Information Theory and the Hydrological Sciences

"Models, Data, Uncertainty and Learning"

Hoshin V Gupta (University of Arizona)

2016 Information Theory Workshop

Schneefernerhaus, Garmisch-Partenkirchen, Germany, April 24-28

© Hoshin Gupta, The University of Arizona

Information Theory & Hydrology

There have been a great many applications … (Singh 1997)*

- Derivation of Probability Distributions & Estimation of Pars
- Flow Forecasting via Maximum Entropy Spectral Analysis
- Basin Geomorphology - Characterization of Landscapes & River networks
- Design of Hydrological Networks for Data Collection
- Reliability of Water Distribution Systems
- Hydraulics
- Water Quality Assessment & Design of Water Quality Networks

* Singh, V. P. (1997), The use of entropy in hydrology and water resources, Hydrological Processes, 11(6)

© Hoshin Gupta, The University of Arizona
CC BY

Incomplete Review of the Hydro. Lit.

Amorocho & Espildor (WRR 1973) - **Entropy In Assessment Of Uncertainty In Hydrologic Systems And Models**

In modeling catchments, a wide choice exists among models of varying completeness and sophistication → The concept of Entropy (via Transinformation) provides an objective criterion for model selection.

© Hoshin Gupta, The University of Arizona

The Recent Revival ...

Amorocho & Espildor (WRR 1973)	**Entropy In Assessment Of Uncertainty In Hydrologic Systems And Models**
• • •	
Abebe & Price (HSJ 2003)	**Managing Uncertainty In Hyd. Models Using Complementary Models**
Gupta et al (HP 2008)	**Reconciling Theory With Observations: Elements Of A Diagnostic Approach To Model Evaluation**
Pokhrel and Gupta (WRR 2011)	**On The Ability To Infer Spatial Catchment Variability Using Streamflow Hydrographs**
Weijs et al (HESS 2010)	**Why Hydrological Predictions Should Be Evaluated Using Info. Theory**
Pan et al (JoH 2012)	**Scale Effects On Info Theory-based Measures Applied To Streamflow Patterns In Two Rural Watersheds**
Weijs et al (HESS 2013)	**Data Compression To Define Info Content Of Hydrological Time Series**
Gong et al (WRR 2013)	**Estimating Epistemic and Aleatory Uncertainties during Hydrologic Modeling: An Info. Theoretic Approach**
Weijs & van de Giesen (JoH 2013)	**An Information-theoretical Perspective On Weighted Ensemble Forecasts**
Nearing et al (JoH 2013)	**Information Loss In Approximately Bayesian Estimation Techniques: A Comparison Of Generative And Discriminative Approaches ...**
Weijs et al (HESS 2013)	**Data Compression To Define Info Content Of Hydrological Time Series**
Nearing et al (WRR 2013)	**An Approach To Quantifying The Efficiency Of A Bayesian Filter**

© Hoshin Gupta, The University of Arizona

The Recent Revival ...

Gong et al (WRR 2014)	**Estimating Information Entropy For Hydrological Data: One-dim Case**
Sharma & Mehrotra (WRR 2014)	**An Info Theoretic Alternative To Model A Natural System Using Obs Info Alone**
Nearing & Gupta (WRR 2015)	**The Quantity And Quality Of Information In Hydrologic Models**
Gupta & Nearing (WRR 2015)	**Using Models And Data To Learn: A Systems Theoretic Perspective On The Future Of Hydrological Sciences (WRR Debates)**

-
-
-

© Hoshin Gupta, The University of Arizona

Some Themes Emerge

- Inherently Probabilistic Nature of Hydrological System Modeling
- Assessing Information Content of Data
- Improving Model Predictions
- Model Selection & Calibration
- Model Benchmarking "Best Achievable Performance"
- Assessing Power of Data Assimilation Strategies
- Evaluating & Improving Model Structures (Learning)

© Hoshin Gupta, The University of Arizona

1 - Inherently Probabilistic Nature of Hydrological System Modeling

Singh (HP 1997): *Environmental systems are inherently spatial and complex, and our understanding is less than complete. Many are either fully stochastic, or part-stochastic/ deterministic, due to randomness in:*

a) *System Structure (geometry)*
b) *System Dynamics*
c) *Forcings (sources and sinks)*
d) *Initial and Boundary Conditions*

A stochastic description is needed → the Principle of Maximum Entropy (POME) enables:

1) *Development of such a description*
2) *Determination of the least-biased PDF of a random variable, subject to available information.*
3) *Suggests whether available information is adequate, and if not then additional information should be sought.*

In this way it brings a model and its modeler closer.

© Hoshin Gupta, The University of Arizona
cc BY

2 - Assessing Information Content of Data

Pan et al (JoH 2008): *"Information Content" of precipitation time series is higher than streamflow, but the "Complexity" is higher for the latter → watersheds act as filters of precipitation info, which results in the observed additional complexity. Temporal effects of information are important and must be considered in model evaluation and comparison.*

Weijs et al (HESS 2013): *Data Compression can be used to Quantify Info Content of hydrological time series → "How much can potentially be learned using a data set"). This requires first answering:*

(a) Information about what?

(b) What is the current state of knowledge/belief about that?

Quantification is closely linked to problems of (i) Separating Aleatoric and Epistemic uncertainties (ii) Quantifying Best Achievable Model Performance.

Gong et al (WRR 2014): *Discusses how to estimate Entropy of hydrologic (rainfall and runoff data) while dealing with practical problems of the (a) Zero Effect (discrete-continuous hybrid distributions), (b) Bin-Width Selection, (c) Finite Precision of Measurements, and (d) Skewness in the PDF.*

© Hoshin Gupta, The University of Arizona
CC BY

3 - Improving Model Predictions

Abebe & Price (HSJ 2003): *Data-driven (ANN) approach based in Mutual Information can be used to model the prediction errors of a conceptual model, and added on top to improve the forecasts.*

Weijs & van de Giesen (JoH 2013): *Present method for Weighting Ensemble Forecasts, based on an extension of POME, to ensure that no more information is added to the ensemble than is present in the forecast. Demonstrate that all other methods result in weights that add either too little or too much (i.e. fictitious) information to the ensemble.*

© Hoshin Gupta, The University of Arizona

4 - Model Selection & Calibration

Amorocho & Espildor (WRR 1973): *In modeling catchments, a wide choice exists among models of varying completeness and sophistication → The concept of Entropy (via Transinformation) provides an objective criterion for model selection.*

Weijs et al (HESS 2010): *All models should be explicitly probabilistic, and all hydrological predictions should be evaluated using Information Theory.*

In practice, <u>all Deterministic Forecasts are interpreted as uncertain</u> by the user.

Models should be calibrated using information-theoretical scores, to:

a) *Maximize the information they provide*
b) *Extract all information from the observations*
c) *<u>Avoid learning from information that is not there</u>.*

© Hoshin Gupta, The University of Arizona

5 - Model Benchmarking: Best Achievable Performance

Gong et al (WRR 2013): *Suggests a way to quantify "Best Achievable Performance" for a model via data-driven modeling, thereby characterizing Model Structure Adequacy.*

Presents way to estimate Model Adequacy in terms of its Aleatory Uncertainty (that cannot be diminished) and its Epistemic Uncertainty (that can be resolved by improving the model).

Sharma & Mehrotra (WRR 2014): *Propose an Info Theoretic approach to model natural systems using observational information alone, as an alternative to empirical or semi-empirical approaches.*

Test their approach using synthetically generated data sets from known linear, nonlinear, and high-dimensional dynamic yet chaotic systems

© Hoshin Gupta, The University of Arizona

6 – Assessing Power of Data Assimilation Strategies

Nearing et al (JoH 2013): Uses *Shannon's theory to measure the information assimilated into models from observations and to characterize:*

(a) Missing Information

(b) Used Information

(c) Bad Information.

Shows that discriminative modeling (regression) to be more efficient than generative modeling (data assimilation) in extracting info from obs (so better suited for many practical problems).

Nearing et al (WRR 2013): *Uses metrics based on Discrete Shannon Entropy to quantify how much of the uncertainty in a Posterior PDF is due to:*

(a) Observation operator

(b) Observation error

(c) Approximations of Bayes' Law.

Makes possible to analyze the efficiency of a proposed observation system and data assimilation strategy (e.g., EnKF does not use all of the info in soil moisture observations).

© Hoshin Gupta, The University of Arizona

7 - Evaluating & Improving Model Structures (Learning)

Gupta et al (HP 2008): *Info Theory can be used to characterize "system-relevant" information in a data set (signature properties), and to develop a "Diagnostic Evaluation" approach to reconciling environmental theory/models with observations.*

Nearing & Gupta (WRR 2015): *Quantify intuition that "models provide information". Demonstrate that dynamical models use induction to assimilate and store information from hypotheses and data. Show how this stored information can be directly measured.*

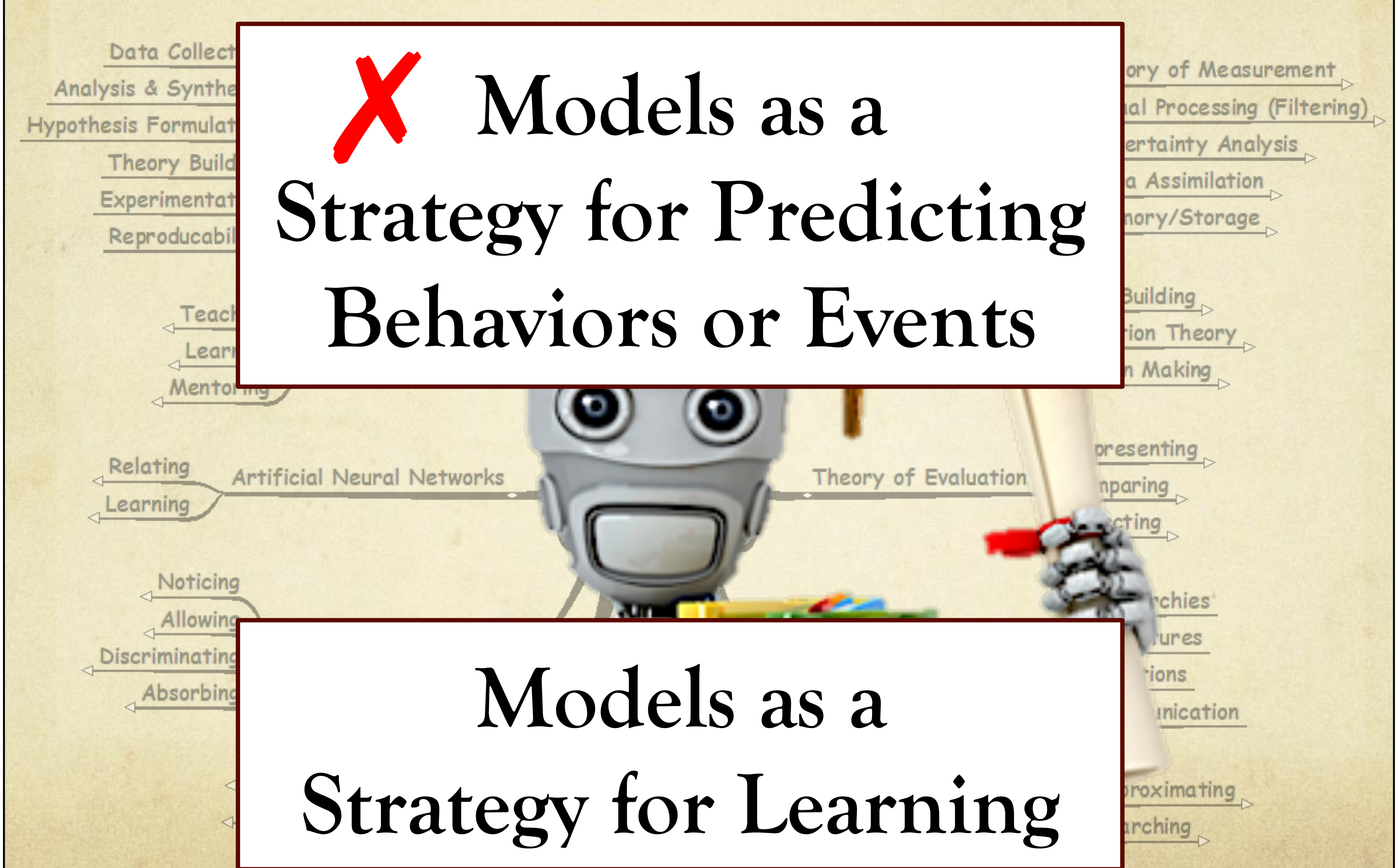
Gupta & Nearing (WRR 2015): *A perspective based in Info Theory can improve our ability to learn from the juxtaposition of models and data.*

> *Build upon the "Step-wise Characterization of Model Structure" proposed by Gupta et al (WRR 2012 - Towards A Comprehensive Assessment Of Model Structural Adequacy).*

Argue that we should give more emphasis to Process Modeling & System Architecture (i.e., incorporating knowledge from Physics) rather than on improving (what are generally semi-empirical) System Parameterization Equations and on Parameter Estimation.

© Hoshin Gupta, The University of Arizona
CC BY

My Interest is the "Learning" Problem



✗ Models as a Strategy for Predicting Behaviors or Events

Models as a Strategy for Learning

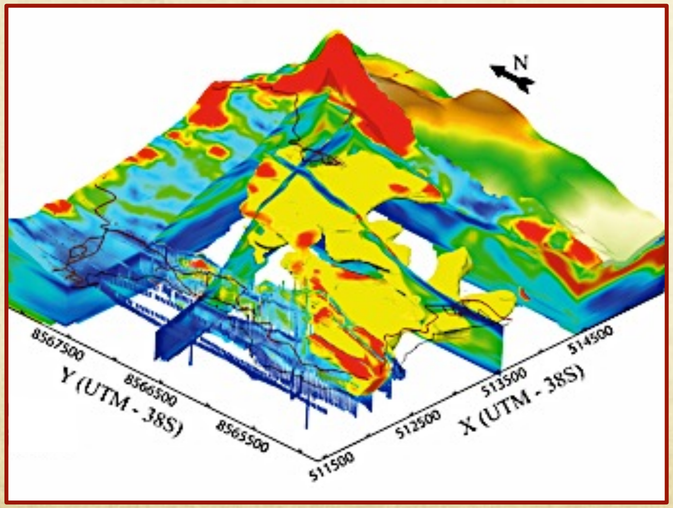
© Hoshin Gupta, The University of Arizona

Dynamical Environmental Systems Models

Working Definition

A **Dynamical Environmental Systems Model (DESM)** is a *simplified representation* of the structure & function of a dynamical system that:

1. Enables (a) *Reasoning* within an idealized framework
   (b) *Testable predictions* under new circumstances.
2. By Encoding (a) Knowledge of *Physics* (conservation, thermodynamics)
   (b) Knowledge of *System Geometry & Material Properties*
   (c) Knowledge of *What we Know that we Do Not Know*



Why Simplified?

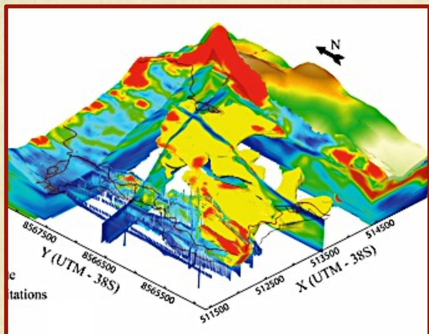
a) Knowledge is *Incomplete* & *Uncertain*
b) Real system is *Infinite Dimensional*
c) Need to "compute" in *Finite Time* using *Finite Resources*

© Hoshin Gupta, The University of Arizona

We Use DESM for Scientific Investigation

Intuitively

We understand that "**Models**" & "**Data**" codify **Knowledge** about the World … in the form of *Information*

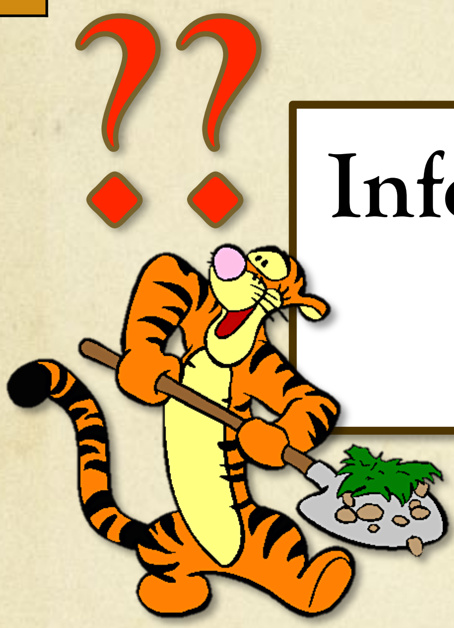


How ?

How can I use this insight to improve my models ?

How does "Learning" happen ?

What is The Nature of Information ?



Info is the answer to questions such as:

When, What, Where, How, Why …

& How Much … etc.

© Hoshin Gupta, The University of Arizona

What is The Nature of Information ?

Info is always "about" something

Context Matters

DATA is not Info … until viewed in context

3.14

© Hoshin Gupta, The University of Arizona

What is The Nature of Information ?

Info is always "<u>about</u>" something

Context Matters

DATA is not Info ... until viewed in context

3.14

Streamflow

(mt^3/sec)

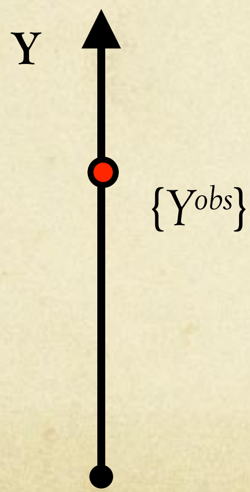
© Hoshin Gupta, The University of Arizona

Information is always "ABOUT" Something

In the Context of Model-Data Learning

Info is always *about*

- Values (Y)
- Relationships (R: X➔Y)
- Constraints (C) … *Assumptions are a kind of Constraint*



A single DATA Point encodes Info about:

A Value of "Something"

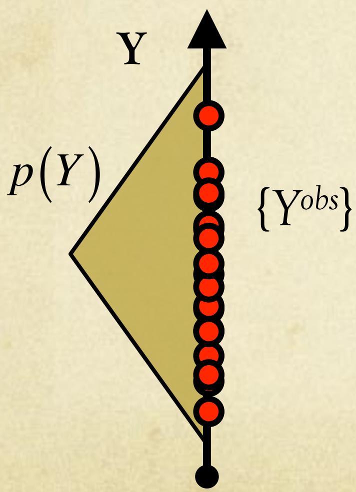
© Hoshin Gupta, The University of Arizona

Information is always "ABOUT" Something

In the Context of Model-Data Learning

Info is always *about*

- Values (Y)
- Relationships (R: X➔Y)
- Constraints (C) … *Assumptions are a kind of Constraint*



A set of DATA Points encodes Info about:

The Distribution of Values

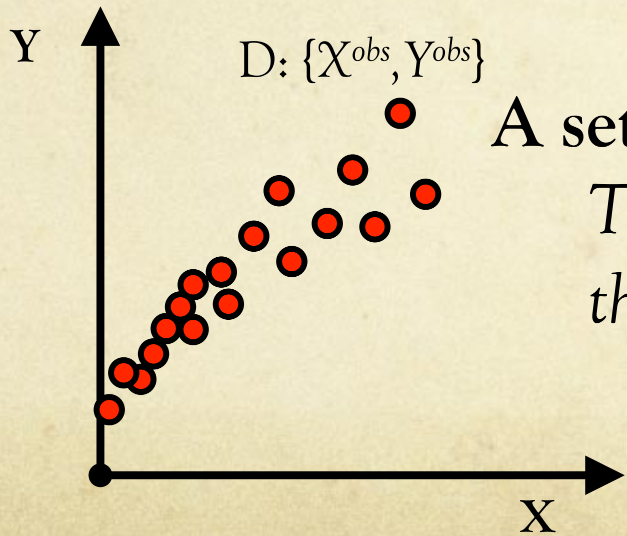
© Hoshin Gupta, The University of Arizona

Information is always "ABOUT" Something

In the Context of Model-Data Learning

Info is always *about*

- Values (Y)
- Relationships (R: X➔Y)
- Constraints (C) ... *Assumptions are a kind of Constraint*



A set of DATA Points encodes Info about:

The space-time-ordered Relationships among those values

© Hoshin Gupta, The University of Arizona

1

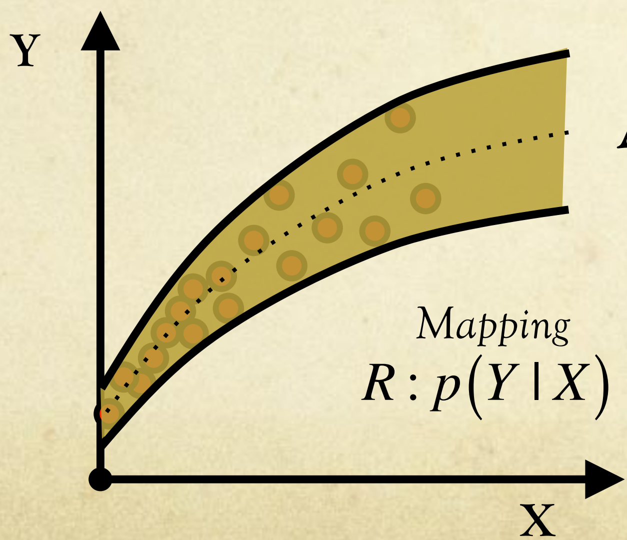
2

Information is always "ABOUT" Something

In the Context of Model-Data Learning

Info is always *about*

- Values (Y)
- Relationships (R: X➔Y)
- Constraints (C) … *Assumptions are a kind of Constraint*



A MODEL encodes Info about:

The space-time-ordered Relationships between variables 'behind' those values

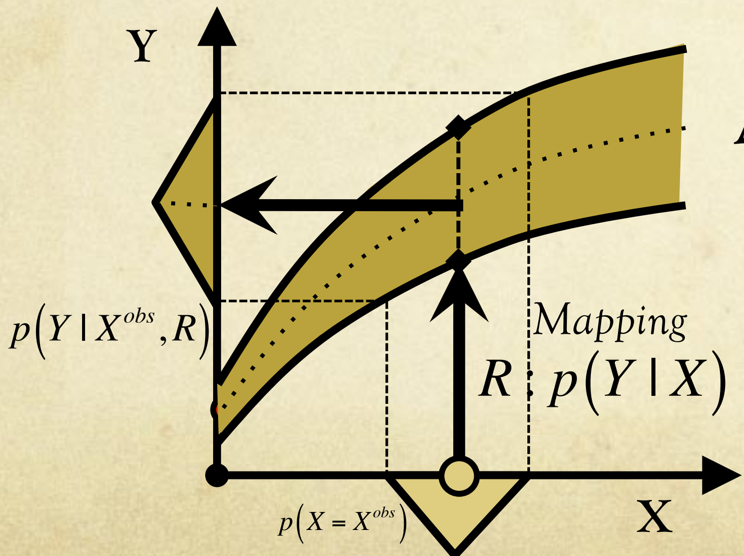
© Hoshin Gupta, The University of Arizona

Information is always "ABOUT" Something

In the Context of Model-Data Learning

Info is always *about*

- Values (Y)
- Relationships (R: X→Y)
- Constraints (C) ... *Assumptions are a kind of Constraint*



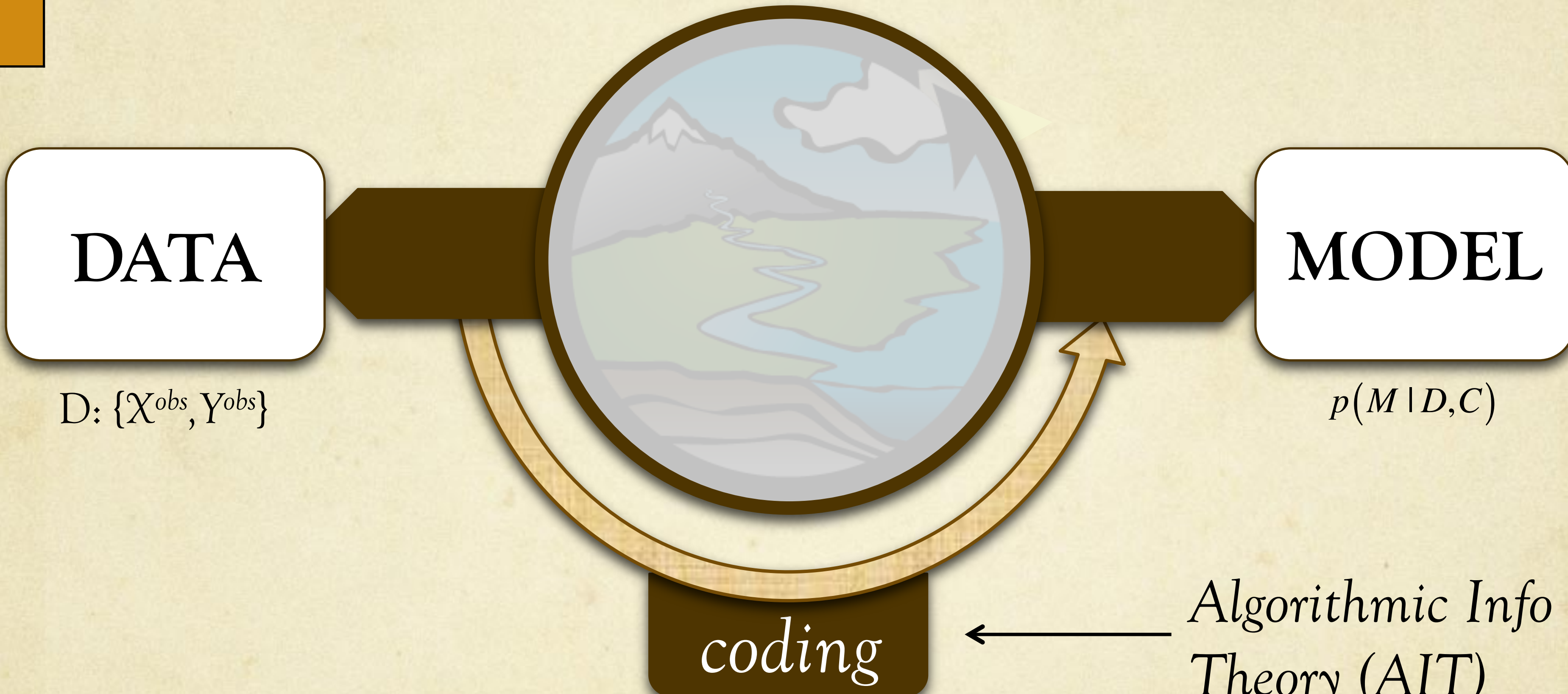
A MODEL encodes Info about:

Values p(Y | X,M) conditional on Relationships and Data

© Hoshin Gupta, The University of Arizona

So … Learning Involves

Converting Data-Info → Model-Info



D: $\{X^{obs}, Y^{obs}\}$

$p(M \mid D, C)$

Algorithmic Info Theory (AIT)

** *Info is added by the "Conversion Process"*

** *The conditioning role of Assumptions (A) is very strong*

© Hoshin Gupta, The University of Arizona

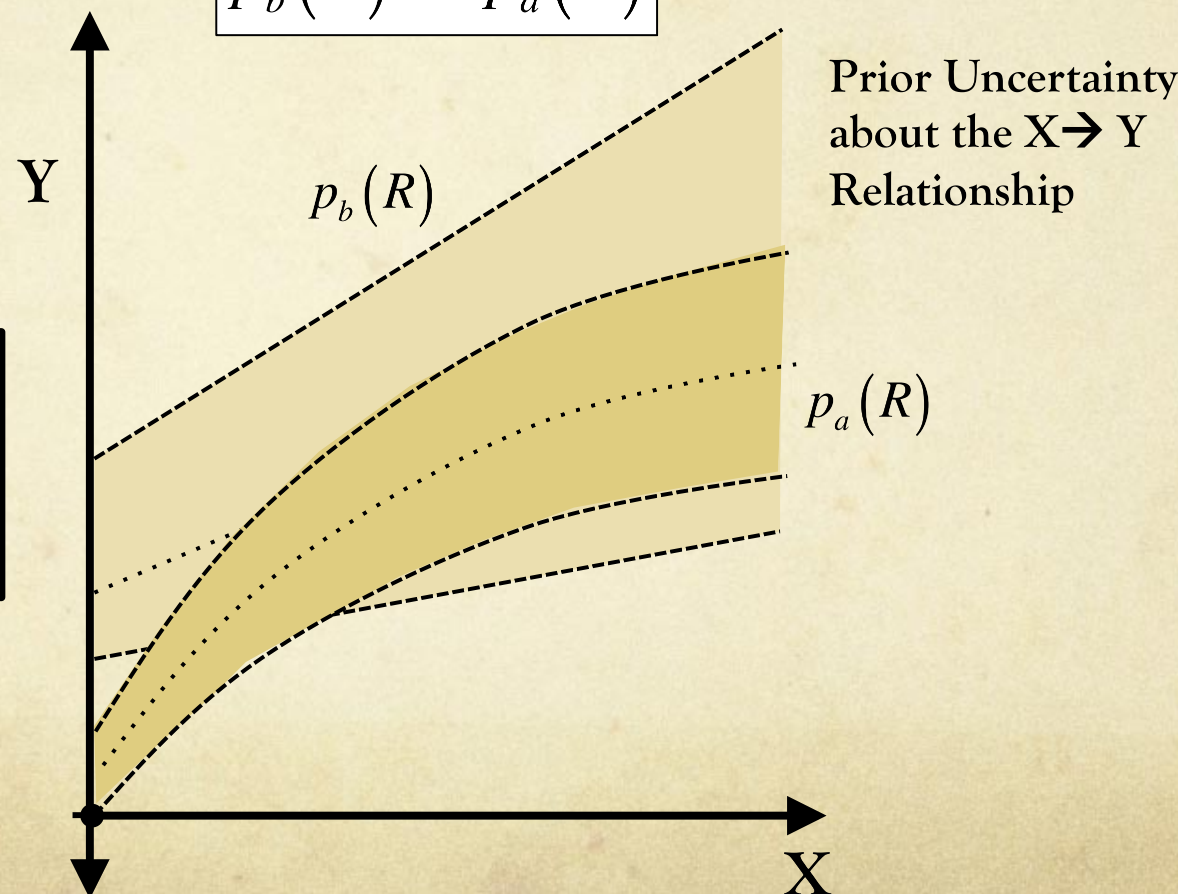
1
2

And Occurs When …

Our **Prior Uncertainty is Changed**

due to the Assimilation of New Information

$$\boxed{p_b(R) \Rightarrow p_a(R)}$$

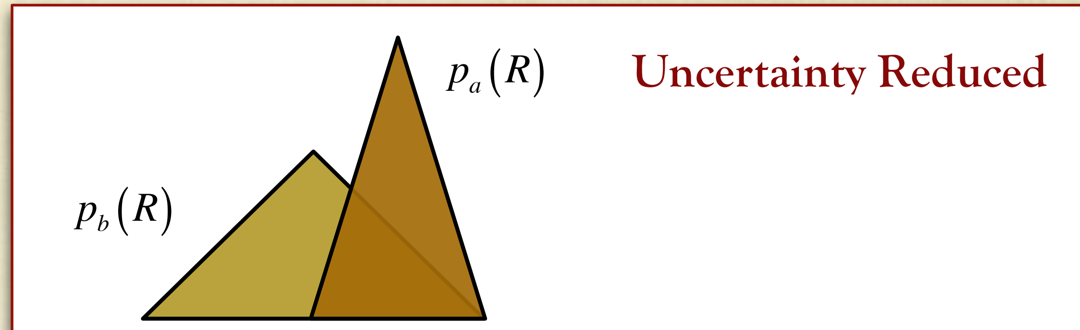


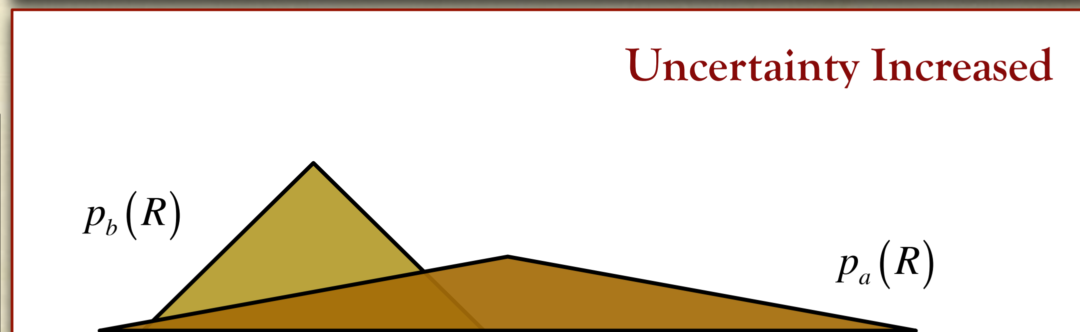
“Changed” …
not Reduced

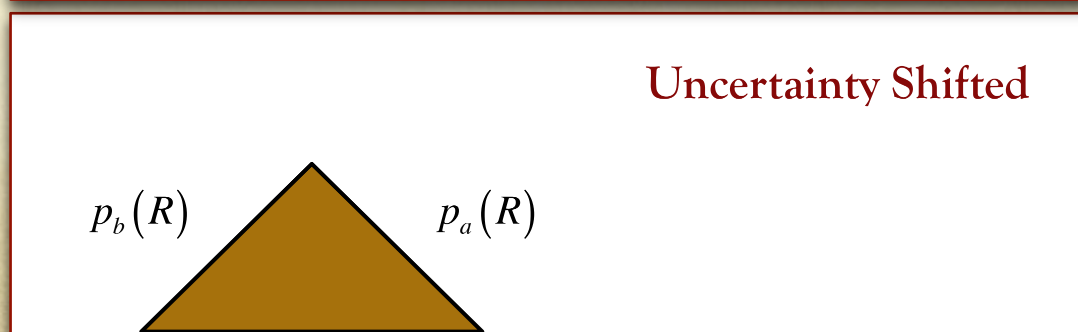
© Hoshin Gupta, The University of Arizona

So … How to Characterize Learning?

Learning occurs when our **<u>Prior Uncertainty is Changed</u>** *due to the Assimilation of New Information*







"<u>Changed</u>" …
not Reduced

© Hoshin Gupta, The University of Arizona
CC BY

1

2

But … Information Can Be

GOOD

BAD

MIXED

(Partially Good & Bad)

Dealing with this is a

MAJOR CHALLENGE to ESTIMATION THEORY

Nearing GS, HV Gupta and W Crow (2013), Information Loss in Approximately Bayesian Data Assimilation: A Comparison of Generative and Discriminative Approaches to Estimating Agricultural Yield, Journal of Hydrology, 507, pp. 163-173

Nearing GS, HV Gupta, WT Crow and Wei G (2013), An Approach to Quantifying the Efficiency of a Bayesian Filter, Water Resources Research, 49, 1–10, doi:10.1002/wrcr.20177

© Hoshin Gupta, The University of Arizona

How do DESM's Encode Information ?
As a Hierarchical Sequence of Decisions

1. *Control Volume, Physics, Processes to Include, System Geometry & Material Properties*

2. *Scale, Dimension & 3D Spatial Structure*

3. *Process Relationships*

4. *Uncertainty*

5. *Solution Methodology*

© Hoshin Gupta, The University of Arizona

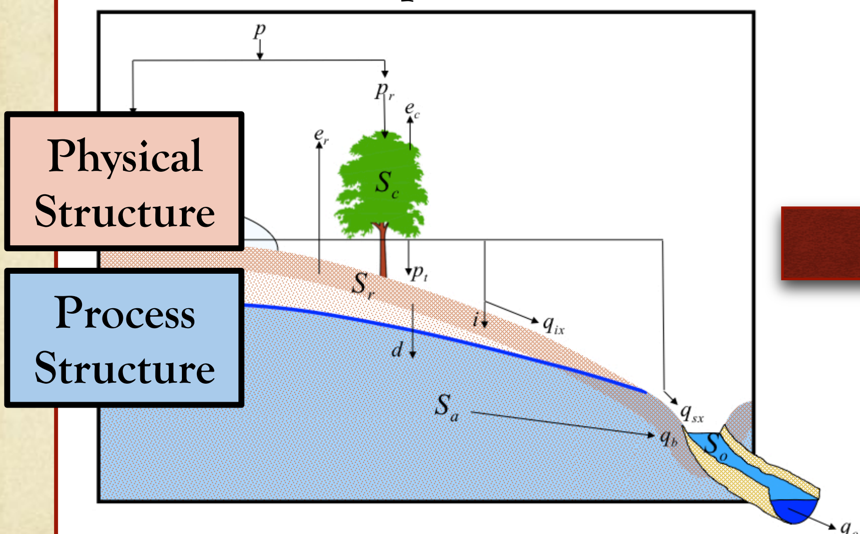
Step One – Conservation Laws

Information About:

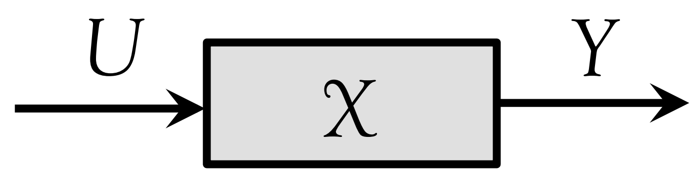
1. Physics (conservation, *thermodynamics*), Physical Processes, System Geometry & Material Properties to include

Question: *What knowledge to encode and what to ignore?*

Conceptual Model



System Diagram



$$\frac{dX_t}{dt} = U_t - Y_t$$

Conservation

$$\{U_t, X_t, Y_t\} \geq 0$$

$$X_t \geq Y_t$$

Constraints

Result: *A System Diagram and Conservation Law Hypothesis - H^{CL}*

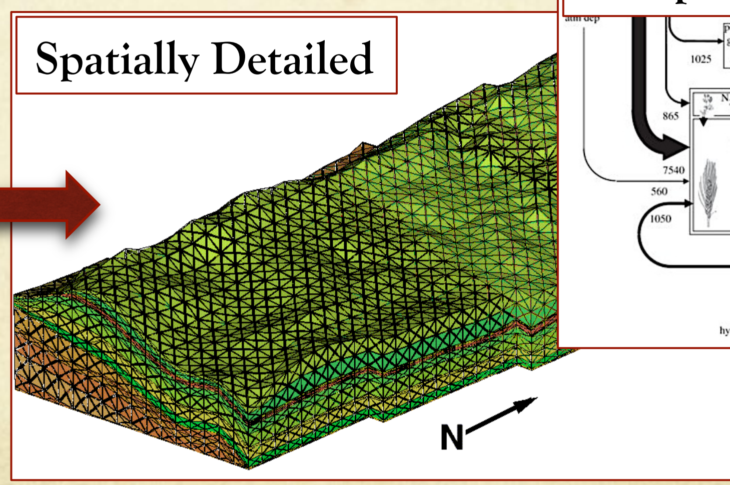
© Hoshin Gupta, The University of Arizona

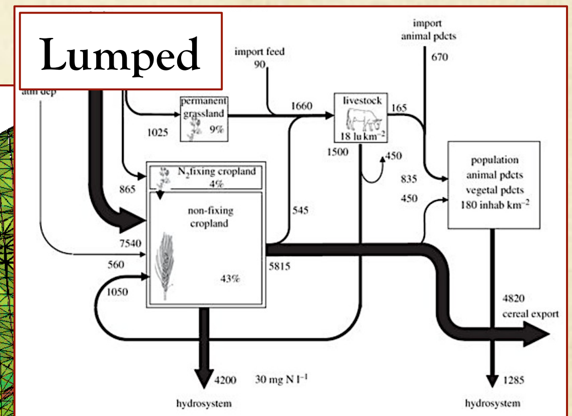
Step Two – System Architecture

<u>Information About:</u>

2. Scale, Dimension and *3D* Spatial Structure of the *State-Space* (elements), to enable finite computation

Question: *What is a sufficiently complex, <u>finite dimensional</u>, spatially organized representation of sub-system architecture?*





Result: *A System Architecture Hypothesis - H^{SA}*

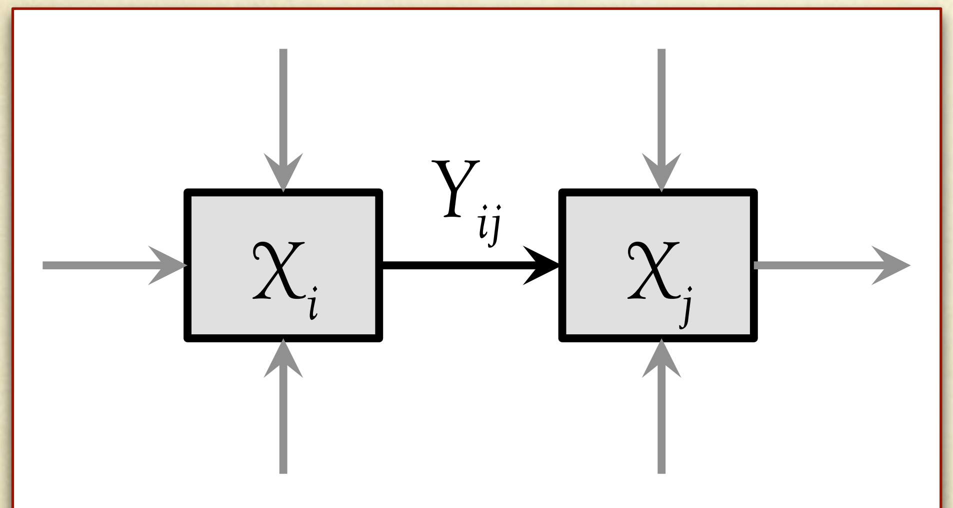
© Hoshin Gupta, The University of Arizona

Step Three – Process Parameterization

Information About:

3. Process Relationships via *Equations*, that account for Sub-Element Process & Material Heterogeneity

Question: *What mathematical forms to use for the Process Parameterization equations, at the architectural scale of interest?*



$$Y_{ij} = \overset{?}{f}_{xy}\left(\Delta X_{ij} \mid \theta_{xy}\right)$$

Parameters are artifacts of equation choice

Result: *A Process Parameterization Hypothesis - H^{PP}*

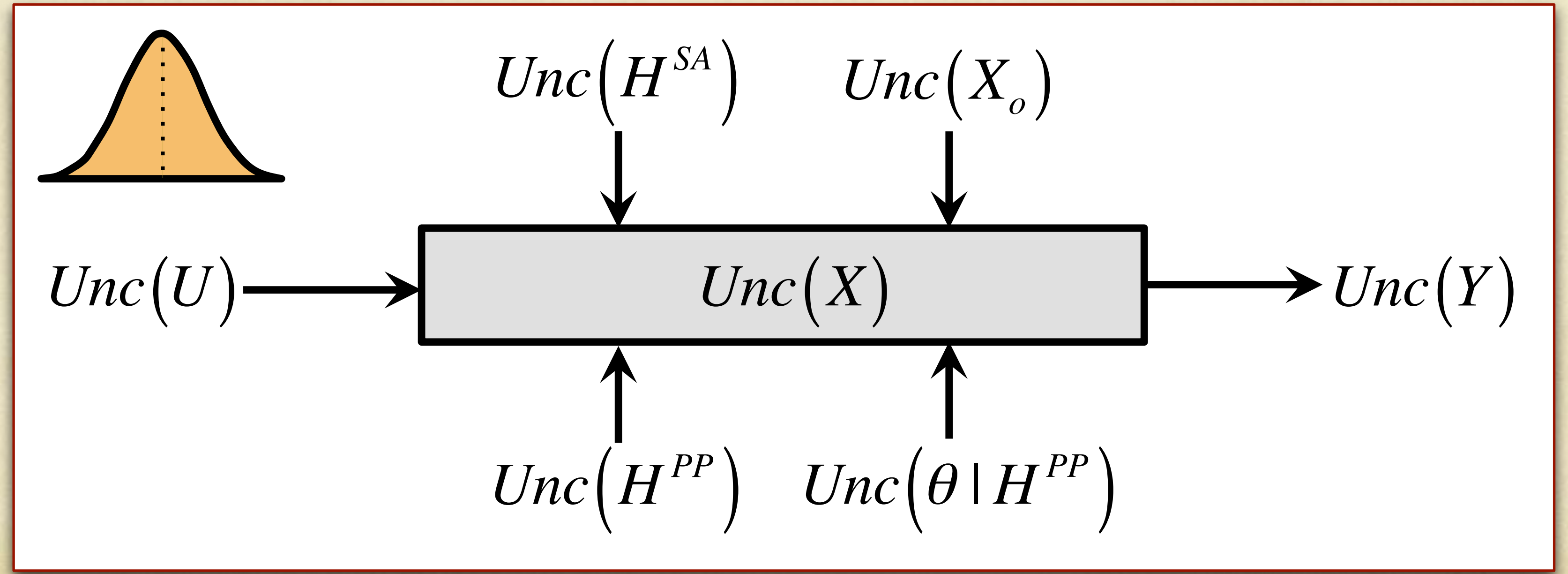
© Hoshin Gupta, The University of Arizona

Step Four – Uncertainty

<u>Information About:</u>

4. Uncertainties to be Accounted For

Question: *What uncertainties are important, and how to represent them mathematically?*



Result: *An Uncertainty Hypothesis - H^{UN}*

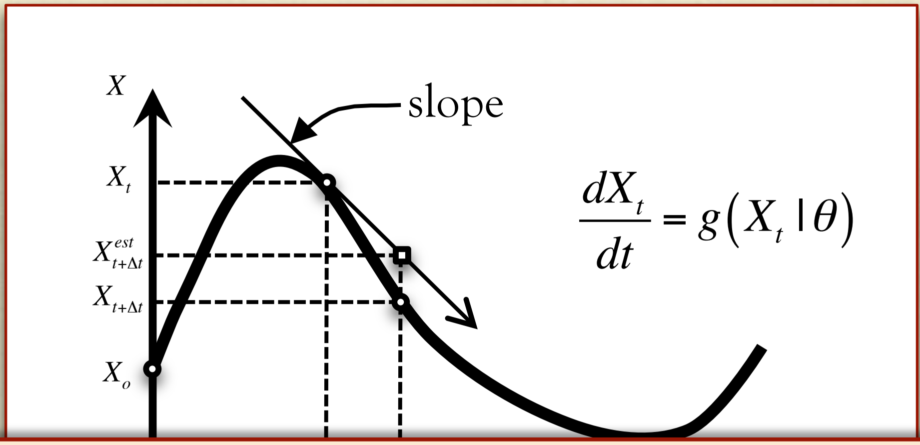
© Hoshin Gupta, The University of Arizona

Step Five – Solution Procedure

Information About:

5. Procedure for 'Solving' the resulting *Mathematical Model*

Question: *How to Integrate (in space & time) the resulting system of (stochastic) differential equations?*



Result: *A Computational Model*

→ Practical manifestation of the Overall System Hypothesis - H^{OS}

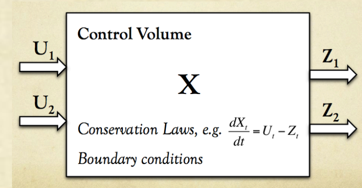
→ Structured hierarchy of Conservation Law, System Architecture, Process Parameterization, and Uncertainty Hypotheses - $H^{OS} = \{H^{UN} | H^{PP} | H^{SA} | H^{CL}\}$

© Hoshin Gupta, The University of Arizona

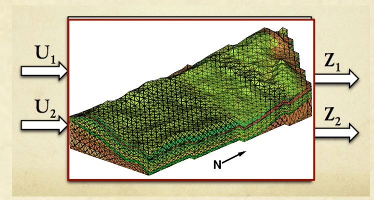
Information is Added at Each Step

(Uncertainty is Changed)

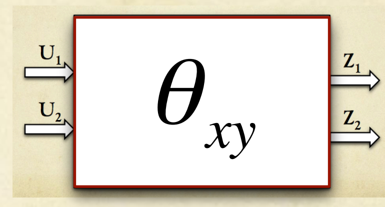
1. Conservation Laws *restrict possible* **U-X-Y** *trajectories*



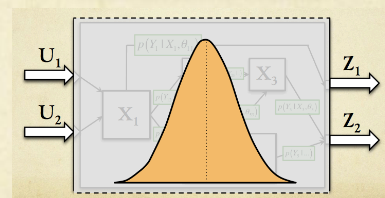
2. System Architecture (a) *further restricts trajectories* & (b) *determines spatial variability*



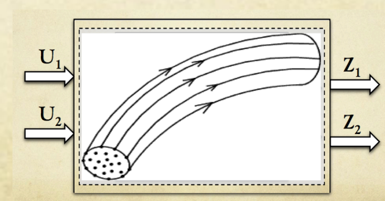
3. Process Parameterization (a) *further restricts trajectories* & (b) *introduces "tunable" parameters*



4. Specification of Uncertainty *characterizes and quantifies "known unknowns"*



5. Solution Procedure *converts Model Info & Input Info into specific (uncertain)* **X-Y** *trajectories*



© Hoshin Gupta, The University of Arizona

What Can Go Wrong?

1. Problem Becomes Over-Constrained

Due to Hypotheses that are Unjustifiably Strong

a) Neglect Heterogeneity that is important

b) Over-simplify the System Architecture

c) Incorrect Process Equations forms

d) Deterministic Process Parameterizations (instead of Stochastic)

2. Problem Becomes Under-Constrained

Due to Lack of Knowledge

a) Do not know Process Physics at the scale of system elements

b) Do not know Heterogeneity of Material Properties and Geometry at scale of system elements

c) Do not know (& account for) Heterogeneity of Material Properties and Geometry *at scales smaller than the system elements*

© Hoshin Gupta, The University of Arizona

1
2
3
4

What Can Go Wrong?

1. **Problem Becomes Over-Constrained**

   Due to Hypotheses that are Unjustifiably Strong

Usually some Combination of Both

2. **Problem Becomes Under-Constrained**

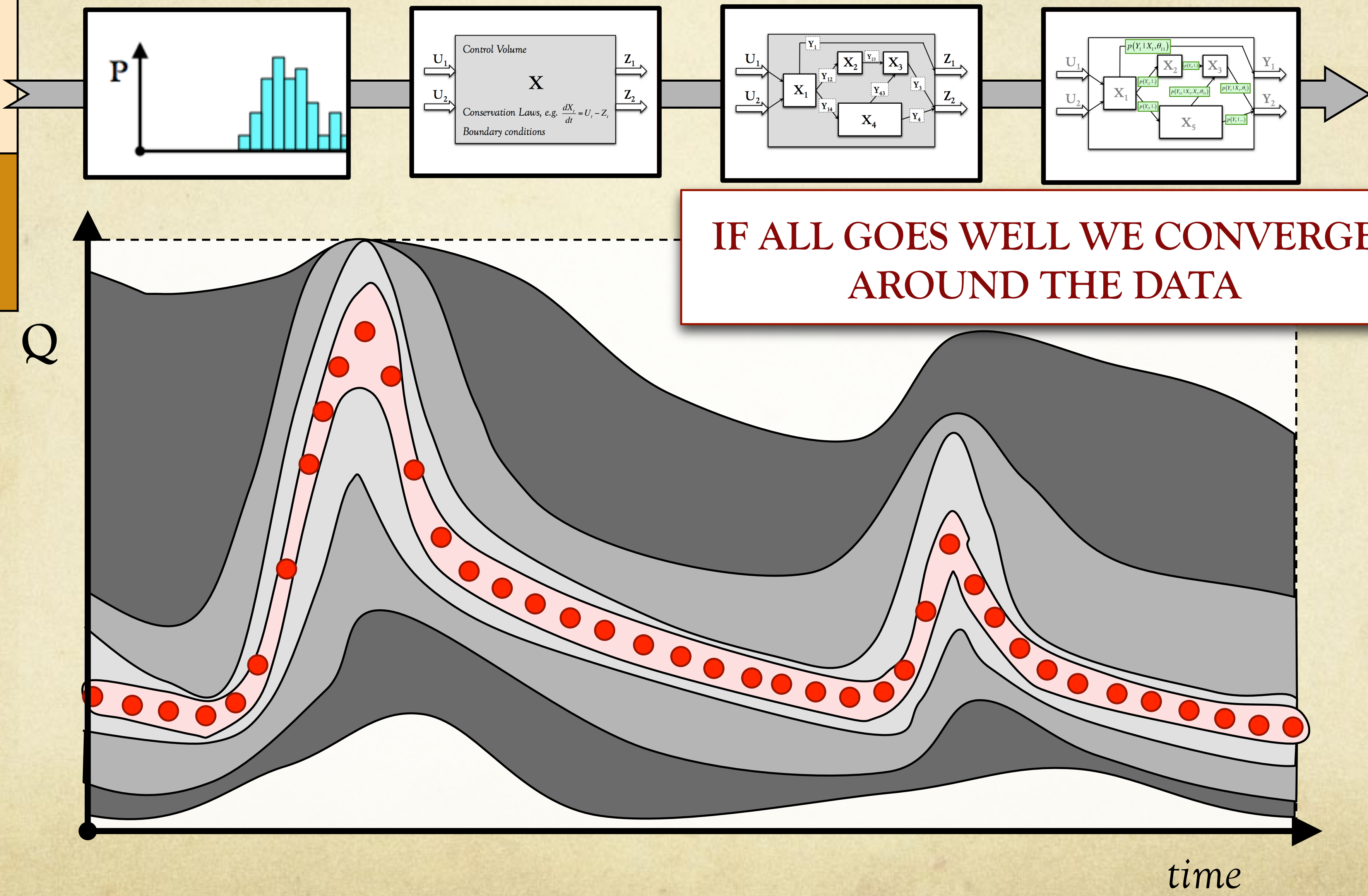
   Due to Lack of Knowledge

© Hoshin Gupta, The University of Arizona
cc BY

1
2
3

39

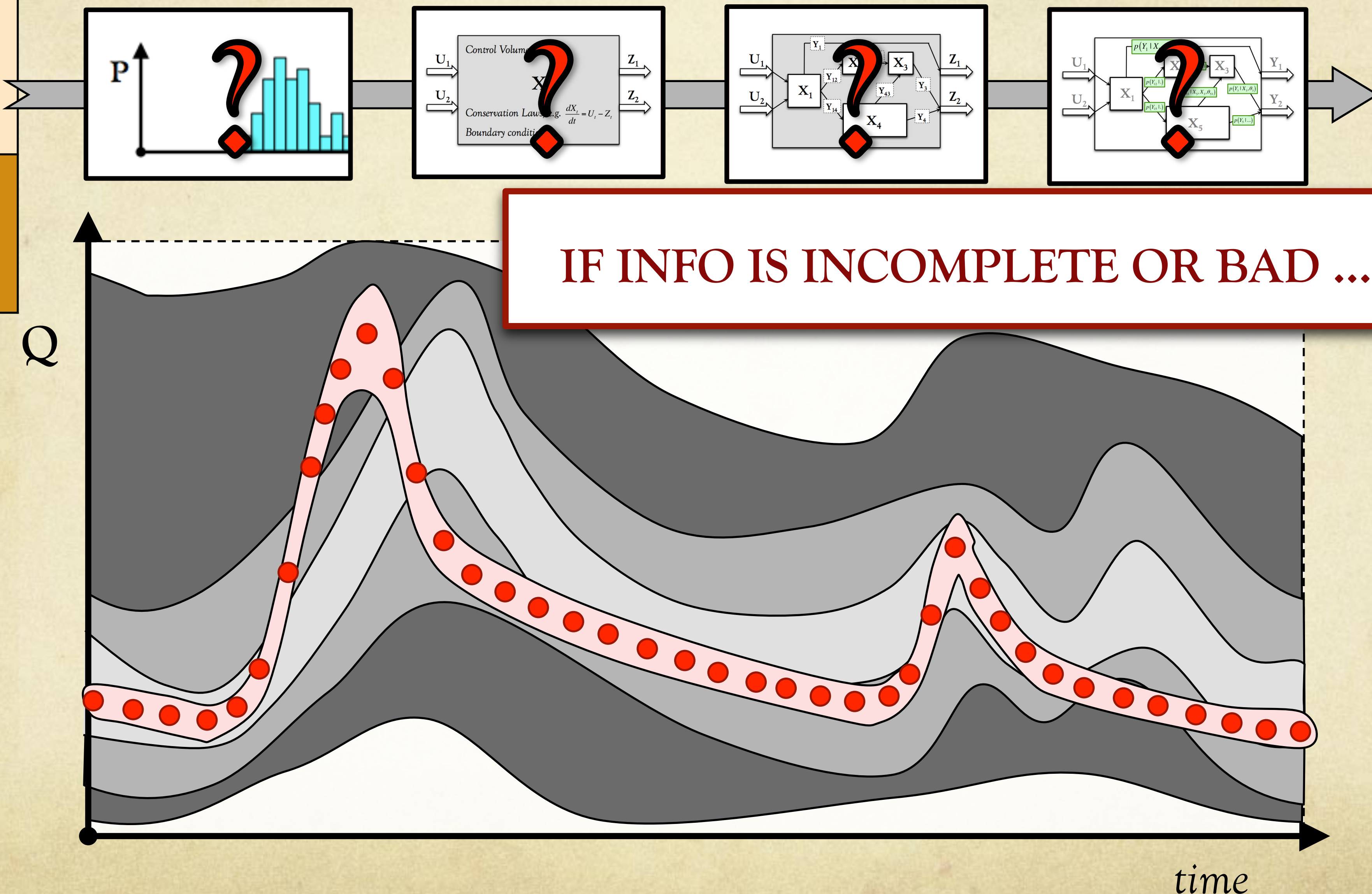
Conceptual Illustration



IF ALL GOES WELL WE CONVERGE AROUND THE DATA

© Hoshin Gupta, The University of Arizona

Conceptual Illustration

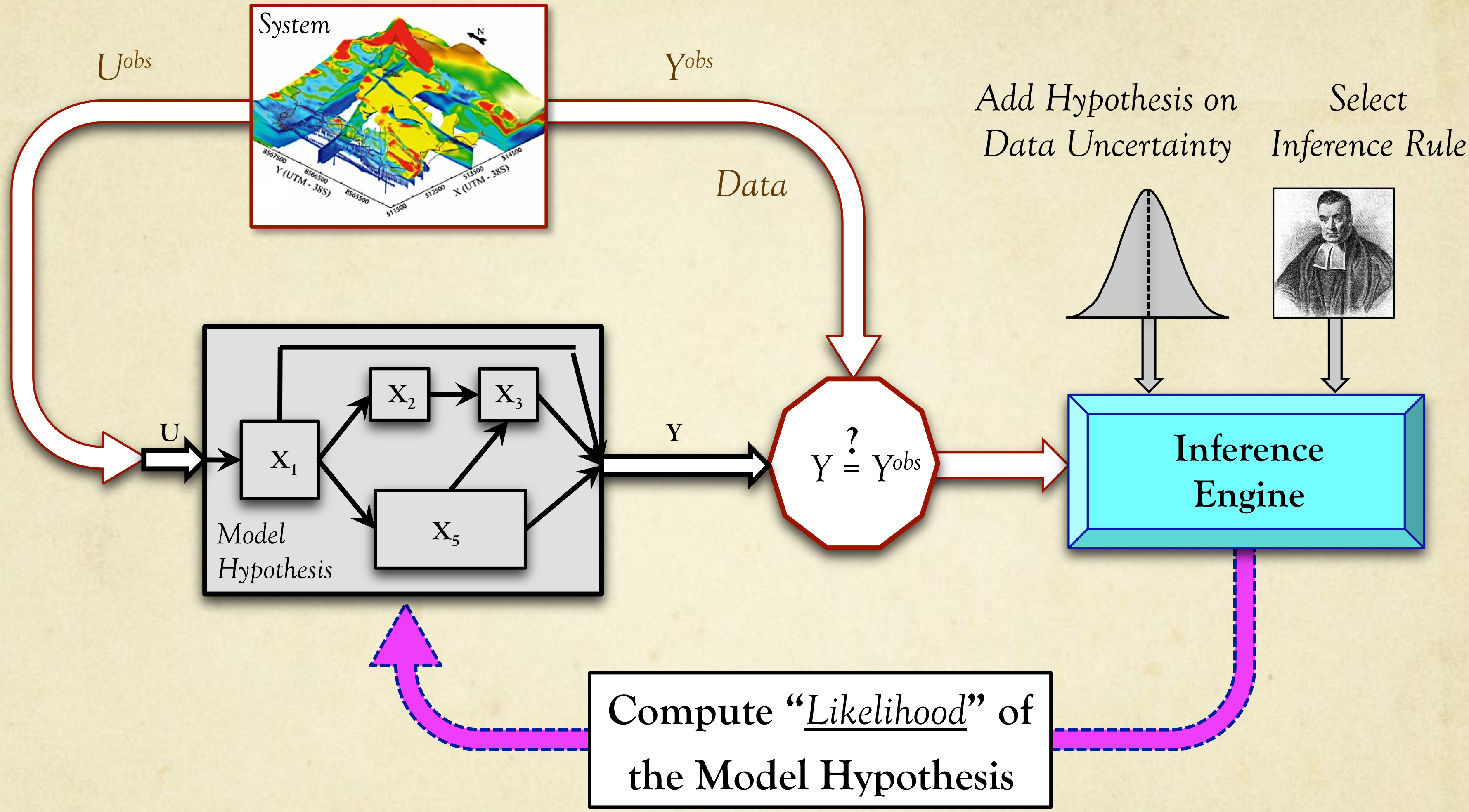


IF INFO IS INCOMPLETE OR BAD ...

© Hoshin Gupta, The University of Arizona

1
2
3
4

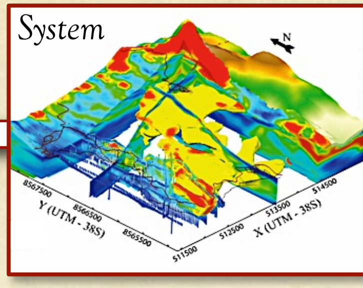
We Try to "Fix" the Model Hypothesis Via Inference



© Hoshin Gupta, The University of Arizona

Parameter / State Estimation

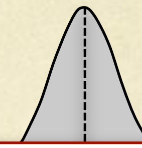
U^{obs}



Y^{obs}

Data

Add Hypothesis on Data Uncertainty



Select Inference Rule

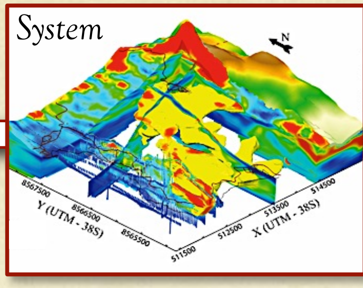
Search Parameter and/or State Space

$$\textbf{Maximize}\ L\left(X, \theta \mid \underbrace{H^{UN}, H^{PP}, H^{SA}, H^{CL}, \textbf{Data}}_{\textbf{Assumed Correct}}\right)$$

© Hoshin Gupta, The University of Arizona

Process Parameterization Estimation

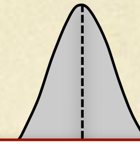
U^{obs}



Y^{obs}

Data

Add Hypothesis on Data Uncertainty



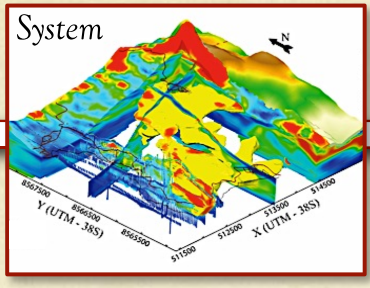
Select Inference Rule

Search Equation Forms

$$\textbf{Maximize} \;\; L\left(H^{PP} \,\middle|\, \underbrace{H^{UN}, H^{SA}, H^{CL}, \textbf{Data}}_{\textbf{Assumed Correct}}\right)$$

© Hoshin Gupta, The University of Arizona
cc BY

System Architecture Estimation

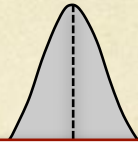


U^{obs}

Y^{obs}

Data

Add Hypothesis on Data Uncertainty



Select Inference Rule

Search System Architectures

$$\textbf{Maximize} \;\; L\left(H^{SA}, H^{PP} \,\middle|\, \underbrace{H^{UN}, H^{CL}, \textbf{Data}}_{\text{Assumed Correct}}\right)$$

© Hoshin Gupta, The University of Arizona

But … ?

U_1

Is the Problem in

Architecture **or**

Parameterization?

Z_1

"Architecture" & "Parameterization" are generally treated together

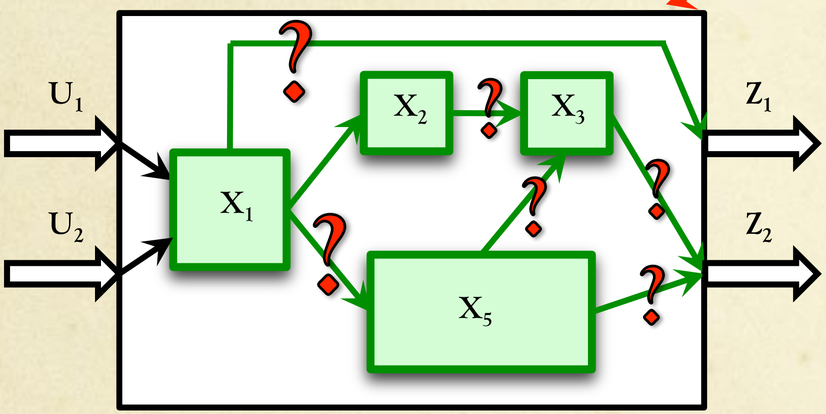
$$\textbf{Maximize } L\,(H^{SA}, H^{PP} \,|\, H^{UN}, H^{CL}, \textbf{Data})$$

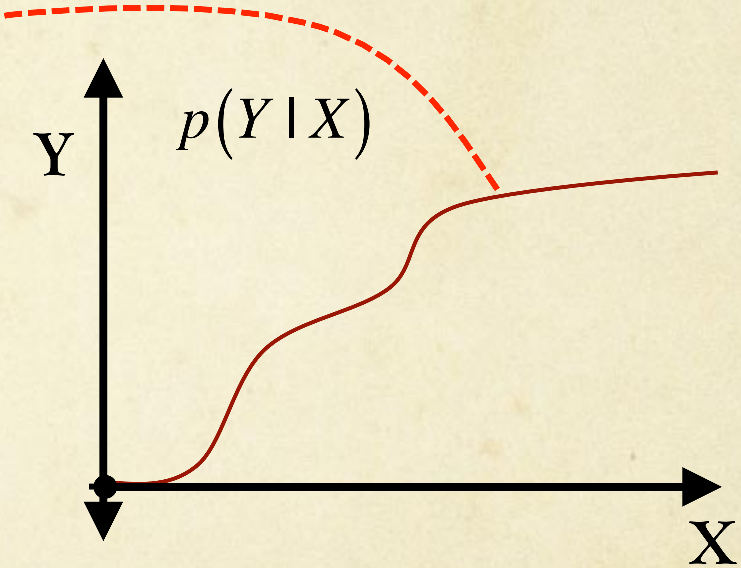
© Hoshin Gupta, The University of Arizona

How to Infer System Architecture ?

Use a *Maximum-Entropy Approach*

Process Parameterization Equations that impose only the Minimal Information required by Physics





© Hoshin Gupta, The University of Arizona

What is A Maximum-Entropy Parameterization ?

Typical Flux Parameterization →
$$Y_{ij} \stackrel{?}{=} f_{xy}\left(X_{ij} \mid \theta_{xy}\right)$$

Can be written as →
$$Y_{ij} = \frac{f_{xy}\left(X_{ij} \mid \theta_{xy}\right)}{X_{ij}} \cdot X_{ij}$$

Or →
$$Y_{ij} = K_{xy}\left(X_{ij} \mid \theta_{xy}\right) \cdot X_{ij}$$

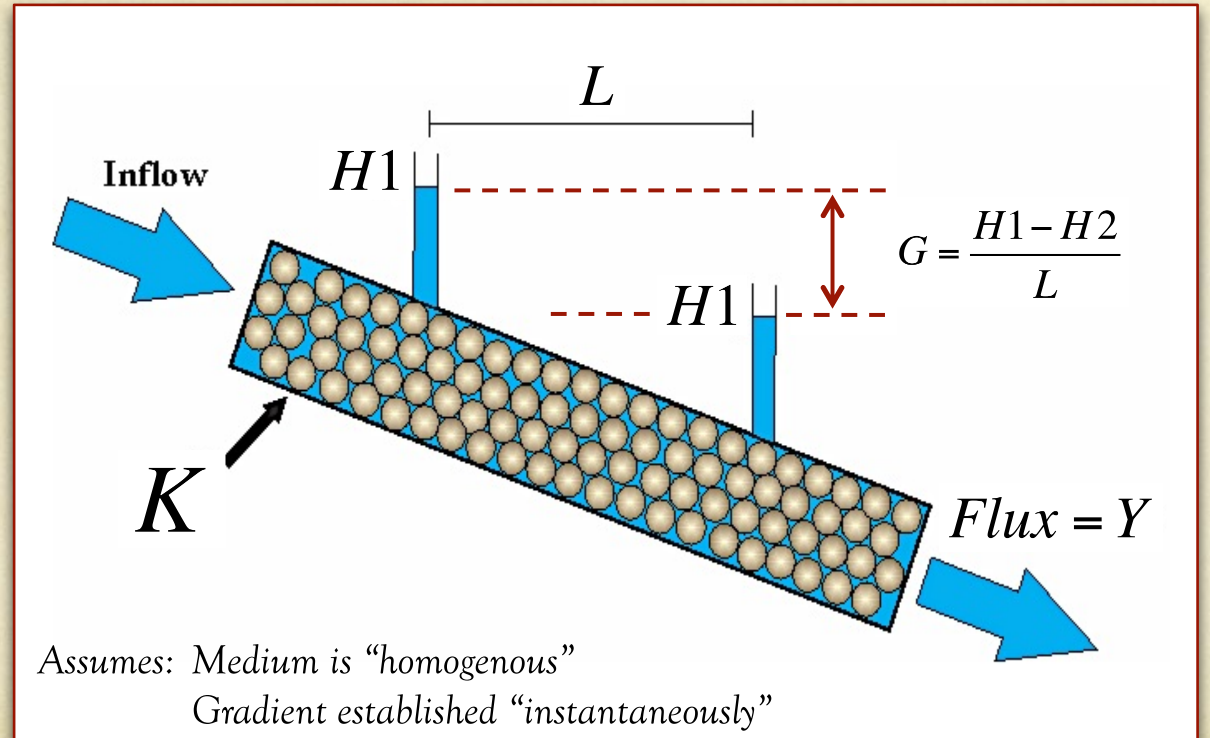
© Hoshin Gupta, The University of Arizona
CC BY

What is A Maximum-Entropy Parameterization ?

1. A Flux Parameterization must have the general form

$$Y_t = K_t^{xy} \cdot X_t$$

where X_t is the gradient to be dispersed
and K_t^{xy} is the conductivity of the medium



Basic Principle of Thermodynamics

© Hoshin Gupta, The University of Arizona

1
2
3
4
5

What is A Maximum-Entropy Parameterization ?

1. A Flux Parameterization must have the general form

$$\boxed{Y_t = K_t^{xy} \cdot X_t}$$

where X_t is the *gradient* to be dispersed and K_t^{xy} is the *conductivity* of the medium

2. Condition $0 \le Y_t \le X_t$ must hold to preserve *mass balance*, implying that

$$\boxed{0 \le K_t^{xy} \le 1}$$

© Hoshin Gupta, The University of Arizona

1
2
3
4
5

What is A Maximum-Entropy Parameterization ?

50

1. A Flux Parameterization must have the general form

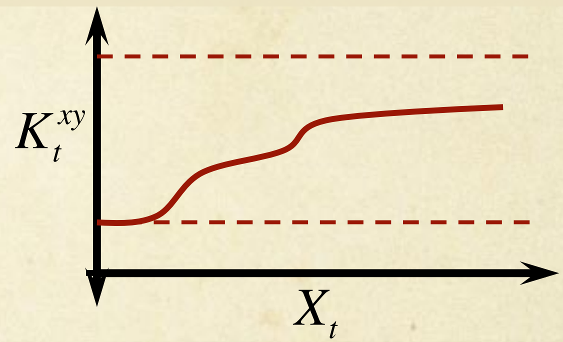
$$\boxed{Y_t = K_t^{xy} \cdot X_t}$$

where X_t is the *gradient* to be dispersed and K_t^{xy} is the *conductivity* of the medium

2. Condition $0 \le Y_t \le X_t$ must hold to preserve *mass balance*, implying that

$$\boxed{0 \le K_t^{xy} \le 1}$$

3. K_t^{xy} is a *monotonic non-decreasing or constant* function of X_t



Consistent with physical principle that

Larger gradients → Larger fluxes

© Hoshin Gupta, The University of Arizona

What is A Maximum-Entropy Parameterization ?

1. A Flux Parameterization must have the general form

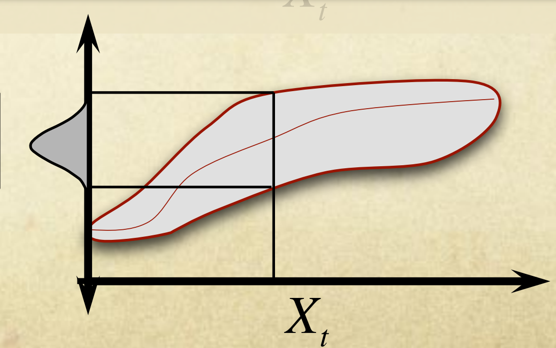
$$Y_t = K_t^{xy} \cdot X_t$$

where X_t is the *gradient* to be dispersed and K_t^{xy} is the *conductivity* of the medium

Because sub-element conditions will generally be different (in an unknown manner) each time the gradient ΔX is applied

4. K_t^{xy} is a Probabilistic function of X_t

$$K_t^{xy} \sim p\left(K_t^{xy} \mid X_t\right)$$



© Hoshin Gupta, The University of Arizona

What is A Maximum-Entropy Parameterization ?

1. A Flux Parameterization must have the general form

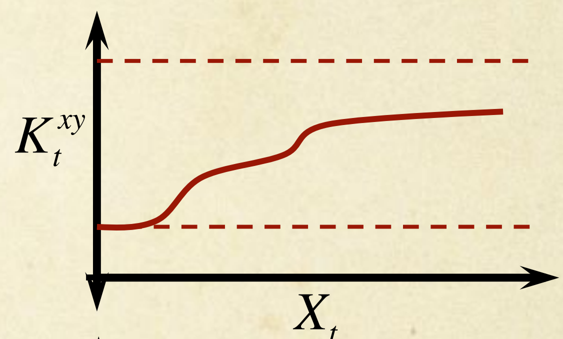
$$\boxed{Y_t = K_t^{xy} \cdot X_t}$$

where X_t is the *gradient* to be dispersed and K_t^{xy} is the *conductivity* of the medium

2. Condition $0 \le Y_t \le X_t$ must hold to preserve *mass balance*, implying that

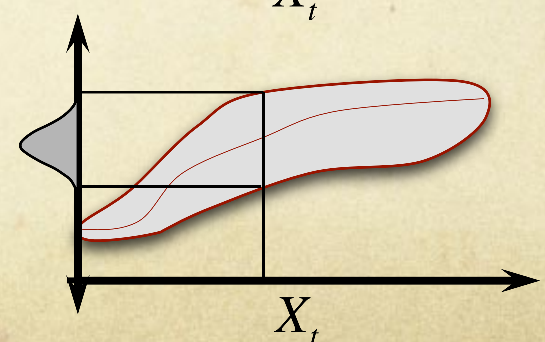
$$\boxed{0 \le K_t^{xy} \le 1}$$

3. K_t^{xy} is a *monotonic non-decreasing or constant* function of X_t



4. K_t^{xy} is a Probabilistic function of X_t

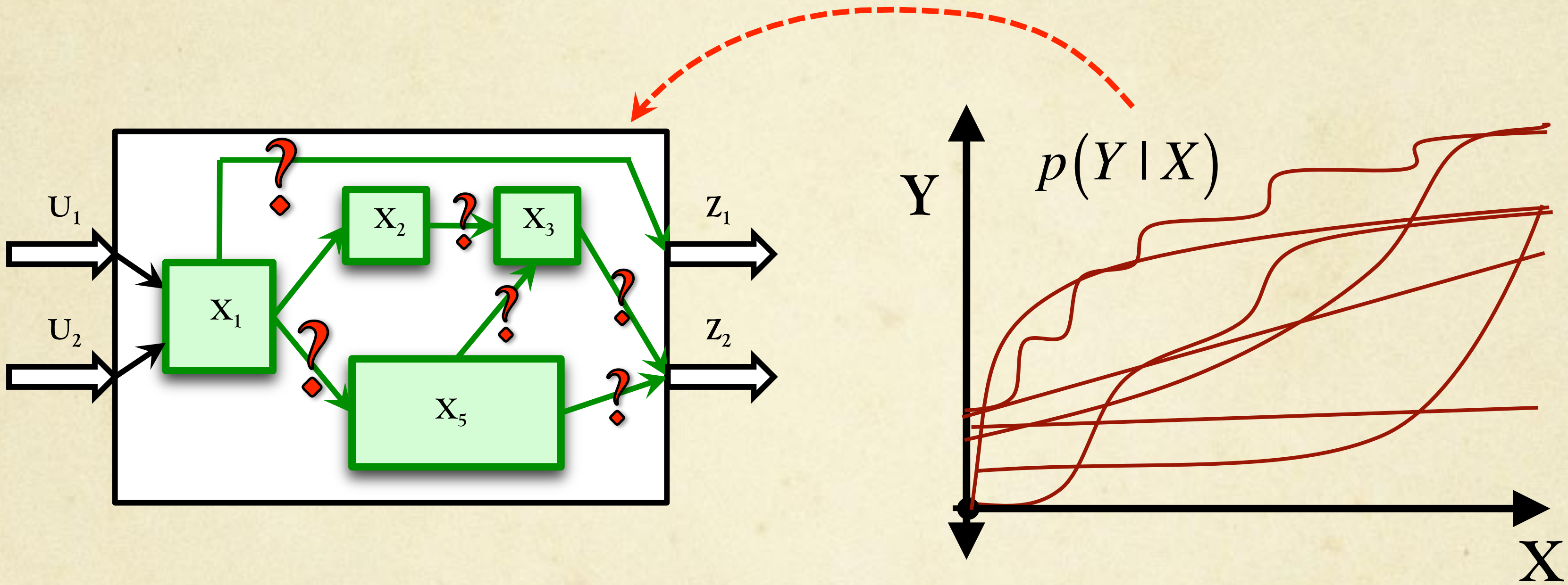
$$\boxed{K_t^{xy} \sim p\left(K_t^{xy} \mid X_t\right)}$$



© Hoshin Gupta, The University of Arizona

Maximum-Entropy Parameterization

Imposes only the minimal information required by Physics

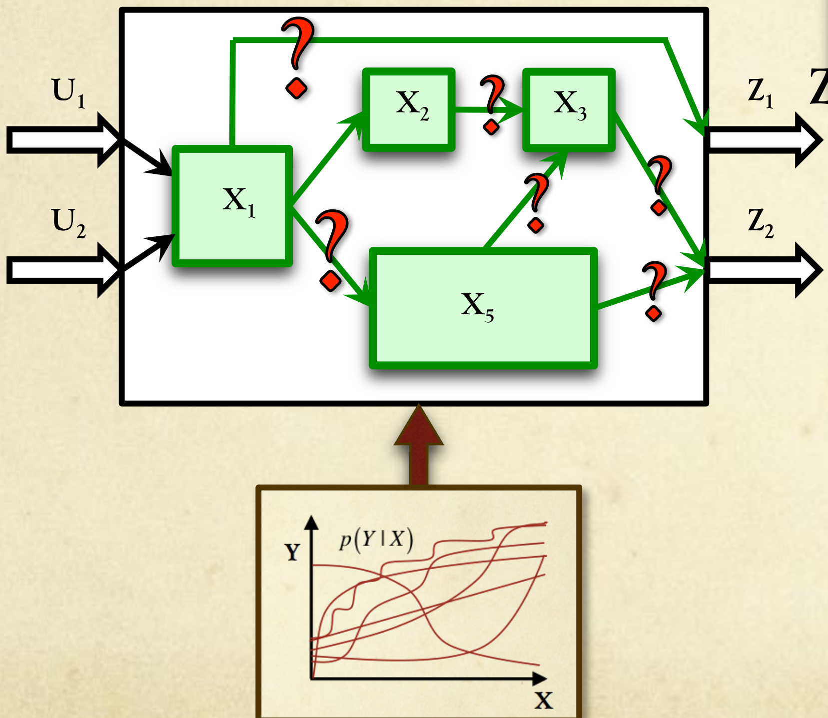


Randomly selected from a Maximum-Entropy Distribution of Random Functions

© Hoshin Gupta, The University of Arizona
cc BY

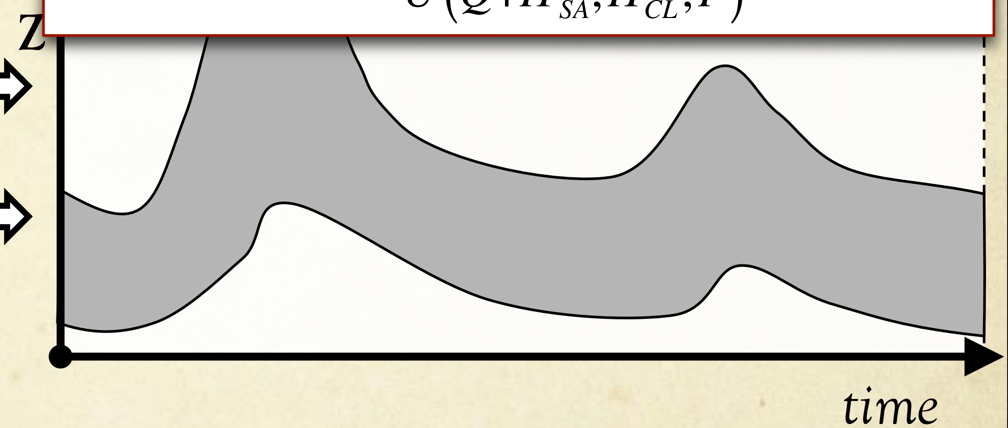
Inferring System Architecture

Using Process Parameterizations



Uncertainty after adding System Architecture

$U(Q|H_{SA},H_{CL},P)$

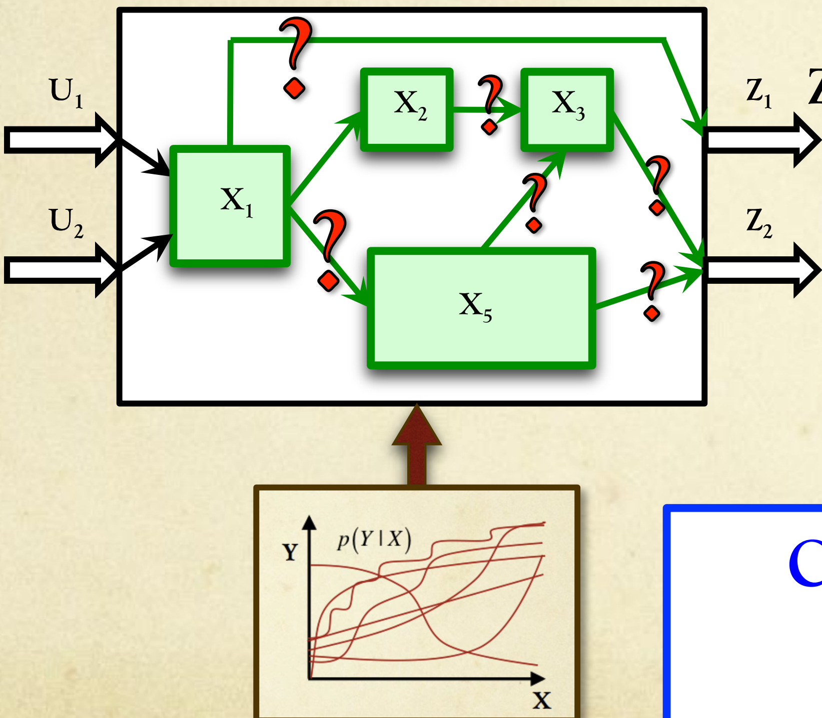


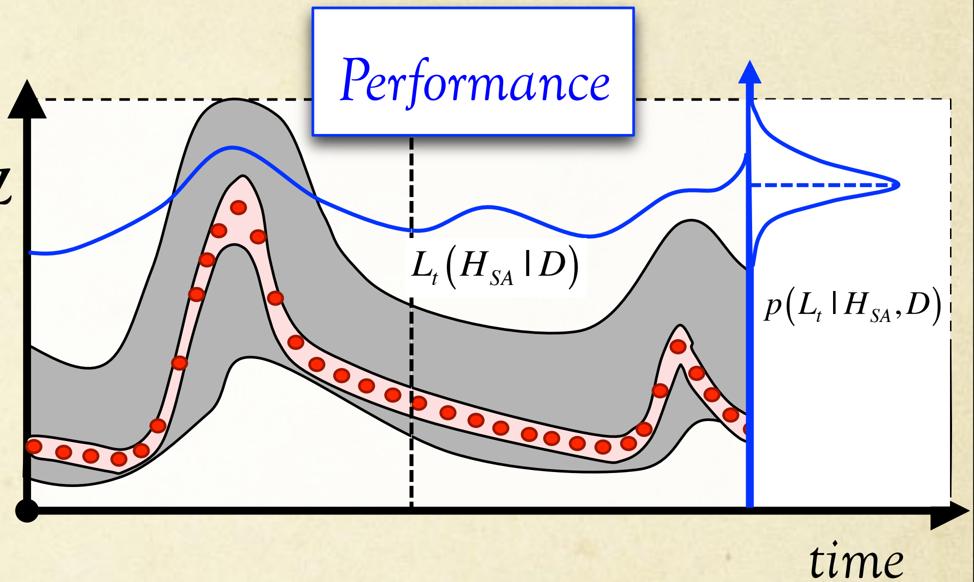
Run e.g., 10,000+ random cases (Monte Carlo on "Functions")

© Hoshin Gupta, The University of Arizona

Inferring System Architecture

Using Process Parameterizations





Run e.g., 10,000+ random

Compute the Likelihood of the Hypothesis given the Data

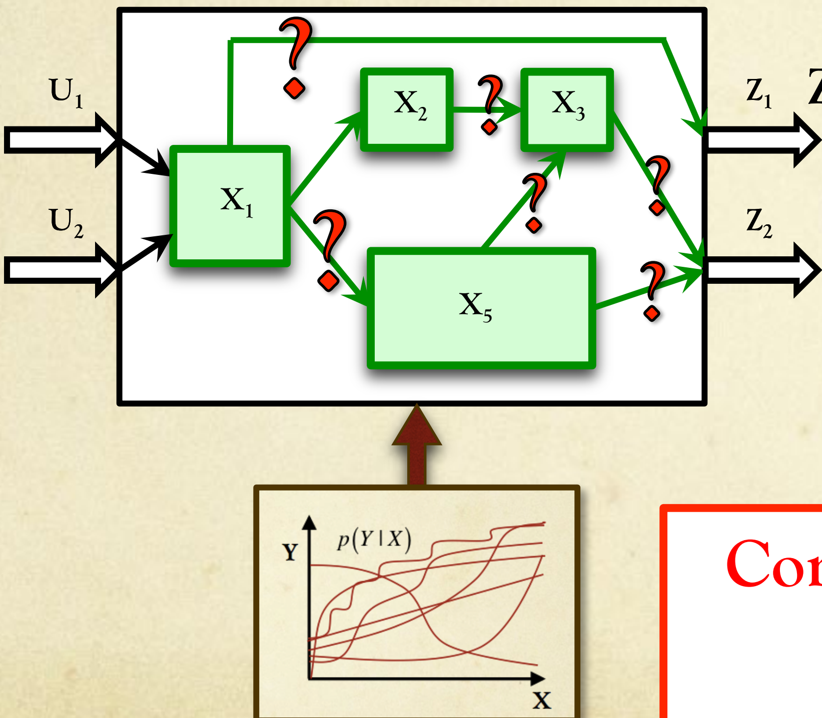
$$L(H_{SA} | D)$$

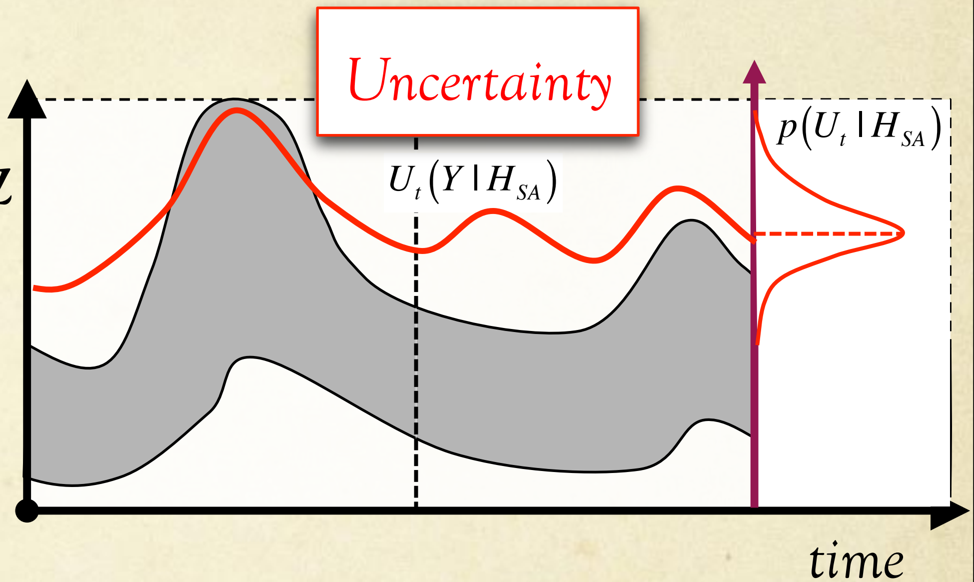
© Hoshin Gupta, The University of Arizona

1
2
3
4
5

Inferring System Architecture

Using Process Parameterizations





Run e.g., 10,000+ random

Compute the Entropy of the Model Ensemble Simulations

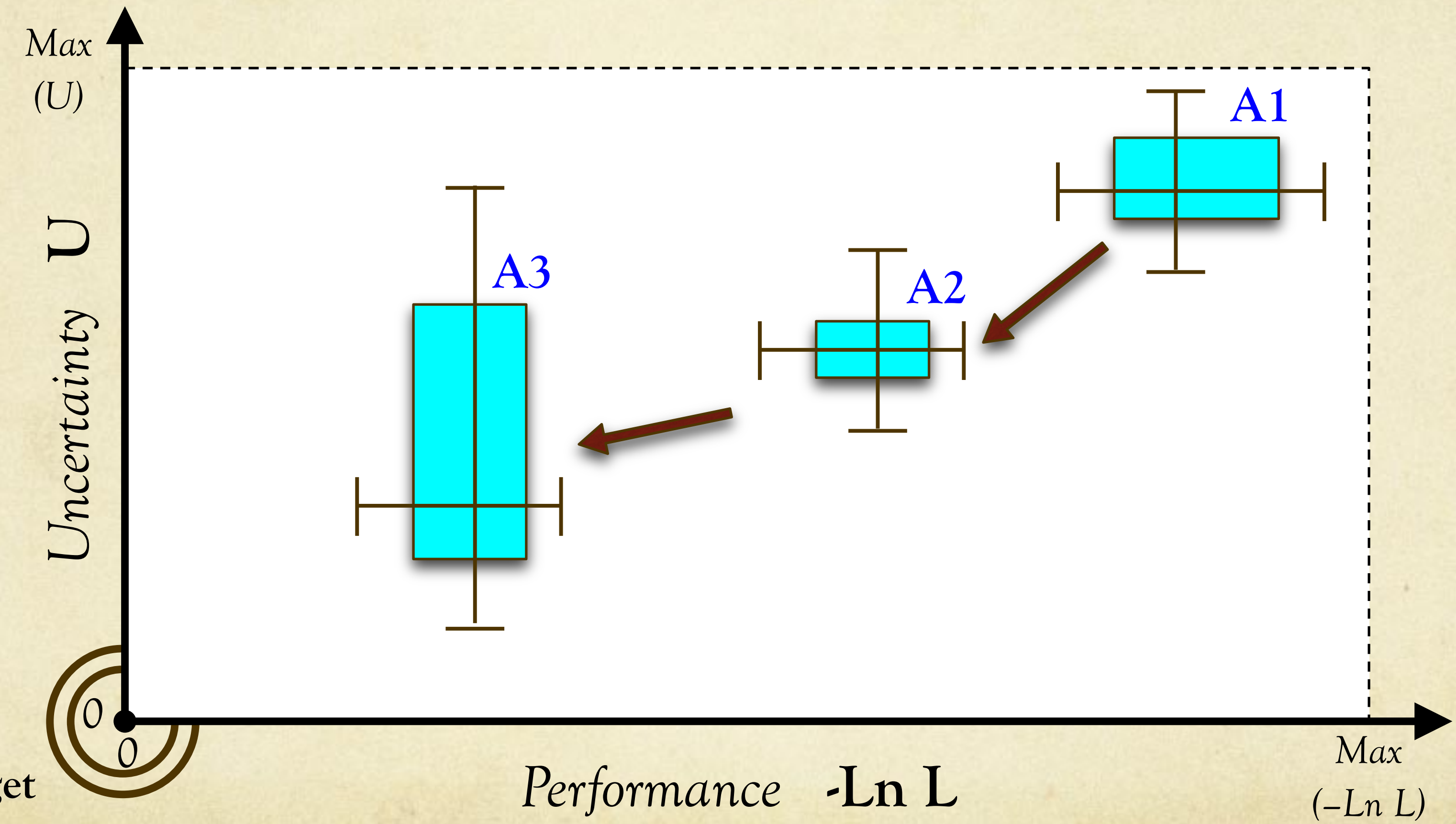
$$U(Y \mid H_{SA})$$

© Hoshin Gupta, The University of Arizona

Inferring System Architecture

57

Bootstrap & Plot the Results
For different System Architectures



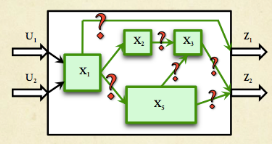
1
2
3
4
5

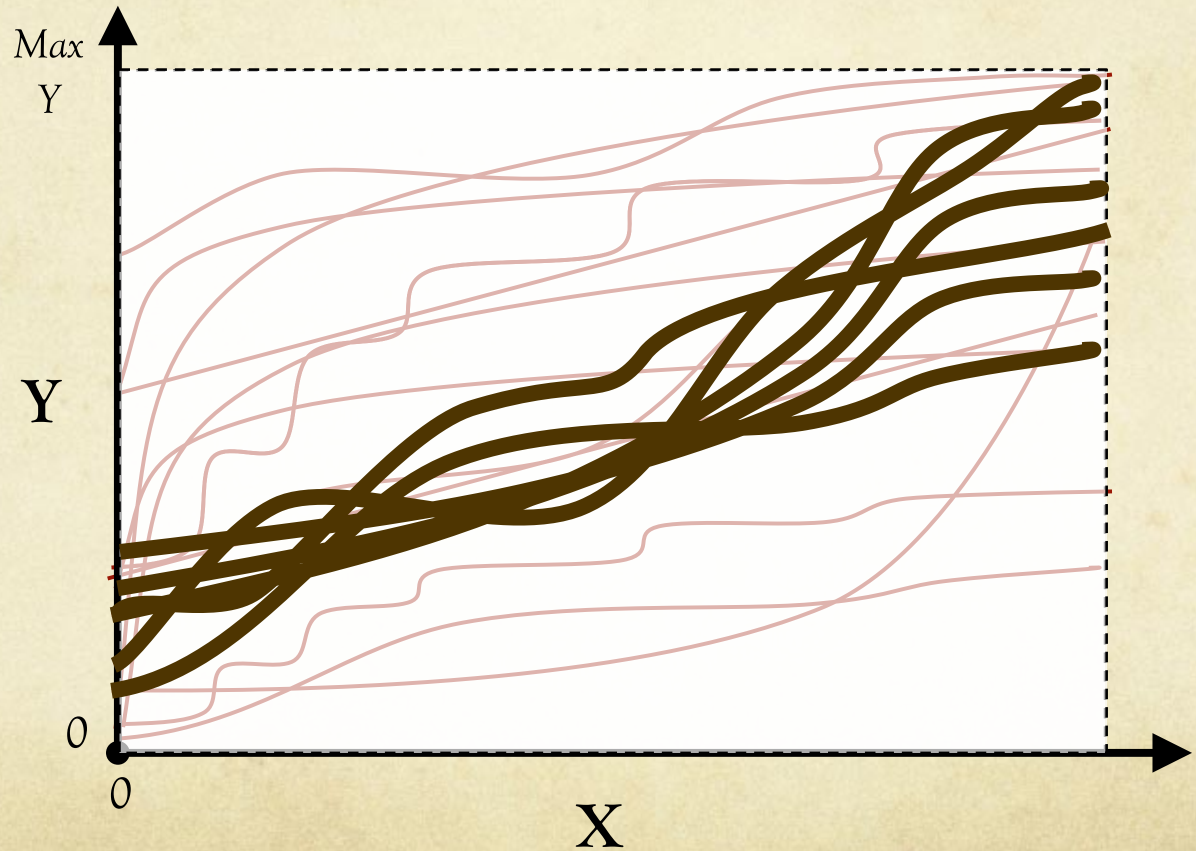
© Hoshin Gupta, The University of Arizona

Inferring Form of Process Parameterization

Use Bayesian Updating to Examine Parameterization Forms with Highest Performance

(conditional on a selected Architecture)



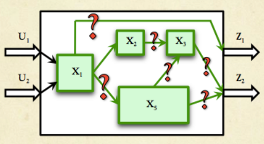


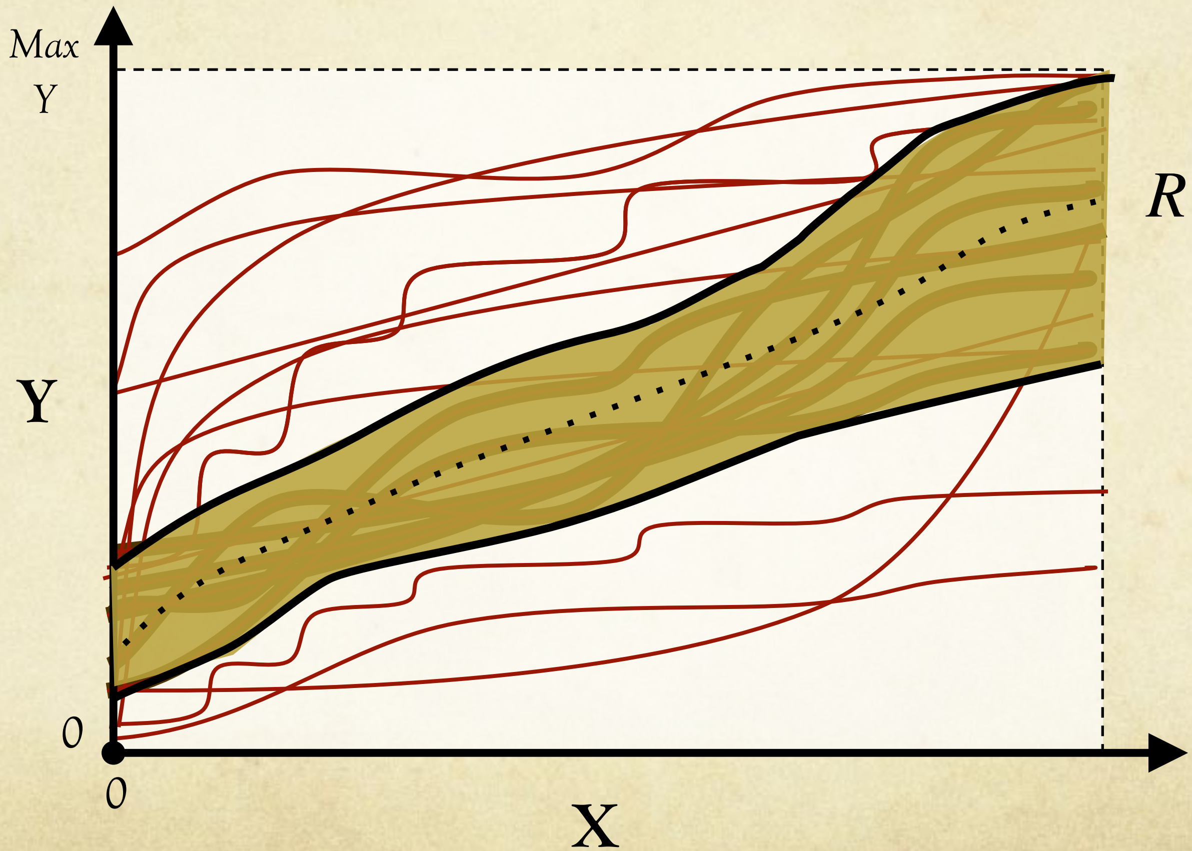
© Hoshin Gupta, The University of Arizona

Inferring Form of Process Parameterization

Propose Mathematical Forms for Parameterization Equations

(conditional on a selected Architecture)





$$R : p\left(Y \mid X, \theta\right)$$

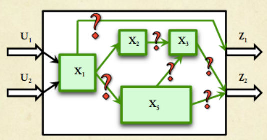
Mapping

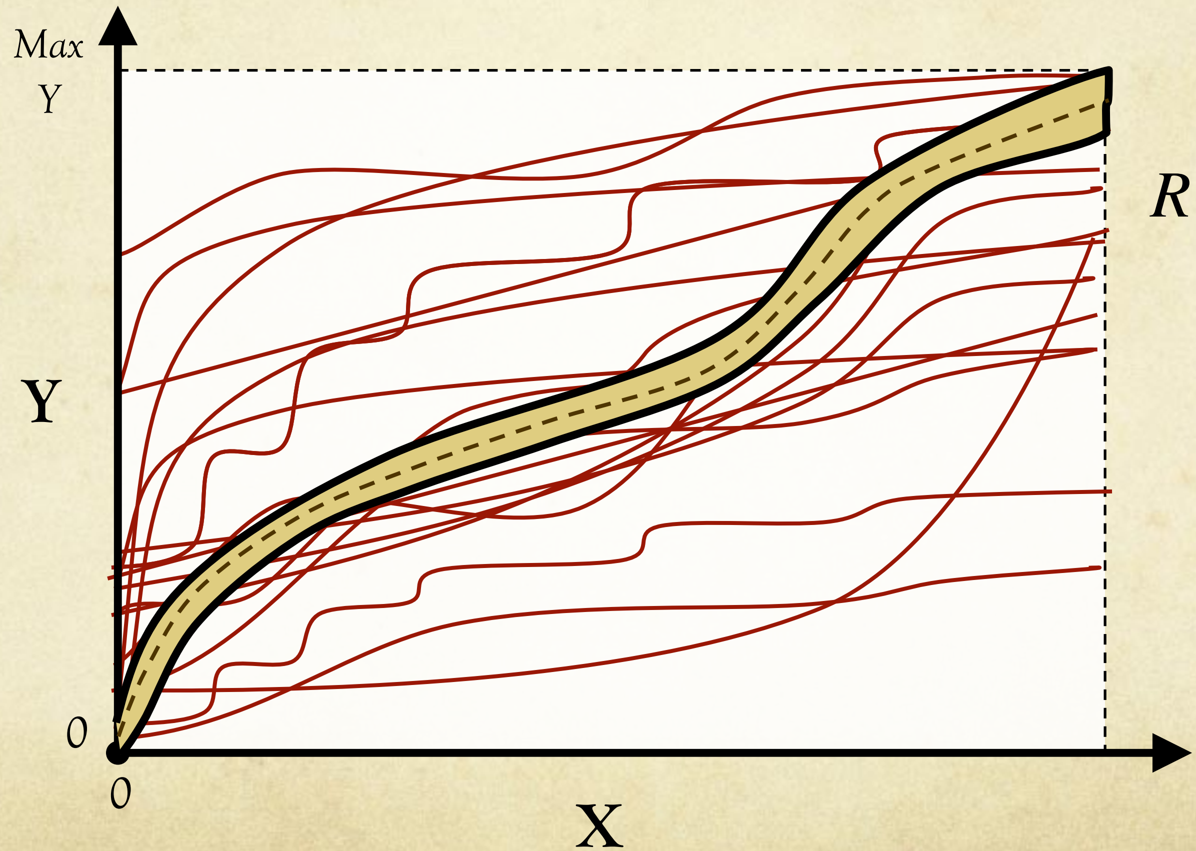
© Hoshin Gupta, The University of Arizona

Inferring Form of Process Parameterization

Proceed with Parameter Estimation

(conditional on a selected Architecture & Parameterization Form)





$R : p\left(Y \mid X, \theta\right)$

Mapping

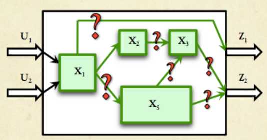
© Hoshin Gupta, The University of Arizona

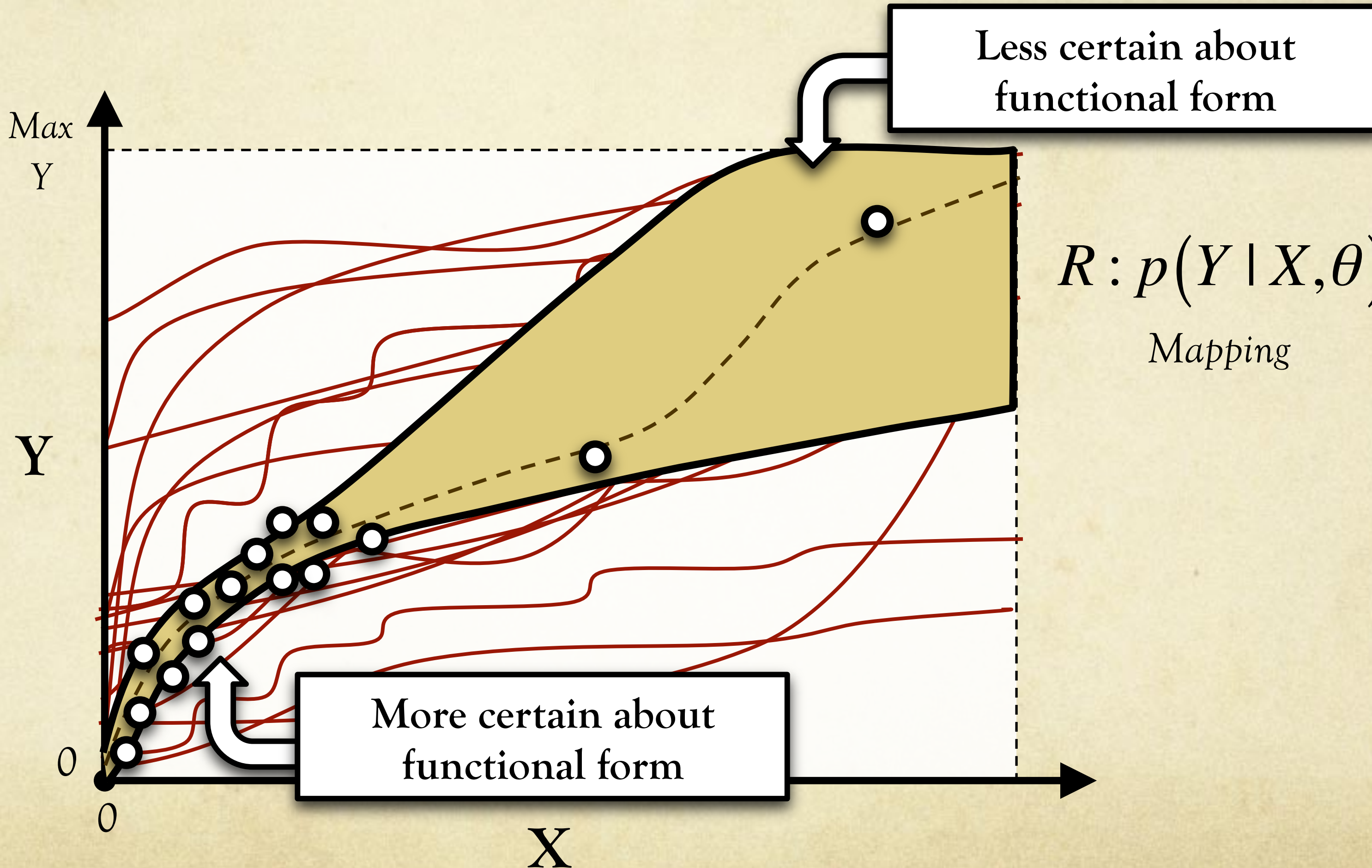
Inferring Form of Process Parameterization

1
2
3
4
5
6

Proceed with Parameter Estimation

(conditional on a selected Architecture & Parameterization Form)





$$R : p(Y \mid X, \theta)$$

Mapping

© Hoshin Gupta, The University of Arizona

1
2
3
4
5
6
7

The Result

Strategy to investigate System Architecture

Without the need to make Strong Assumptions
Regarding Process Parameterizations (Equations)

In Principle a similar approach could be
used to investigate value of different
Conservation Laws

© Hoshin Gupta, The University of Arizona
cc BY

1
2
3
4
5
6
7

More Generally

Bring more "Honesty/Rigour" into the Model Building Process

1) *Build Into the Model Clarity Regarding What We Feel Certain/Uncertain About*

2) *Be Clear about "What is Known" versus "What is Hypothesis / Assumption"*

"Maximum Entropy Approach" To Model Building

© Hoshin Gupta, The University of Arizona
cc BY

Some Comments in Conclusion

1. Models & Data codify *Information* about the world

2. Information implies *Change in Uncertainty* about Something

3. Models are *Hierarchical Assemblages of Hypotheses*
   1. *Conservation Laws*
   2. *System Architecture*
   3. *Process Parameterization*
   4. *Uncertainty*

4. Model Hypotheses can be:
   1. *Over-Constrained* by un-justifiably strong hypotheses
   2. *Under-Constrained* by lack of knowledge about
      a) Scale-dependence of process relationships
      b) Sub-element heterogeneity

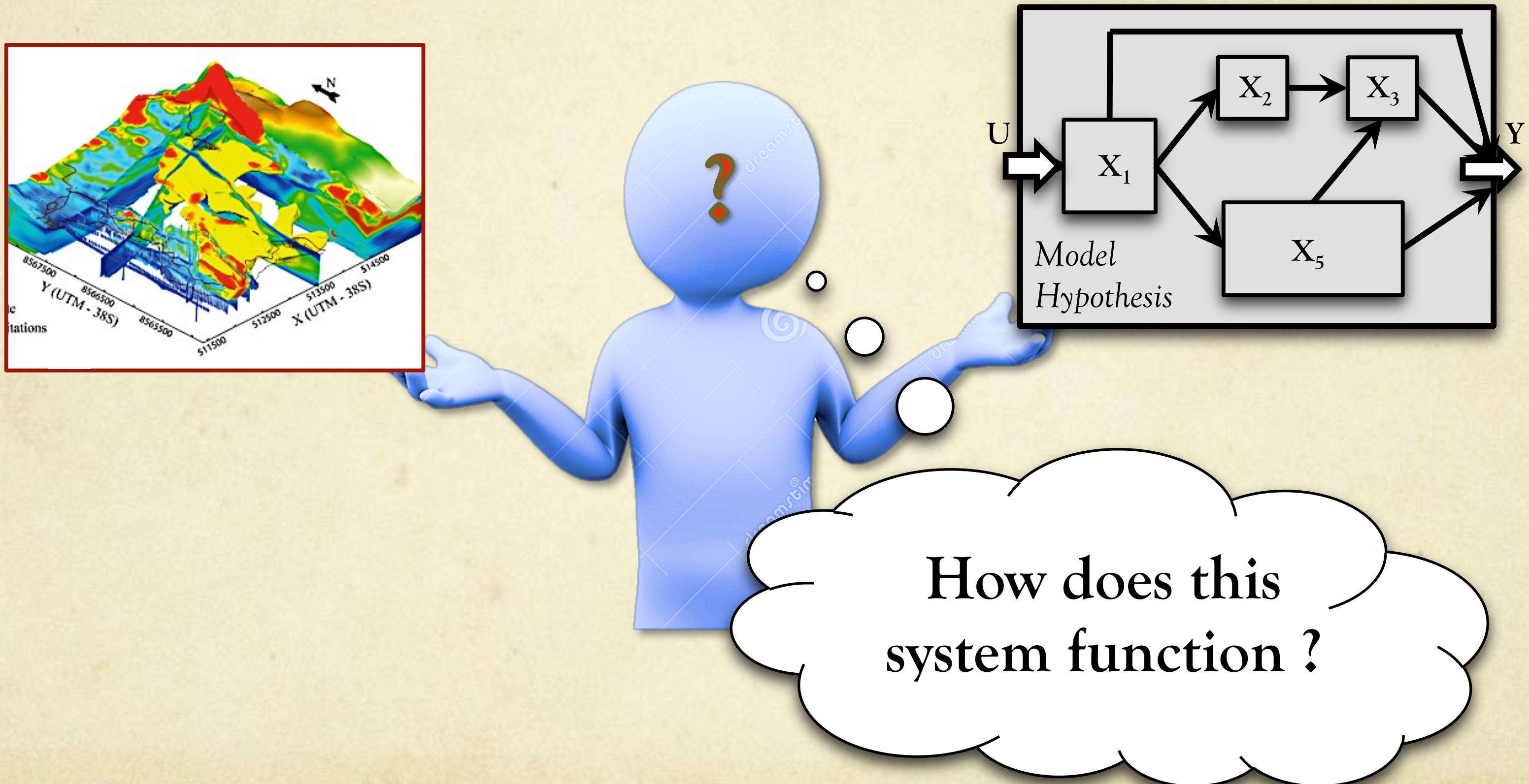
5. System Architecture Inference can be done using *Max-Entropy* PP's

6. Process Parameterization Inference can be done by Application of Bayes' Law

© Hoshin Gupta, The University of Arizona

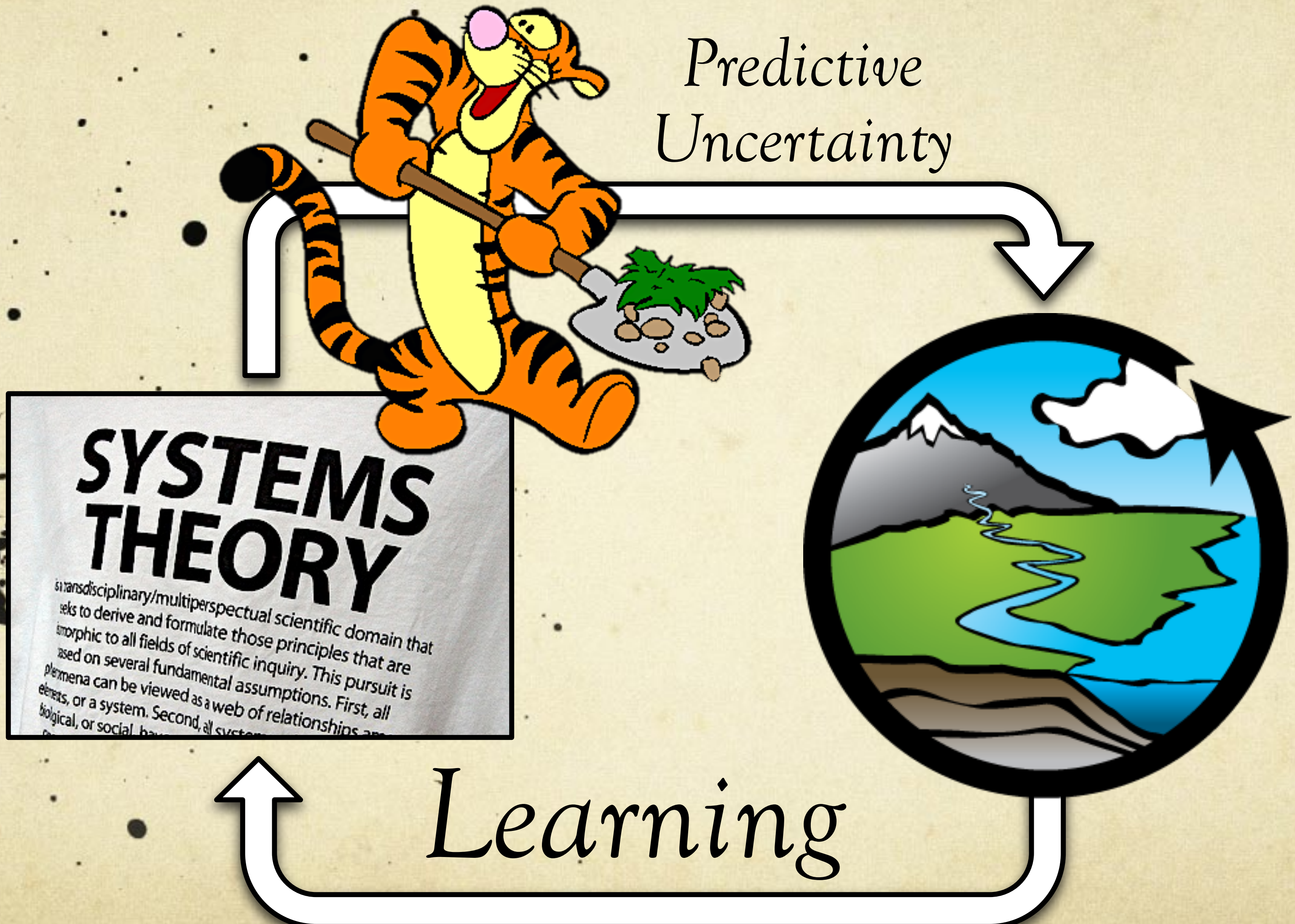
And Finally …

An ***Information Theory*** perspective can help improve the way we use models as Hypotheses for ***Scientific Investigation***



© Hoshin Gupta, The University of Arizona

Thank You



© Hoshin Gupta, The University of Arizona