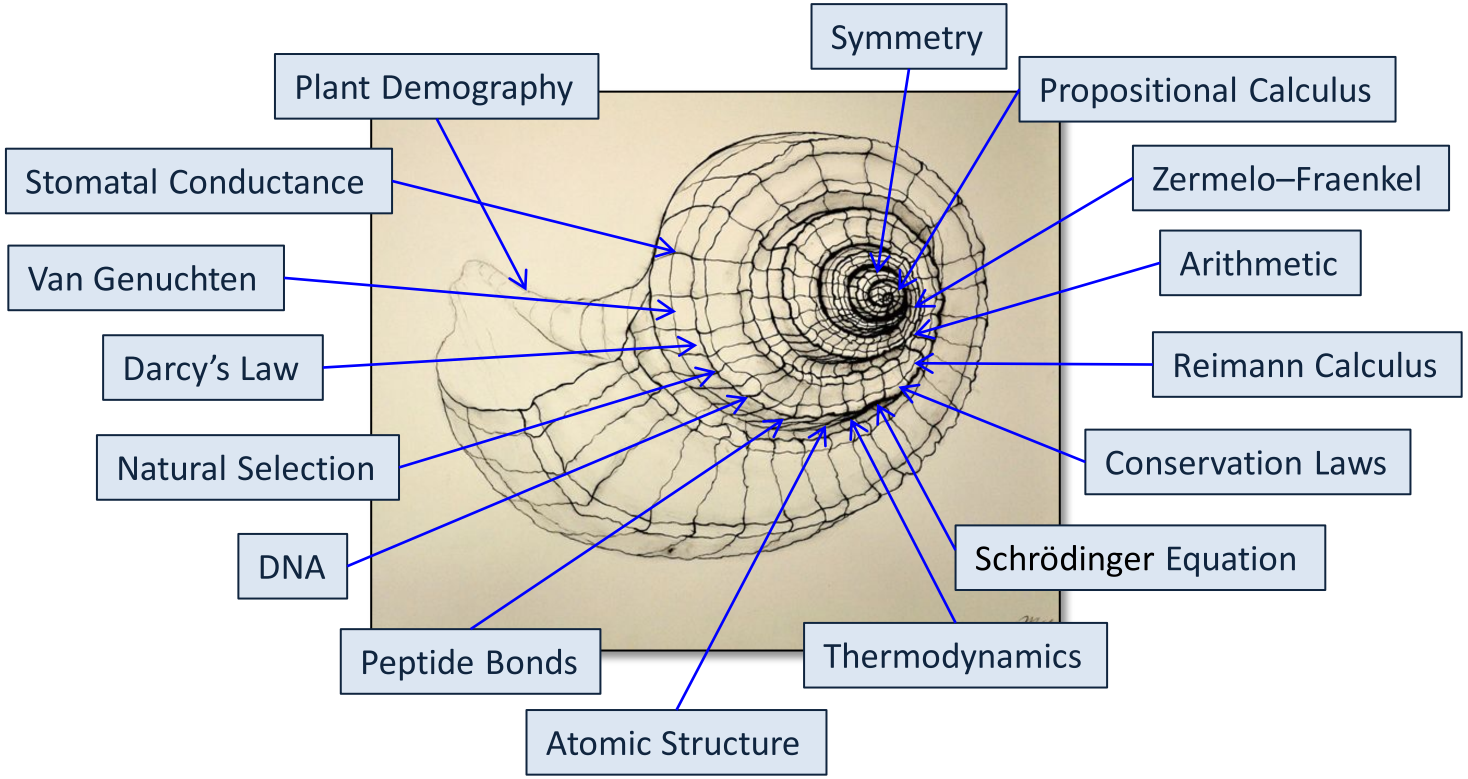
Information in Models and Data

Grey Nearing

National Center for Atmospheric Research
NASA Goddard Space Flight Center
University of Maryland Baltimore County

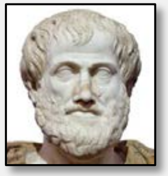
What is a Model?

“The totality of our so-called knowledge or beliefs, from the most casual matters of geography and history to the profoundest laws of atomic physics or even of pure mathematics and logic, is a man-made fabric which impinges on experience only along the edges.” - Van Quine (1951)



Ontology & Epistemology <u>in</u> Geophysical Models

$$p(M|D) \propto p(D|M)\,p(M)$$



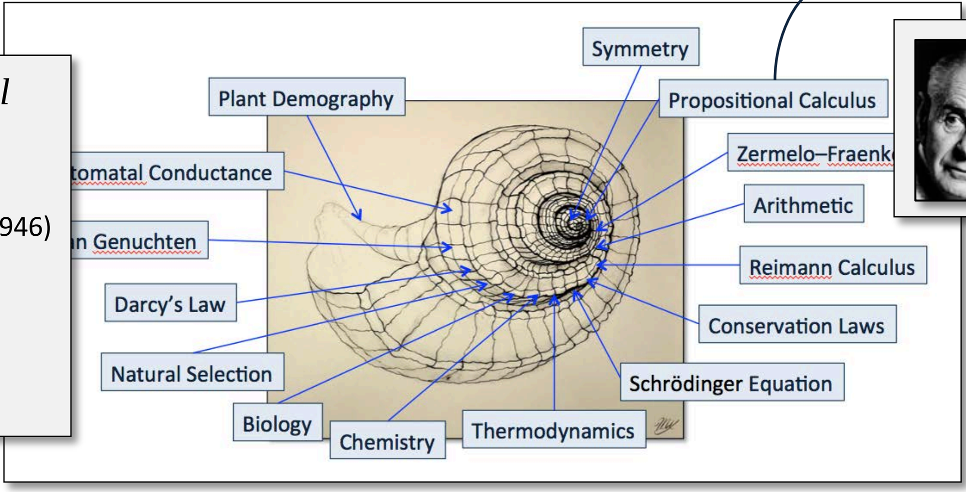
Modus Tollens:

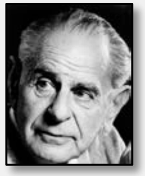
$$((P \rightarrow Q) \wedge \sim Q) \rightarrow \sim P$$

"the belief system of any rational agent must obey the standard axioms of probability"

– Richard Cox (1946)

- $p(P) \geq 0$
- $p(\Omega) = 1$
- $p(\cup P_i) = \sum P_i$



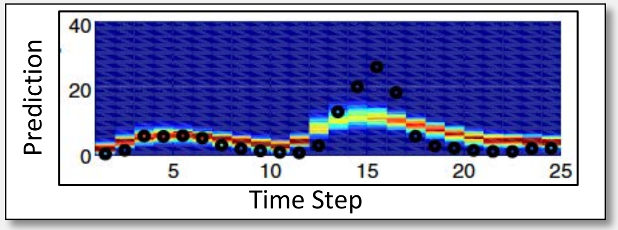


"In so far as a scientific statement speaks about reality, it must be falsifiable: and in so far as it is not falsifiable, it does not speak about reality."

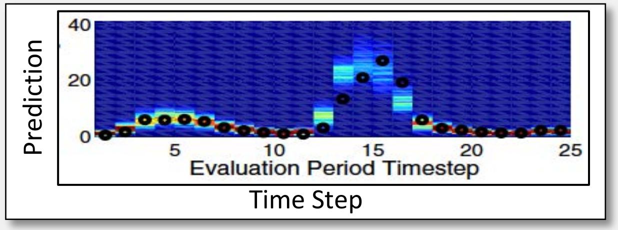
- Karl Popper (1959)

"With enough data – and often only a fairly moderate amount – any analyst could reject any model"

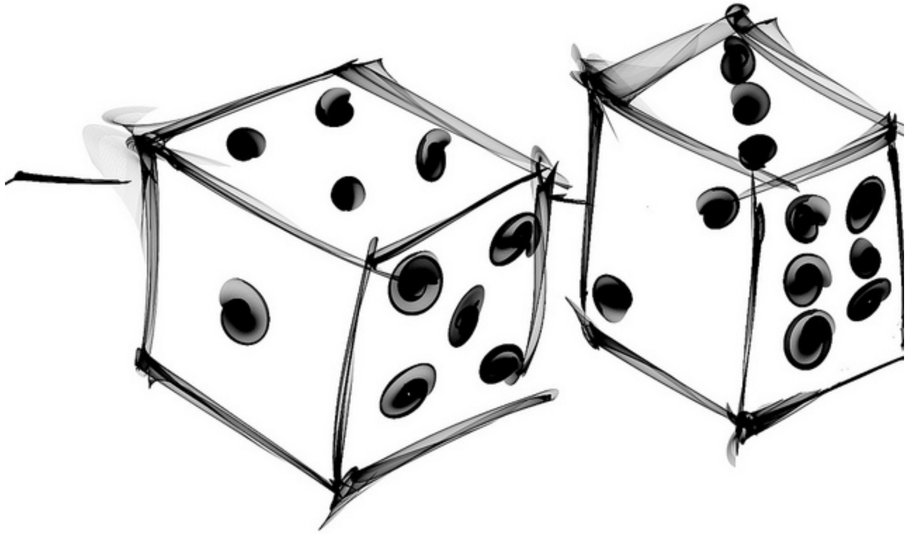
- Andrew Gelman (2010)



$$((P \rightarrow Q) \wedge Q) \nRightarrow P$$



The Description of an Experiment



Before running the experiment: $p(D_1 \wedge D_2) = p(D_2 | D_1) \times p(D_1)$

After running the experiment: $d(D_1 \wedge D_2) = d(D_2) + d(D_1) - d(D_1 \vee D_2)$

The Open Problem

The Demarcation Problem: Science is either well-defined and impractical or un-defined and practical.

- **Hume:** Induction cannot be rigorously supported.
- **Popper:** Therefore, science must be deductive.
- **Salmon:** Falsification is not practical because few (all) models are falsified.
- **Jaynes:** Bayesianism is at least consistent with the axioms of deductive logic, however fundamentally inductive.

To reconcile the falsification criteria with Bayesian model evaluation:

Does the model contain as much information as the observations?

How well are we doing right now?

Our best physically-based surface hydrology models cannot beat linear regressions that have *no state memory*, and which are *trained out-of-sample* and *extrapolated globally*.

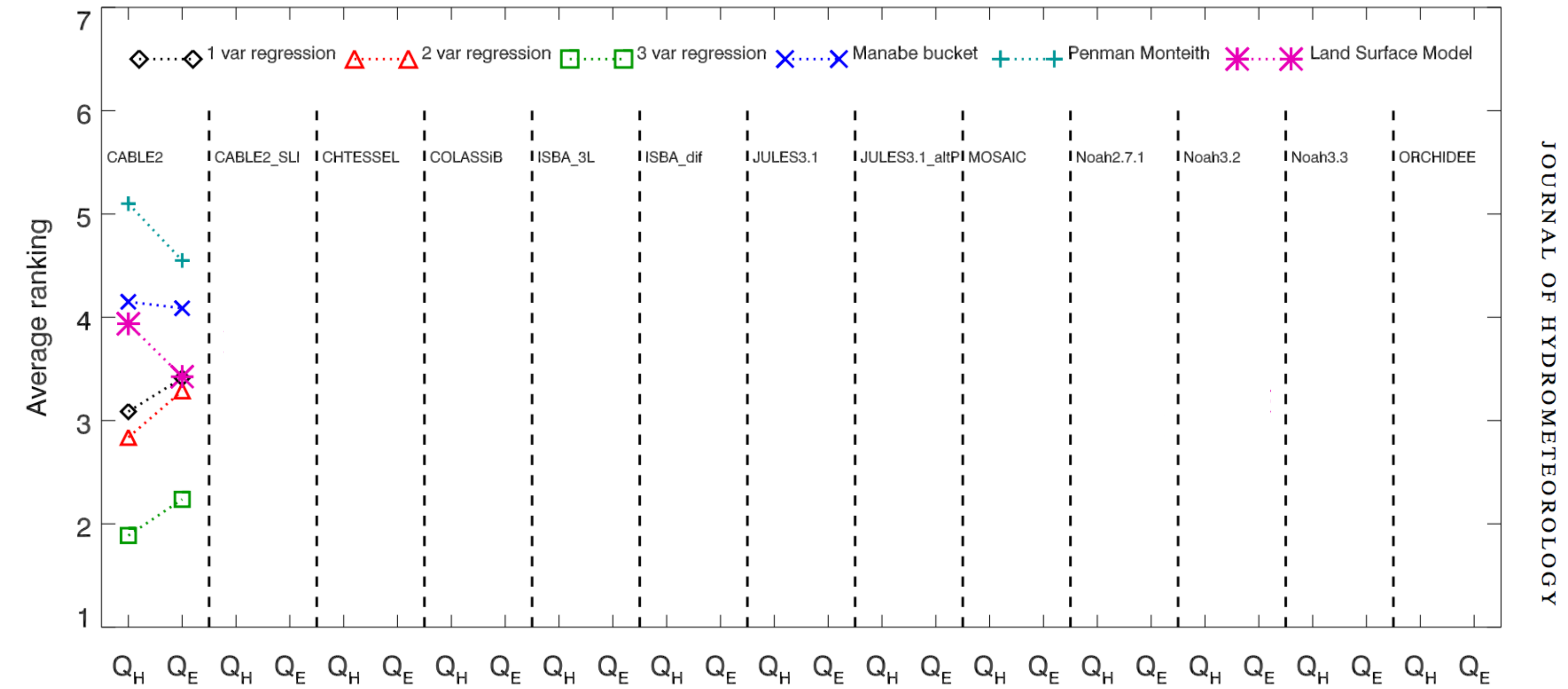
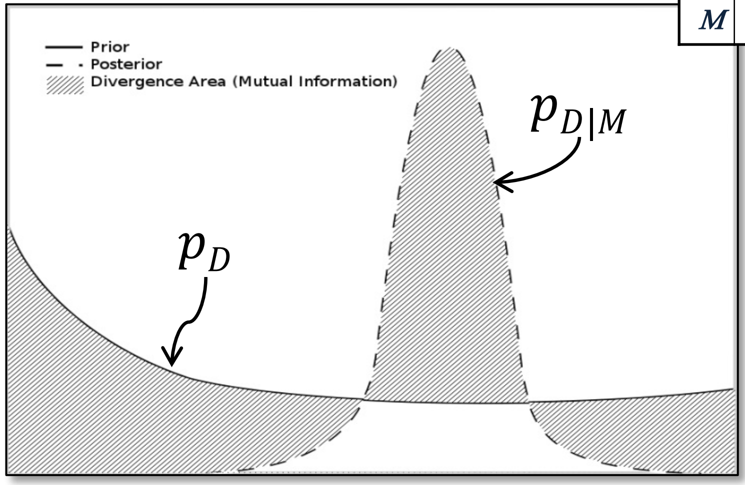


FIG. 4. Ranking of benchmarks and each model for the standard statistics (MBE, NME, SD, *r*) across all 20 sites. A ranking of 1 corresponds to the best performance. The dotted lines are a visual guide and have no scientific relevance.

M. Best et al. (2015) "The Plumbing of Land Surface Models" *Journal of Hydrometeorology*

Information Theory – Basic Principles

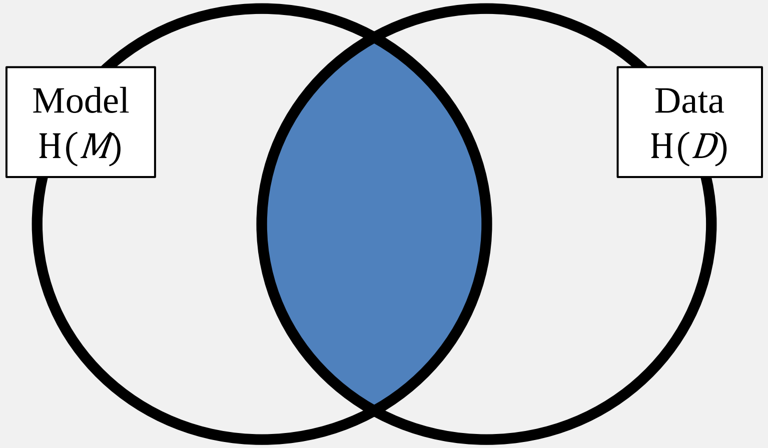
D	Evaluation Data
U	Model Inputs
M	Model



Information is the ability of signal to effect a change in a probability distribution.

General Definition	$I(D;M) = E\left[f\left(\frac{p_{D\|M}}{p_D}\right)\right]$
Specific Definition:	$f(\xi) = -\ln(\xi)$ $I(D;M) = E\left[\ln\left(p_{D\|M}\right)\right] - E[\ln(p_D)]$
Linearity Property:	$\boldsymbol{I(D;M) = H(D) - H(D\|M)}$

<u>Consequence 1</u>: Linearity

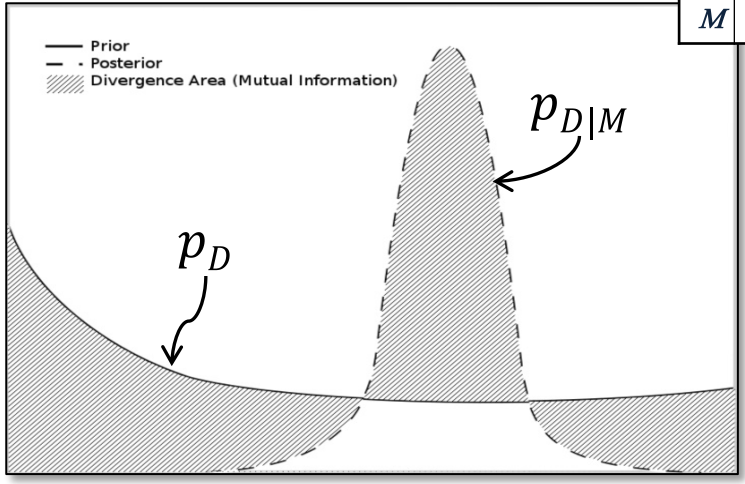


<u>Consequence 2</u>: Bounded under transformations.

$$\boldsymbol{I(D;U) \geq I\big(D;M(U)\big)}$$

Measuring Information

D	**Evaluation Data**
U	**Model Inputs**
$\mathcal{M}$	**Model**



Information is the ability of signal to effect a change in a probability distribution.

General Definition	$I(D;M) = E\left[f\left(\frac{p_{D\|M}}{p_D}\right)\right]$
Specific Definition:	$f(\xi) = -\ln(\xi)$ $I(D;M) = E\left[\ln\left(p_{D\|M}\right)\right] - E\left[\ln\left(p_D\right)\right]$
Linearity Property:	$\boldsymbol{I(D;M) = H(D) - H(D\|M)}$

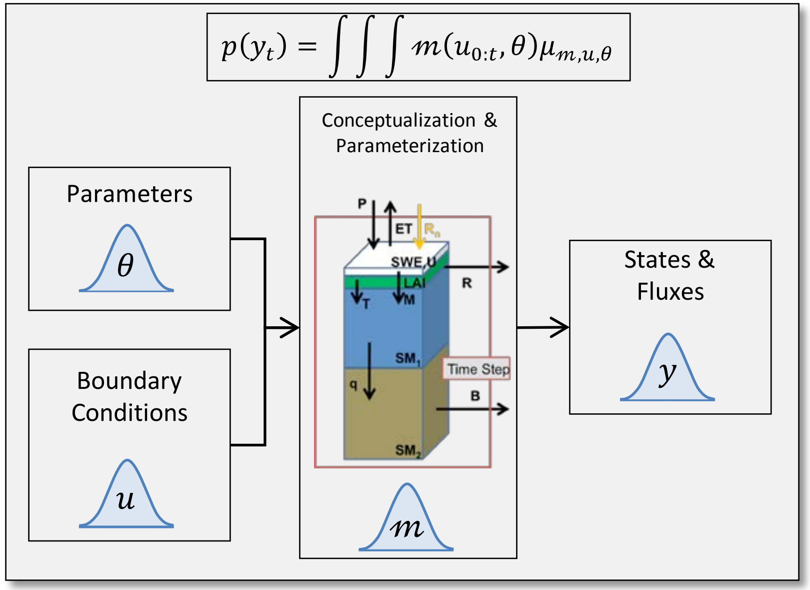
$$U \longrightarrow \boxed{\mathcal{M}} \longrightarrow Y$$

Premise: No system is isolated. A scientist perturbs a system and measures it's response

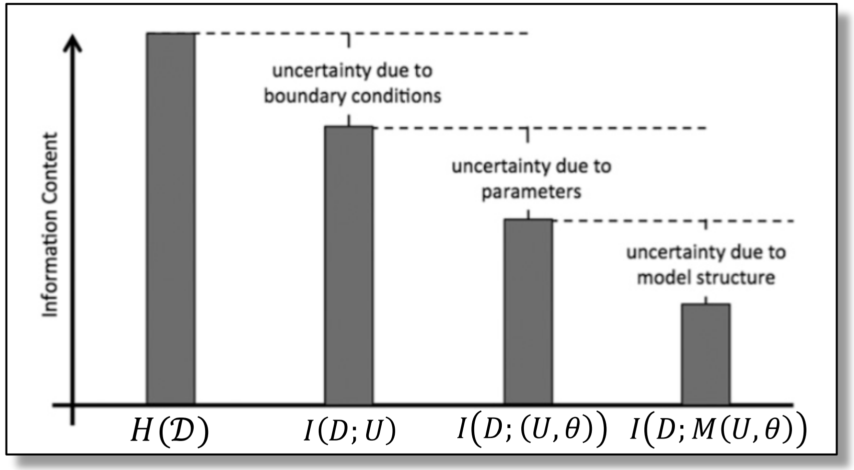
Definition: The information content of data is defined by our ability to derive asymptotic relationships between measured perturbations and responses.

$$D = \mathcal{R}_e(U)$$

Example 1: Uncertainty Segregation

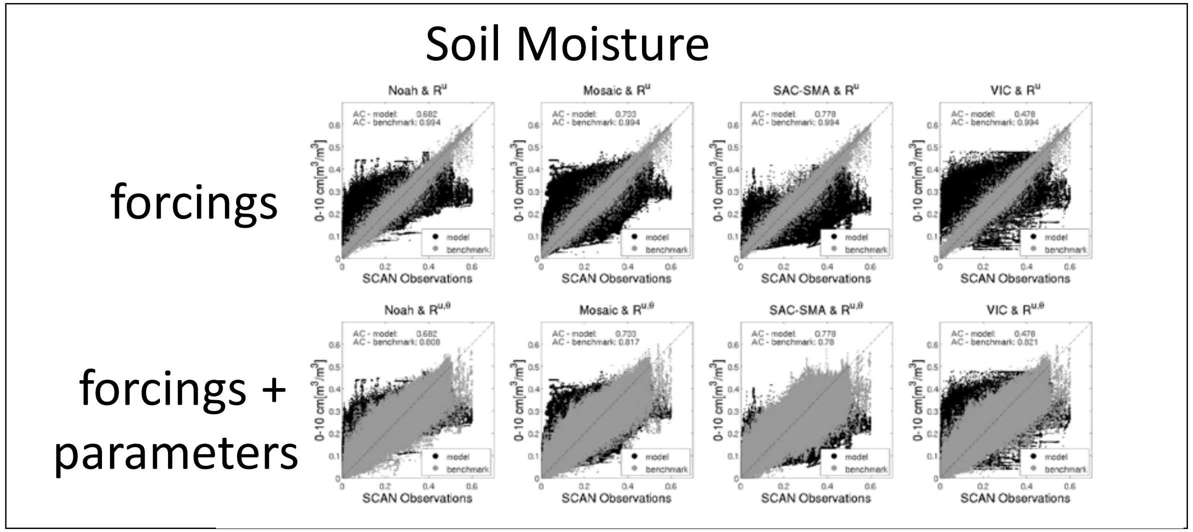


$$I(D;M) = H(D) - H(D|M)$$

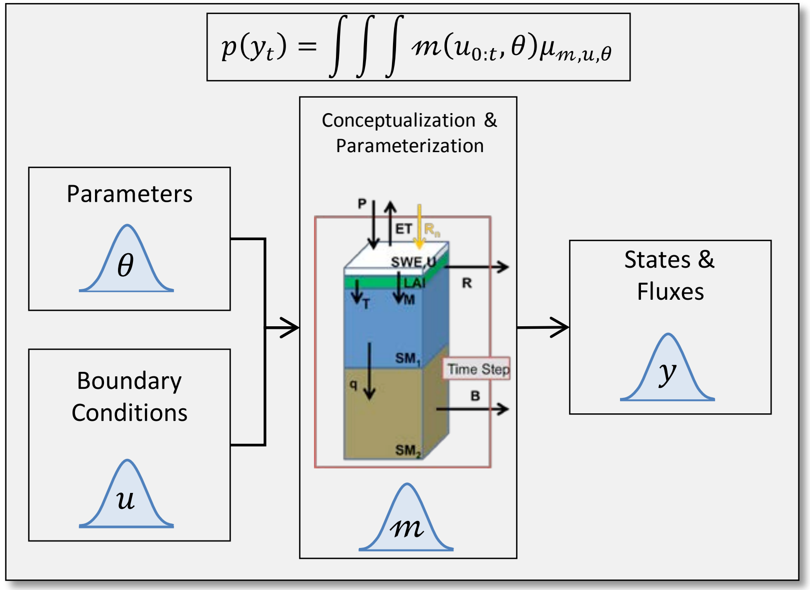


Definition: The information content of data is defined by our ability to derive asymptotic relationships between measured perturbations and responses.

$$D = \mathcal{R}_e(U)$$



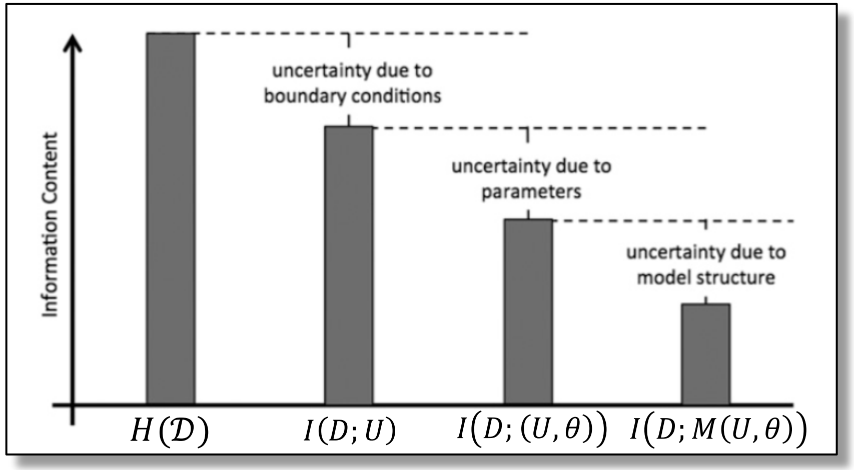
Example 1: Uncertainty Segregation

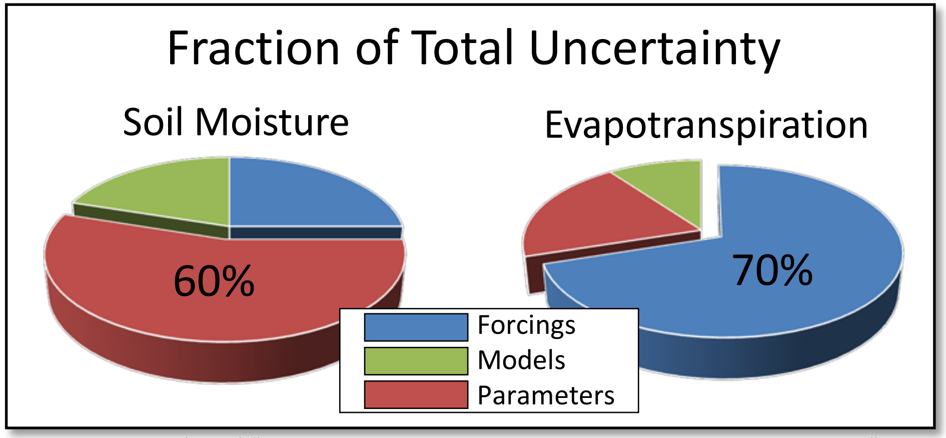


Definition: The information content of data is defined by our ability to derive asymptotic relationships between measured perturbations and responses.

$$D = \mathcal{R}_e(U)$$

$$\boldsymbol{I(D; M) = H(D) - H(D|M)}$$

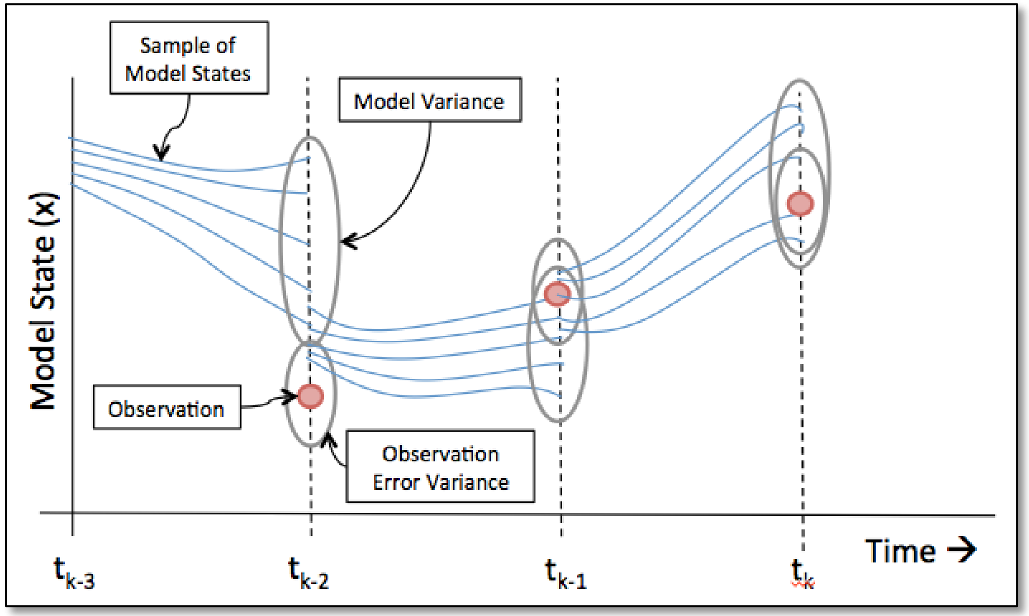


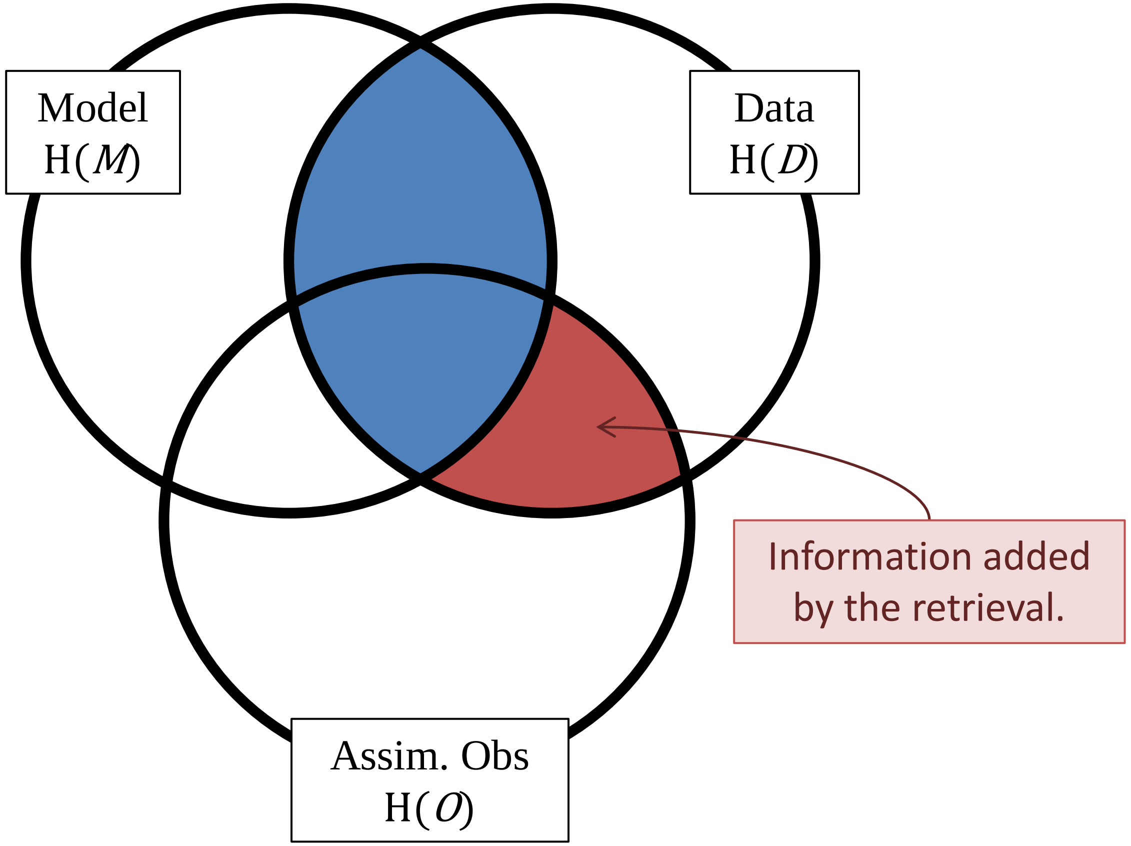


G. Nearing et al. (2016) “Benchmarking NLDAS-2 to Separate Uncertainty Contributions” *JHM*

Example 2: Data Assimilation

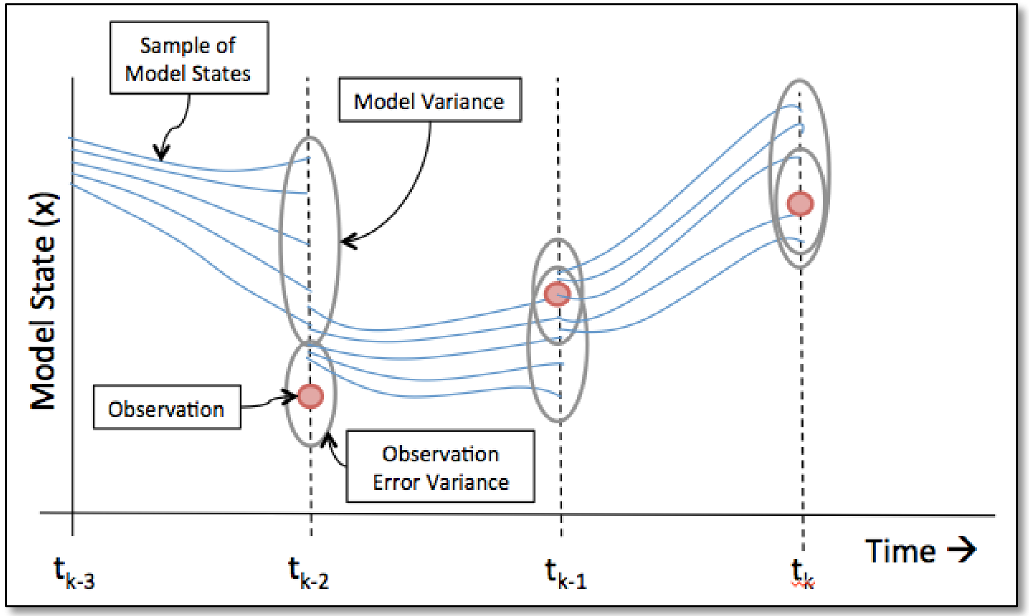
Model:	$dx = \mu(\boldsymbol{x},\boldsymbol{u})dt + \sigma(\boldsymbol{x},\boldsymbol{u})dB_t$
Data Assimilation:	$p(\boldsymbol{x}_t\vert\boldsymbol{y}_t) \propto h(\boldsymbol{y}_t\vert\boldsymbol{x}_t)m(\boldsymbol{x}_t\vert\boldsymbol{u}_{1:t})$

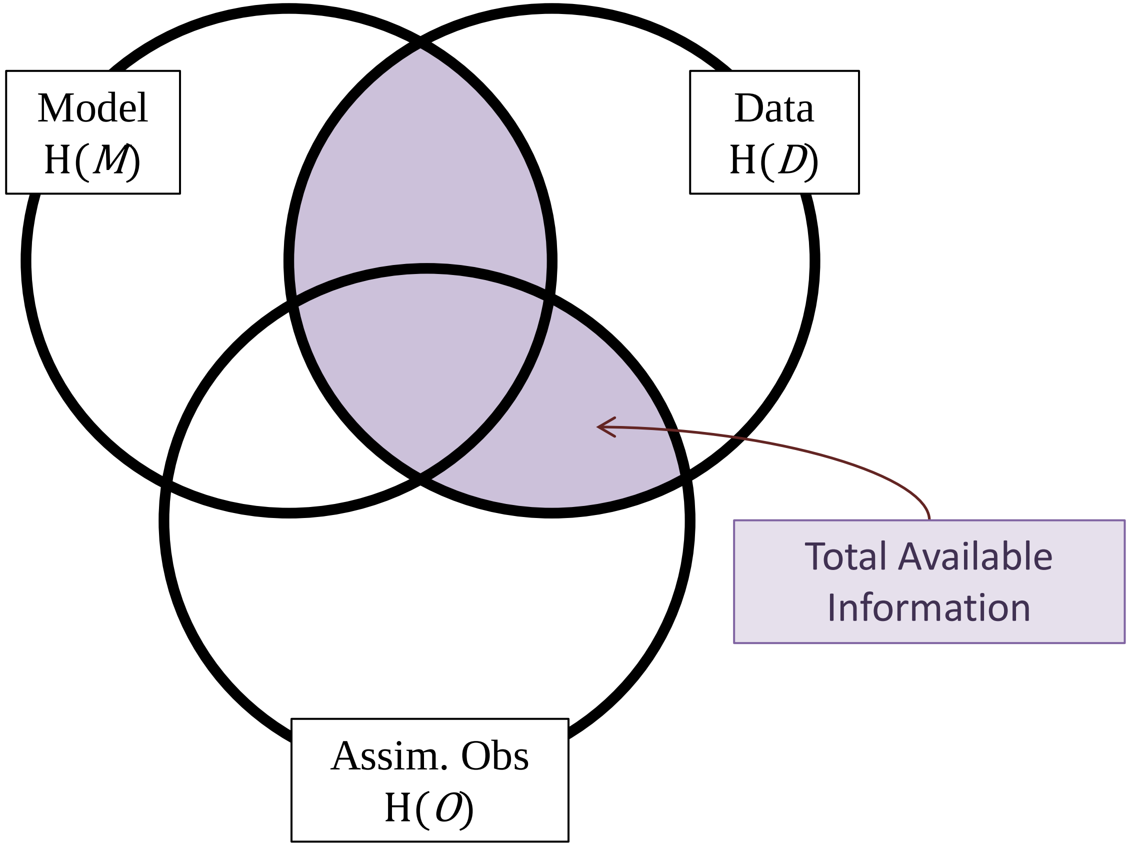




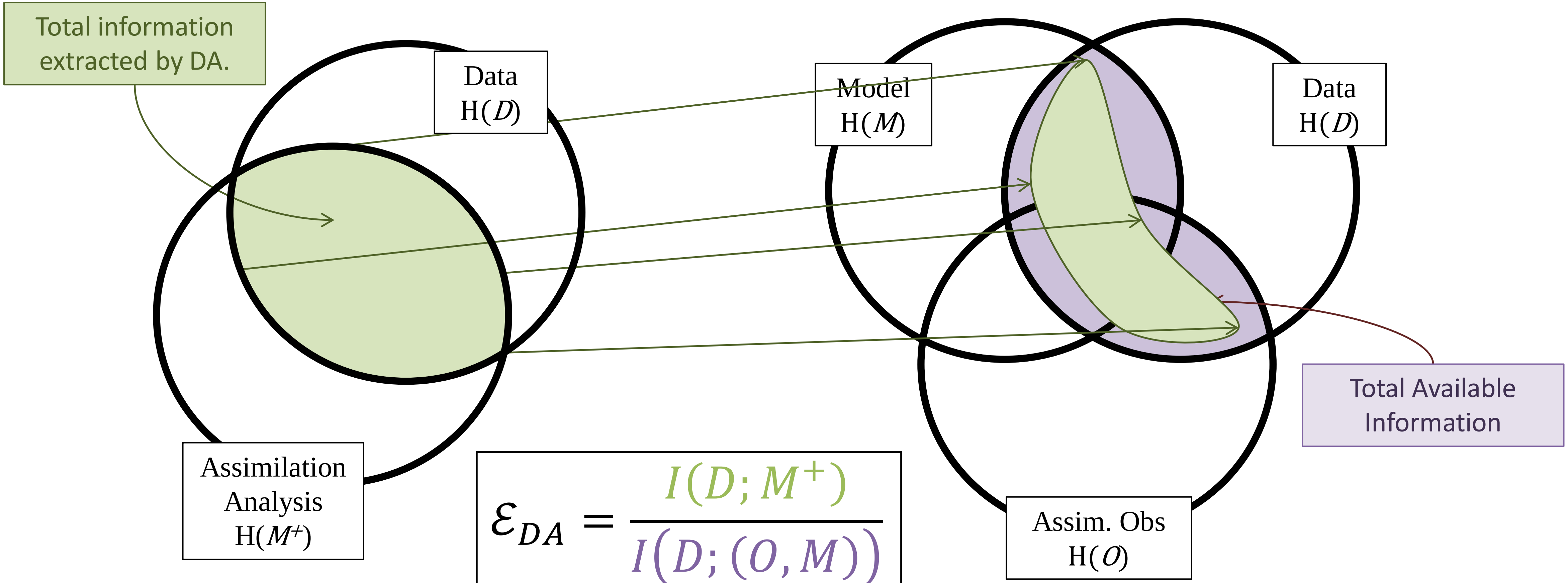
Example 2: Data Assimilation

Model:	$d\boldsymbol{x} = \mu(\boldsymbol{x}, \boldsymbol{u})dt + \sigma(\boldsymbol{x}, \boldsymbol{u})dB_t$
Data Assimilation:	$p(\boldsymbol{x}_t \| \boldsymbol{y}_t) \propto h(\boldsymbol{y}_t \| \boldsymbol{x}_t)m(\boldsymbol{x}_t \| \boldsymbol{u}_{1:t})$





Example 2: Data Assimilation

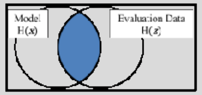
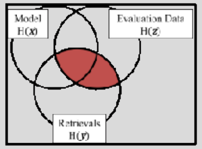
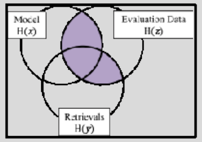
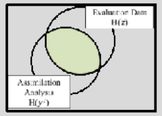
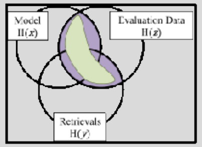


$$\mathcal{E}_{DA} = \frac{I(D;M^{+})}{I\big(D;(O,M)\big)}$$

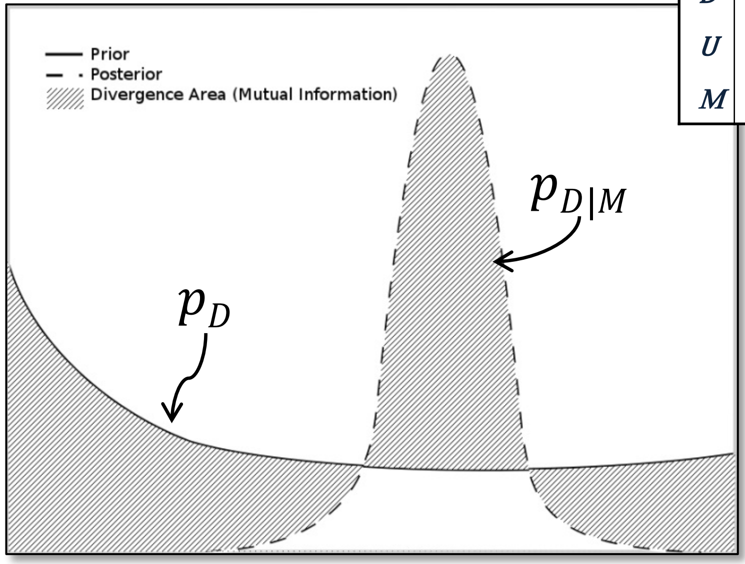
Example 2: Data Assimilation

- AMSR-E Soil Moisture Retrievals
- NOAH-MP Model
- Ensemble Kalman Filter

The Ensemble Kalman Filter is only about 30% efficient in this experiment.

Measurement	Interpretation	Value [nats/nats]
Information in Noah simulations		0.17
Information in LPRM (AMSR-E) retrievals		0.24
Total information in Noah and LPRM (AMSR-E) together		0.61
Information from Data Assimilation		0.18
EnKF Efficiency		0.29

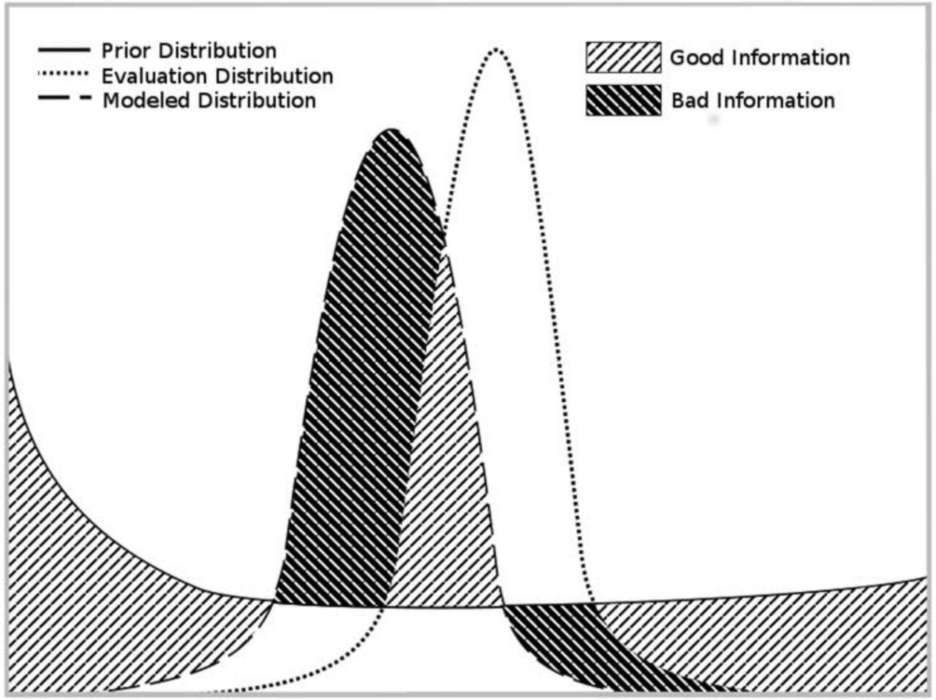
Measuring Information



D	Evaluation Data
U	Model Inputs
M	Model

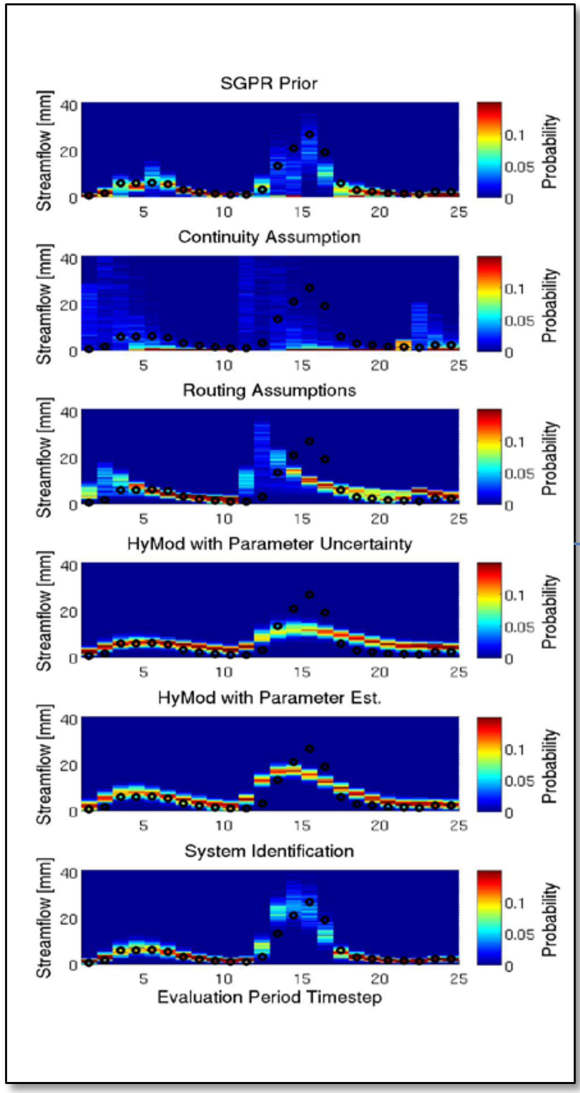
Information is the ability of signal to effect a change in a probability distribution.

General Definition	$I(D;M) = E\left[f\left(\frac{p_{D\|M}}{p_D}\right)\right]$
Specific Definition:	$f(\xi) = -\ln(\xi)$ $I(D;M) = E[\ln(p_{D\|M})] - E[\ln(p_D)]$
Linearity Property:	$\boldsymbol{I(D;M) = H(D) - H(D\|M)}$

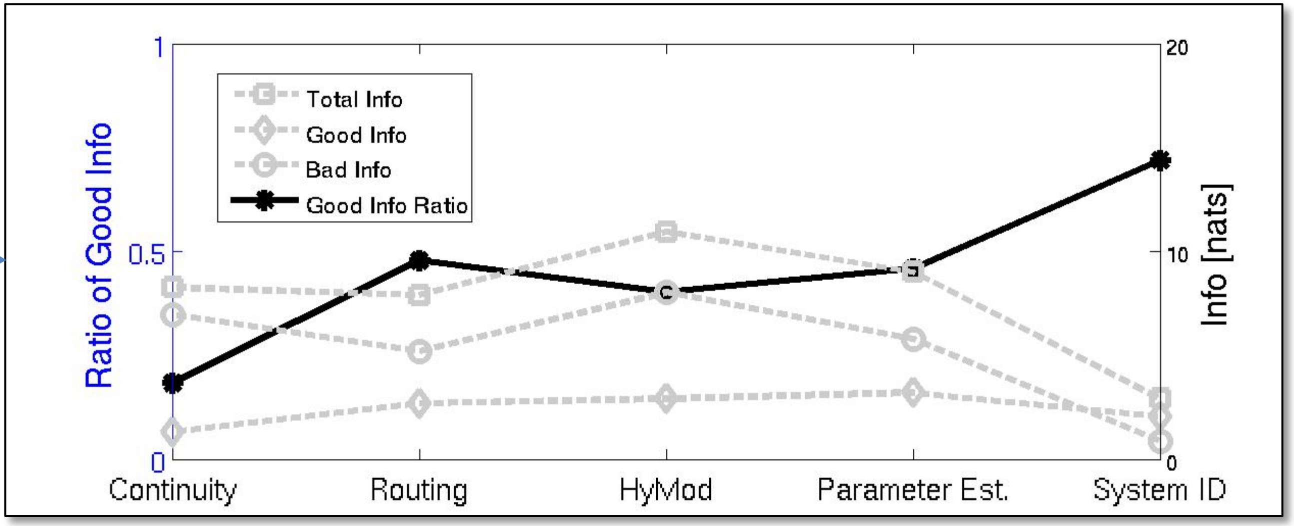


Information *quality* is related to whether the probabilities move in the right direction.

Example 3: Information from Hypotheses

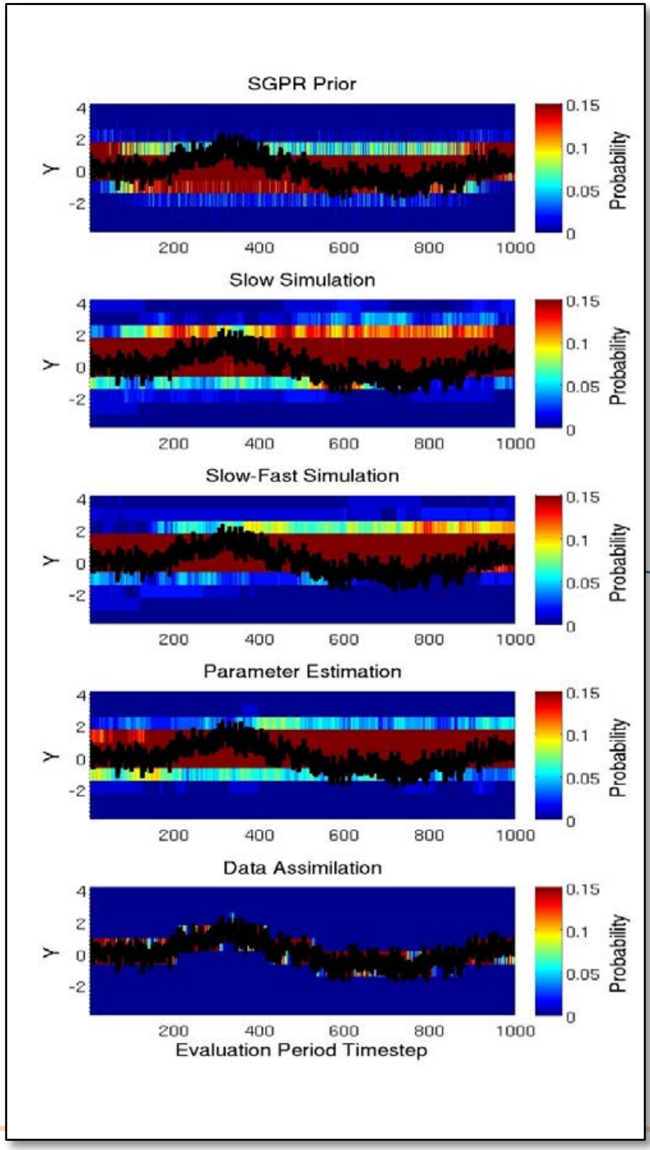


"it must be demonstrated that the model physics actually adds information to the prediction system."
- van den Hurk et al. (2011; BAMS)

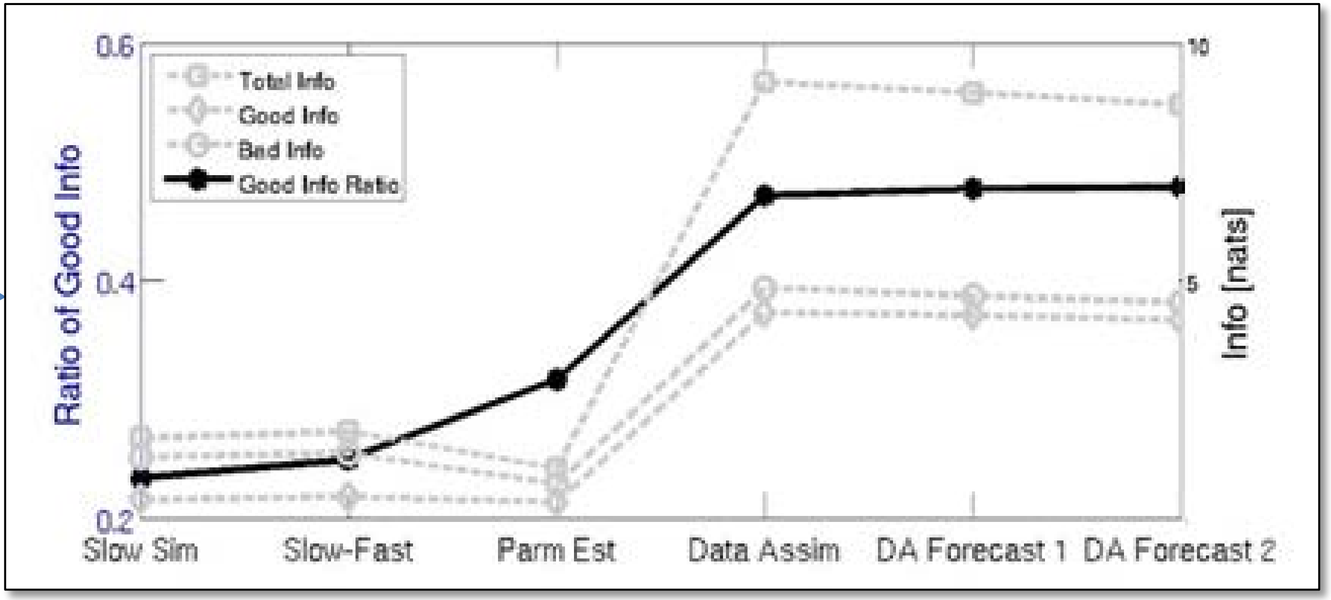


G. Nearing & H. Gupta (2015) "The quantity and quality of information in models" *WRR*

Example 3: Information from Hypotheses



> *"it must be demonstrated that the model physics actually adds information to the prediction system."*
> *- van den Hurk et al. (2011; BAMS)*

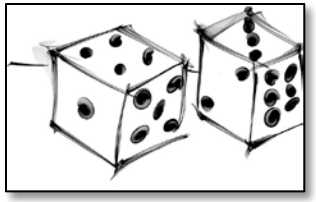


Summary

The ontological model cannot be separated from the epistemological model.

$$U \rightarrow \boxed{\mathcal{M}} \rightarrow Y$$

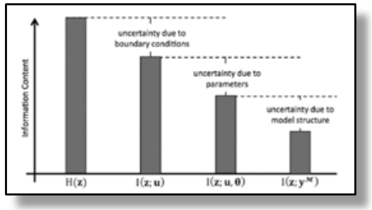
Models translate information.



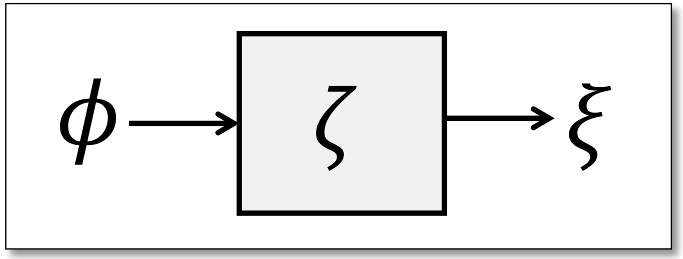
The model of an experiment is a logarithm.

$$I(D;U) >^{?} I\big(D;M(U)\big)$$

This model of an experiment yields a deductive science.



Information is easier to work with than probabilities.


ϕ
ζ
ξ