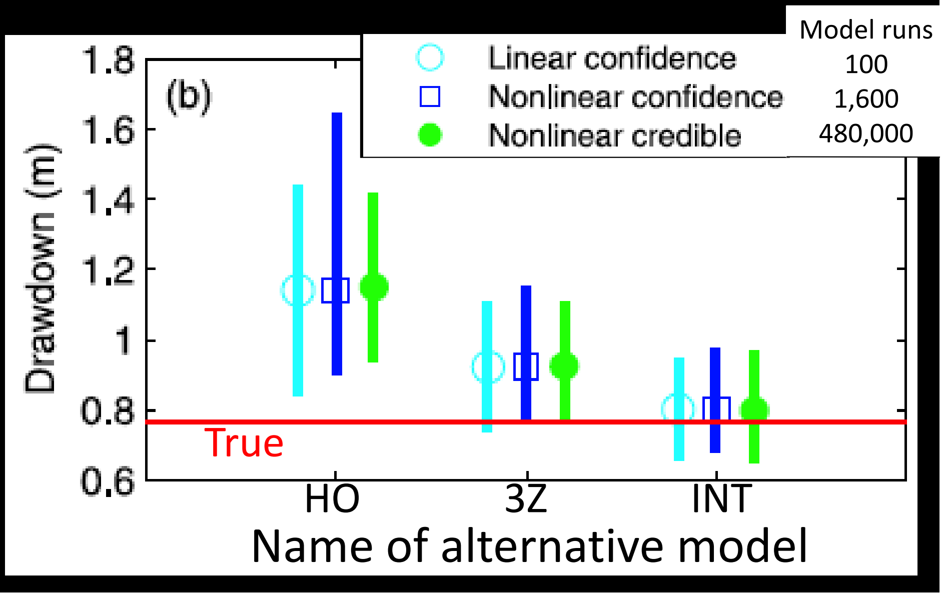
- Methods used to
  - Parameterize system characteristics
  - Compare to data
  - Adjust for data (calibration, tuning)
  - Sensitivity analysis

Foglia et al, in preparation



What do I mean by a modeling framework?

- Methods used to
  - Parameterize system characteristics
  - Compare to data
  - Adjust for data (calibration, tuning)
  - Sensitivity analysis
  - Uncertainty quantification



Lu et al. 2012 WRR

Overview of framework methods

Problems:
-Many method - Tower of Babel?
-Colossal computational demands – Necessary?

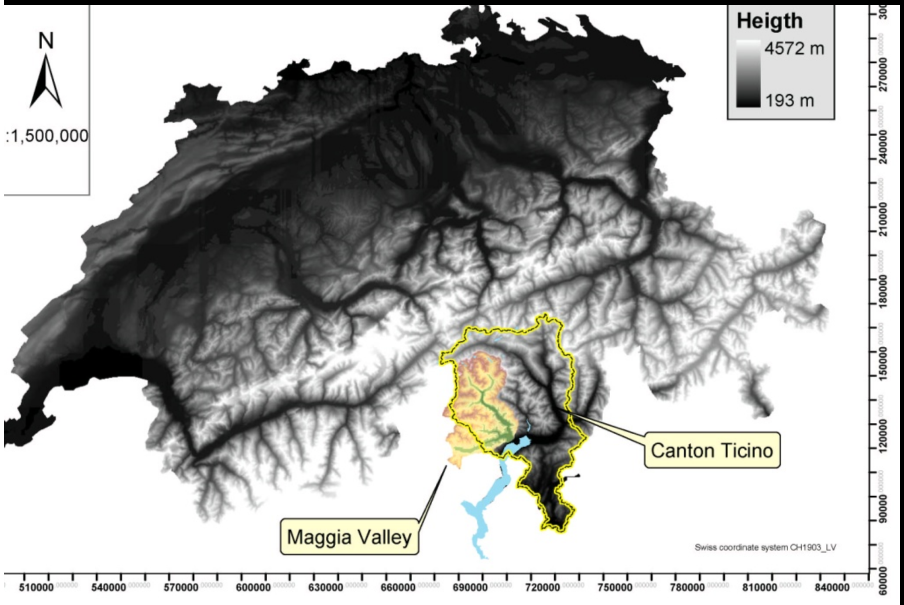
Hill + 2015, Groundwater

Common questions	Frugal methods	Demanding methods
Model Adequacy		
1. How can many data types with variable quality be included?	Error-based weighting and SOO or MAP	MOO, Pareto curve
2. Is model misfit/overfit a problem? Is the fit to prior knowledge and data subsets consistent? Are errors Gaussian?	RMSE, Nash-Sutcliffe, graphs, R^2_N, s_n^2, $s_{(n-p)}^2$ Compare fit to a priori error analysis using s_n^2, $s_{(n-p)}^2$	MOO, Pareto curve
3. How nonlinear is the problem?	Intrinsic nonlinearity, DELSA	DELSA, Explore objective function
Sensitivity and Uncertainty		
Observations (Obs) ←→ Parameters (Pars)		
4. What pars can and cannot be estimated with the obs?	Scaled local stats (CSS, ID, PCC, etc.), SVD, DoE, MoM(OAT, EE)	DoE, MoM(OAT, EE), eFAST, Sobol', RSA
5. Are any parts dominated by one obs and, thus, its error?	Scaled local stats (Leverage, DFBETAS)	Cross validation
6. How certain are the par values?	Par uncertainty intervals	Par uncertainty intervals
7. Which obs are important and unimportant to pars?	Scaled local stats (Leverage, Cook's D)	Cross validation
Parameters (Pars) ←→ Prediction (Preds)		
8. Which pars are important and unimportant to preds?	Scaled local stats (PSS, etc.), DELSA	DELSA, eFAST, Sobol'
9. How certain are the preds?	z/SD_z, Pred uncertainty intervals	Pred uncertainty intervals, multi-model analysis
10. Which pars contribute most and least to the pred uncertainty?	Scaled local stats (PPR VOII)	eFAST, Sobol'
Observations (Obs) ←→ Prediction (Preds)		
11. Which existing and potential obs are important to preds?	Scaled local stats (OPR VOII)	Cross validation
12. For multi-model analysis, which models are likely to produce accurate preds?	Analyze model fit and estimated parameters, AIC, AICc, BIC, KIC	Cross validation
Risk Assessment		
13. What risk is associated with a given decision strategy and set of scenarios?	Combine uncertainty analysis and scenario simulation. Smooth cost function	Combine uncertainty analysis and scenario simulation. Cost function need not be smooth.
14. What are decisions are robust given a set of uncertain scenarios?	Evolutionary multiobjective optimization. Within this demanding method use frugal model analysis methods.	

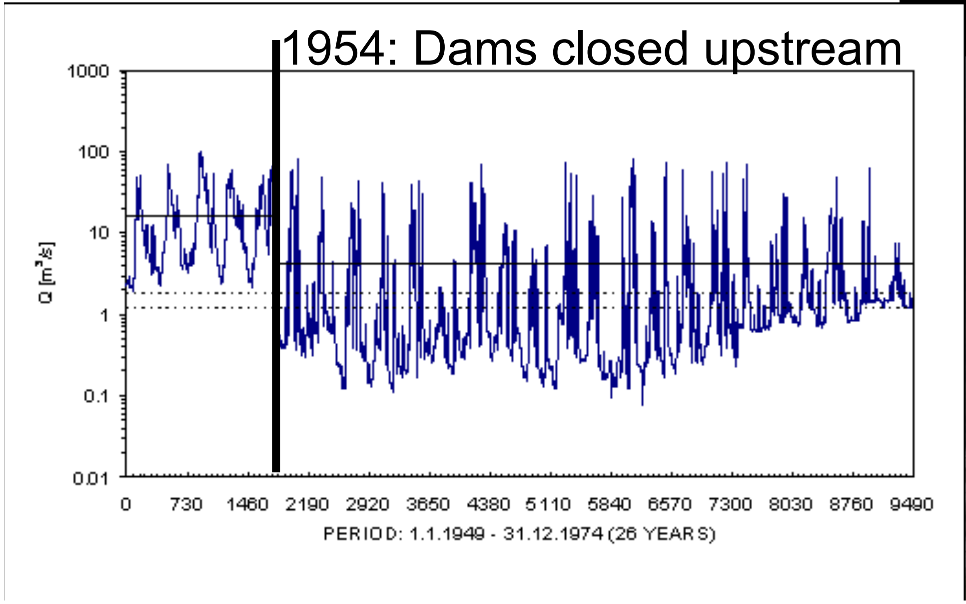
Testing modeling frameworks

Test using cross-validation

Maggia Valley, Switzerland



- Goal: Integrated hydrologic model to help manage ecology of this altered hydrologic system.
- Use component sw and gw models to test methods for the integrated model.



Maggia Valley, southern Switzerland

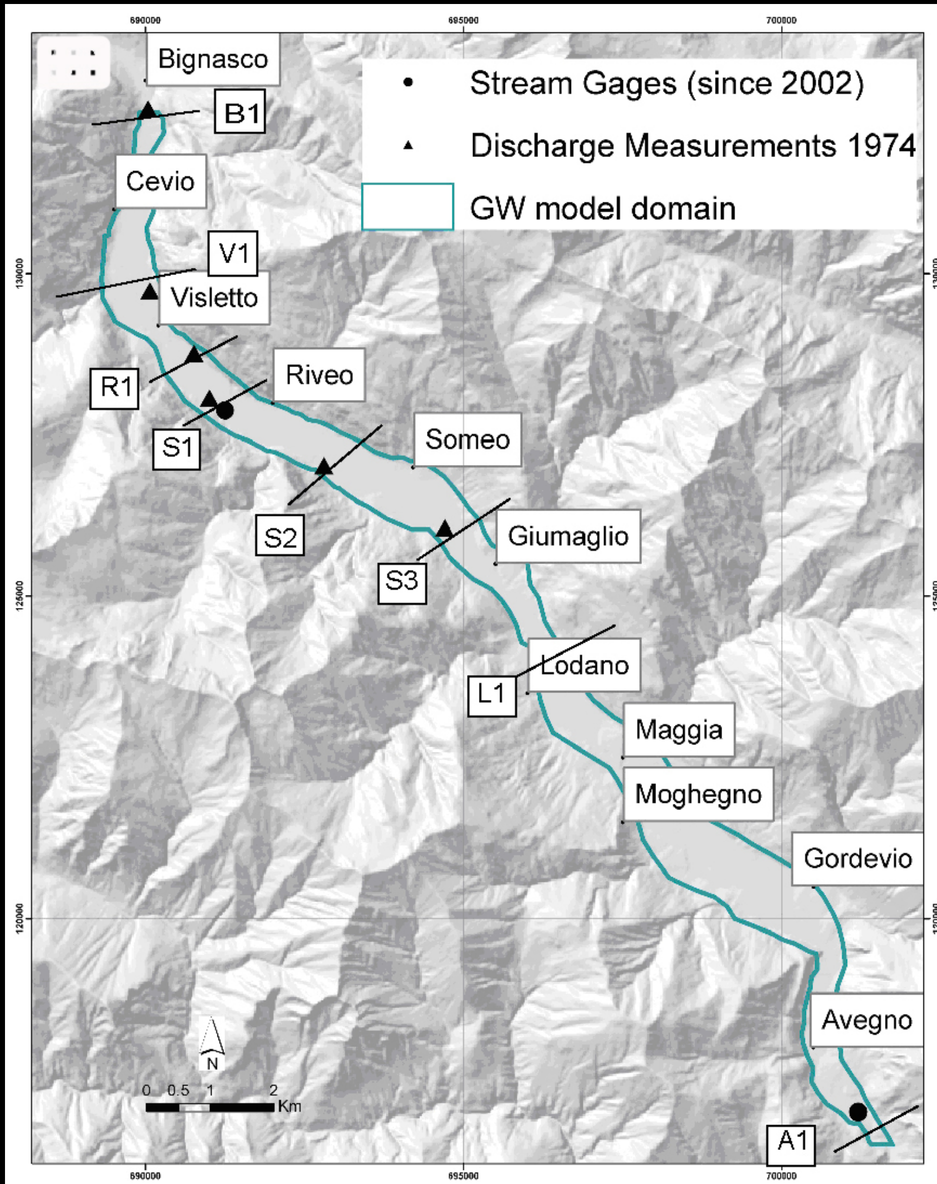
Series of studies to identify and test a computationally frugal protocol for model development

1. Test frugal sensitivity analysis (SA) with cross-validation
   – MODFLOW Model w stream (dry) (Foglia+ 2007 GW)
2. Demonstrate frugal optimal calibration method
   – TOPKAPI Rainfall-Runoff model (Foglia + 2009 WRR)
3. Here - SA and calibration methods enable test of computationally frugal model selection criteria
   MODFLOW GW model (wet) (Foglia + 2013 WRR)
4. Integrated hydrologic model – SW and GW

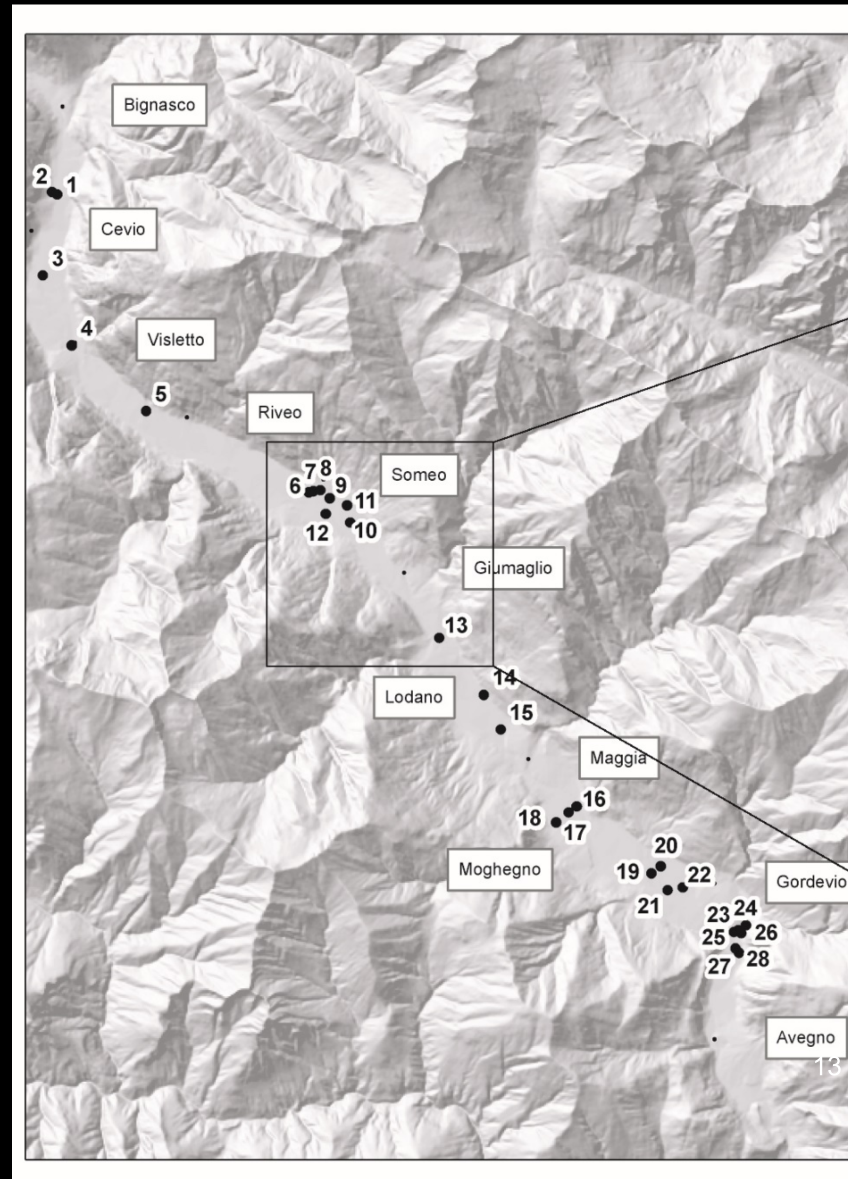
Software: UCODE (Poeter + 2005, 2014)
MMA (MultiModel Analysis) (Poeter and Hill 2007)

Model outline and 35 Obs Locations

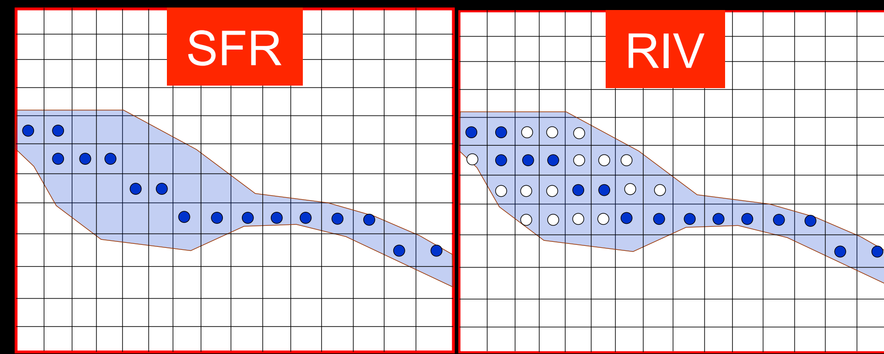
Flow data (use gains/losses)



Head data

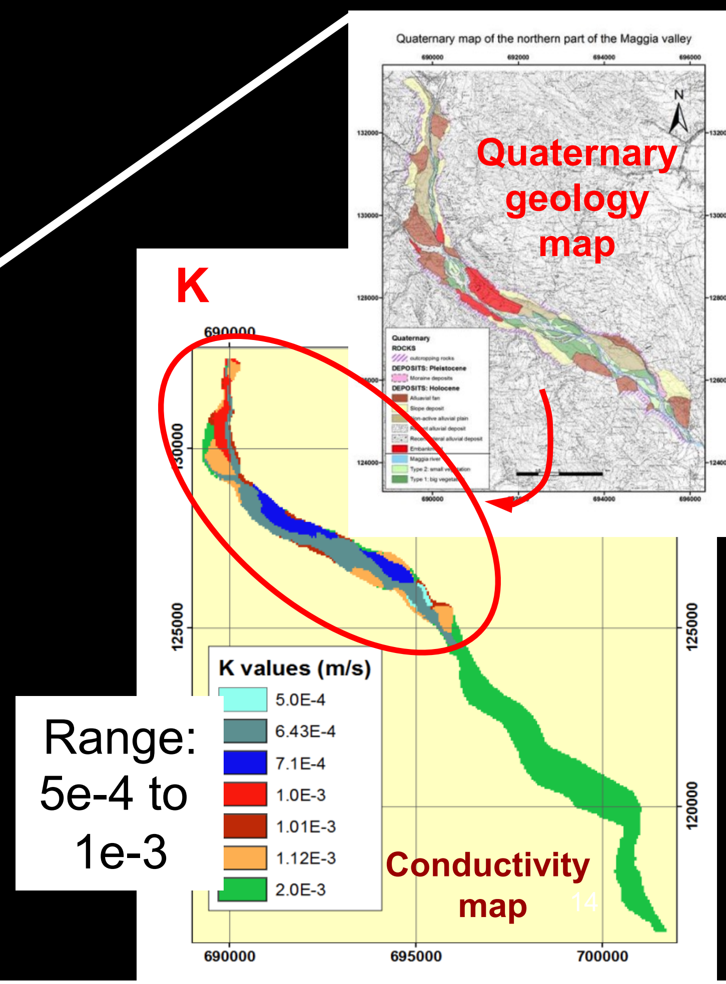


Model structures



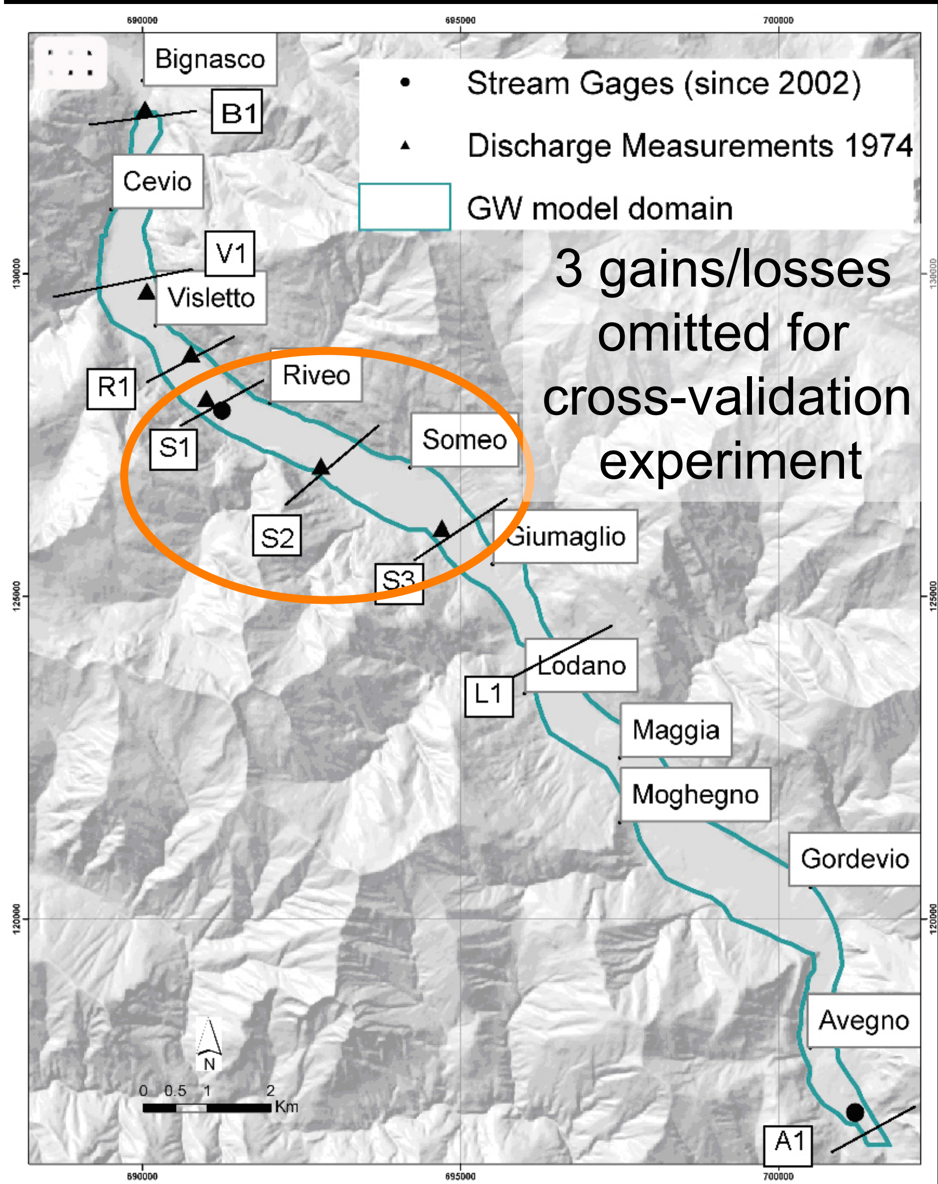
- Streamflow Routing (SFR2) vs. River (RIV)
- Recharge
  - Zero, constant, TOPKAPI
- Bedrock boundary
  - Old & New: Before and after gravimetric survey.
  - Depth in south x 2.
- K
  - Based on map of surficial quaternary geology
  - 5 combinations, from 1 parameter to 6

64 alternative models

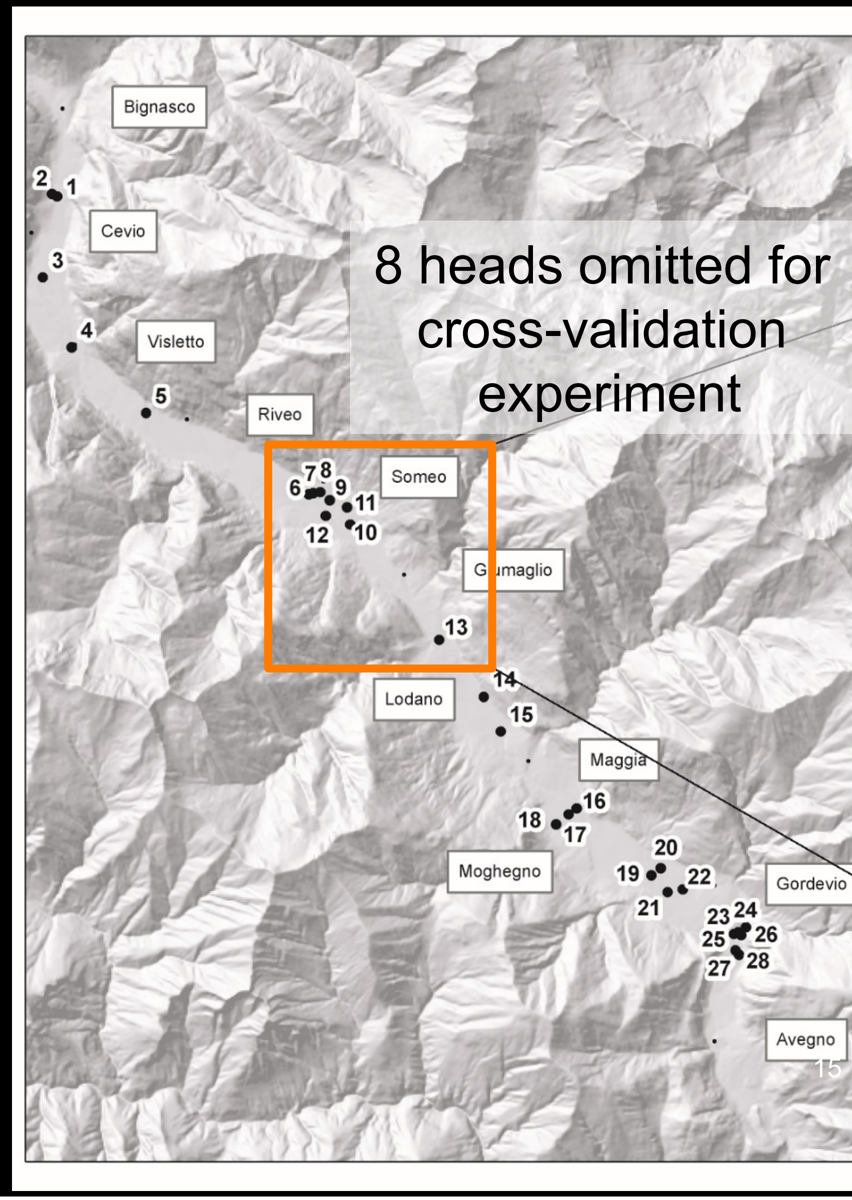


Cross validation experiment

Flow data (use gains/losses)



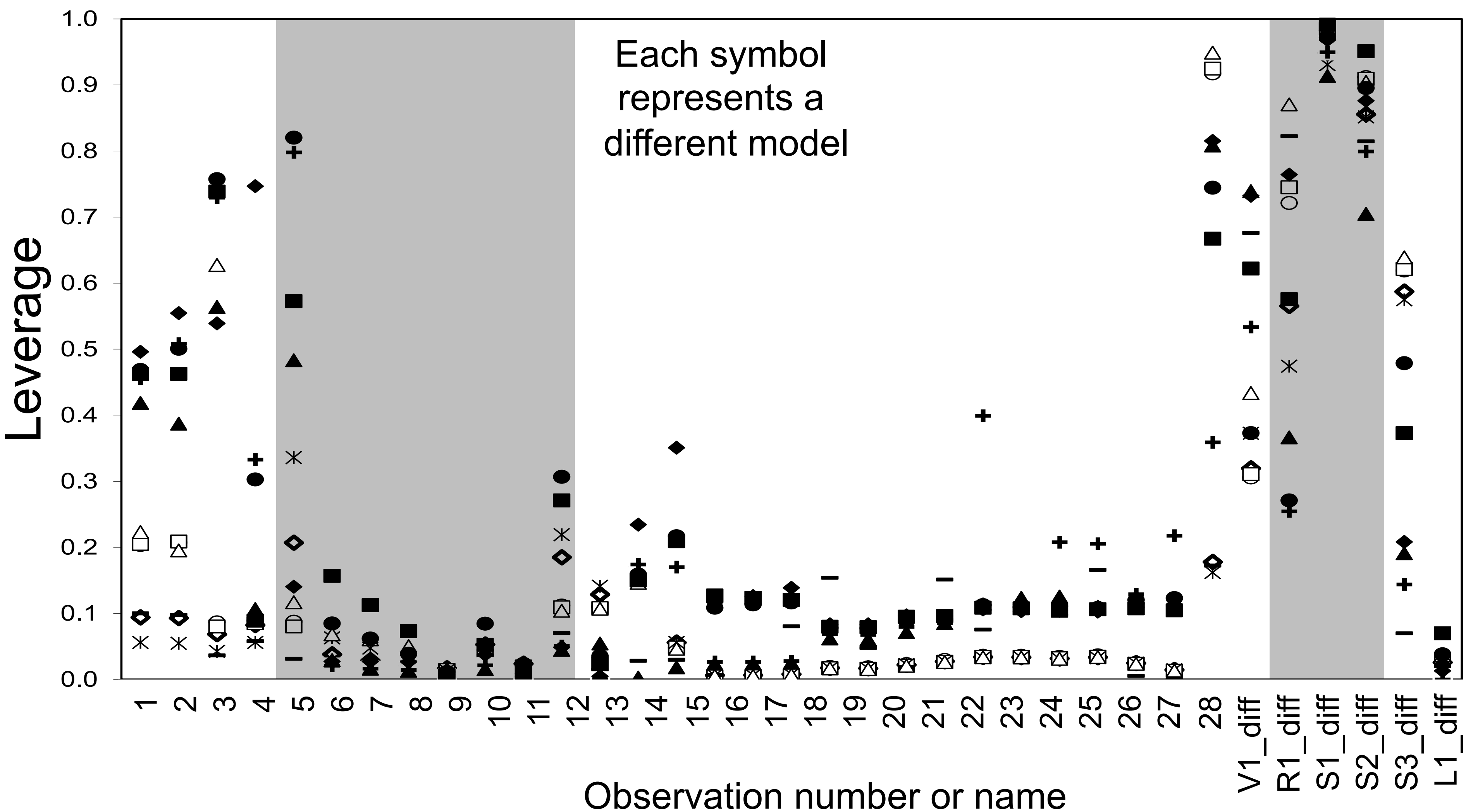
Head data



How important are the omitted observations in the alternative models?

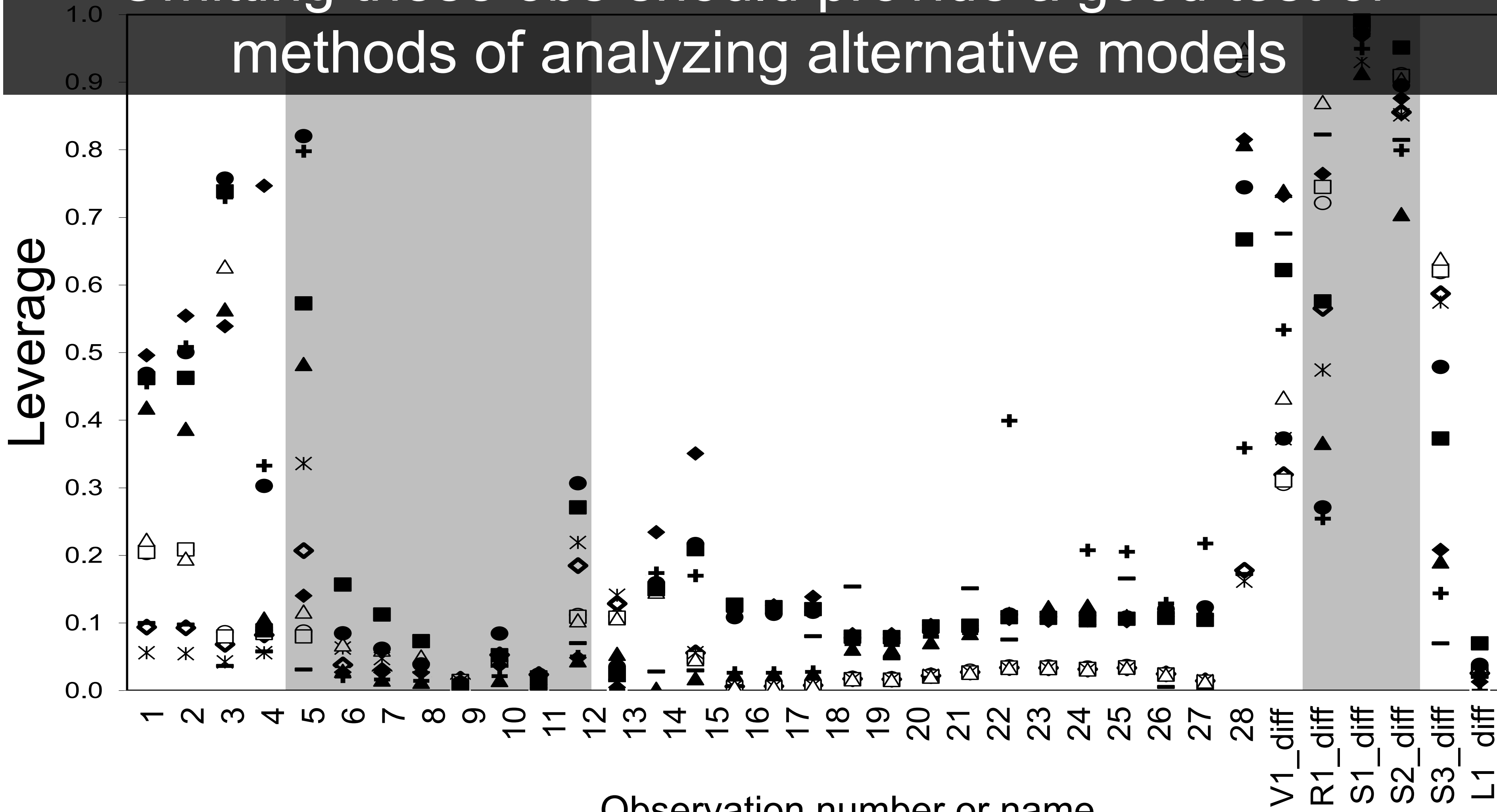
Use SA measure leverage for models calibrated with all 35 obs.

High leverage = important obs.



How important are the omitted observations in the alternative models?

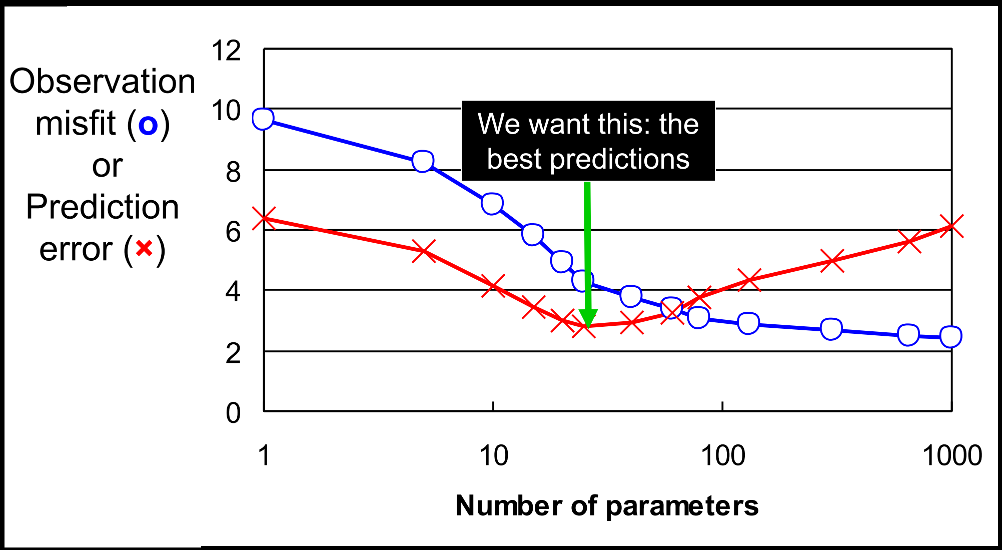
Omitted obs are very important to not important.
Omitting these obs should provide a good test of methods of analyzing alternative models



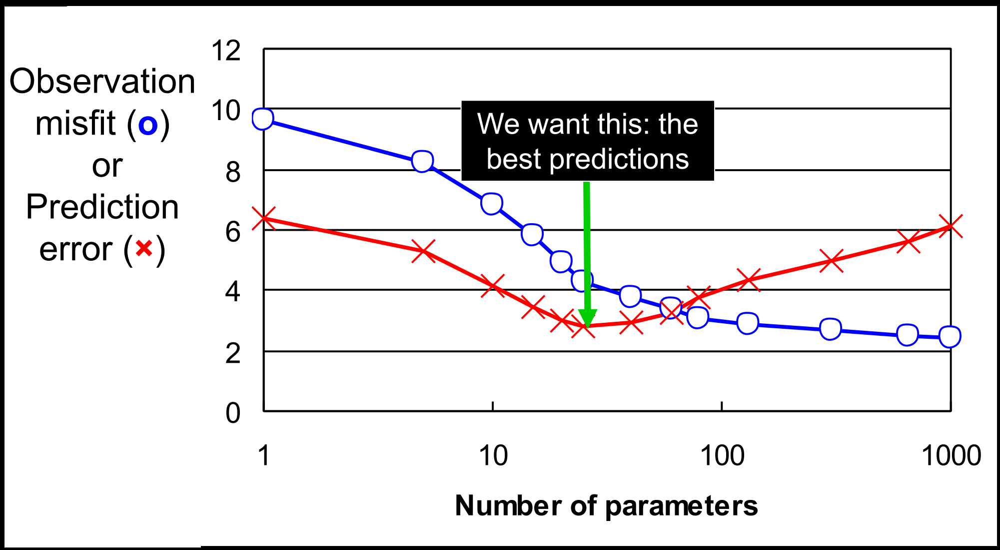
Tests conducted

- Test model selection criteria AIC, BIC, and KIC
  - Forund problems – especially for KIC
- Compared results with a basic premise related to model complexity
  - Led to some specific suggestions about model development
  - Present this part of the analysis here

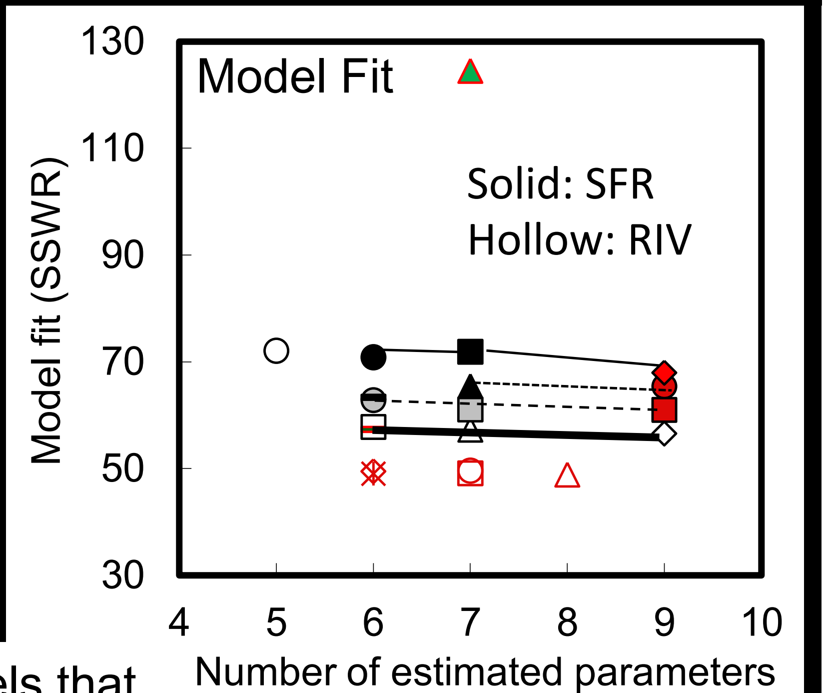
Our results vs basic premise?



Our results vs basic premise?

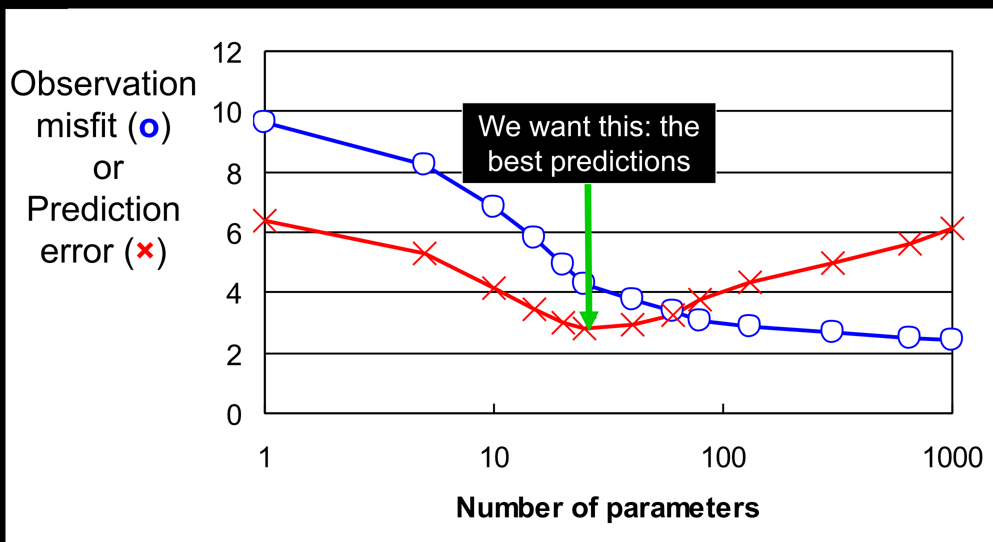


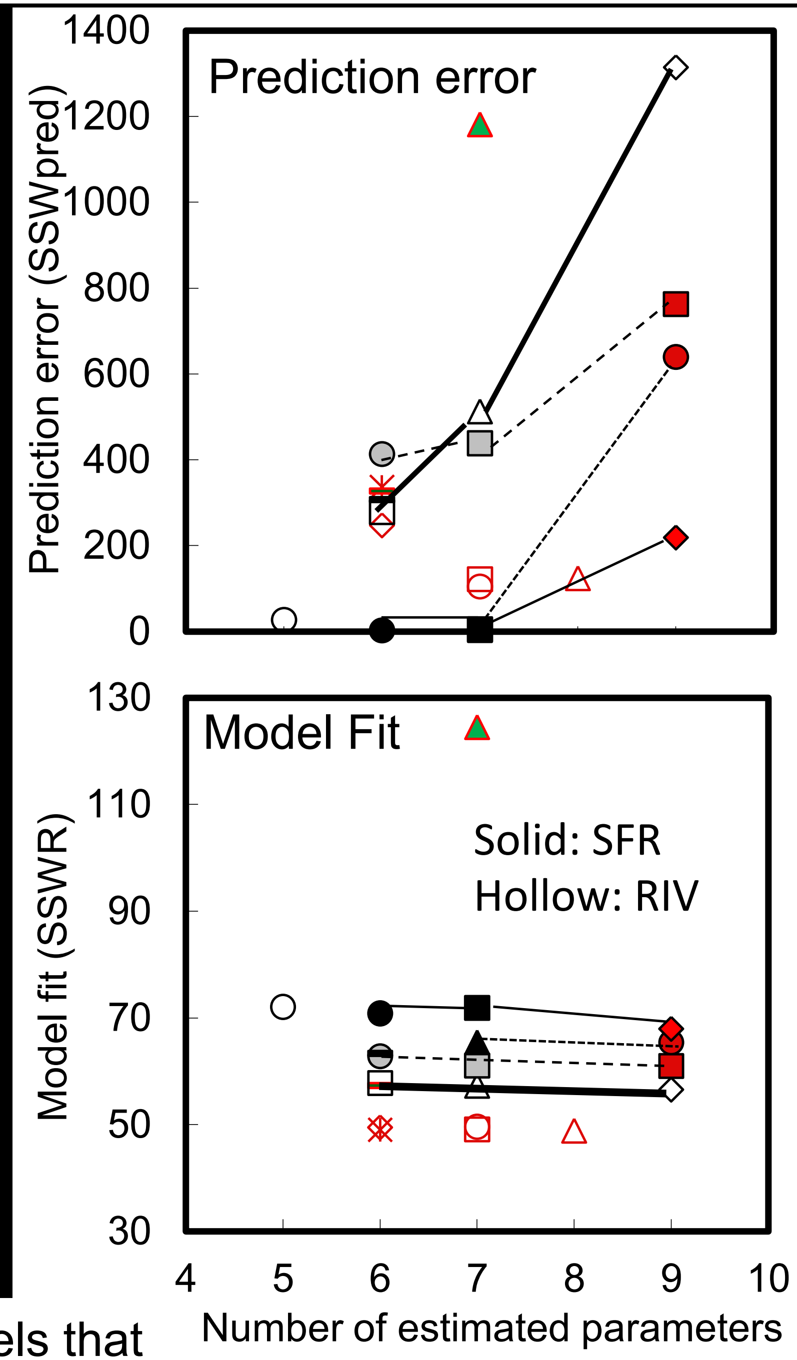
Lack of improvement with added K parameters indicates the K parameterization has little relation to the processes affecting the data



Lines connect models that are the same except for K

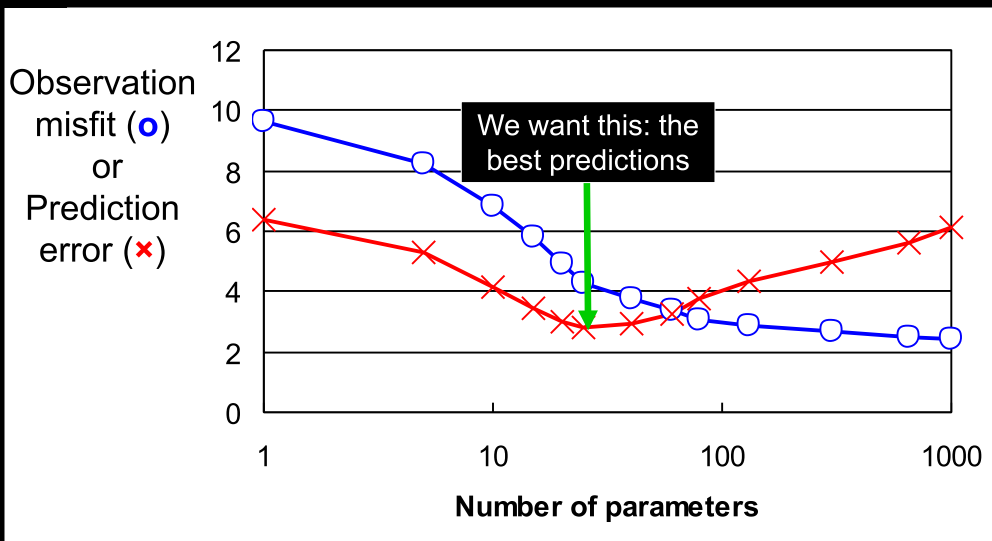
Our results vs basic premise?

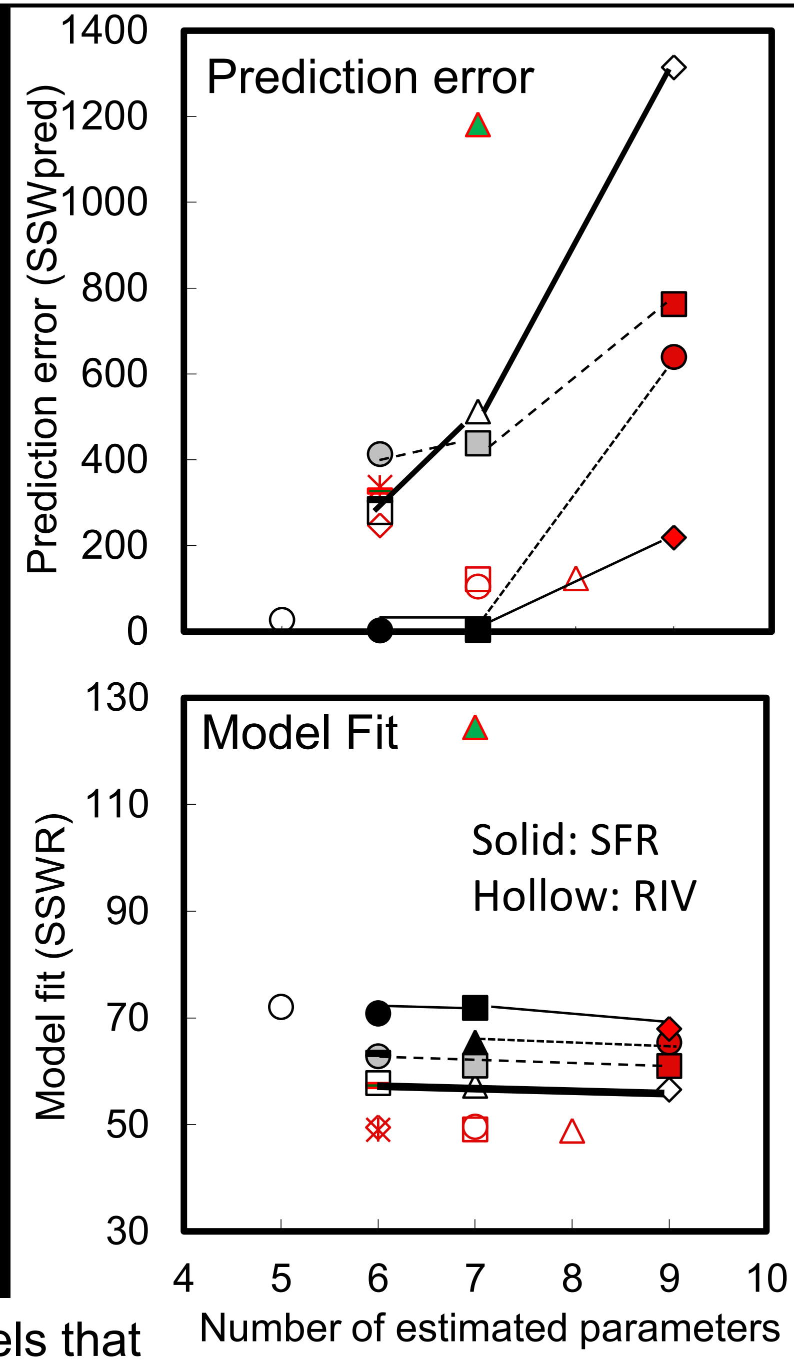




Lines connect models that are the same except for K

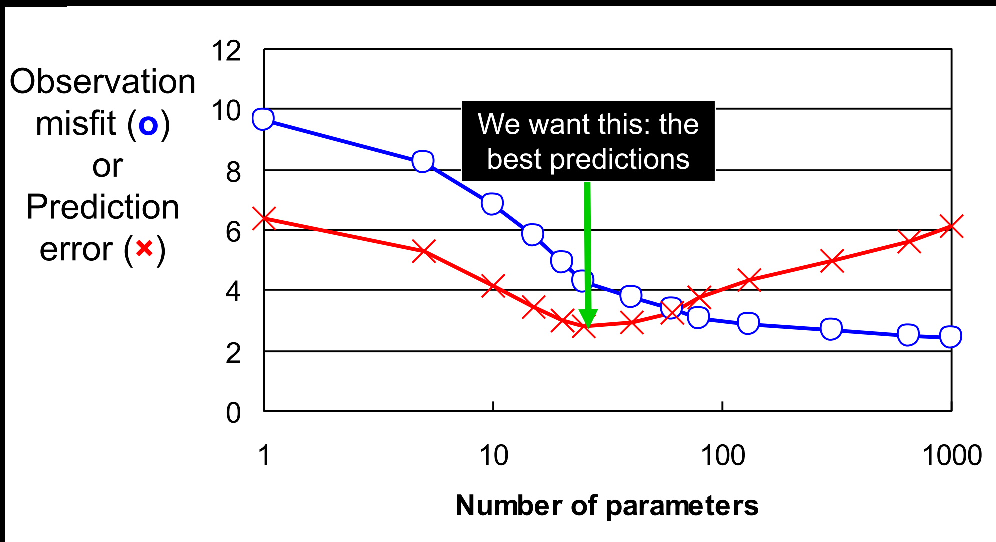
Our results vs basic premise?



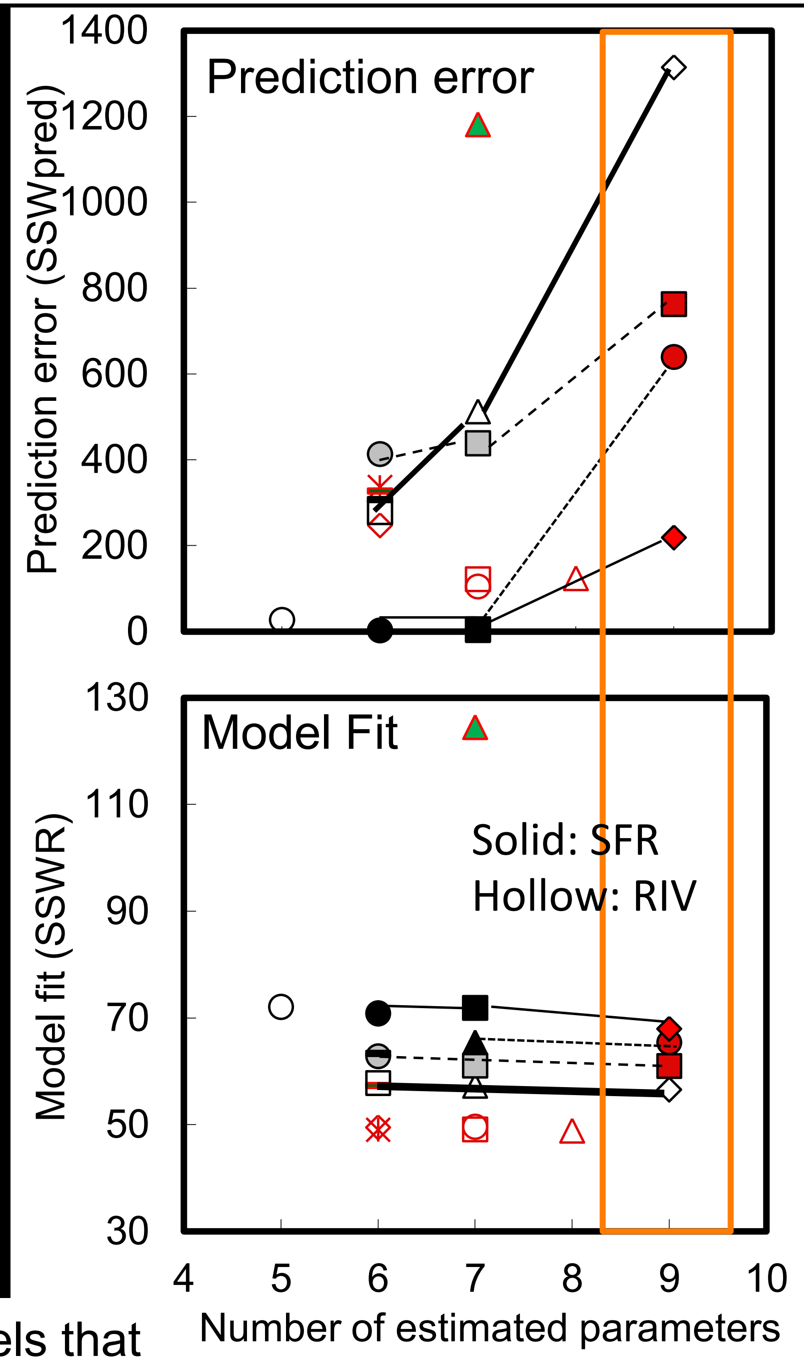


Lines connect models that are the same except for K

Our results vs basic premise?

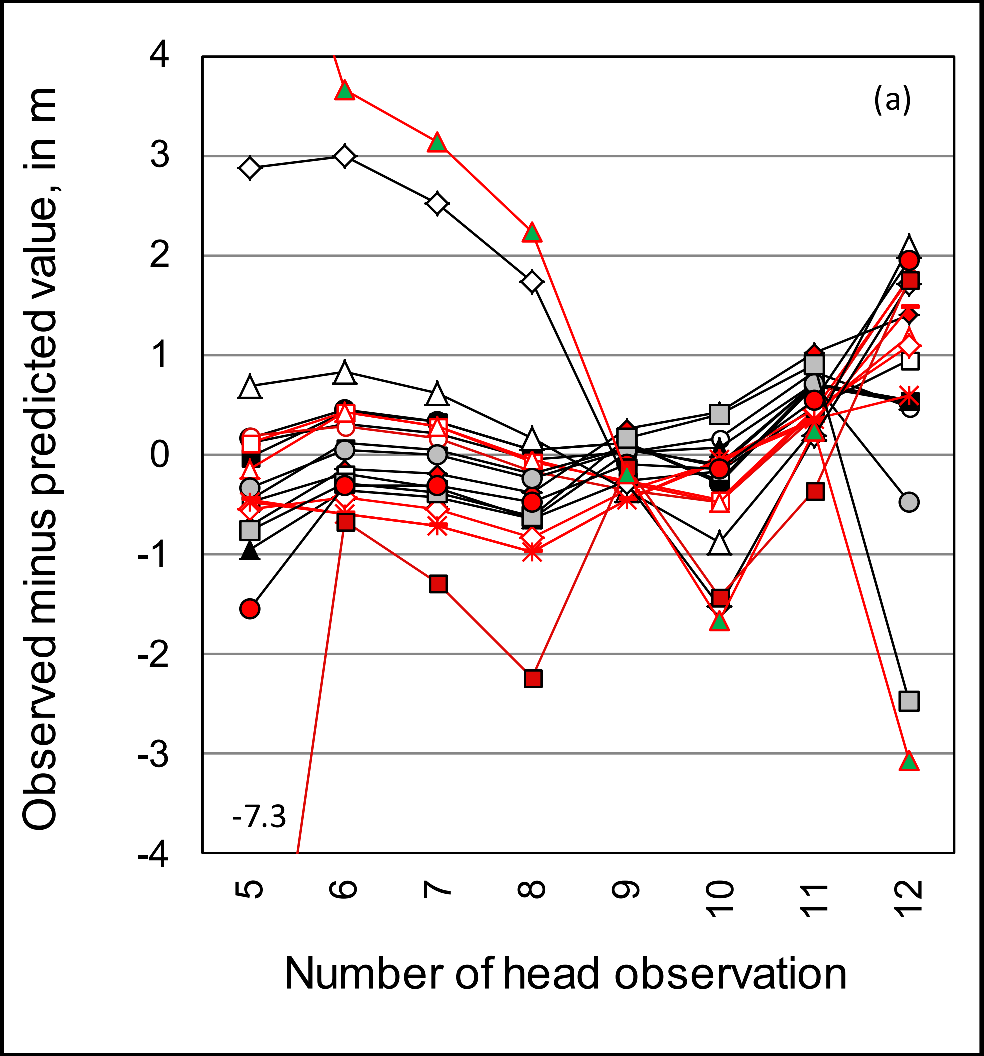


Sensitivity analysis indicated the models with the most parameters were poorly posed. Suggests we can use these methods to detect poor models.



Lines connect models that are the same except for K

What do poor predictions look like?



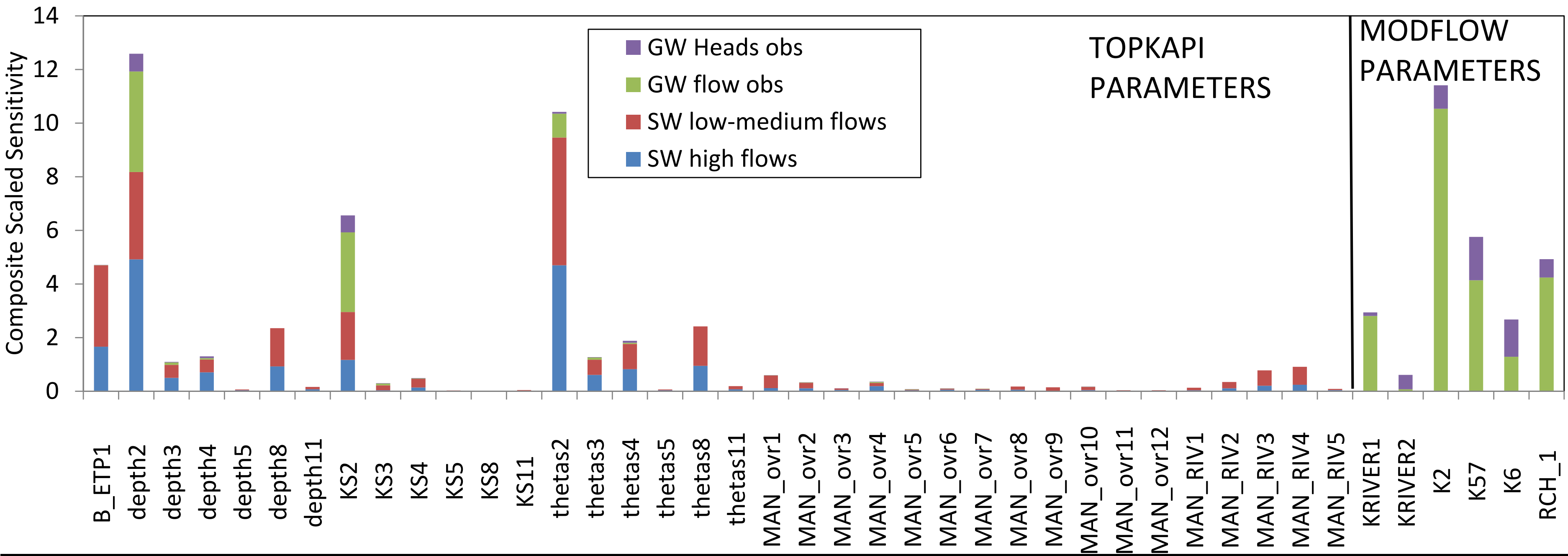
Identify important parameters

Results from the integrated hydrologic model

Groundwater model + distributed hydrologic model + new observations for low flows

	Original R-R model	New integrated model
Observations	120 streamflows	618 streamflows
		28 GW head observations
		7 gain/loss observations
Parameters included in calibration	4 SW parameter types	6 SW parameter types (36 pars)
		3 GW parameter types (6 pars)
Processes	Rainfall-runoff model	Rainfall-runoff model
		Groundwater model
Results	More frequent sampling for low flows is needed (Foglia + 2009 WRR)	Low flows are still as important as high flows for many of the parameters
		Observations related to the groundwater model are important also for the rainfall-runoff model

Identify important parameters

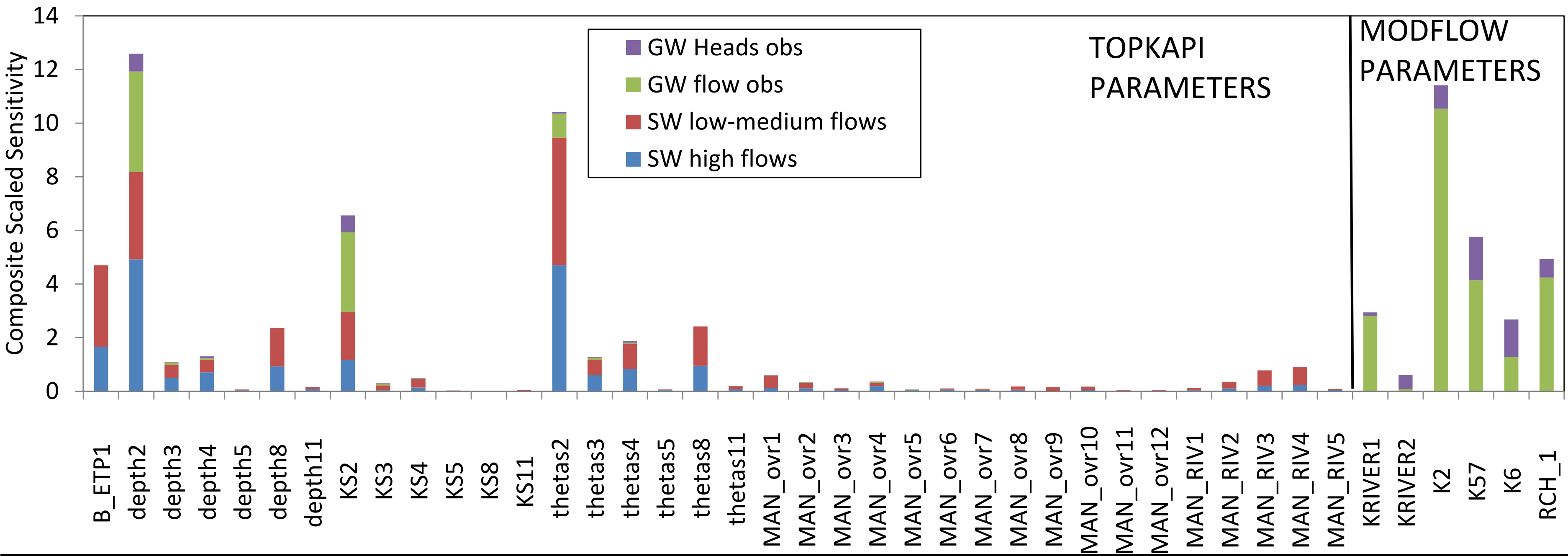


B_ETP1 | Soil depth | Saturated K | Porosity | Overland | Rivers | KRiv | K | Rech

Mannings

Identify important parameters

GW observations are quite important to some of the SW parameters
SW obs are interpreted as GW flow (gain/loss) obs, which are important



B_ETP1 | Soil depth | Saturated K | Porosity | Overland | Rivers | Mannings | KRiv | K | Rech

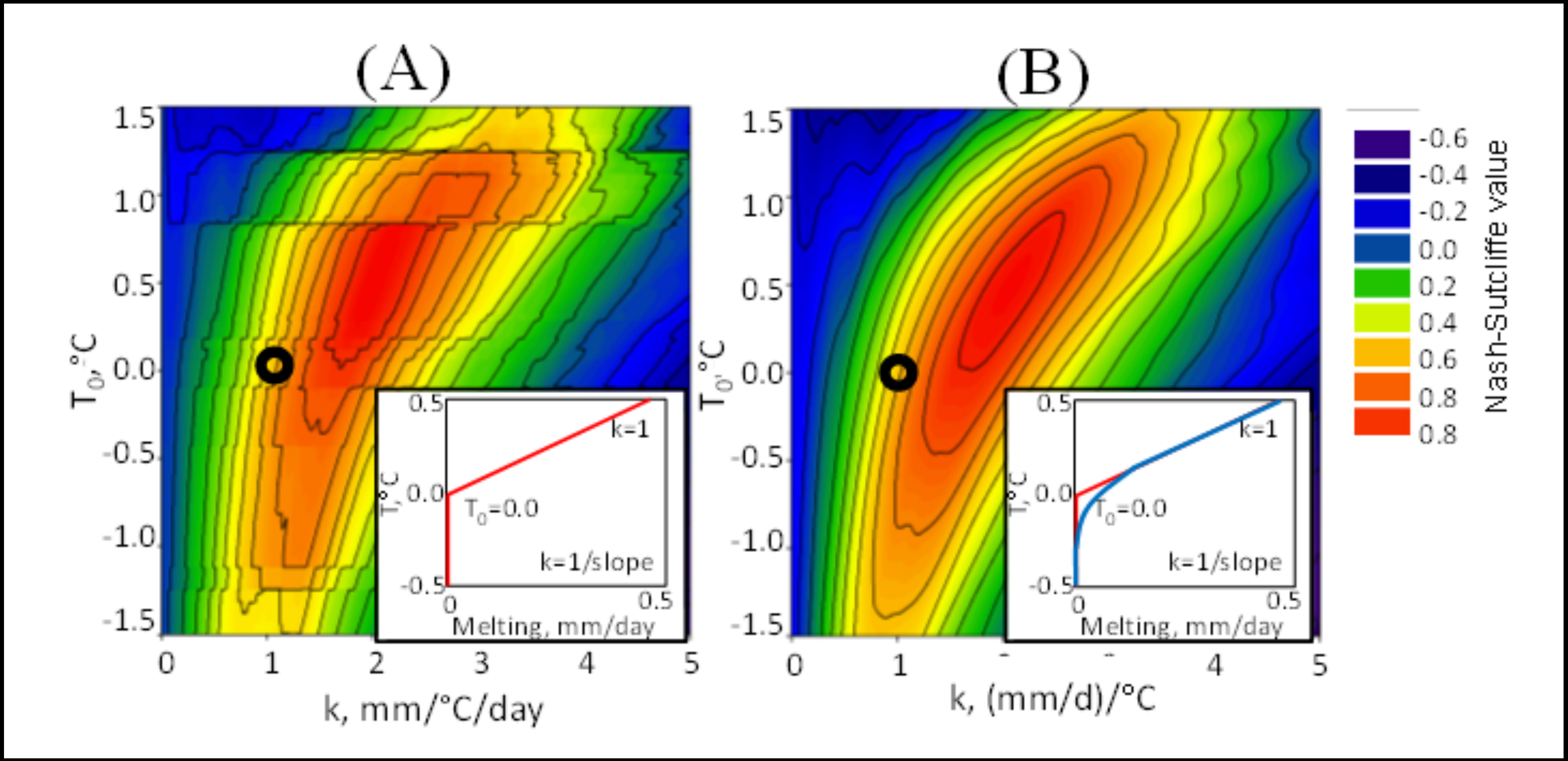
Conclusions from this test

- The surficial geologic mapping did not provide a useful way to parameterize the K
- Sensitivity analysis can be used to determine when the model detail exceeds what can be supported by data.
- Models unsupported by data are likely to have poor predictive ability
- 5 suggestions for model development

Testing modeling frameworks

- Enabled by computationally frugal methods, which are enabled by computationally robust models
- This is where you, the model developers come in!

Do your models have numerical daemons???



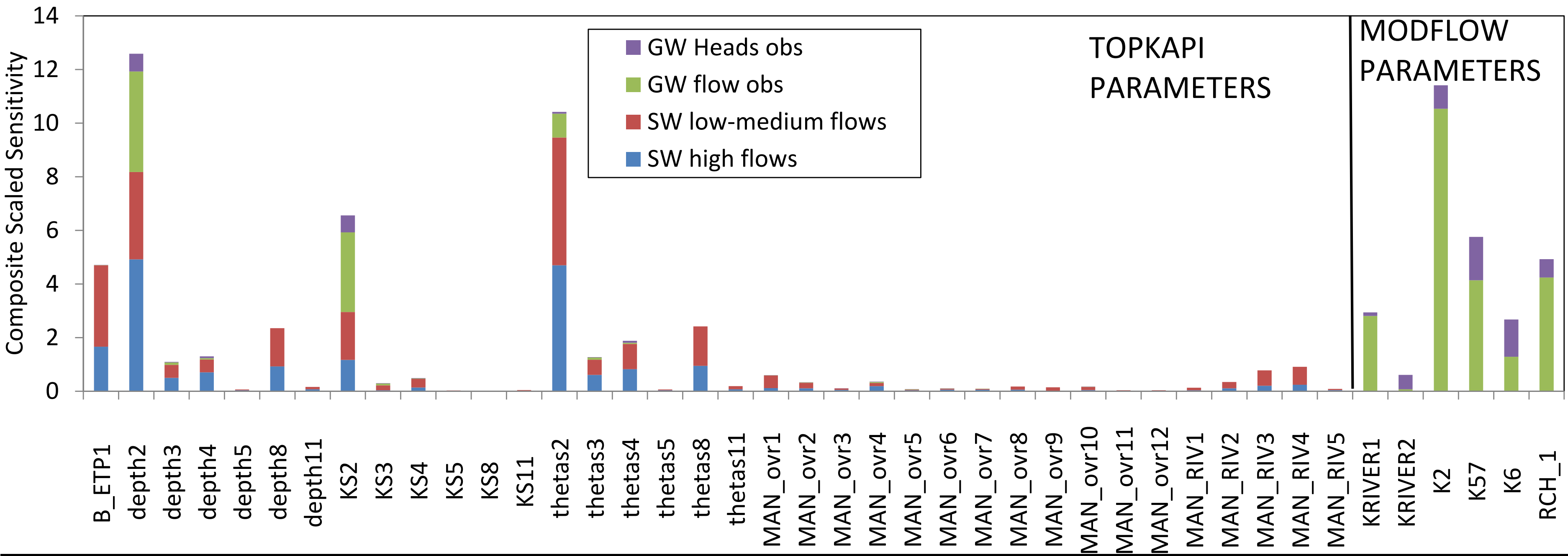
Kavetski Kuczera 2007 WRR; Hill + 2015 GW; test with DELSA Rakovec 2014 WRR

Need important nonlinearities. Keep them!

Get rid of unrealistic, difficult numerics

Identify important parameters

GW observations are quite important to some of the SW parameters
SW obs are interpreted as GW flow (gain/loss) obs, which are important



B_ETP1 | Soil depth | Saturated K | Porosity | Overland | Rivers | Mannings | KRiv | K | Rech

Perspective

- In 100 years, we may very well be thought of as being in the Model T stage of model development
- A prediction: In the future, simply getting a close model fit will not be enough. It will also be necessary to show what obs, pars, and processes are important to the good model fit.
- This increased model transparency is like the safety equipment in automobile development.

I am looking for a test case

- Test a way to detect and diagnose numerical daemons.
- High profile model applications preferred.

AGU: New Technical Committee on Hydrologic Uncertainty

SIAM: New SIAM-UQ section and journal

[redacted]

Quantifying Uncertainty

- Lu Ye Hill 2012 WRR
- Investigate methods that take 10s to 100,000s of model runs (Linear to MCMC)

- New UCODE_2014 with MCMC supports these ways of calculating uncertainty intervals

Test Case

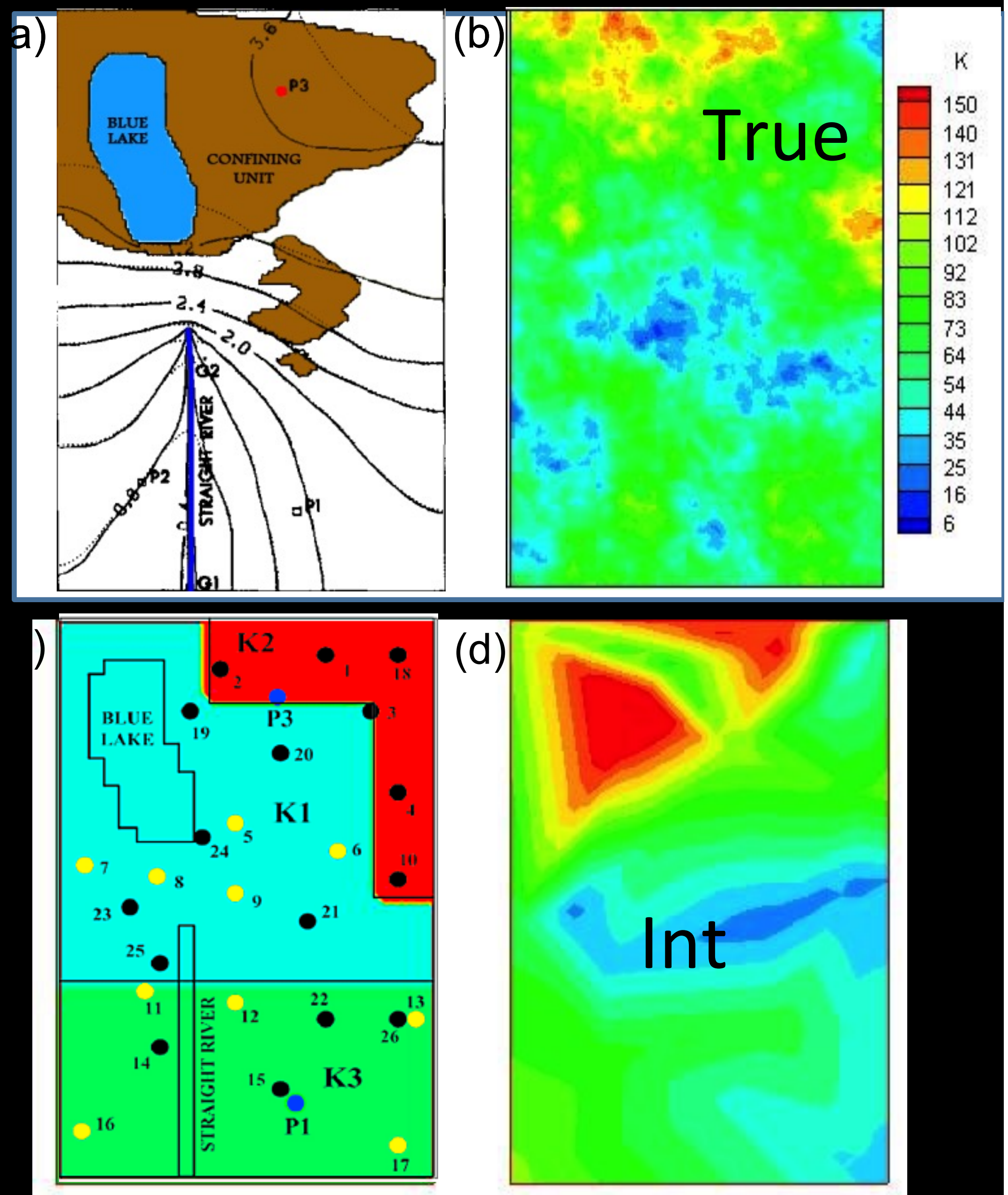
- MODFLOW
- 3.8 km x 6 km x 80m
- Top: free-surface
- No-flow sides, bottom
- Obs: 54 heads, 2 flows, lake level, lake budget
- Parameters:

  5 for

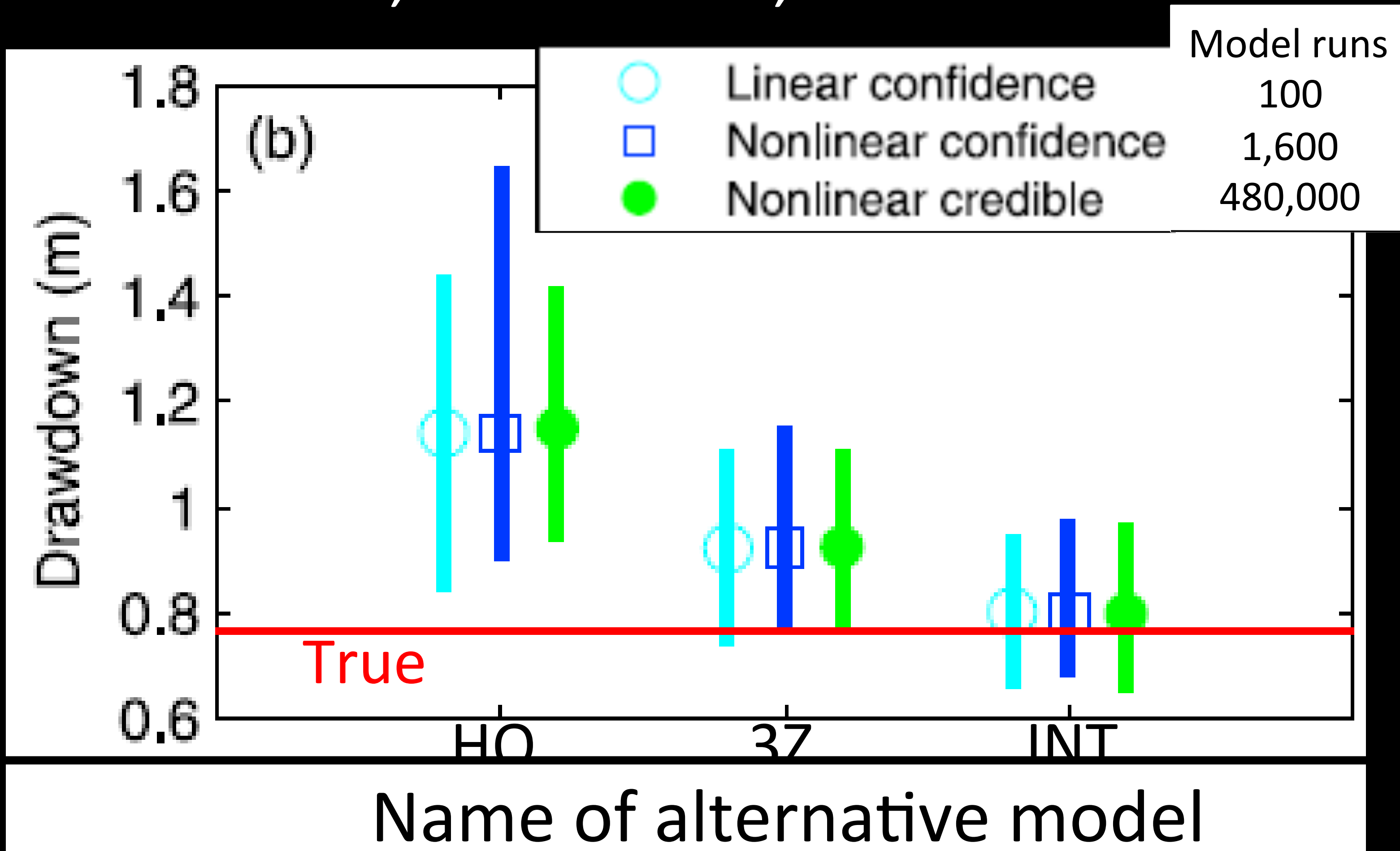
  - recharge
  - KRB
  - confining unit KV
  - net lake recharge,
  - vertical anisotropy

  K : 1 (HO), 3 (3Z), or 21 (INT)

  Data on K at yellow dots



Calibrated model, Predictions, Confidence Intervals



Lesson about uncertainty

- May be more important to consider the uncertainty from many models than use a computationally demanding way to calculate uncertainty for one model