Using Models and Data to Learn:

The Need for a Perspective based in Characterization of Information

Hoshin V Gupta (University of Arizona)

2014 Dalton Medal Lecture

Presented at Meeting of the European Geophysical Union, Vienna, Austria, April 29

Thank You

Guenter Blöschl, Alberto Montanari, Philippe Courtial

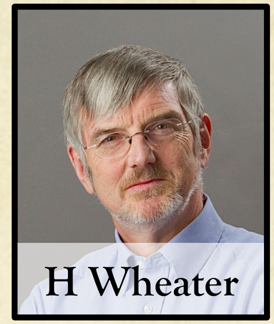
Members of the Medal Committees

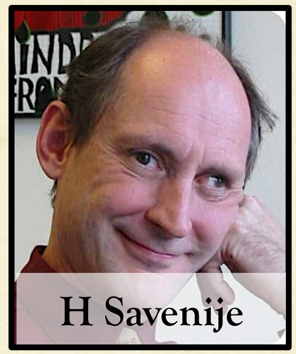
Officers of the Union

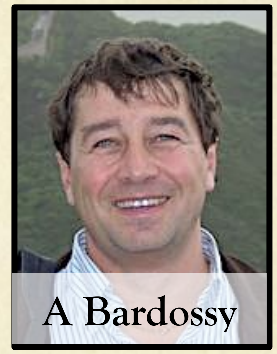
Thank You

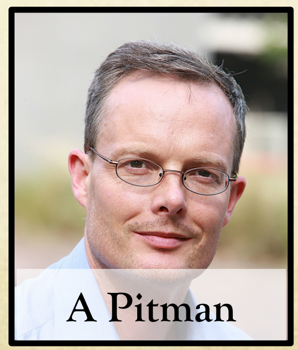

Thorsten Wagener


K Beven


H Wheater


H Savenije


A Bardossy


A Pitman

At Times Like This

Buddhist Saying → “*Do not look around*”

Don’t be Distracted by <u>other people’s</u> Opinions

&

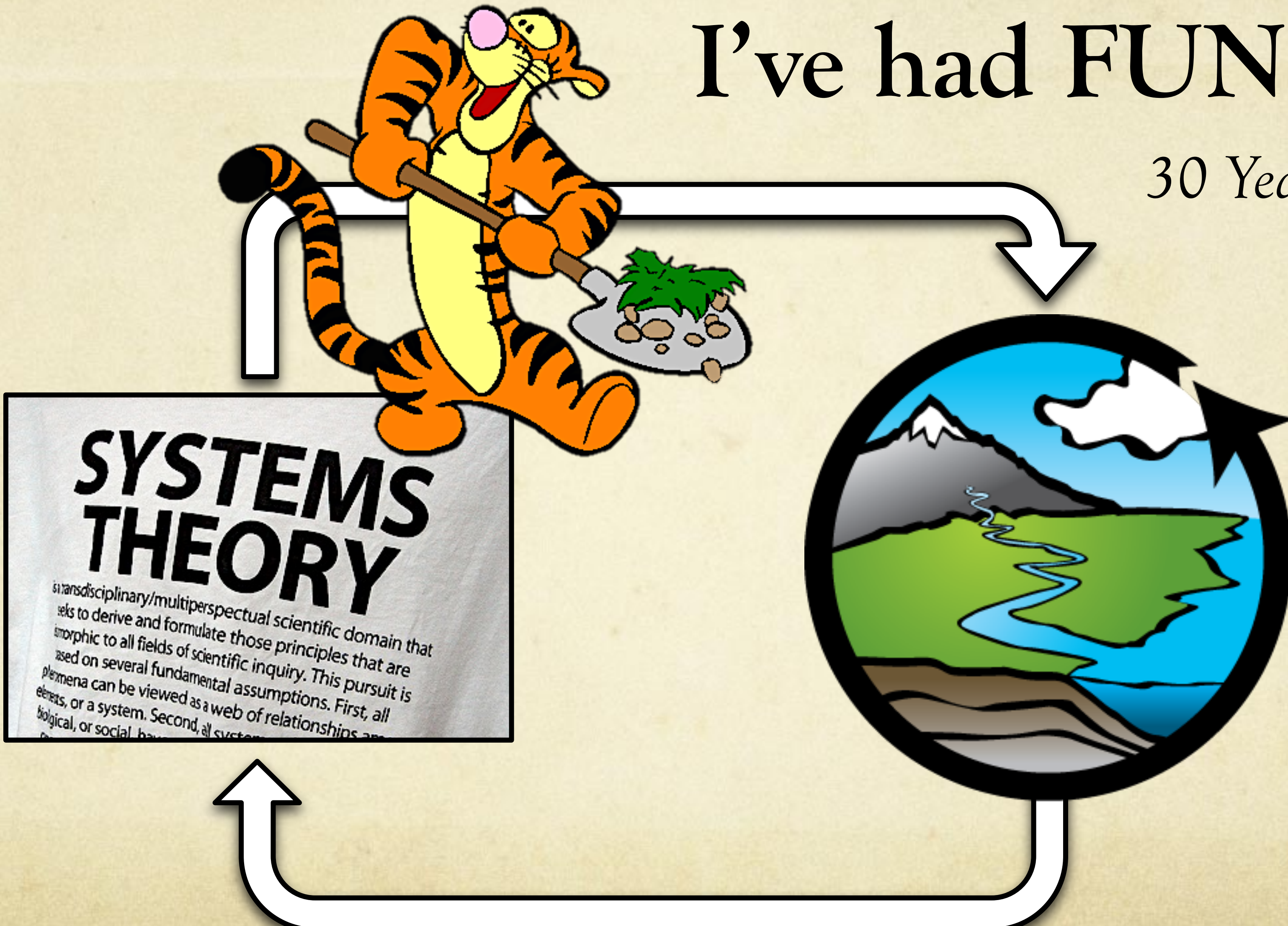
Do not be Distracted by <u>your own</u> Opinions

A Good Motto for Science

What Instead Seems Important - 1

I've had FUN ☺

30 Years



What Instead Seems Important - 2

I've had amazing Mentors

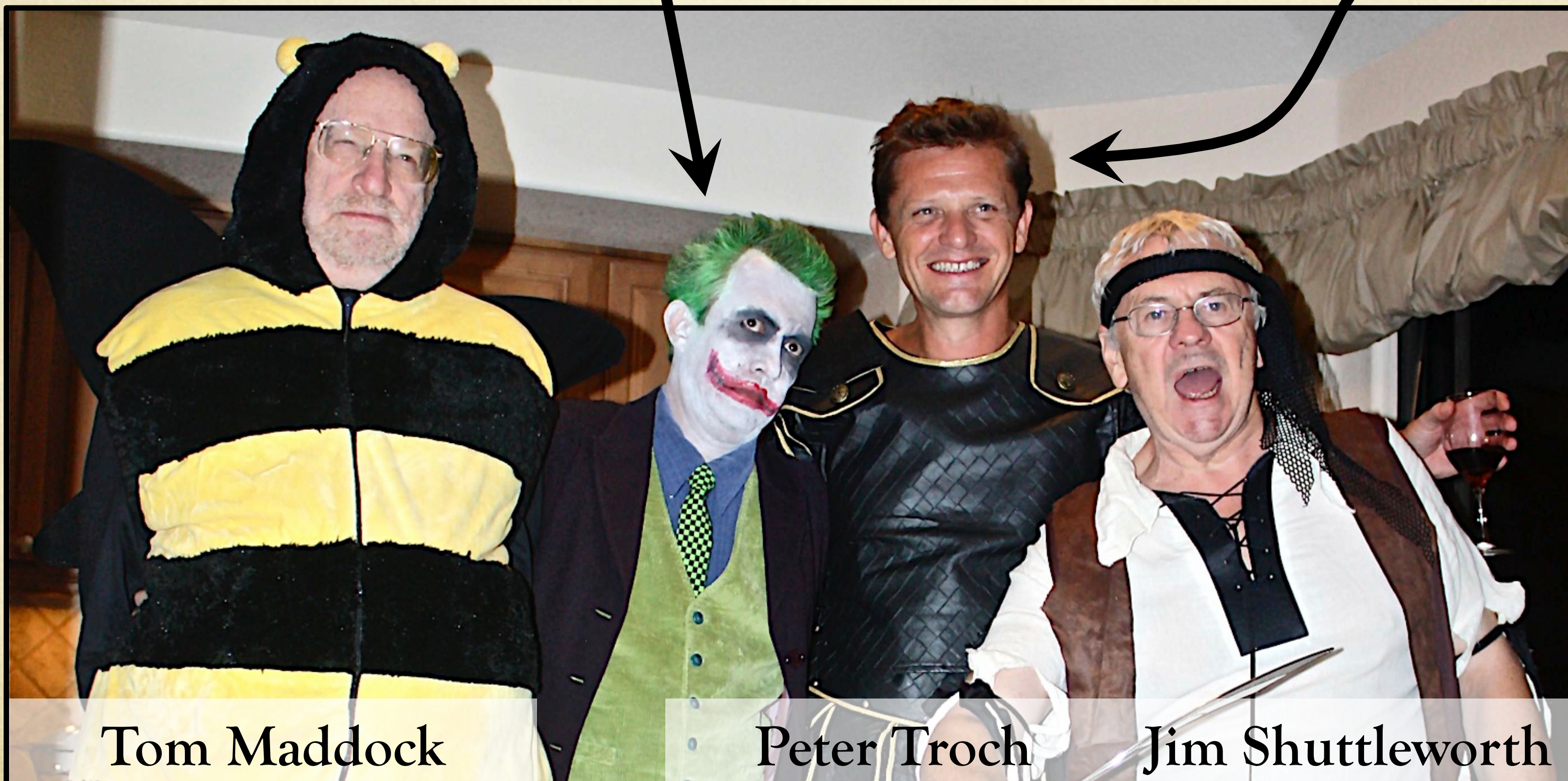
Soroosh & Shirin Sorooshian

What Instead Seems Important - 2

And Colleagues

Me

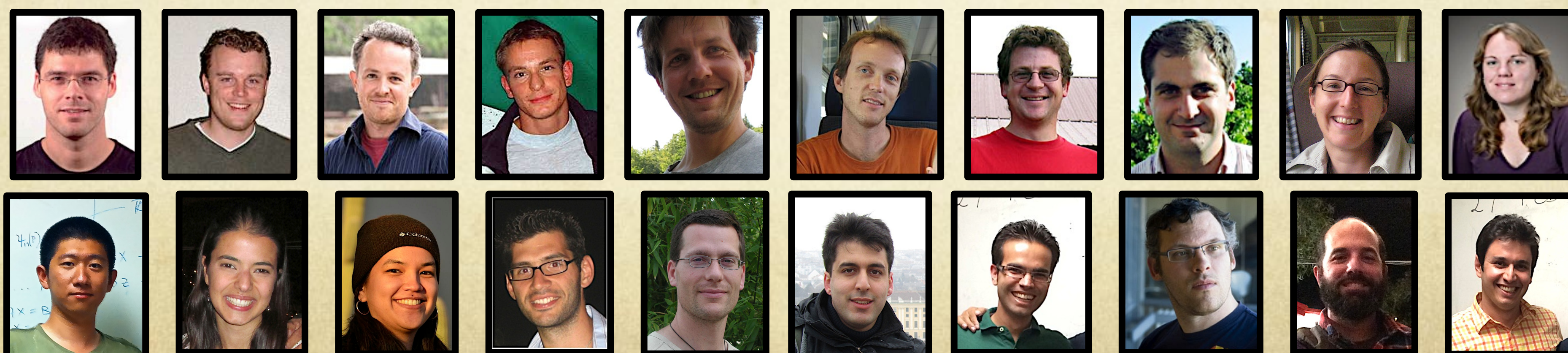
My “Dalton Brother”



What Instead Seems Important - 3

A *great many young & very smart collaborators*

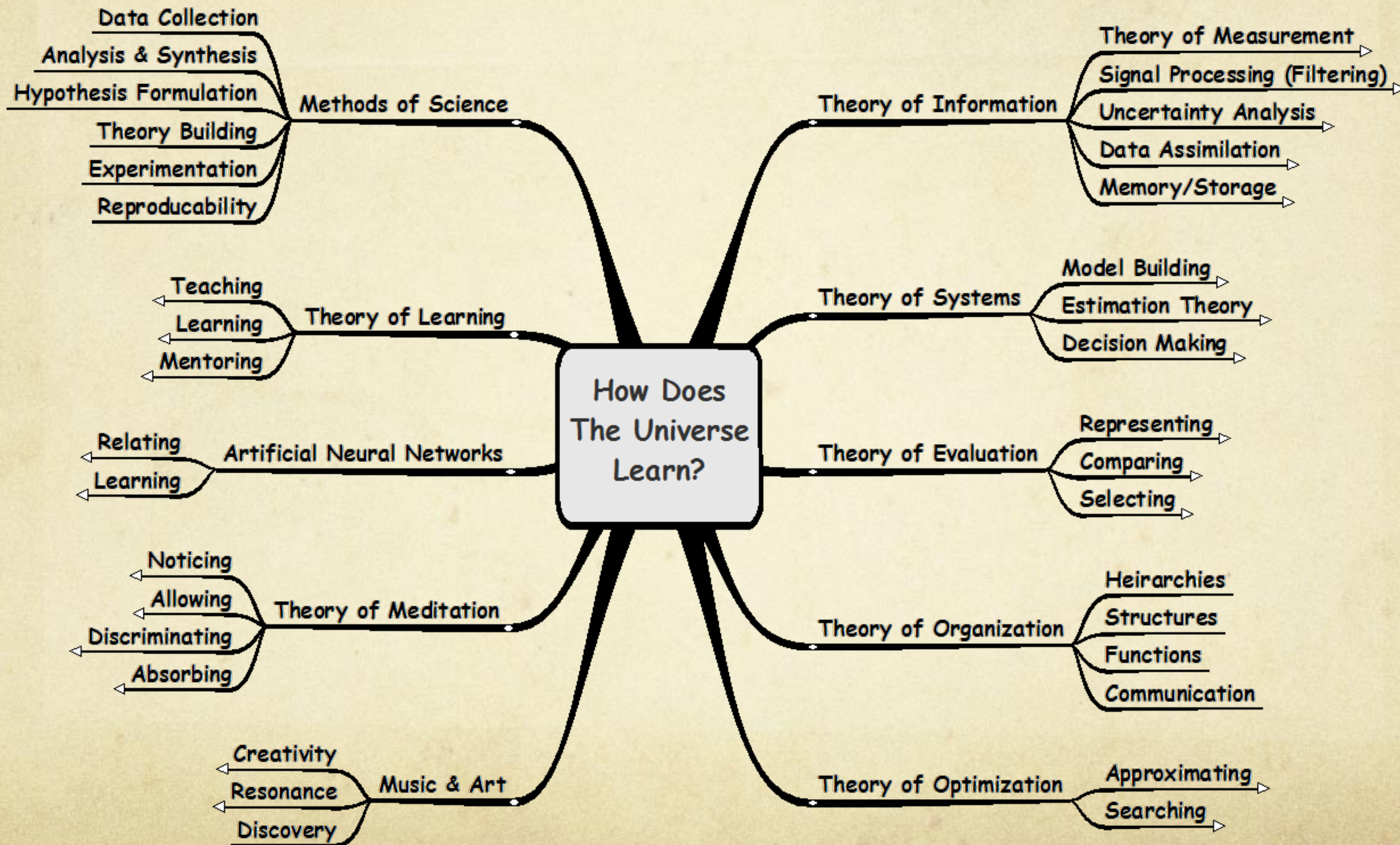
And very many visitors



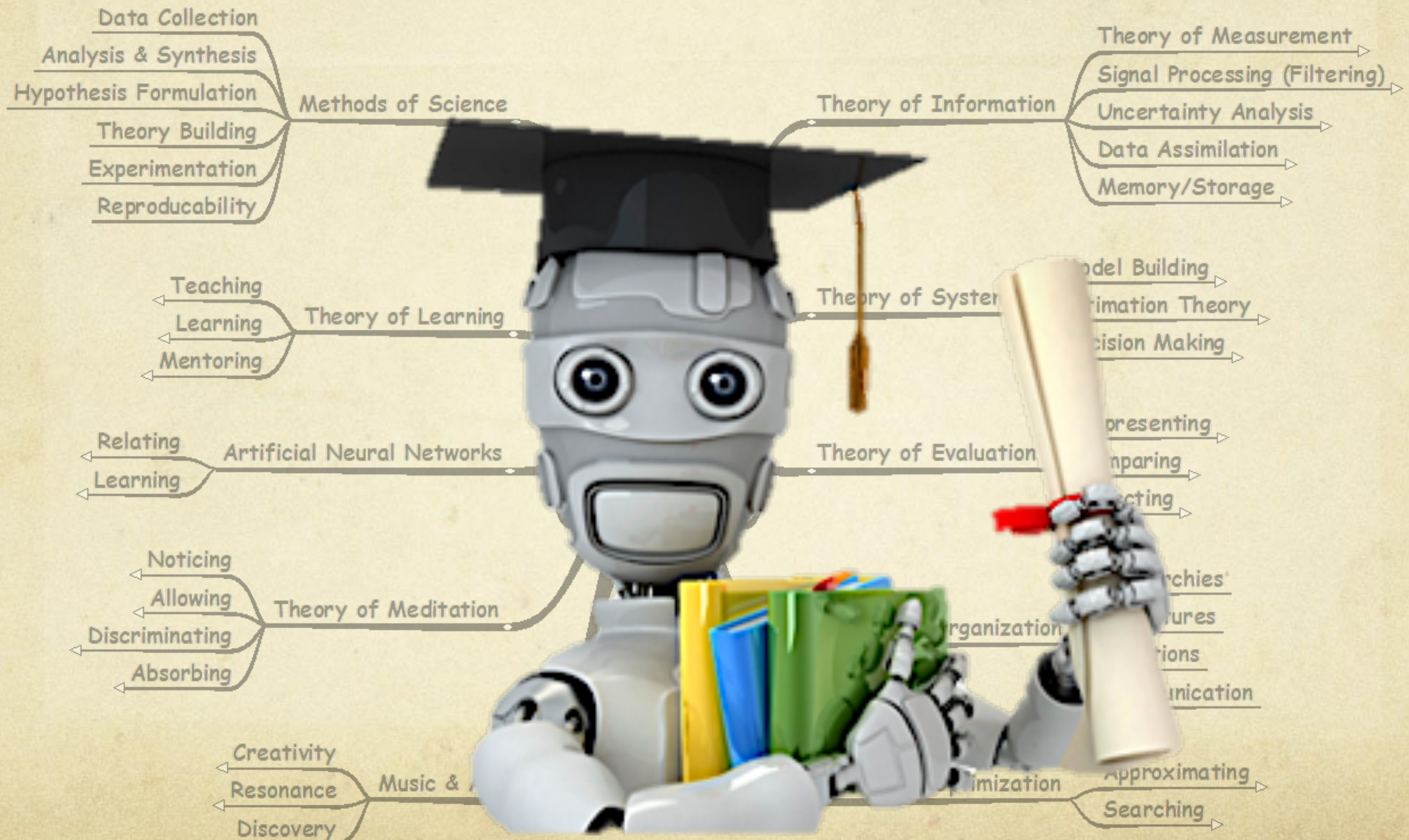
What Instead Seems Important - 3

A great many young & very smart collaborators

The Learning Problem



The Learning Problem



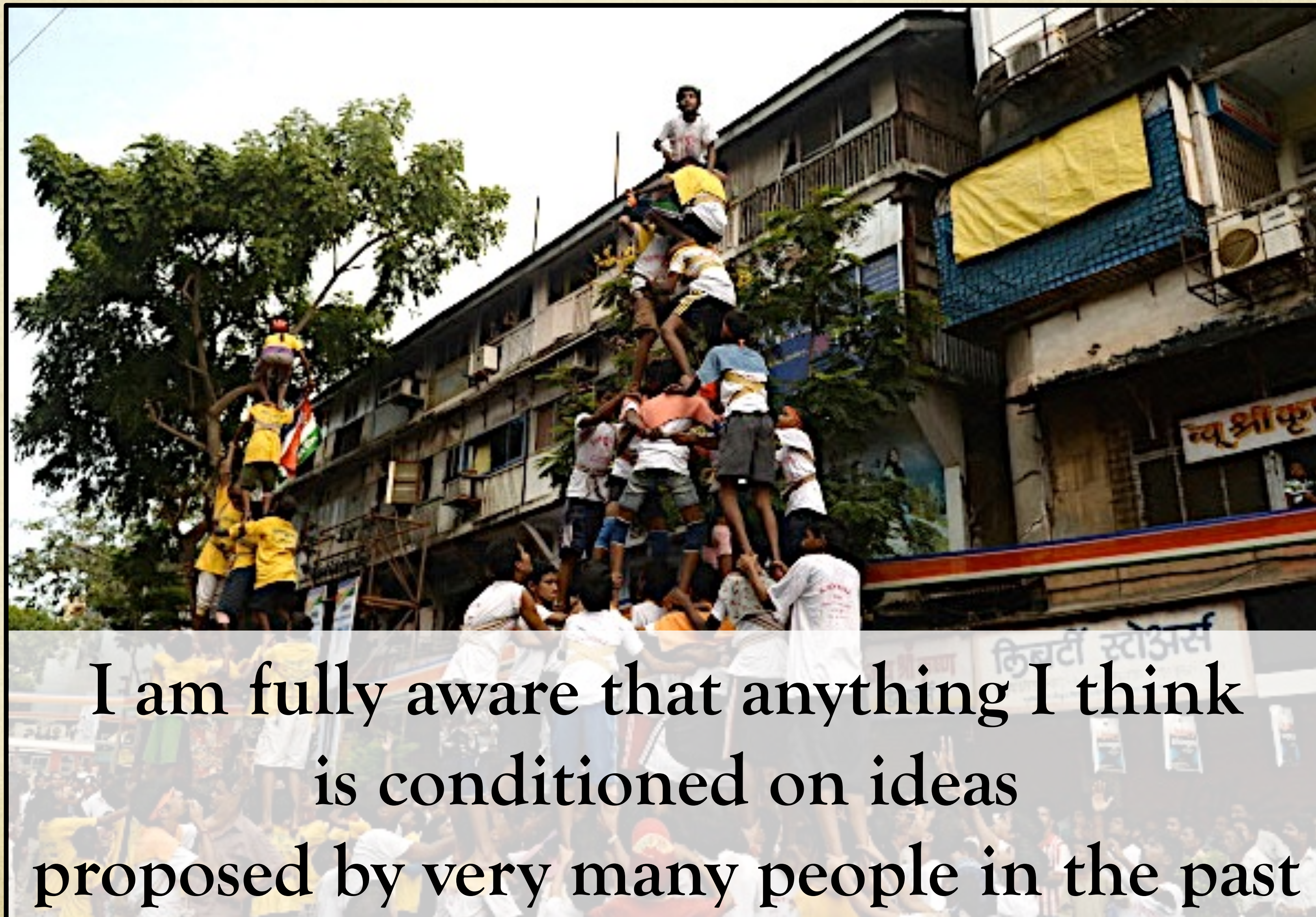
I Never Claim Originality

Original minds are not distinguished by being the first to see a new thing …

But instead by seeing the old, familiar thing that is over-looked as something new

Friedrich Nietzsche

I am a dedicated "Bayesian"



I am fully aware that anything I think
is conditioned on ideas
proposed by very many people in the past

Historical Influences

Richard Ibbitt (1970 Dissertation)

Tested 10 automatic search methods
on 10 catchment models

[Reported Poor Results]

Parameter Optimization for Watershed Models

P. R. JOHN

School of Civil Engineering, Universi

A detailed search for the optimum valu
Simplex and Davidon optimization metho
were readily achieved, but the solutions a
use of different optimization methods and
progress to be made in the search. Much
optimization. These include interdepende
surface and the occurrence of discontinui
types of stores, and the effects of using
analyzed and the only basic assumption in
apply to most watershed models.

Johnston & Pilgrim (1976 WRR)

"A true optimum set of parameters was not found in over 2 years of full time work concentrated on one watershed"

[Reported Failure]

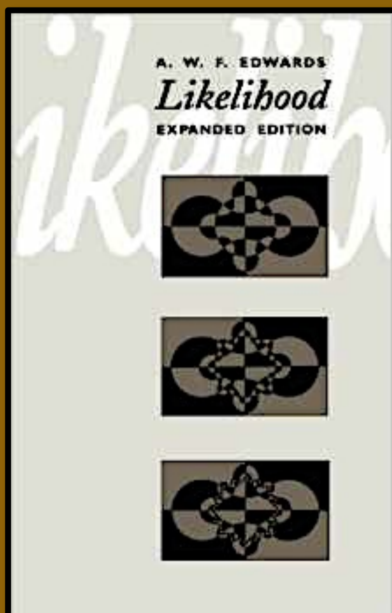
Historical Influences

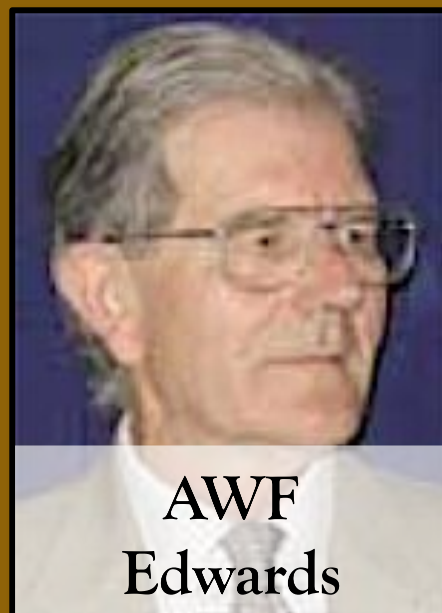
Soroosh Sorooshian(PhD 1978)

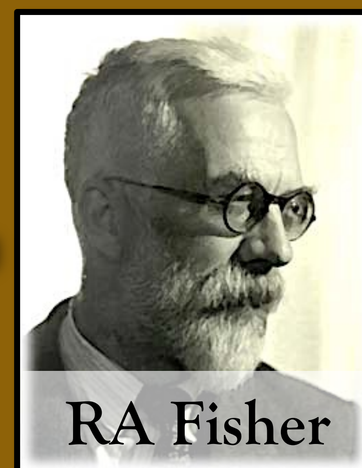
Application of Maximum Likelihood Theory

George Kuczera (PhD 1984)

Application of Bayesian Theory

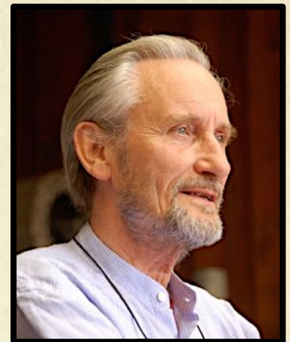






Historical Influences




John Schaake
Hydrologic Models
Large Samples

Yakov Haimes
Multi-Criteria
Risk

Mihalo Mesarovic
Systems Theory

Peter Young
Systems Methods

Ezio Todini
Predictive Uncertainty

Historical Influences

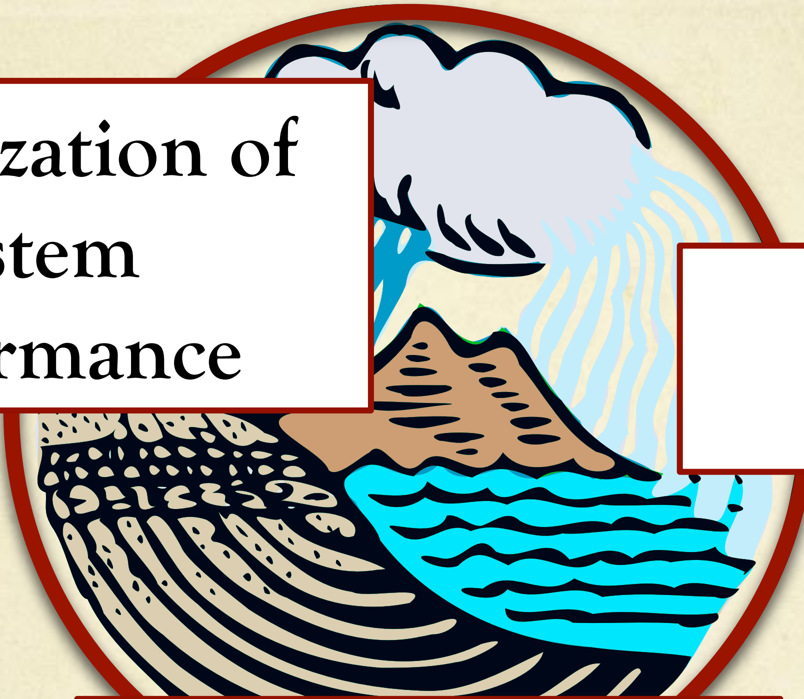
The Universe

Systems

Everything

Jesus Carrera-Ramirez

Early Systems Hydrology



Optimization of System Performance

Decision Analysis

Dynamic Programming

The FAILURES reported by *Ibbitt* and by *Johnston & Pilgrim* intrigued me

Para

School of C

A detailed sear
Simplex and Davi
were readily achi
use of different o
progress to be m
optimization. Th
surface and the o
types of stores, a
analyzed and the
apply to most wa

Failure

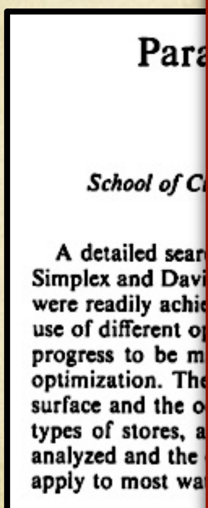
Failure

Failure

Failure

So for many years I plugged away at the problem !!!

The FAILURES reported by *Ibbitt* and by *Johnston & Pilgrim* intrigued me


Para

School of C

A detailed sear
Simplex and Davi
were readily achi
use of different o
progress to be m
optimization. The
surface and the o
types of stores, a
analyzed and the
apply to most wa

My Hair Grew
Very Long

Failure

Failure

Failure

Failure

So for many years I plugged away at the problem !!!

[Reported Failure]

And between 1985 & 1991 (6+ years)...

I did not publish a single paper !

(try doing that today)

Then in 1992 & 1993 ...



WATER RESOURCES RESEARCH, VOL. 28, NO. 4

Effective and Efficient Global Optimization for Conceptual Rainfall-Runoff Models

QINGYUN DUAN, SOROOSH SOROOSHIAN, AND VIJAI GUPTA

Department of Hydrology and Water Resources, University of Arizona, Tucson

The successful application of a conceptual rainfall-runoff (CRR) model depends on how well it is calibrated. Despite the popularity of CRR models, reports in the literature indicate that it is typically difficult, if not impossible, to obtain unique optimal values for their parameters using automatic calibration methods. Unless the best set of parameters associated with a given calibration data set can be found, it is difficult to determine how sensitive the parameter estimates (and hence the model forecasts) are to factors such as input and output data error, model error, quantity and quality of data,

EARCH, VOL. 29, NO. 4, PAGES 1185–1194, APRIL 1993

Calibration of Rainfall-Runoff Models: Application of Global Optimization to the Sacramento Soil Moisture Accounting Model

SOROOSH SOROOSHIAN, QINGYUN DUAN,[1] AND VIJAI KUMAR GUPTA

Department of Hydrology and Water Resources, University of Arizona, Tucson

Conceptual rainfall-runoff models are difficult to calibrate by means of automatic methods; one major reason for this is the inability of conventional procedures to locate the globally optimal set of parameters. This paper investigates the consistency with which two global optimization methods, the shuffled complex evolution (SCE-UA) method (developed by the authors) and the multistart simplex (MSX) method, are able to find the optimal parameter set during calibration of the Sacramento soil moisture accounting model (SAC-SMA) of the National Weather Service River Forecast System (NWSRFS).

Shuffled Complex Evolution

MORAL

Be Patient

Be Stubborn

Don’t be Afraid to work on (seemingly) Difficult Problems

We Have Come a Long Way

Optimality

Model-Data Consistency

Characterizing Uncertainty
& Data Assimilation

The Truth Is

"Predictive Uncertainty"
does not interest me very much

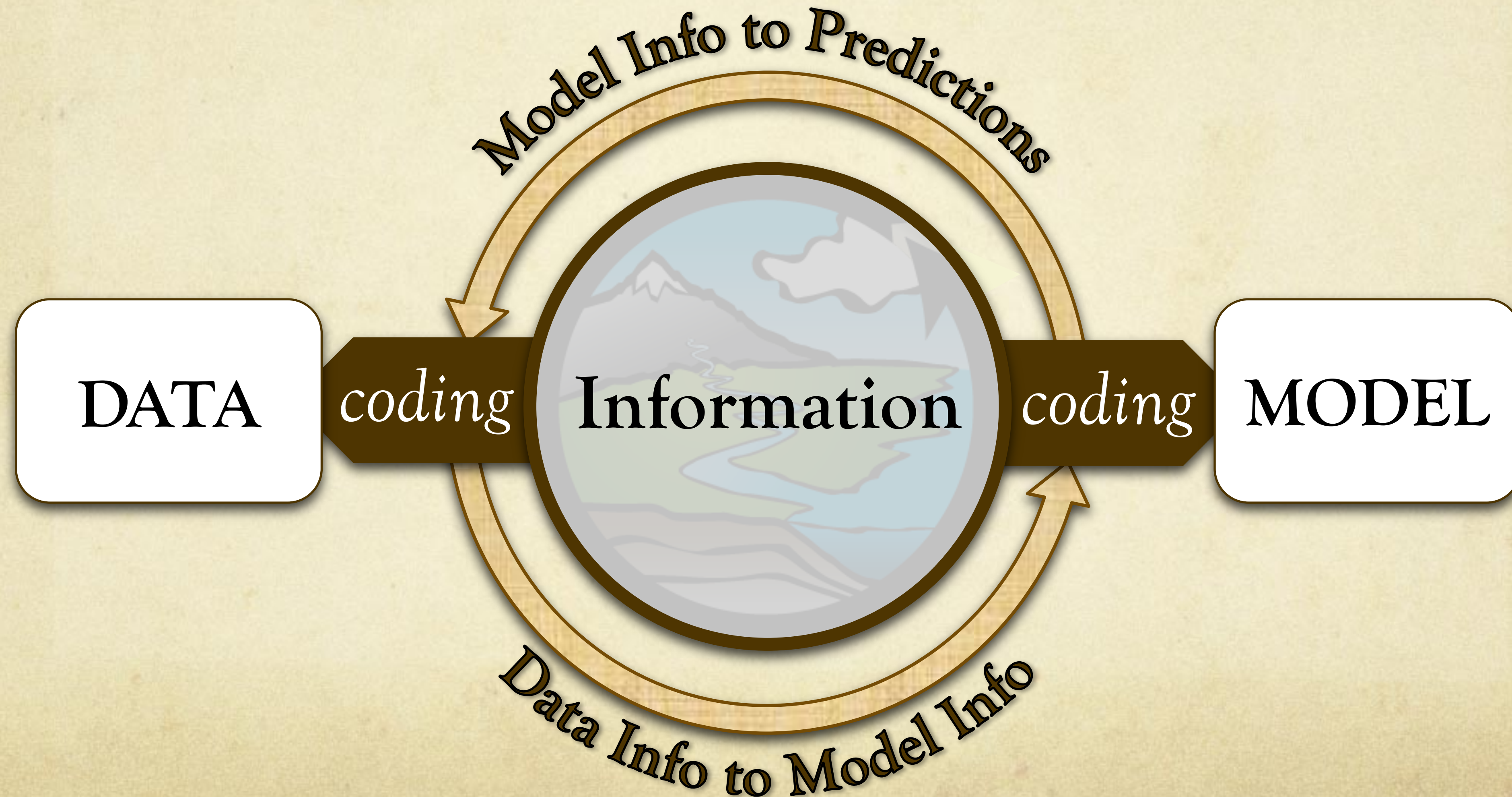
What interests me is the Problem of "Learning"

How to Develop and Codify
Scientific Understanding

How to use
Models & Data to Learn

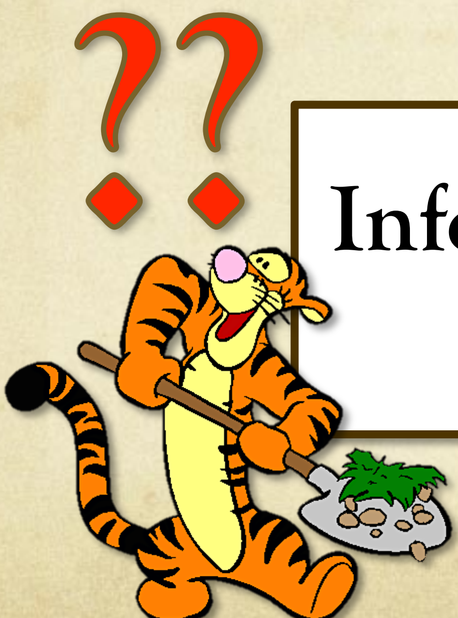
What Is "Information"?

How Information can be Characterized



♫ Lets Start at the Very Beginning ♬

What Is The Nature of “Information”?



Info is the answer to questions such as:
When, What, Where, How, Why …

Information & Context

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

Context Matters

DATA is not Info … until viewed in context

3.14

Information & Context

Info is always "about" something:

Context Matters

DATA is not Info … until viewed in context

3.14

Streamflow

(mt^3/sec)

Information & Context

Statements like

"The Data Contain Information"

"The Model Contains Information"

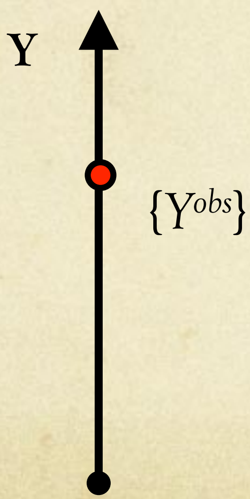
Are not well-formed (*they are ambiguous*)

Info is "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:

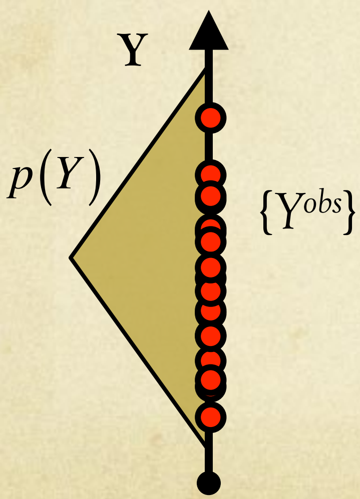
A Value of "Something"

DATA-Info (1)

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:

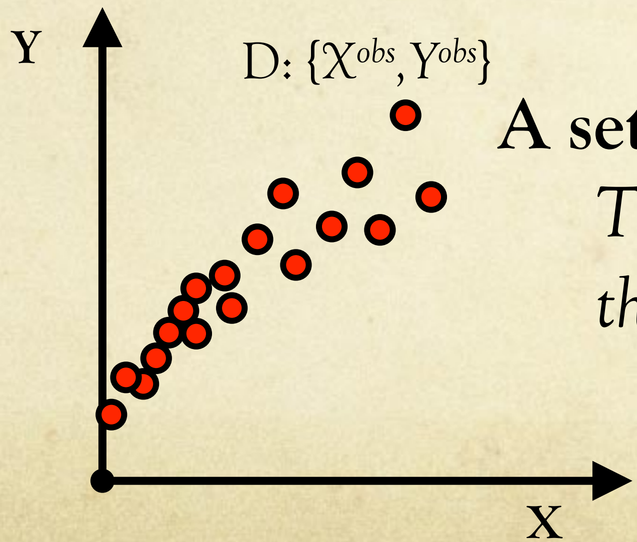
The Distribution of Values

DATA-Info (2)

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:

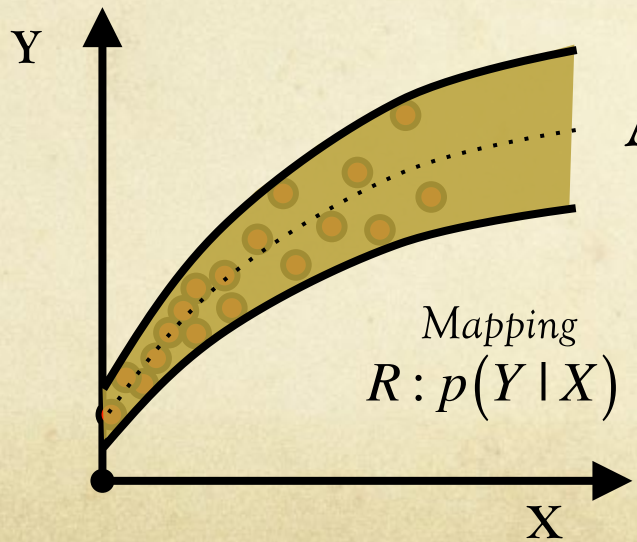
The space-time-ordered Relationships among those values

MODEL-Info (1)

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:

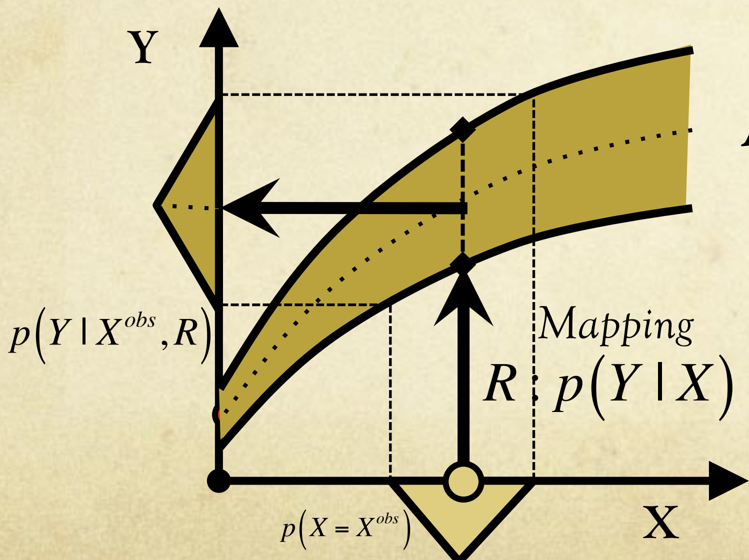
The space-time-ordered Relationships between variables 'behind' those values

MODEL-Info (2)

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*

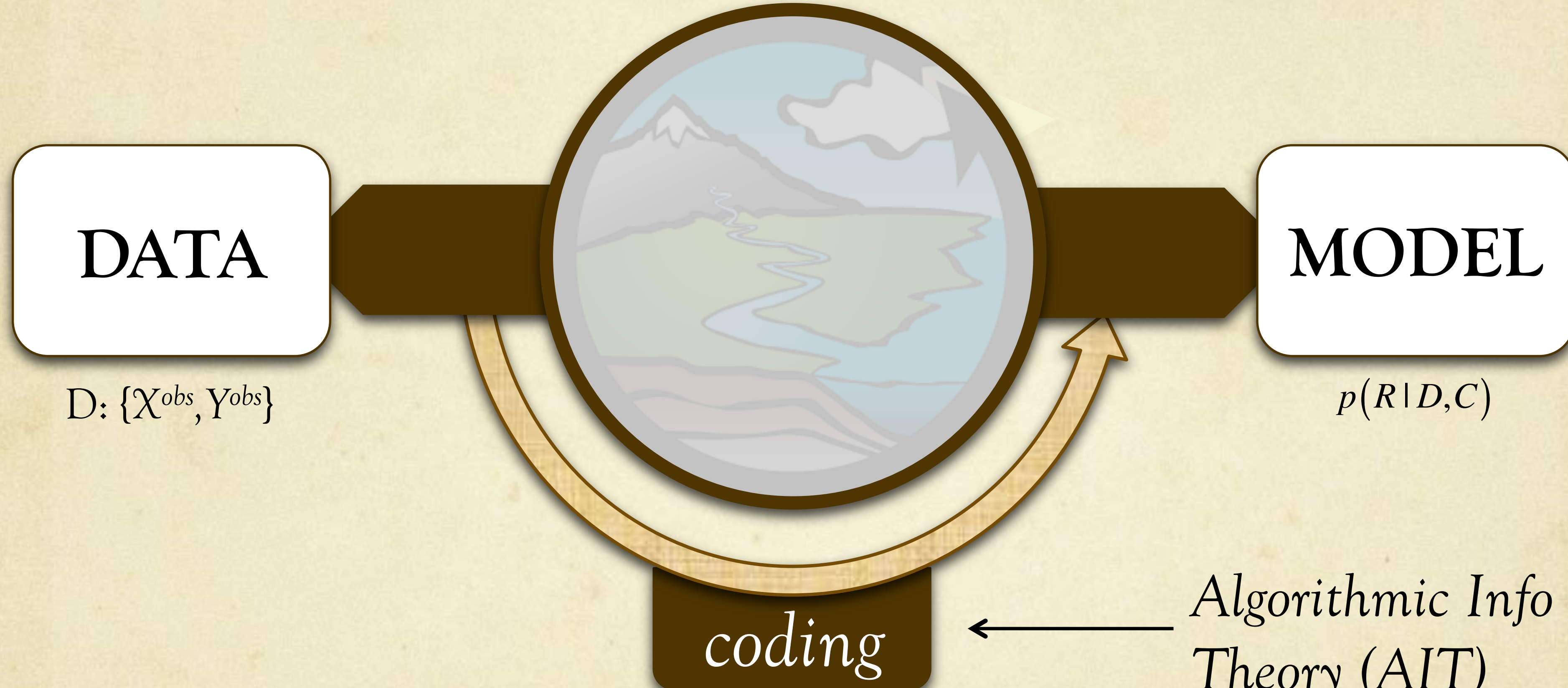


A MODEL encodes Info about:

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

Learning Involves

Converting Data-Info → Model-Info



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

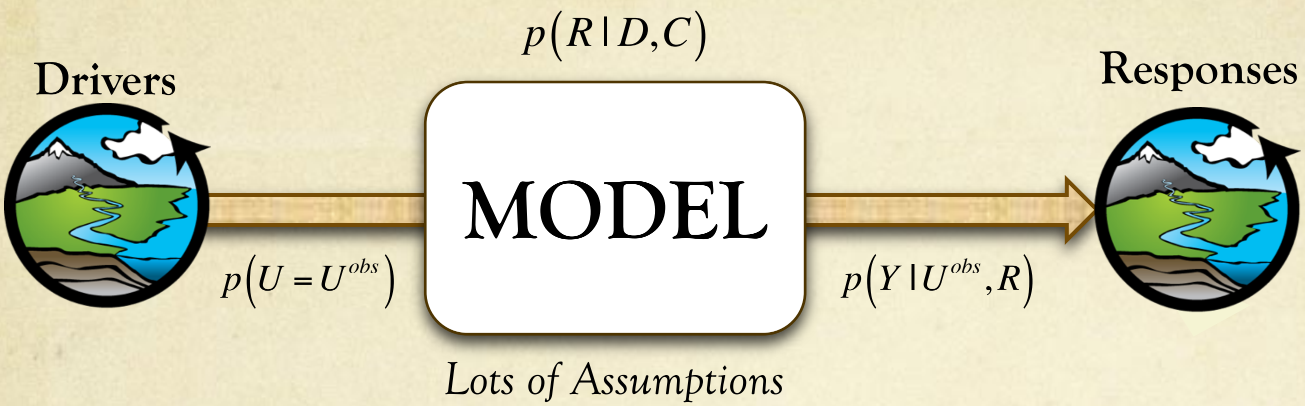
$p(R \mid D,C)$

Algorithmic Info Theory (AIT)

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

***The conditioning role of Assumptions is incredibly strong*

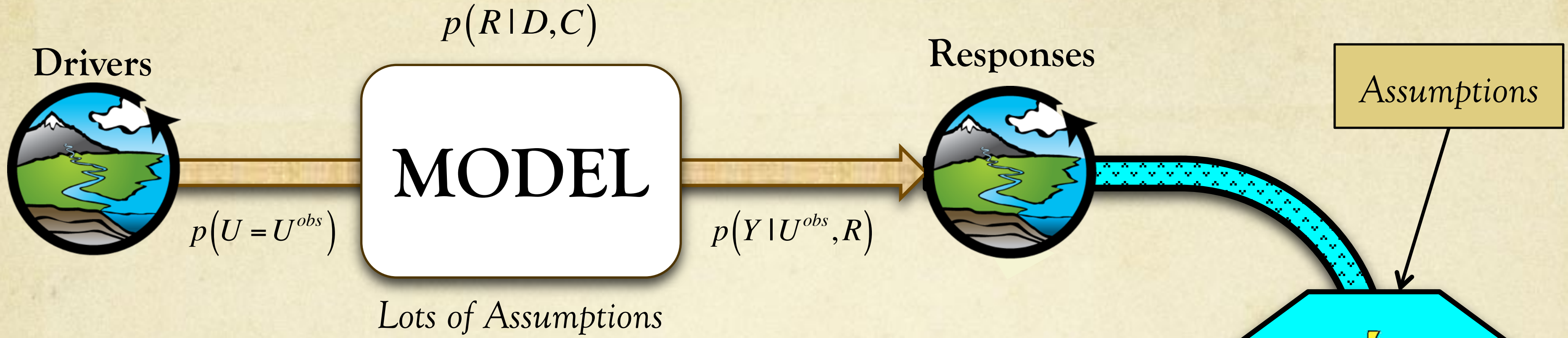
The Model Enables

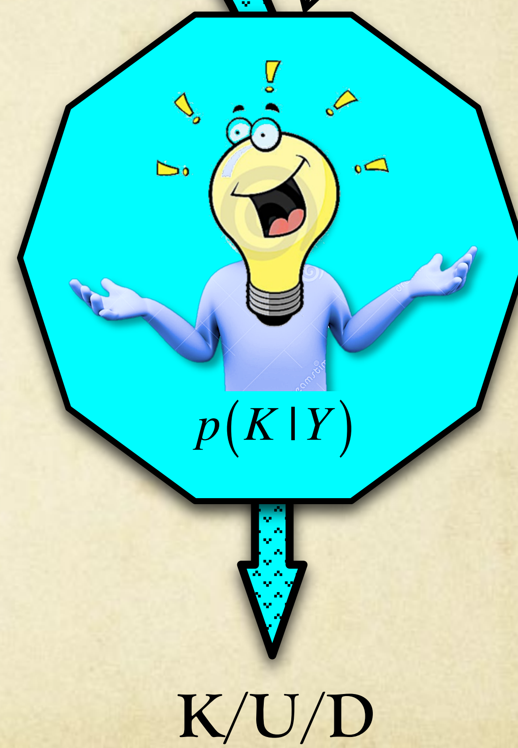


MODEL

Converting one kind of Data-Info into another kind of Data-Info

The Model Enables





Which can be further converted into

Knowledge/Understanding/Decisions

by the recipient (human or otherwise)

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

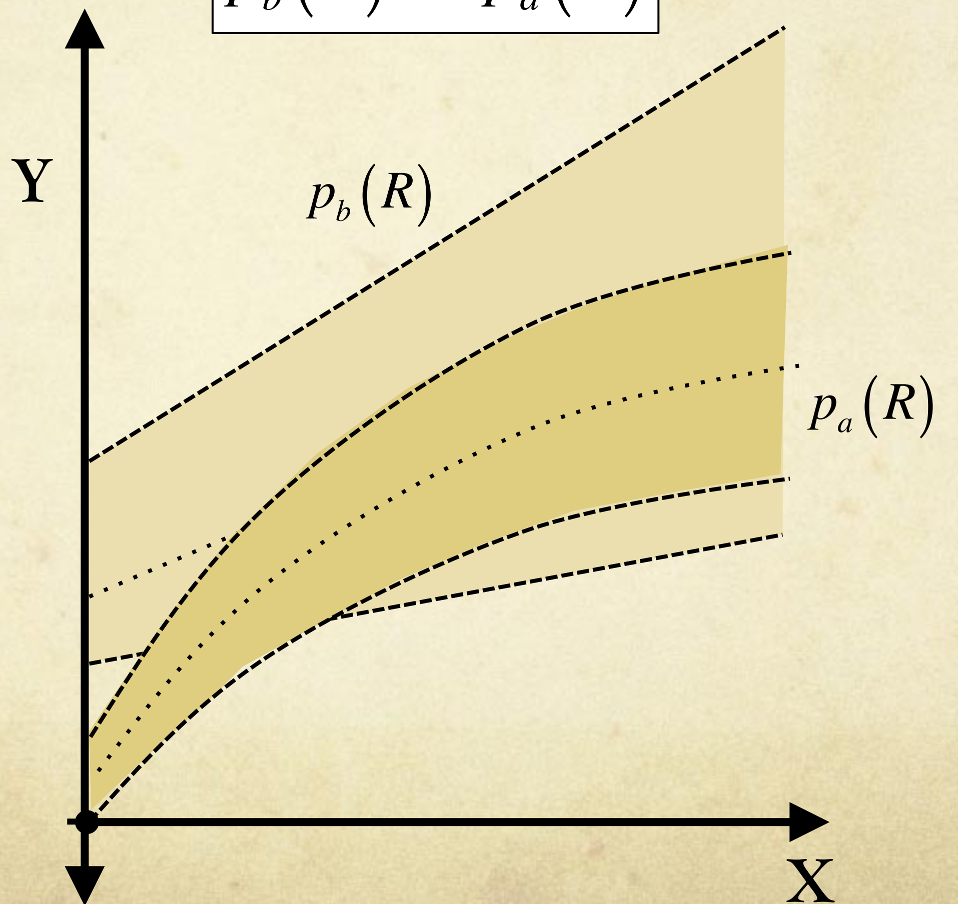
* *The conditioning role of Assumptions is incredibly strong*

So … How Do We Characterize Learning?

- Learning occurs if our Prior UNCERTAINTY is *CHANGED* by the addition of INFO

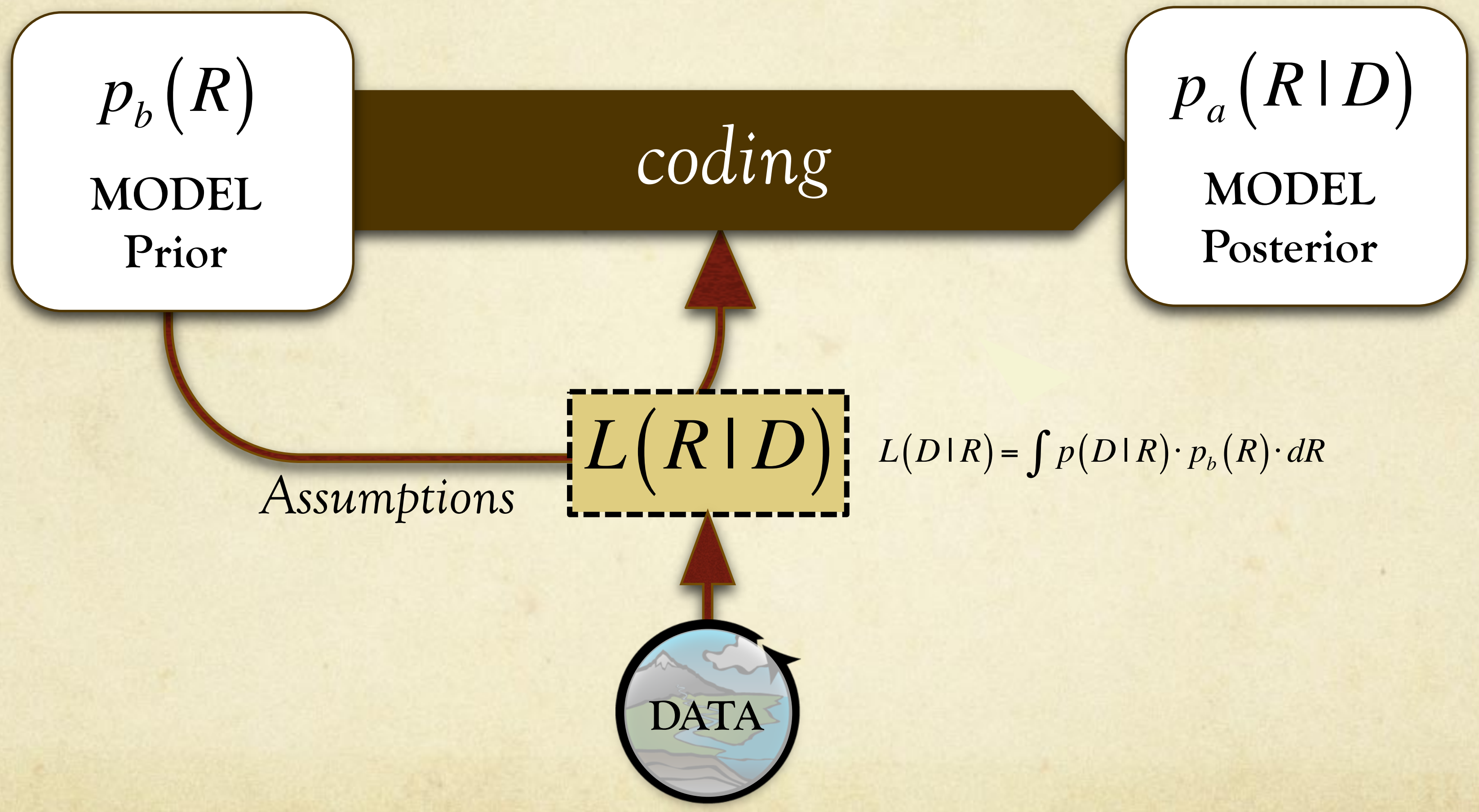
$$\boxed{p_b(R) \Rightarrow p_a(R)}$$

"Changed" …
not Reduced



Bayes Law codifies this Change

$$p_a(R|D) = k \cdot L(R|D) \cdot p_b(R)$$



Under very specific conditions (which ones?)

Information can be

Grey Nearing

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

Prior Work

How Info is coded into Data → Shannon & Others

"Shannon Info" is about coding/decoding info in data

But the context is Communication Theory …

How Info can be extracted from Data → Fisher, Bayes, Bernoulli, Ramsey, Edwards … long list

Many Statisticians have studied this

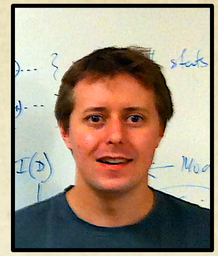
But the context is largely generic …

How Info is encoded in Models → Solomonoff & Others

Many Mathematicians have studied this

Again the context is largely generic …

Our Interest



Grey Nearing

Maximize *Efficiency & Effectiveness* **of "Learning"**

(about dynamical systems)

1. *Decode the Info in Data:* → *Diagnostic Approach (data in context)*
2. *Understand how Info is encoded into Structures of "Physics" Models* → **? ? ?**
3. *Design strategies for learning from data* → *Detect & Correct Model Structural Inadequacies*

So … How is Info coded in Models ?

THREE IMPORTANT STEPS

Each Step Adds (codifies) Info

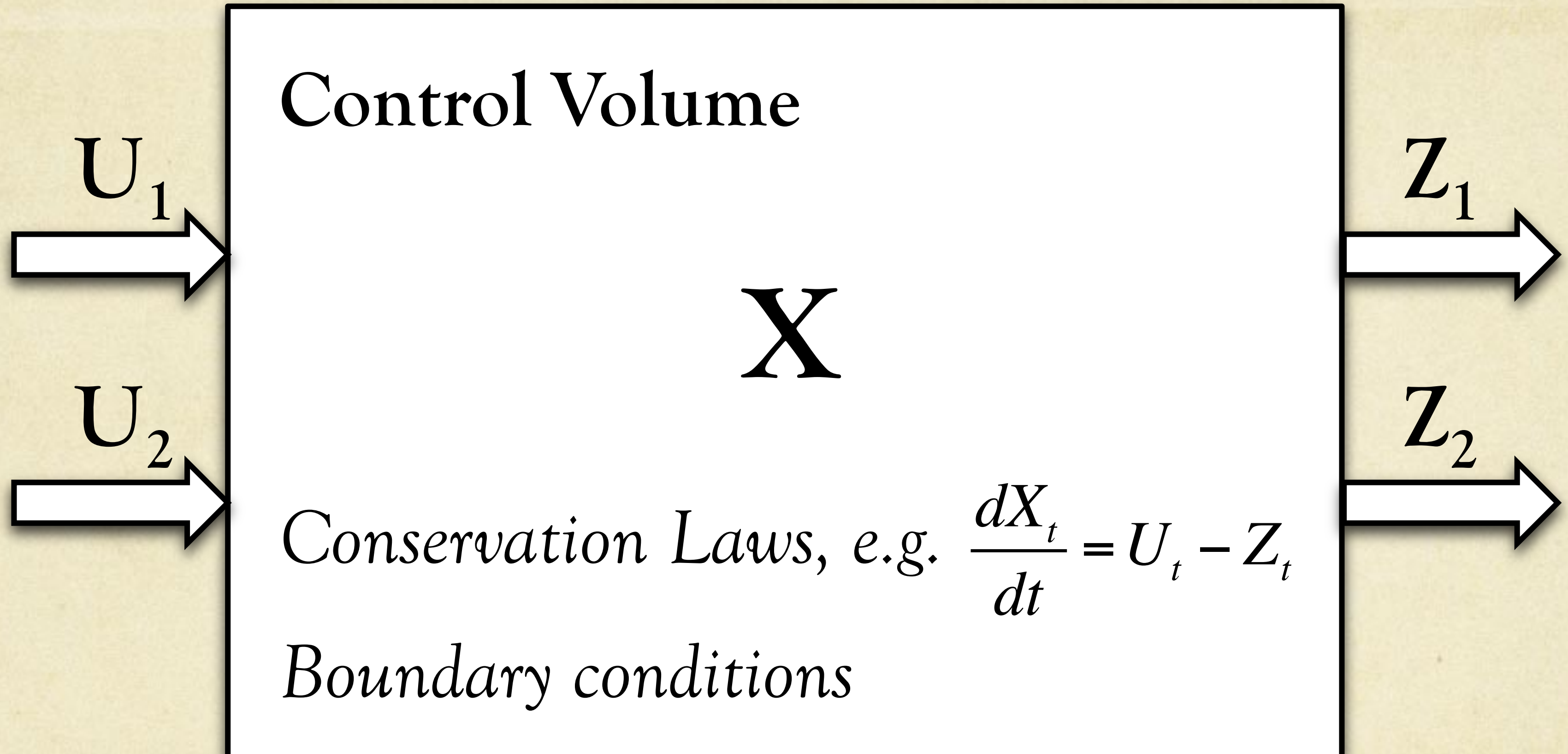
1. **System Diagram** (*Conservation Law Hypothesis*)
2. **Sub-system Architecture** (*Process Model Hypothesis*)
3. **Parameterization** (*Process Equations Hypothesis*)
4. *Computational Implementation*
5. *Inference*

Gupta, Clark, Vrugt, Abramowitz & Ye (WRR 2012)
Towards a Comprehensive Assessment of Model Structural Adequacy

Gupta & Nearing (WRR 2014, Debates on Water Resources)
Using Models and Data to Learn - A Systems Theoretic Perspective on the Future of Hydrological Science

<u>Step (1) – System Diagram</u>

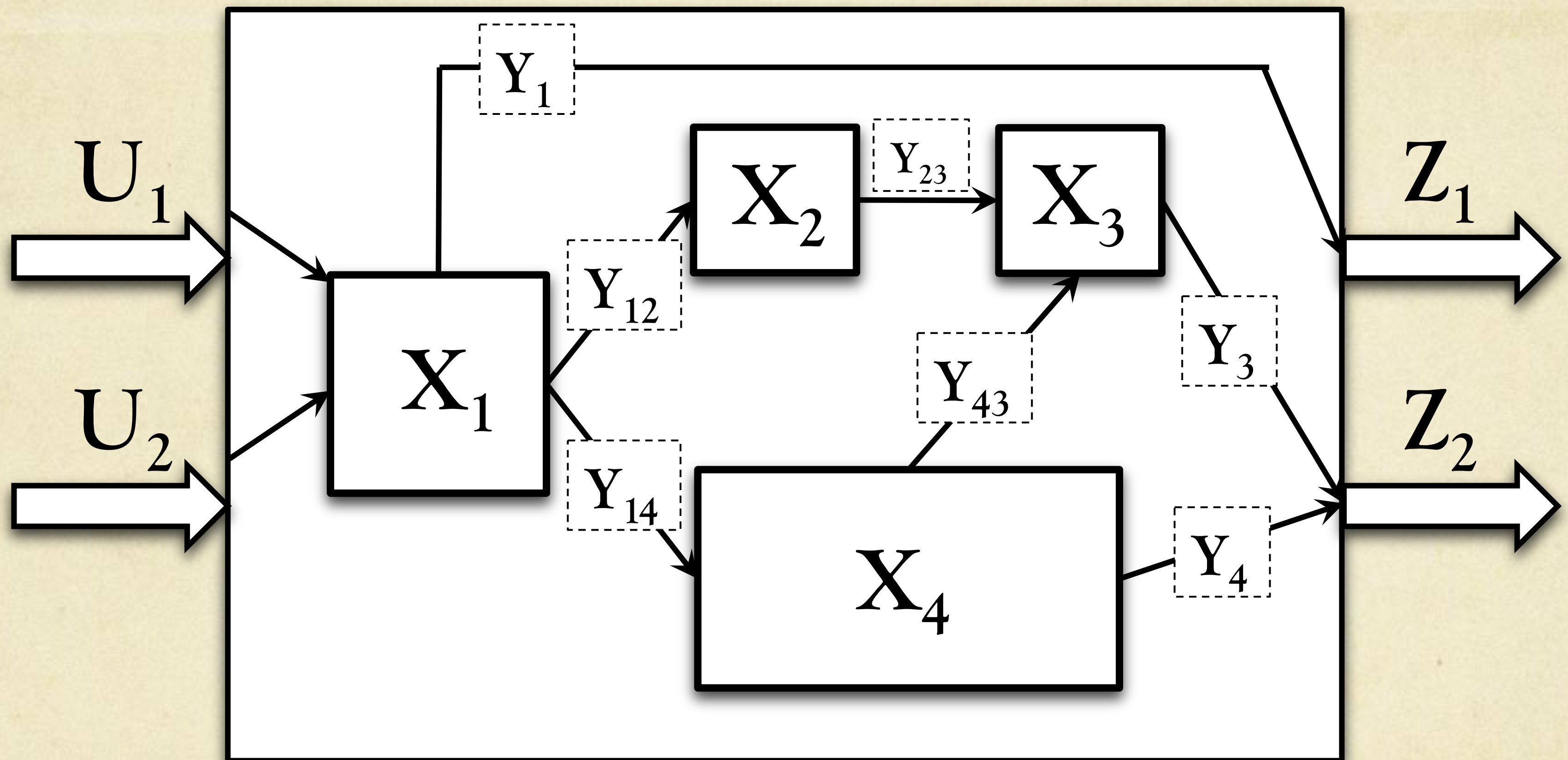
(Conservation Laws)



This *<u>restricts the possible values</u>* **that U, X and Z can jointly take on,**
<u>as well as the possible trajectories</u> **for system evolution.**

Step (2) – Sub-system Architecture

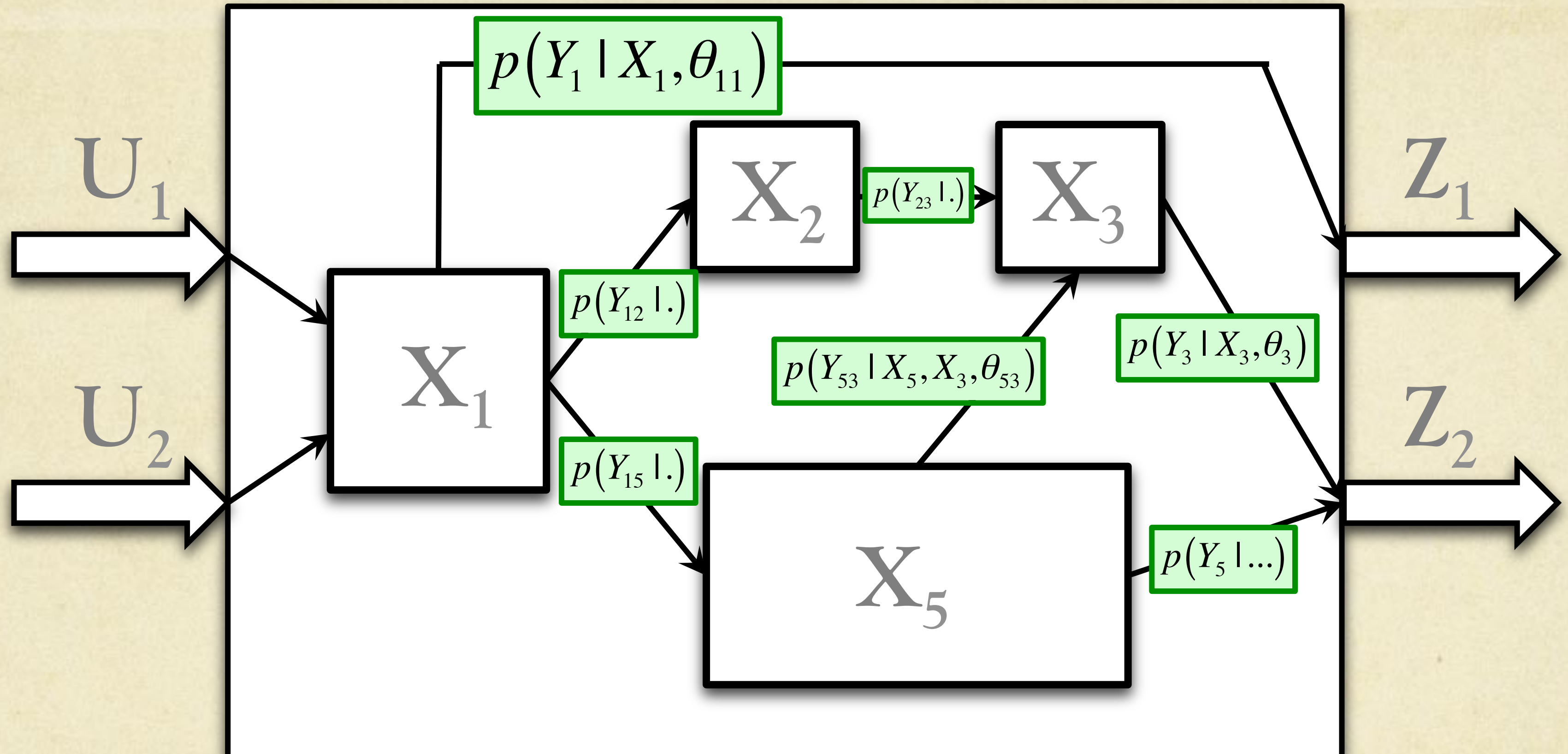
(Process Model | Conservation Laws)



This can further restrict the possible trajectories, while adding information about internal variability.

Step (3) – Process Parameterization

(Process Equations | Process Model | Conservation Laws)



Introduces Parameterized Equations (State-Flux relationships)
That can be "tuned" by adjusting the "parameters".

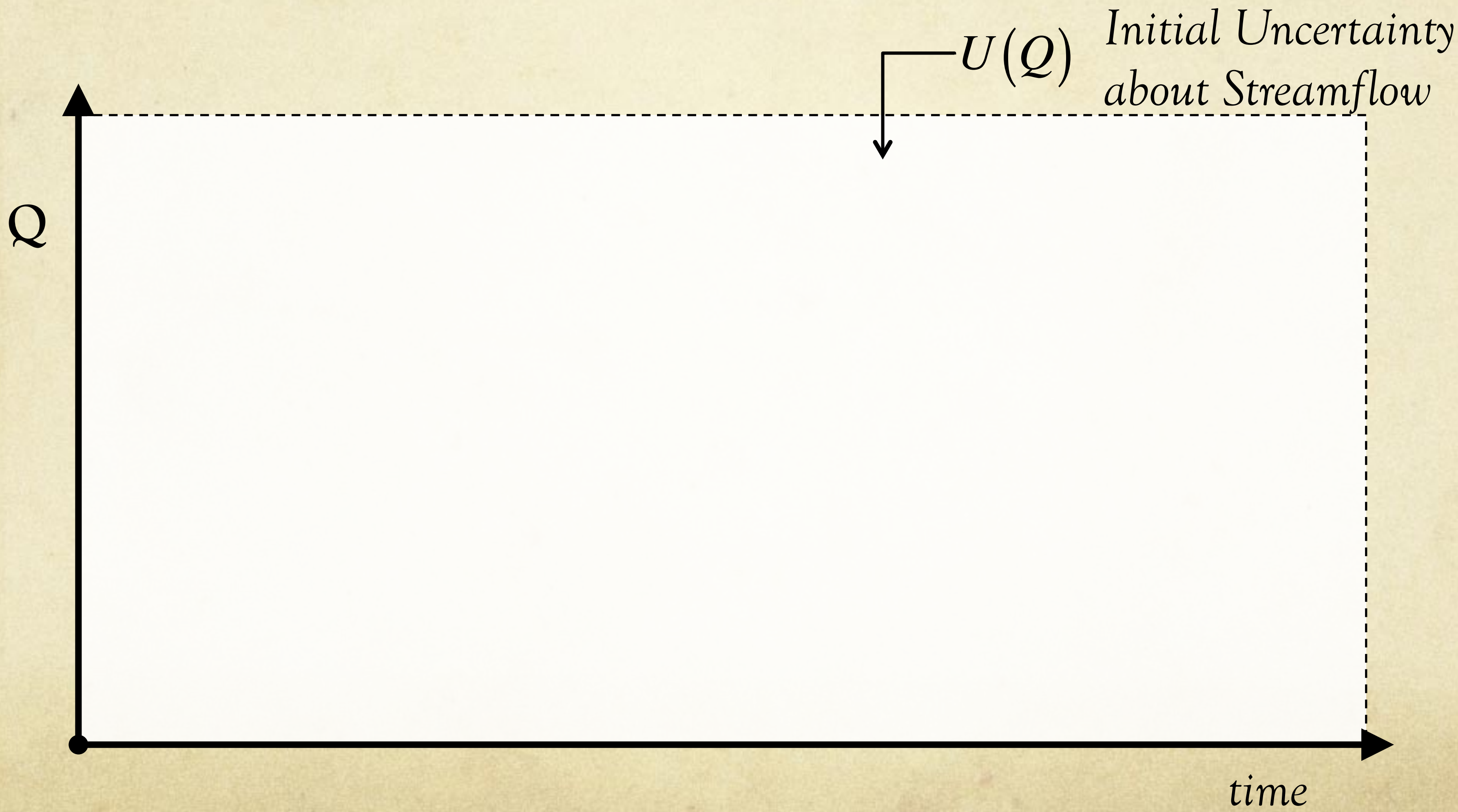
At Each Step

Info is Added (to Model Structure)

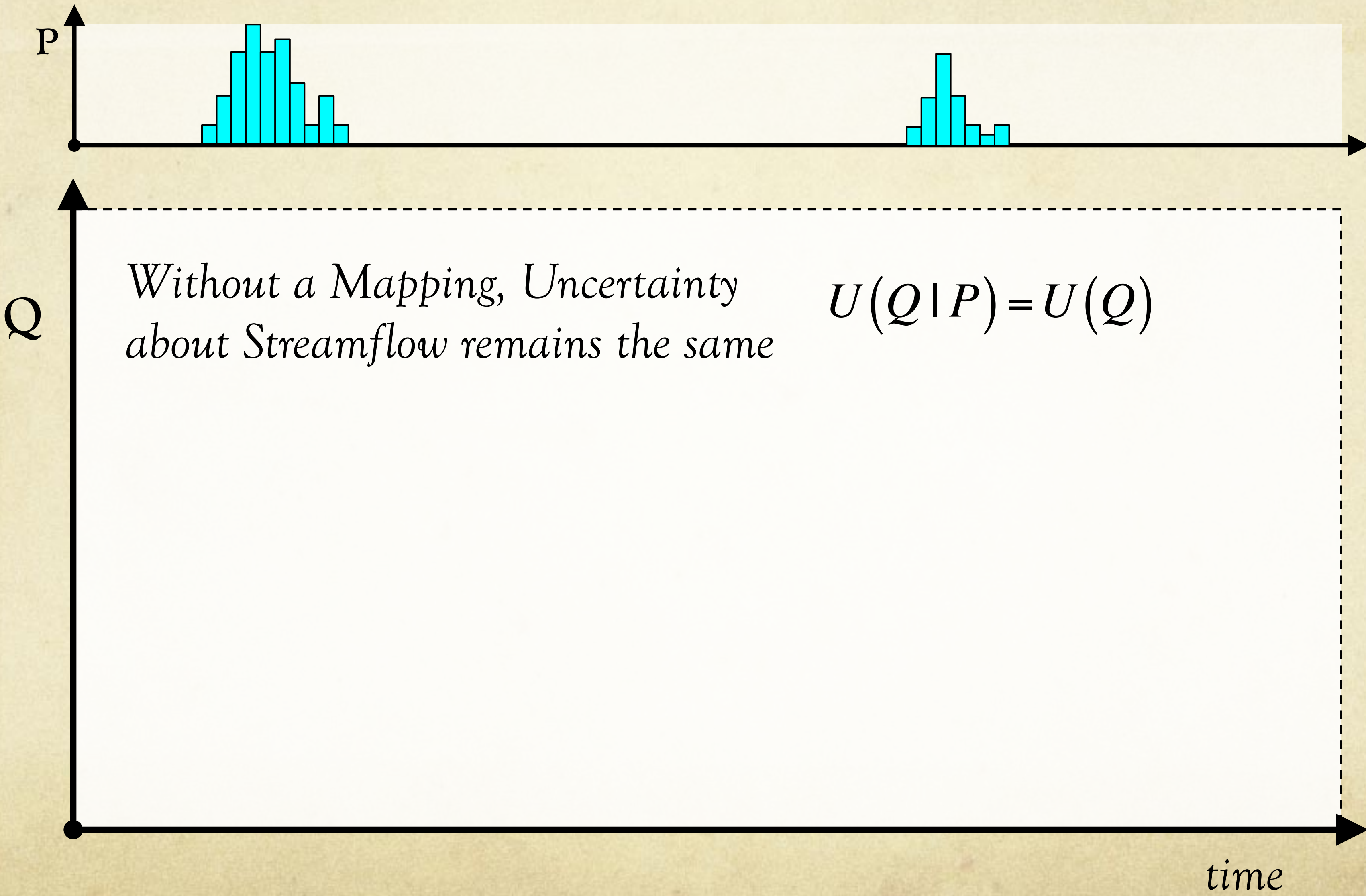
Uncertainty is Changed

Conceptual Illustration

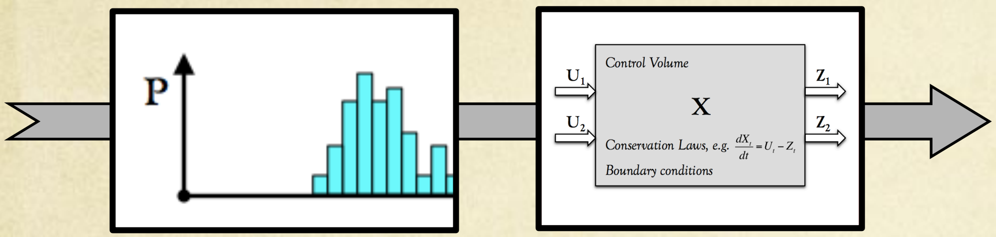
Want to simulate streamflow Q



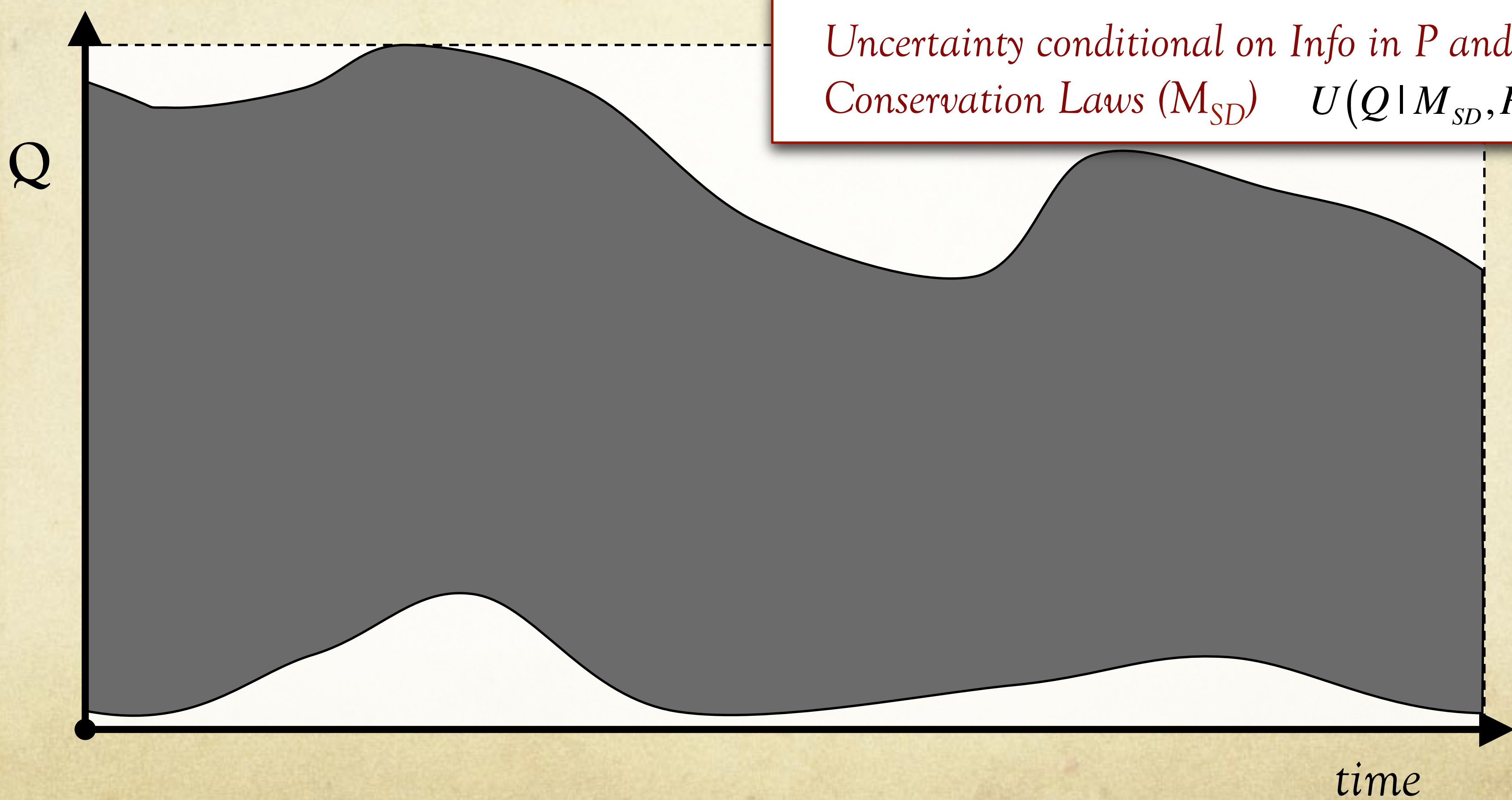
Conceptual Illustration



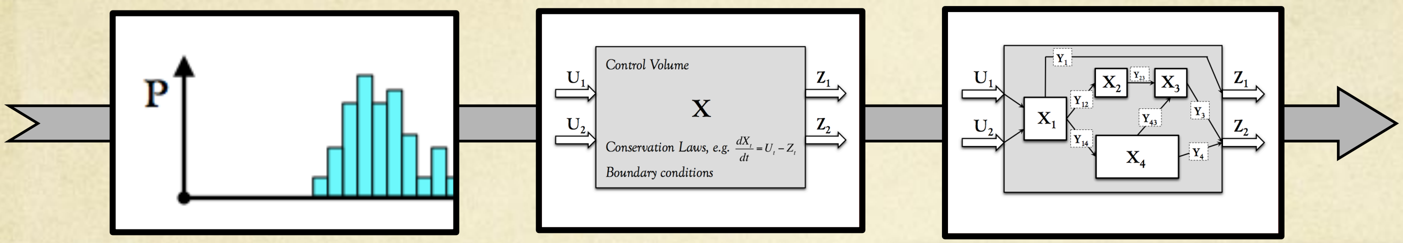
Conceptual Illustration



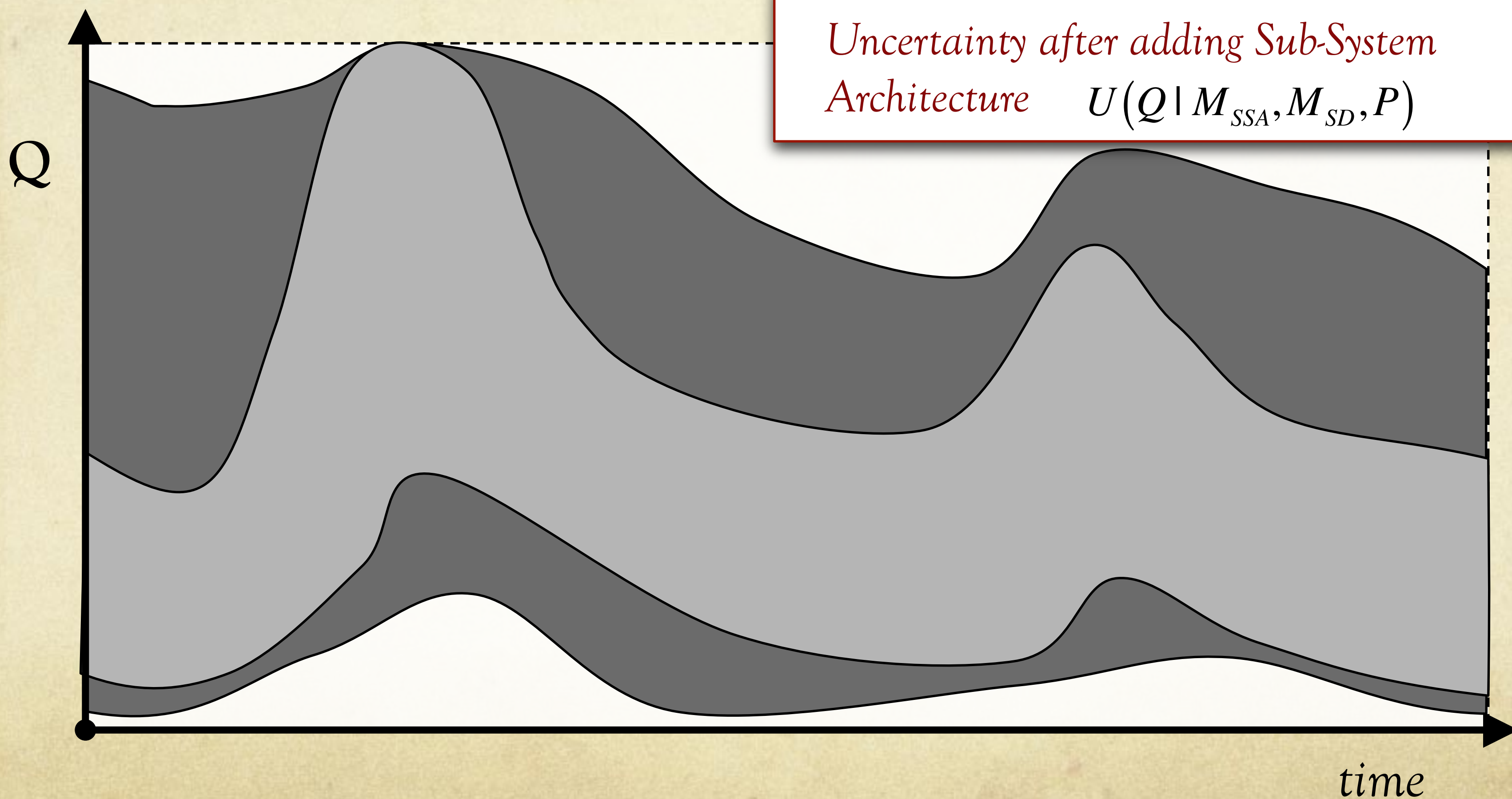
Uncertainty conditional on Info in P and Conservation Laws (M_{SD}) $U(Q \mid M_{SD}, P)$



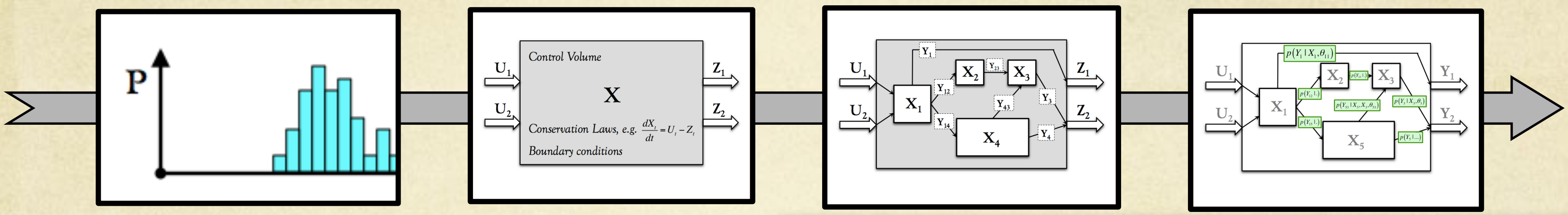
Conceptual Illustration



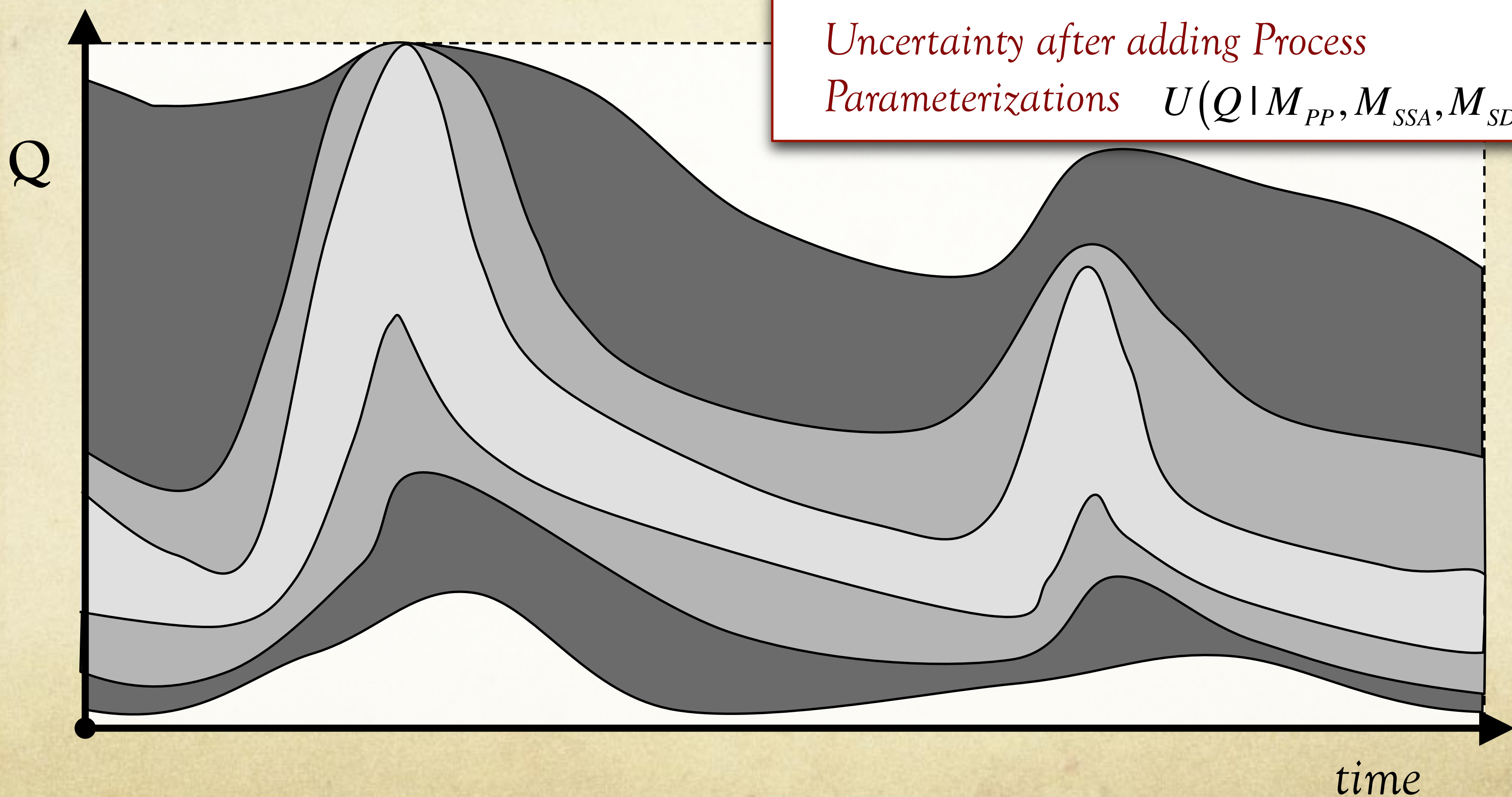
Uncertainty after adding Sub-System Architecture $U(Q \mid M_{SSA}, M_{SD}, P)$



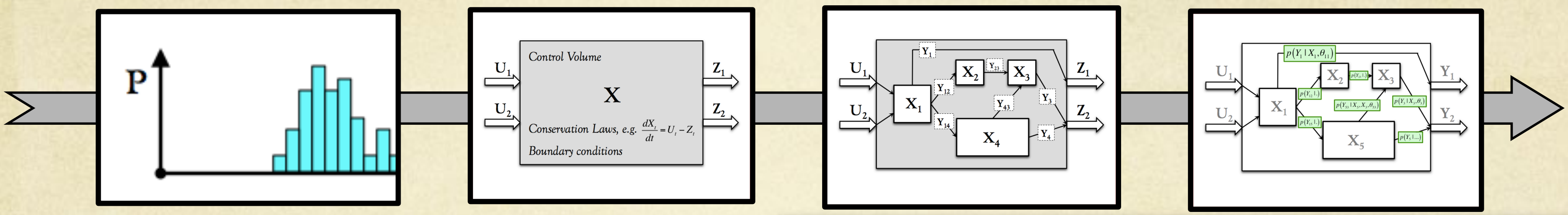
Conceptual Illustration



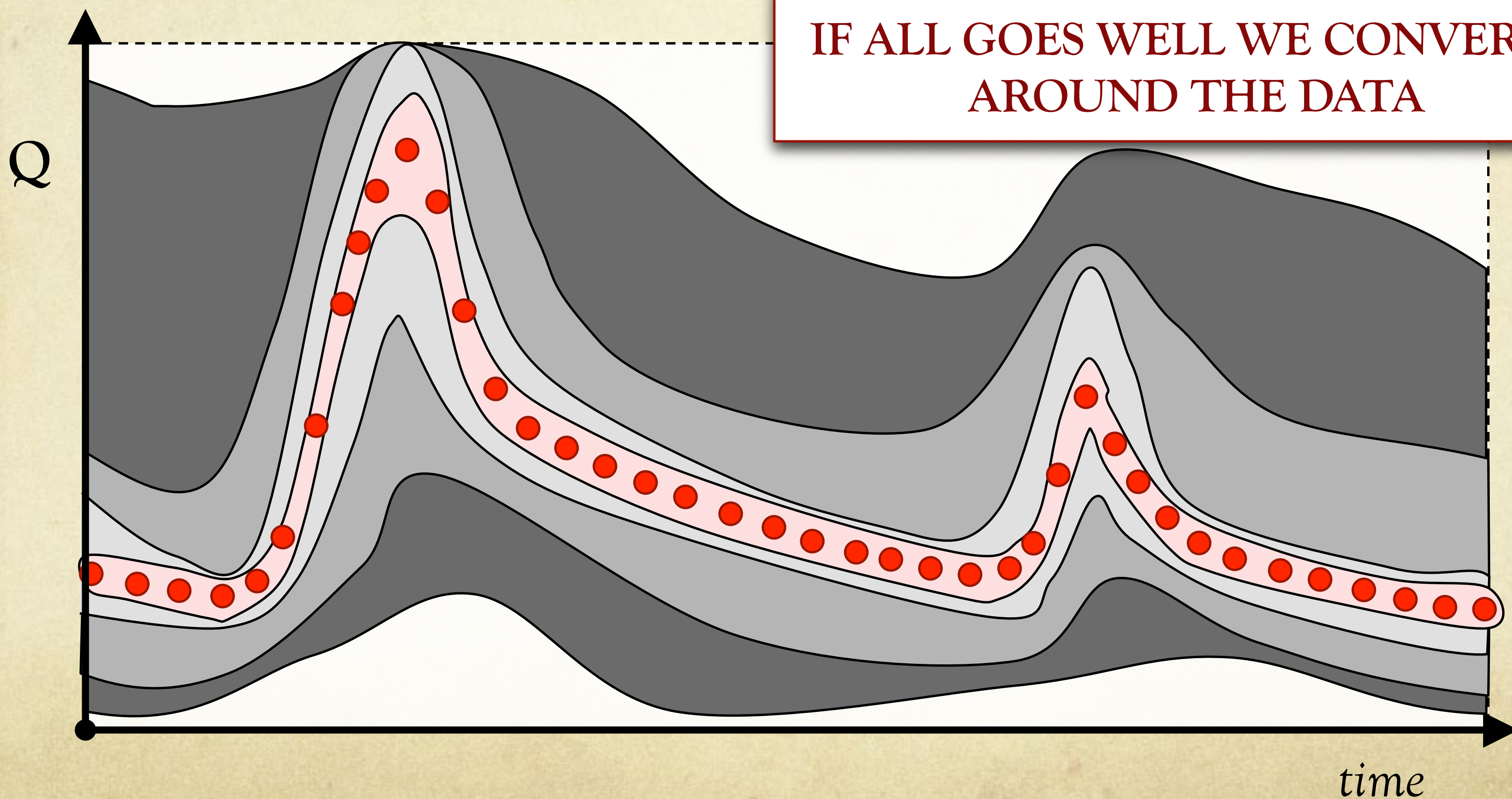
Uncertainty after adding Process Parameterizations $U\left(Q \mid M_{PP}, M_{SSA}, M_{SD}, P\right)$



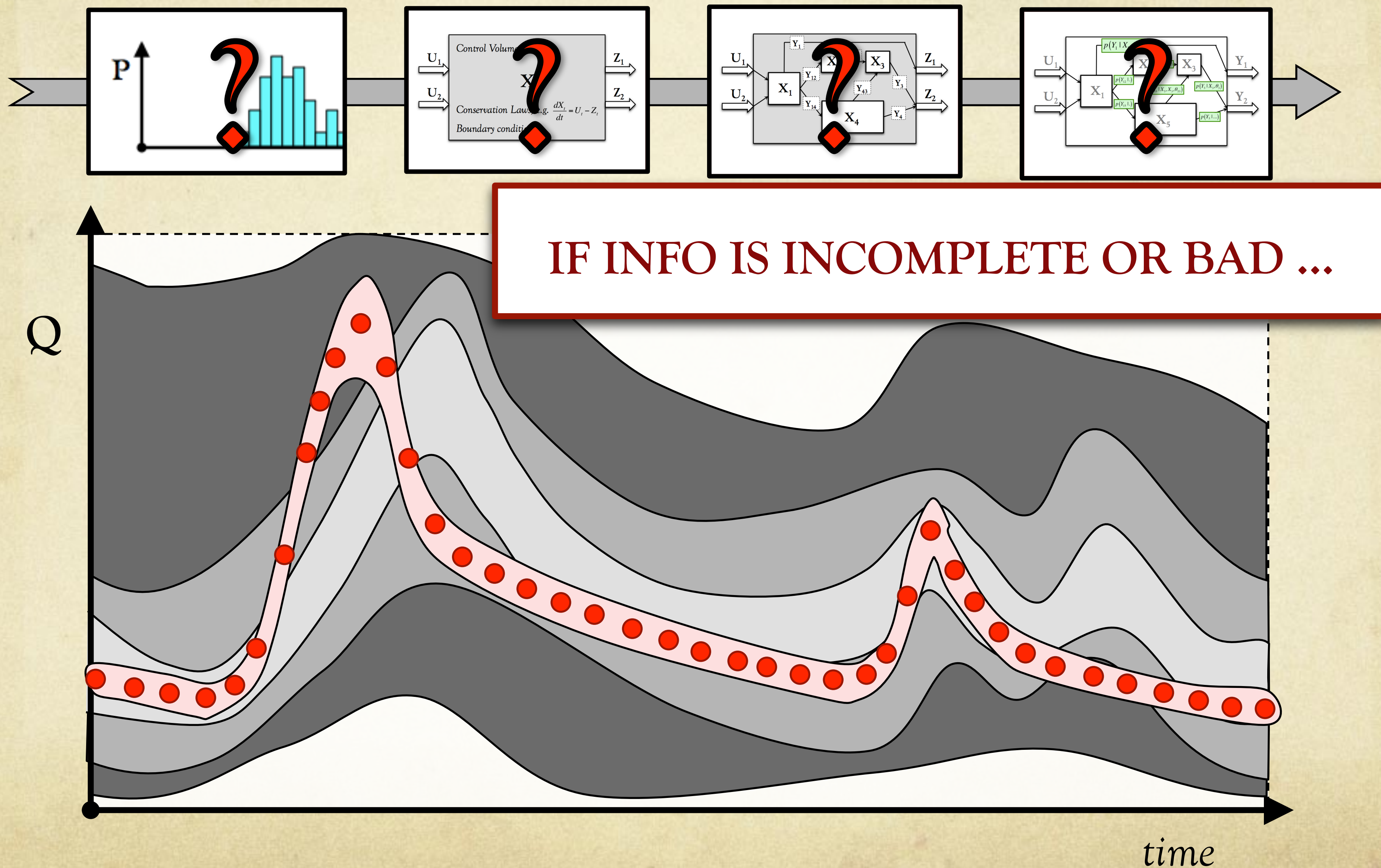
Conceptual Illustration



IF ALL GOES WELL WE CONVERGE AROUND THE DATA



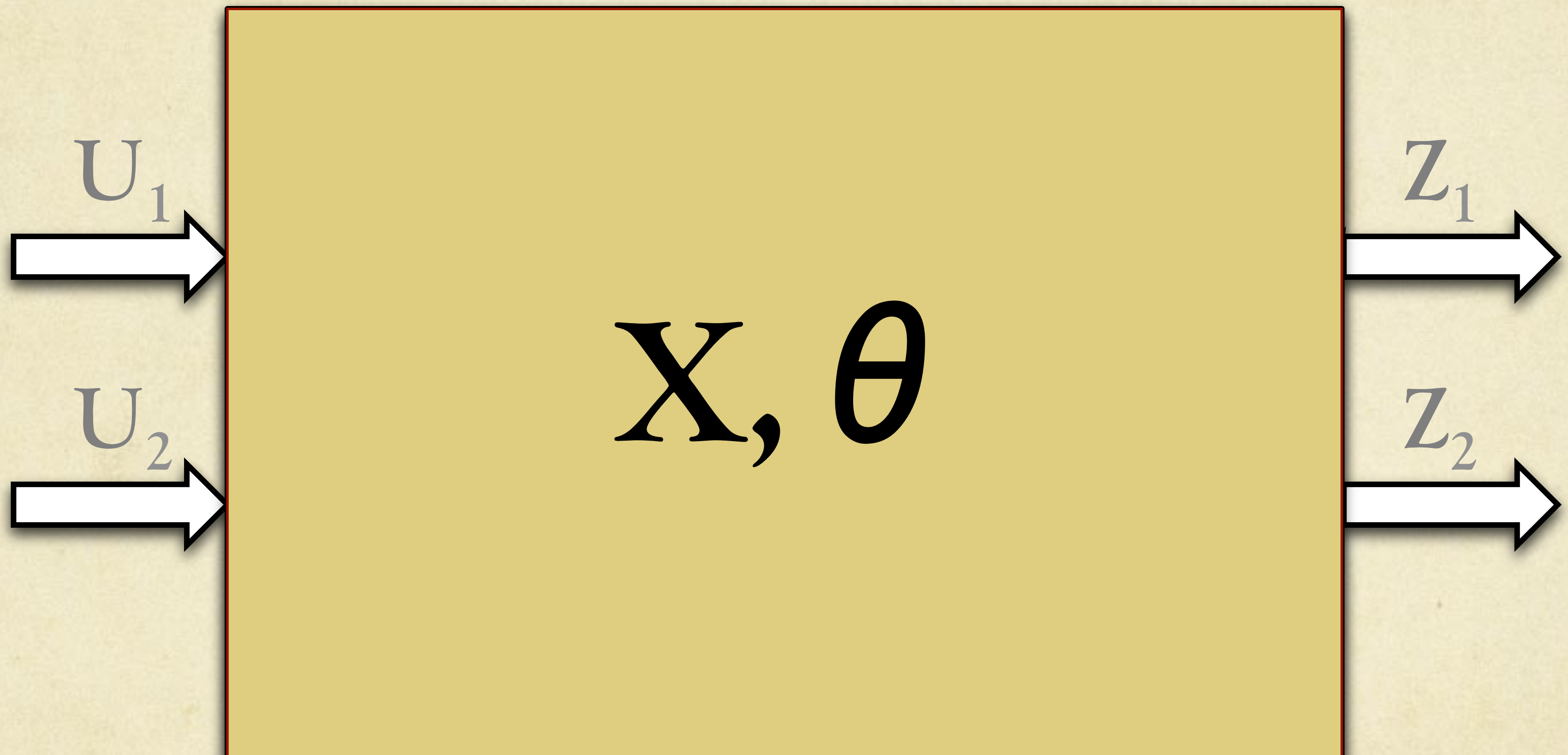
Conceptual Illustration



Historically

People try to fix this by "Parameter/State" Estimation

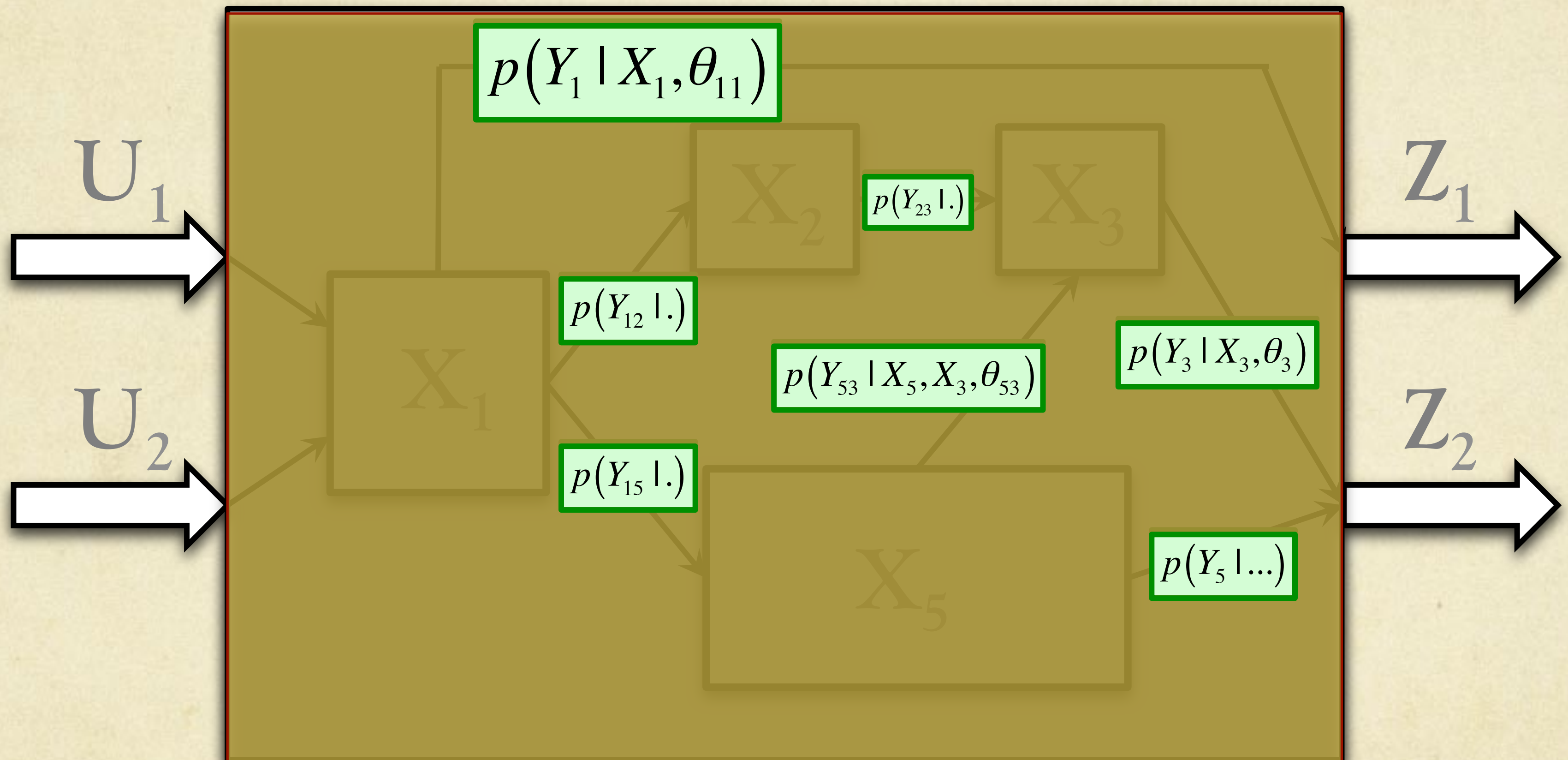
Conditioned on everything else assumed known



$$Max\ L\,(\, X \mid \theta \mid \underbrace{Param\_Hyp. \mid Arch\_Hyp. \mid Syst\text{-}Diag\_Hyp. \mid Data}_{\text{Assumed Known}}\,)$$

Recently

By trying "Parameterization" Estimation

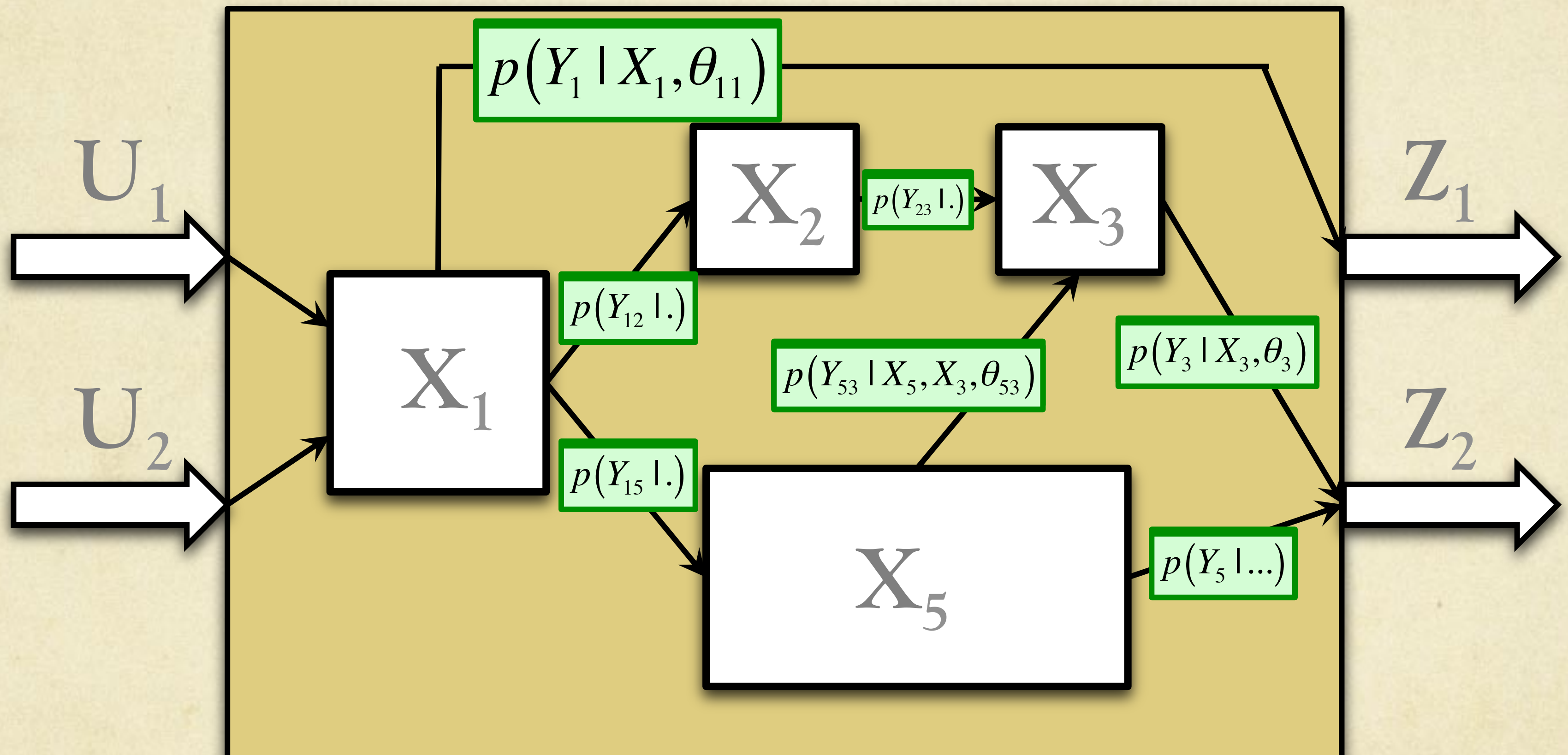


Max L (Param_Hyp. | Arch_Hyp. | Syst-Diag_Hyp. | Data)

Assumed Known

Recently (past few years)

And even "Sub-system Architecture" Estimation

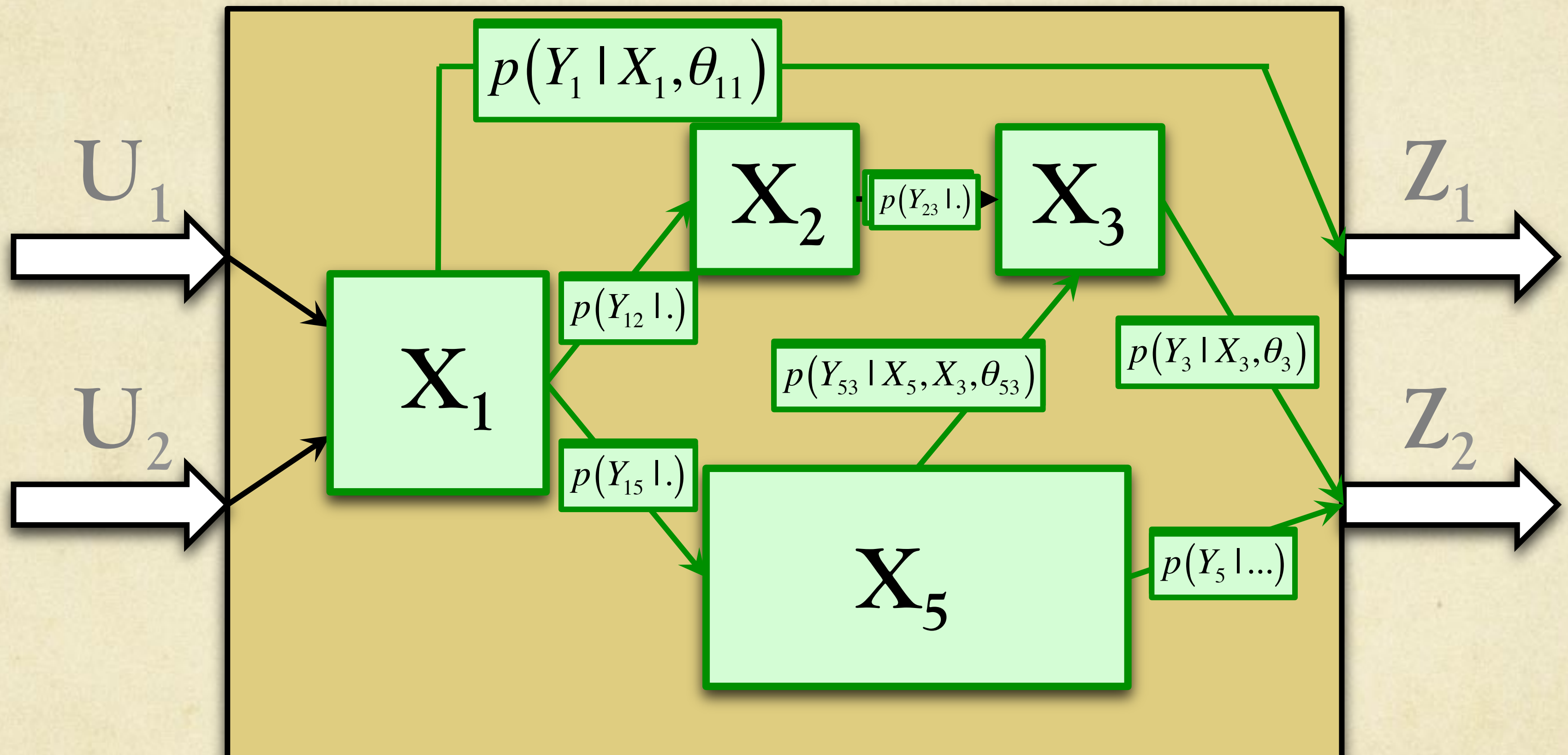


Max L (Param_Hyp. | Arch_Hyp. | Syst-Diag_Hyp. | Data)

Assumed Known

But in All These Efforts

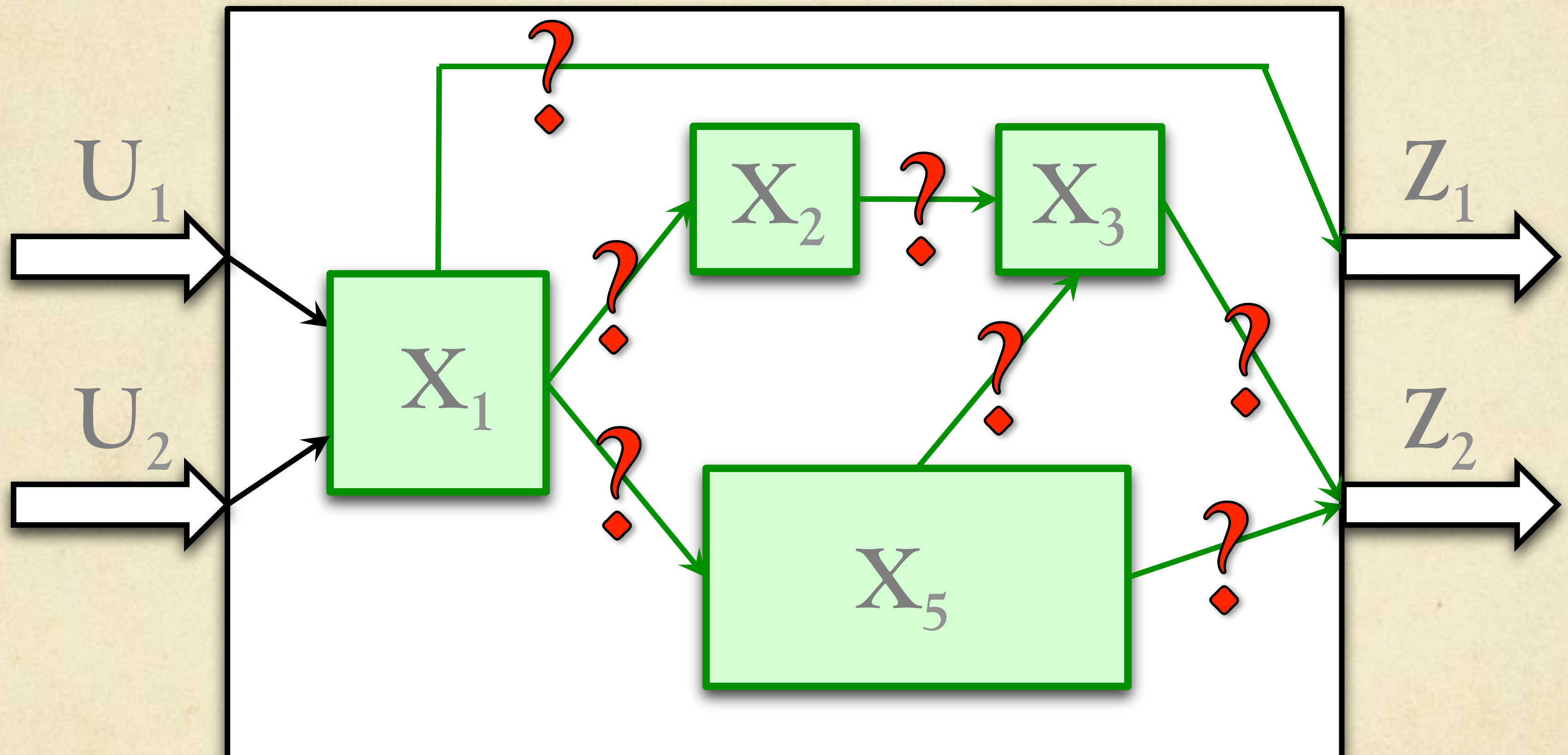
"Architecture" & "Parameterization" were treated together



So it is hard to know if the Model Structure Problem is with the "Architecture?" or with the "Parameterization?"

What if we could investigate "Architecture"

Without making STRONG statements about "Parameterization"?

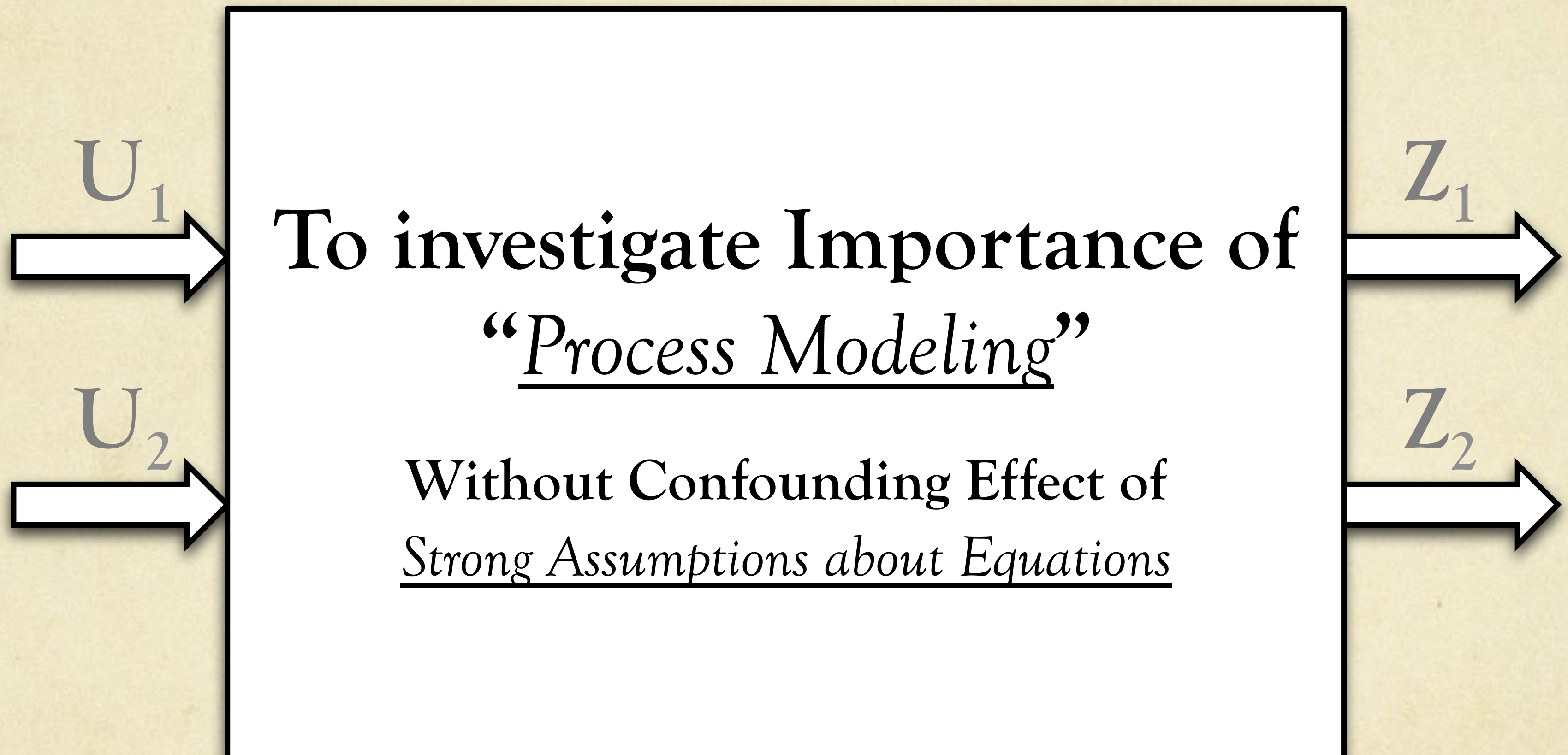


Max L (Arch_Hyp. | Syst-Diag_Hyp. | Data)

Assumed Known

What if we could investigate "Architecture"

Without making STRONG statements about "Parameterization"?



Grey Nearing

Last Six-Months

Shervan Gharari (TU Delft)

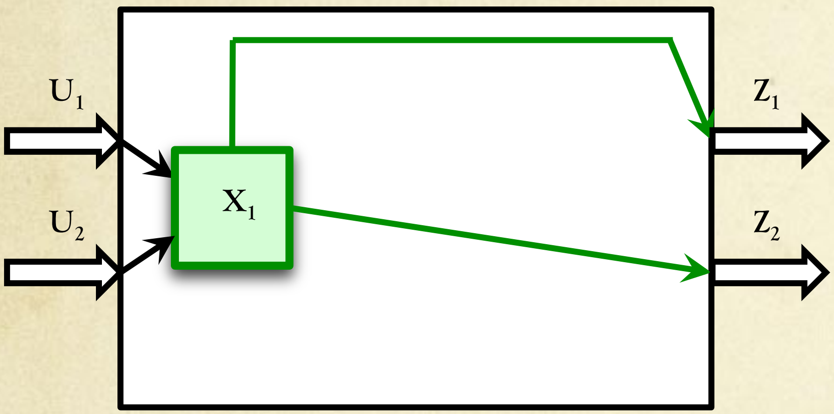
"Sub-system Architecture" Estimation

Grey Nearing

Last Six-Months

Shervan Gharari (TU Delft)

"Sub-system Architecture" Estimation

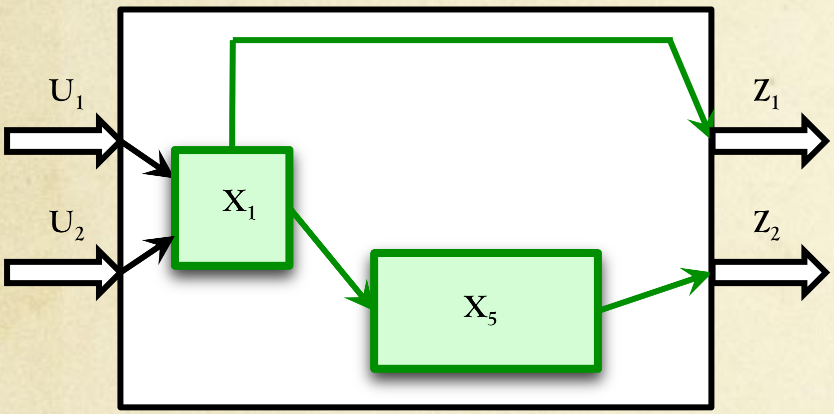


Grey Nearing

Last Six-Months

Shervan Gharari (TU Delft)

"Sub-system Architecture" Estimation

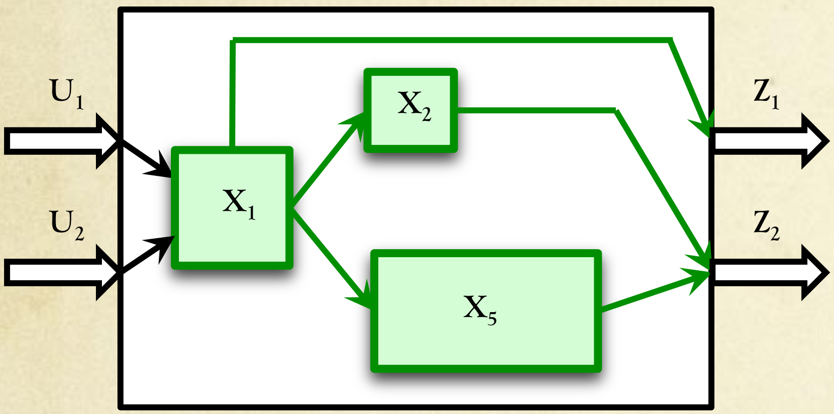


Grey Nearing

Last Six-Months

Shervan Gharari (TU Delft)

"Sub-system Architecture" Estimation

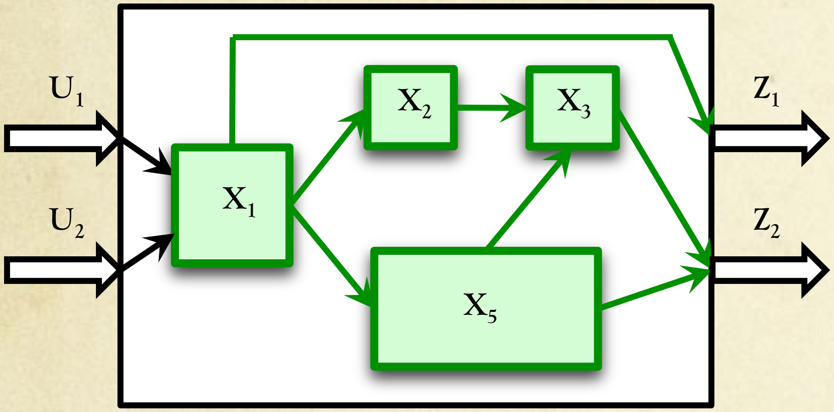


Grey Nearing

Last Six-Months

Shervan Gharari (TU Delft)

"Sub-system Architecture" Estimation

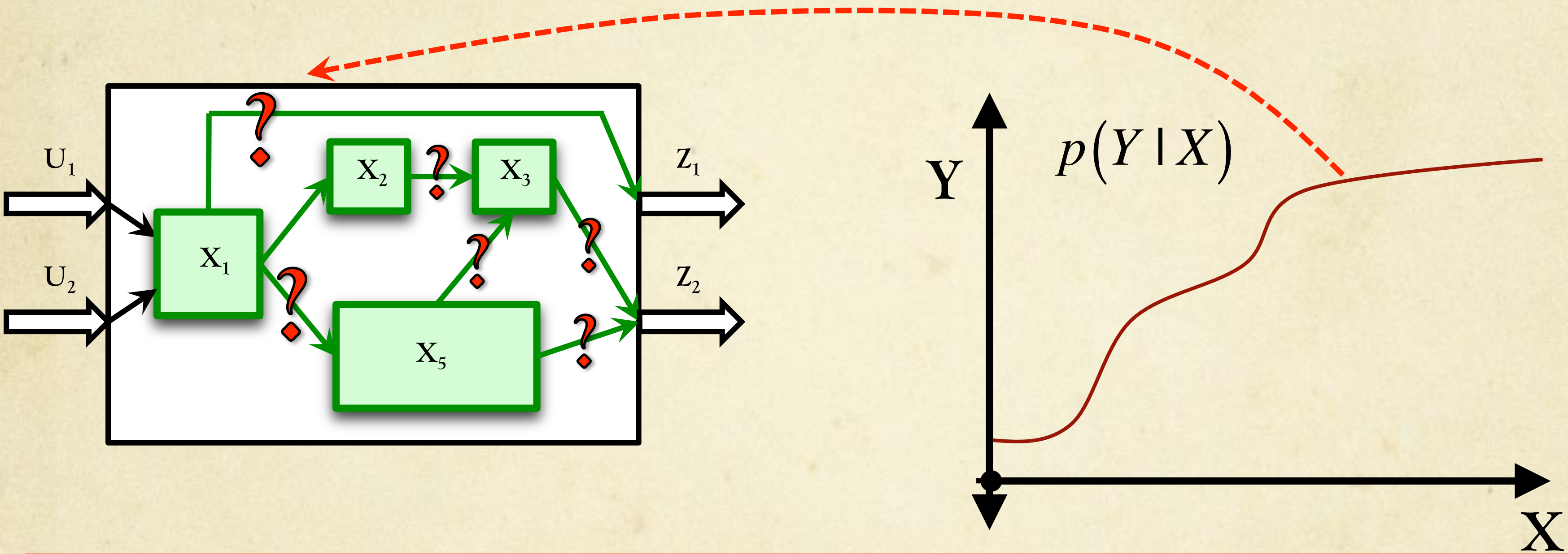


Grey Nearing

Last Six-Months

Shervan Gharari (TU Delft)

“Sub-system Architecture” Estimation



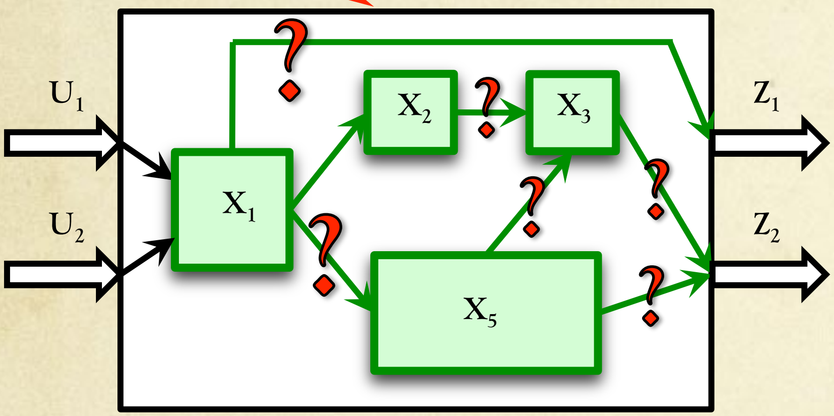
Assuming Only that the Parameterizations are “Monotonic”

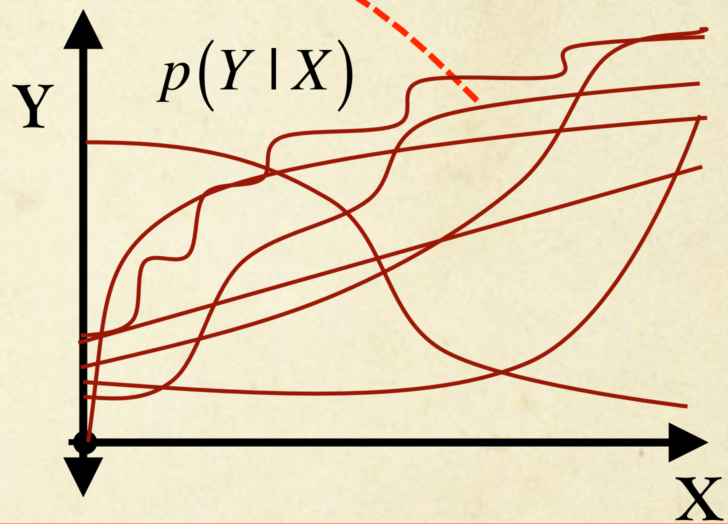
Grey Nearing

Last Six-Months

Shervan Gharari (TU Delft)

"Sub-system Architecture" Estimation





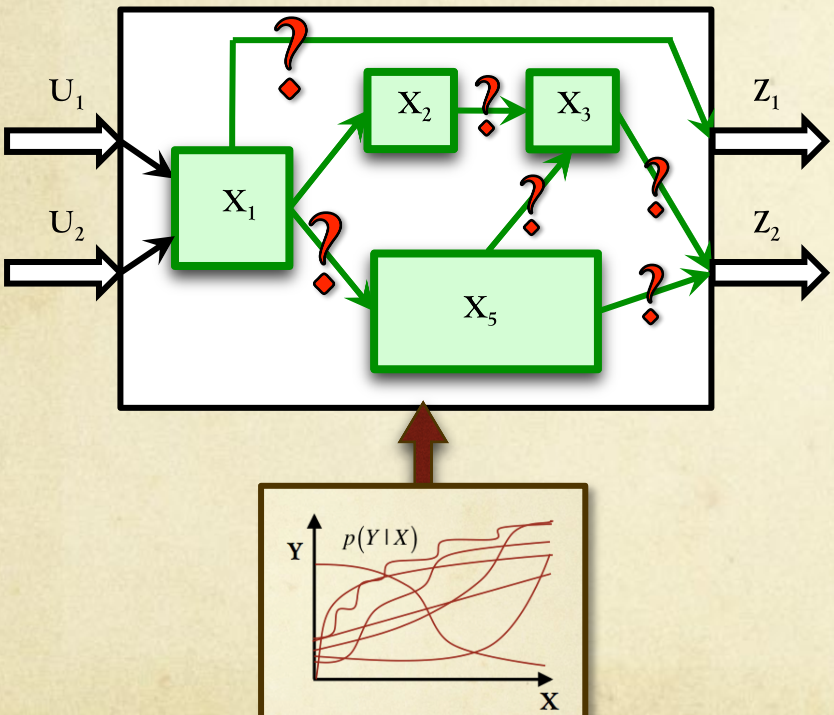
And Randomly Generated

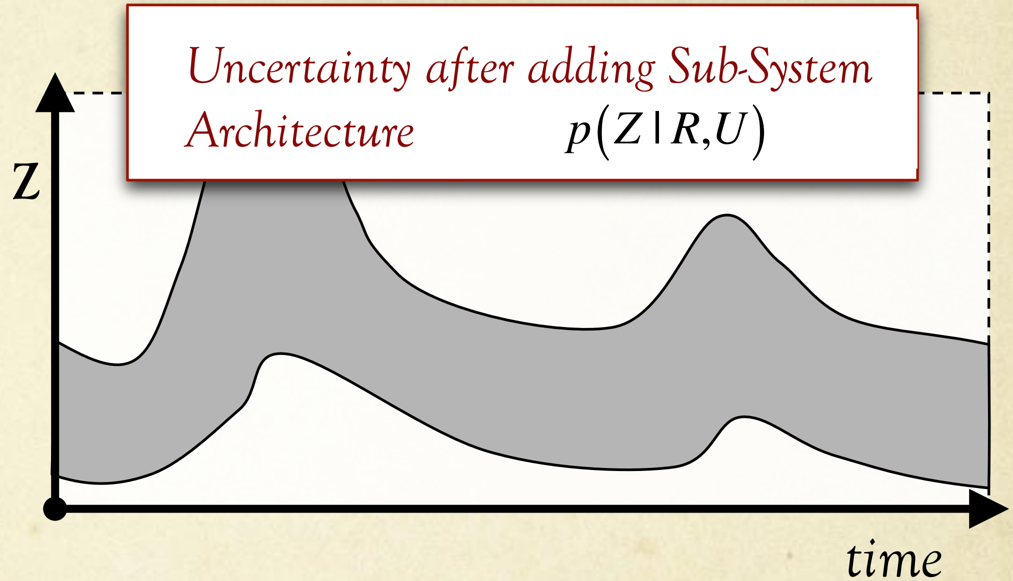
Grey Nearing

Last Six-Months

Shervan Gharari (TU Delft)

"Sub-system Architecture" Estimation





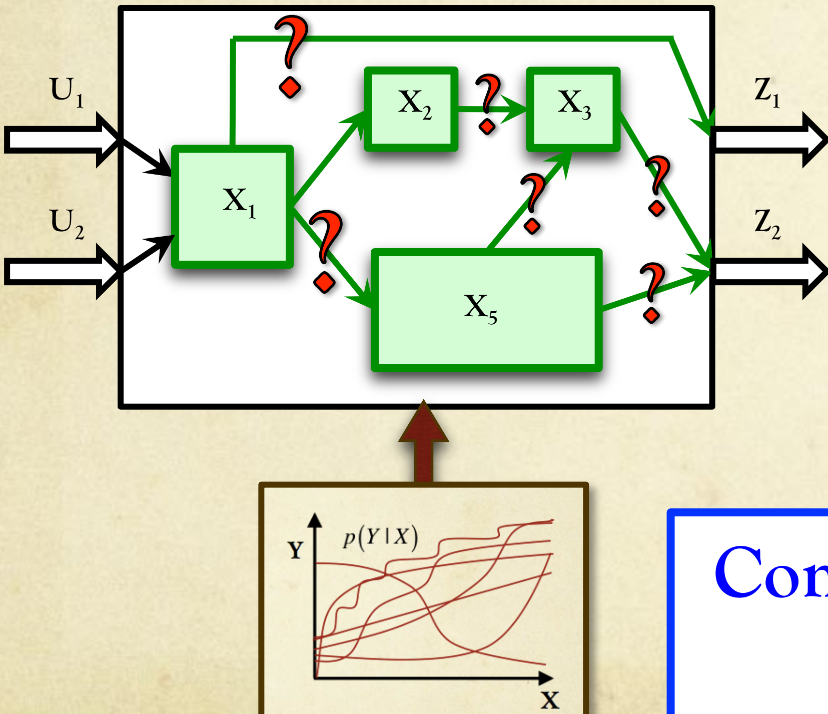
Run 10,000+ random cases

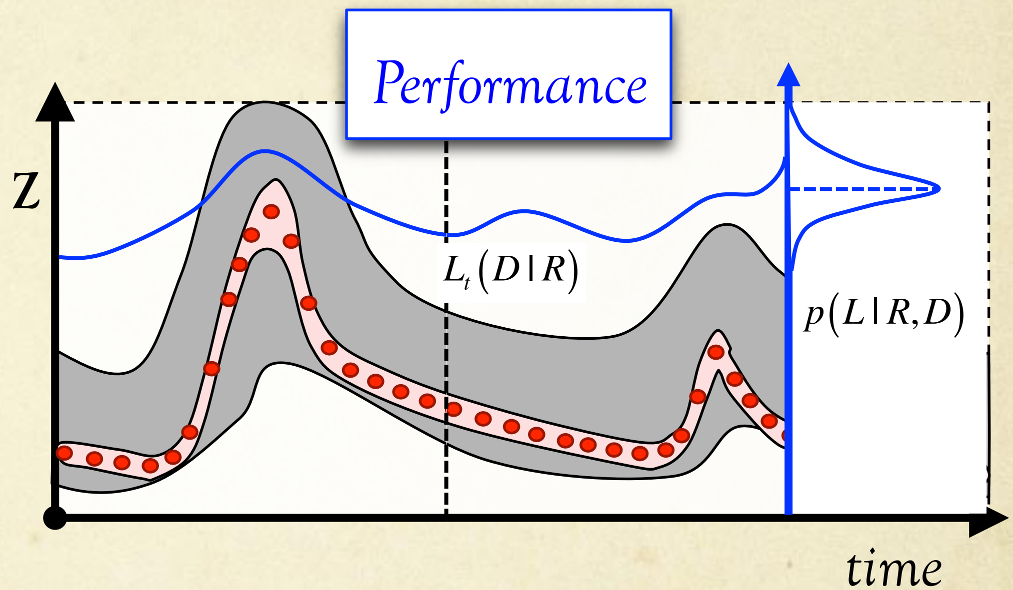
Grey Nearing

Last Six-Months

Shervan Gharari (TU Delft)

“Sub-system Architecture” Estimation





Run 10,000+ random c

Compute the Likelihood of the Data given the Model Ensemble

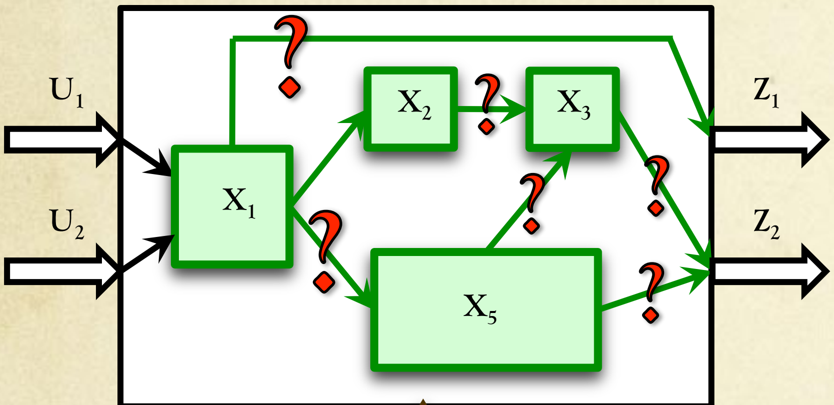
$$L(D|R)$$

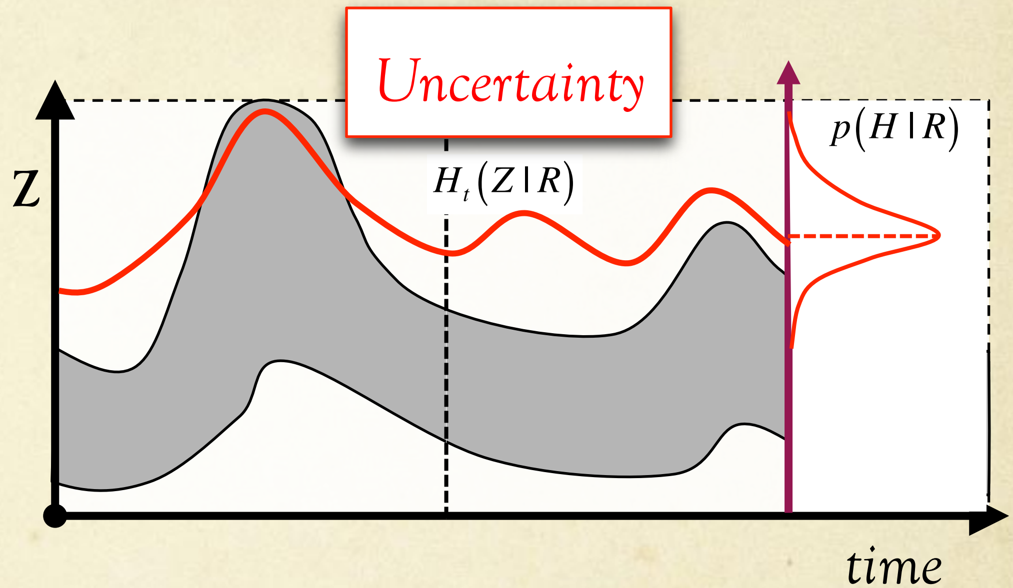
Grey Nearing

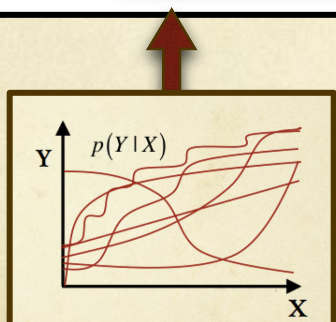
Last Six-Months

Shervan Gharari (TU Delft)

"Sub-system Architecture" Estimation







Run 10,000+ random

Compute the Entropy of the Simulations generated by the Model Ensemble

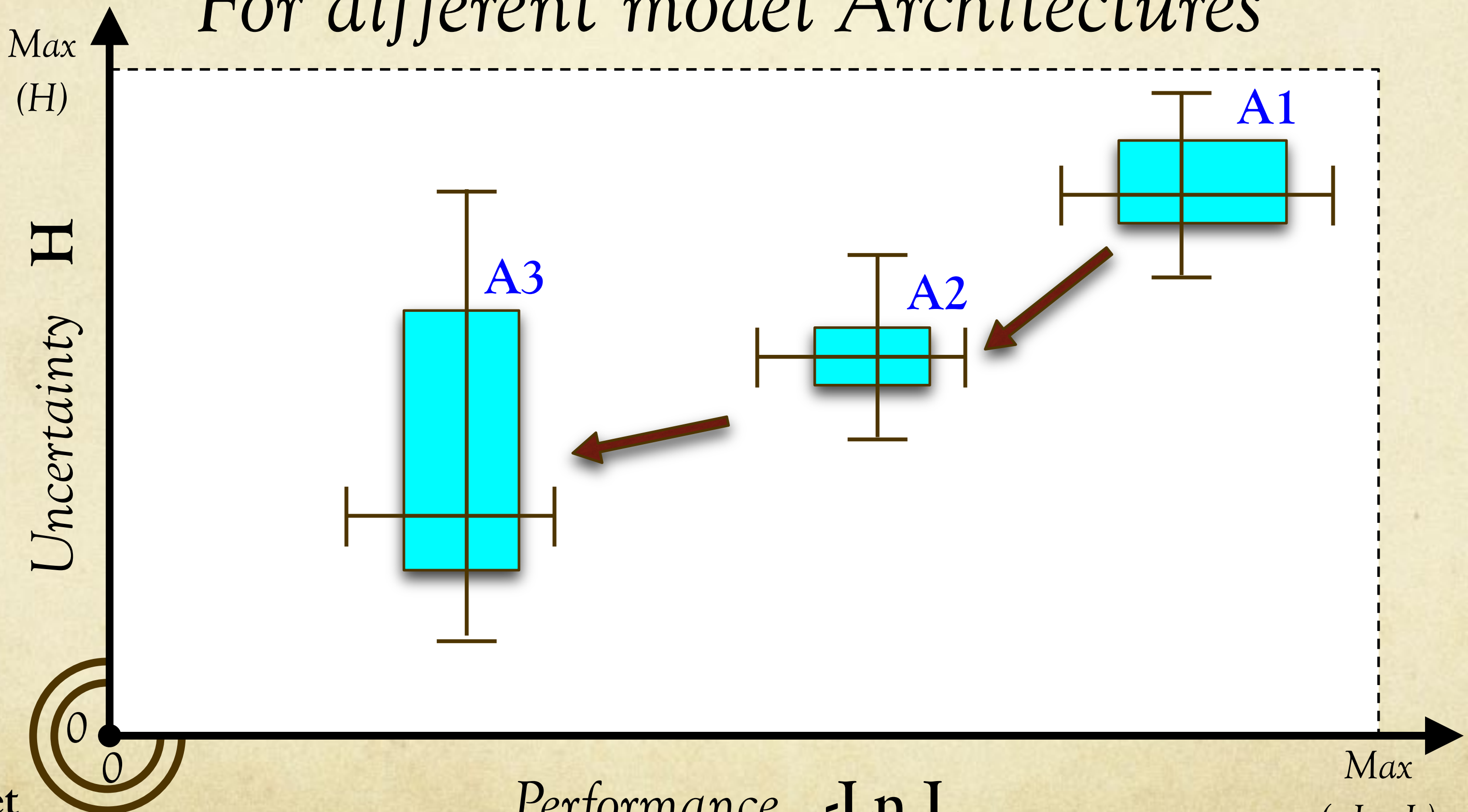
$$H(Z\,|\,R)$$

Grey Nearing

Last Six-Months

Shervan Gharari (TU Delft)

Bootstrap & Plot the Results For different model Architectures



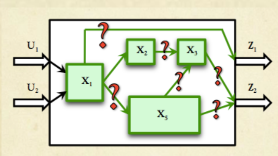
Grey Nearing

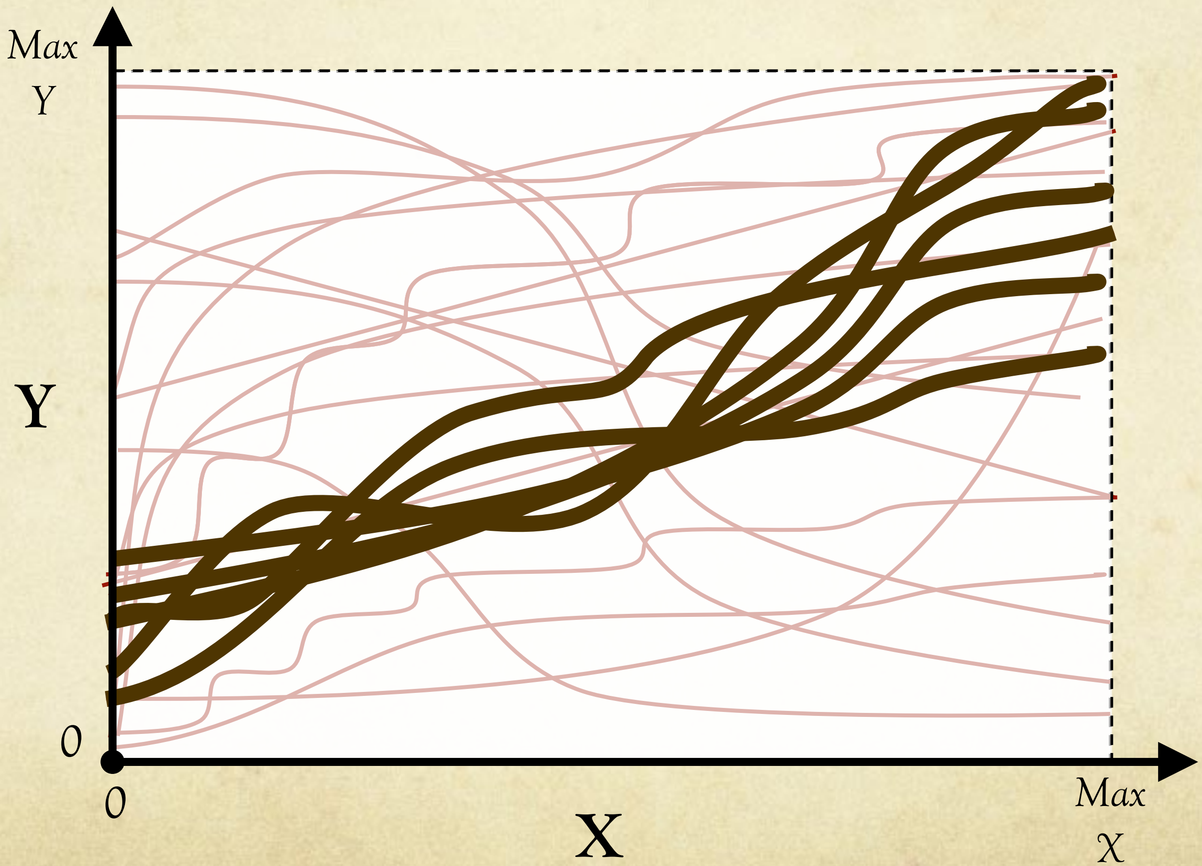
Last Six-Months

Shervan Gharari (TU Delft)

Examine the Parameterizations with Highest Performance

(conditional on a selected Architecture)





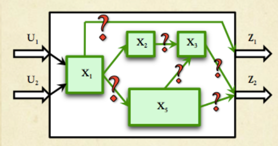
Grey Nearing

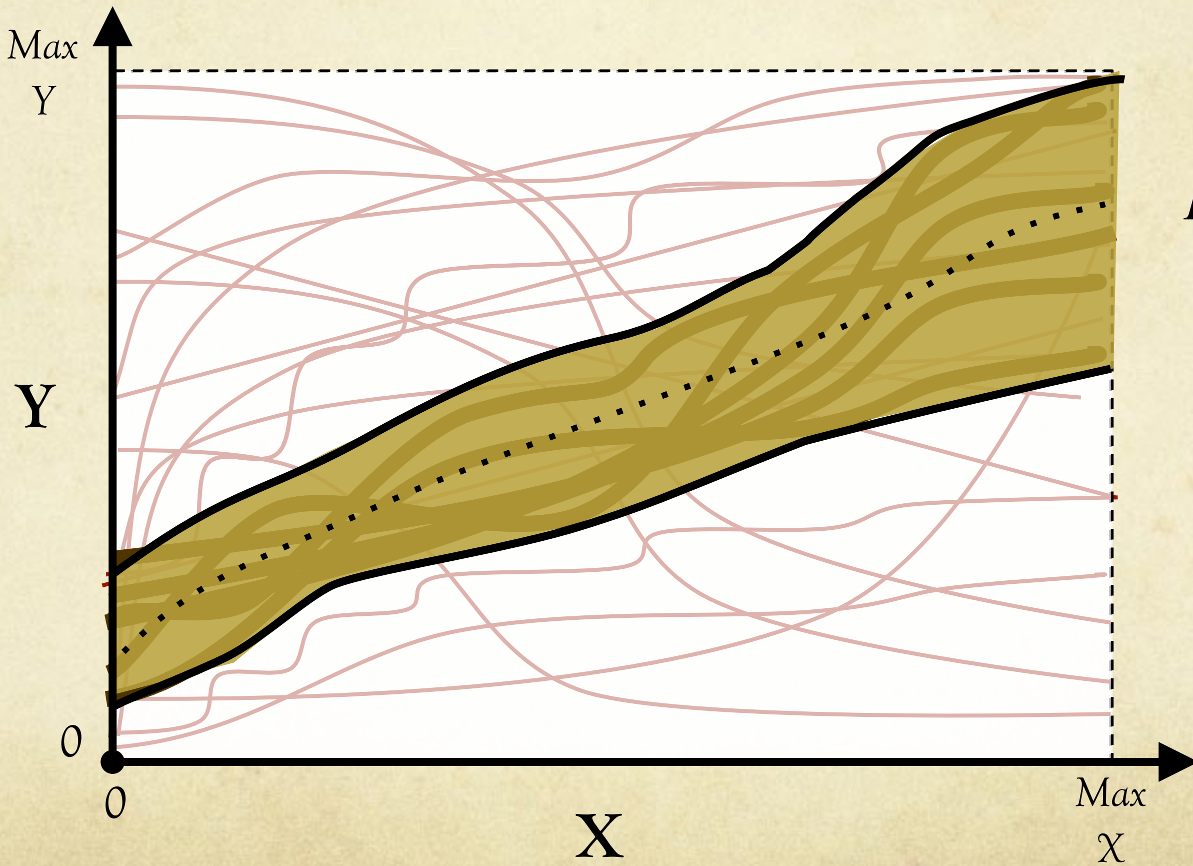
Last Six-Months

Shervan Gharari (TU Delft)

Propose Parameterization Equations

(conditional on a selected Architecture)





$R : p(Y \mid X)$

Mapping

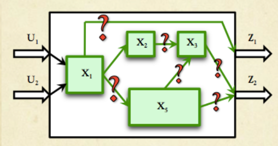
Grey Nearing

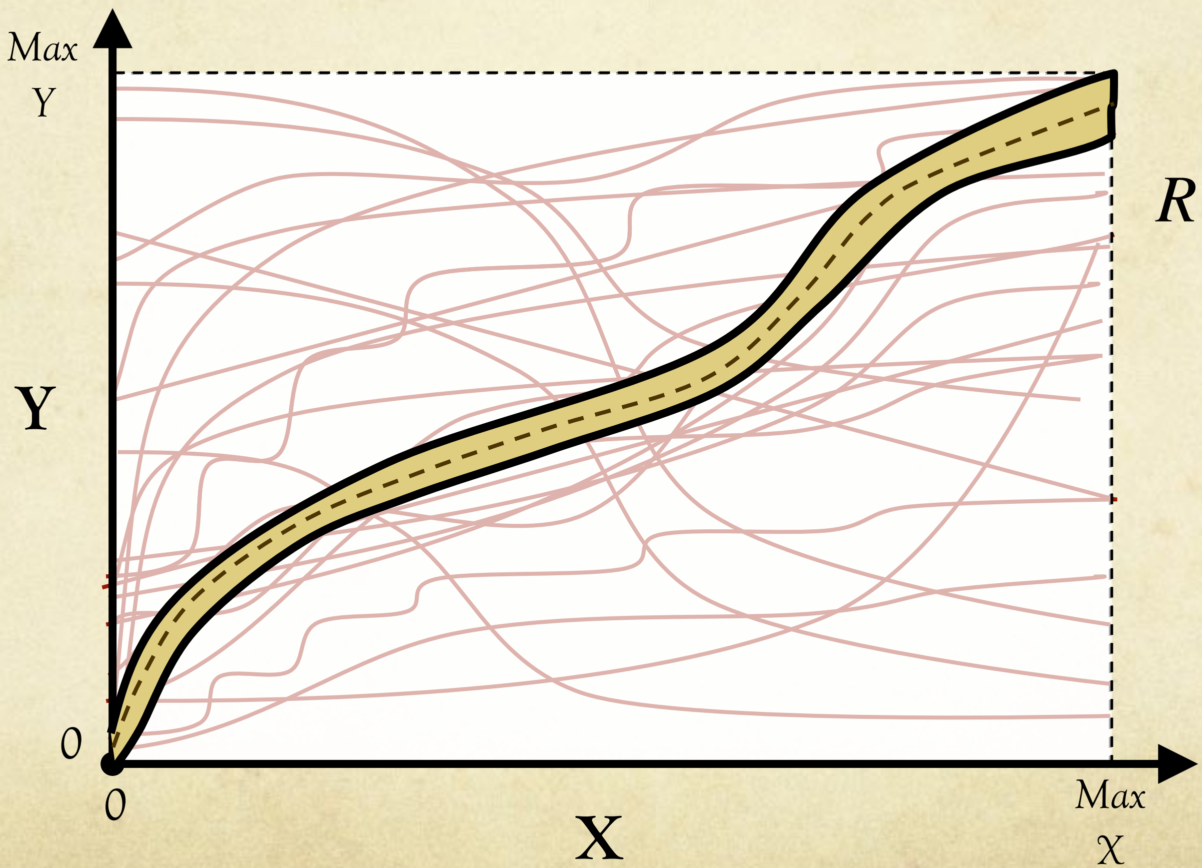
Last Six-Months

Shervan Gharari (TU Delft)

Proceed with Parameter Estimation

(conditional on a selected Architecture & Parameterization Form)





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

Mapping

Grey Nearing

The Result

Shervan Gharari (TU Delft)

Ability to investigate Process Models (Sub-system Architectures)

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

In Principle a similar approach could be used to investigate value of different Conservation Laws (System Diagram)

Grey
Nearing

For Details
please see

*Shervan
Gharari
(TU Delft)*

PICO Session HS1.5/GI1.9

Data & Models ...: Towards a common framework for model building and predictions in the Geosciences

Abstract # EGU2014-4441

Progressive evaluation of incorporating information into a model building process

Gharari et al

Grey Nearing

In Preparation

Shervan Gharari (TU Delft)

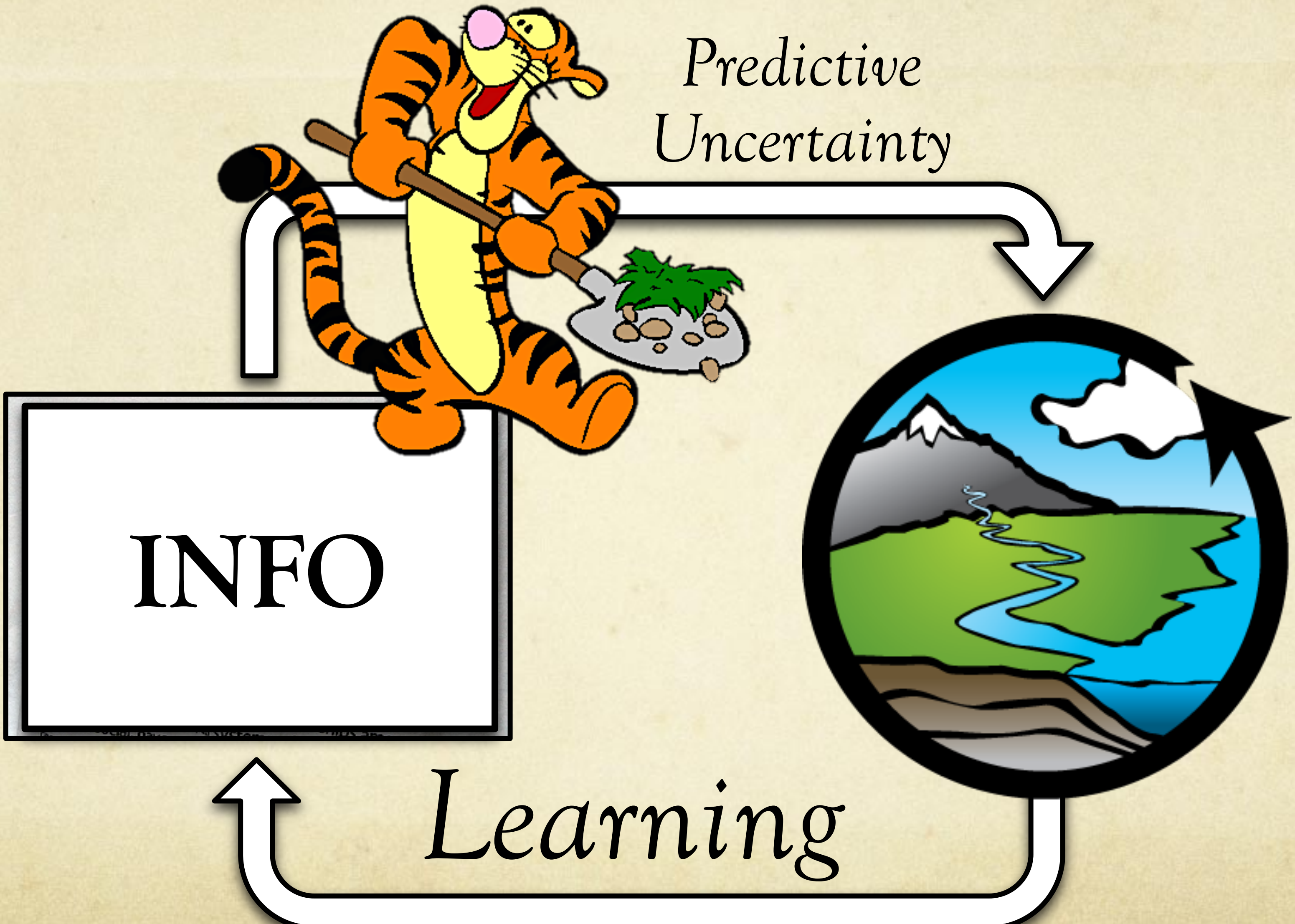
Gupta, Gharari, Nearing (In Prep)

Towards Improved Model Structural Inference: Theory & Methods

Gharari, Gupta et al. (In Prep)

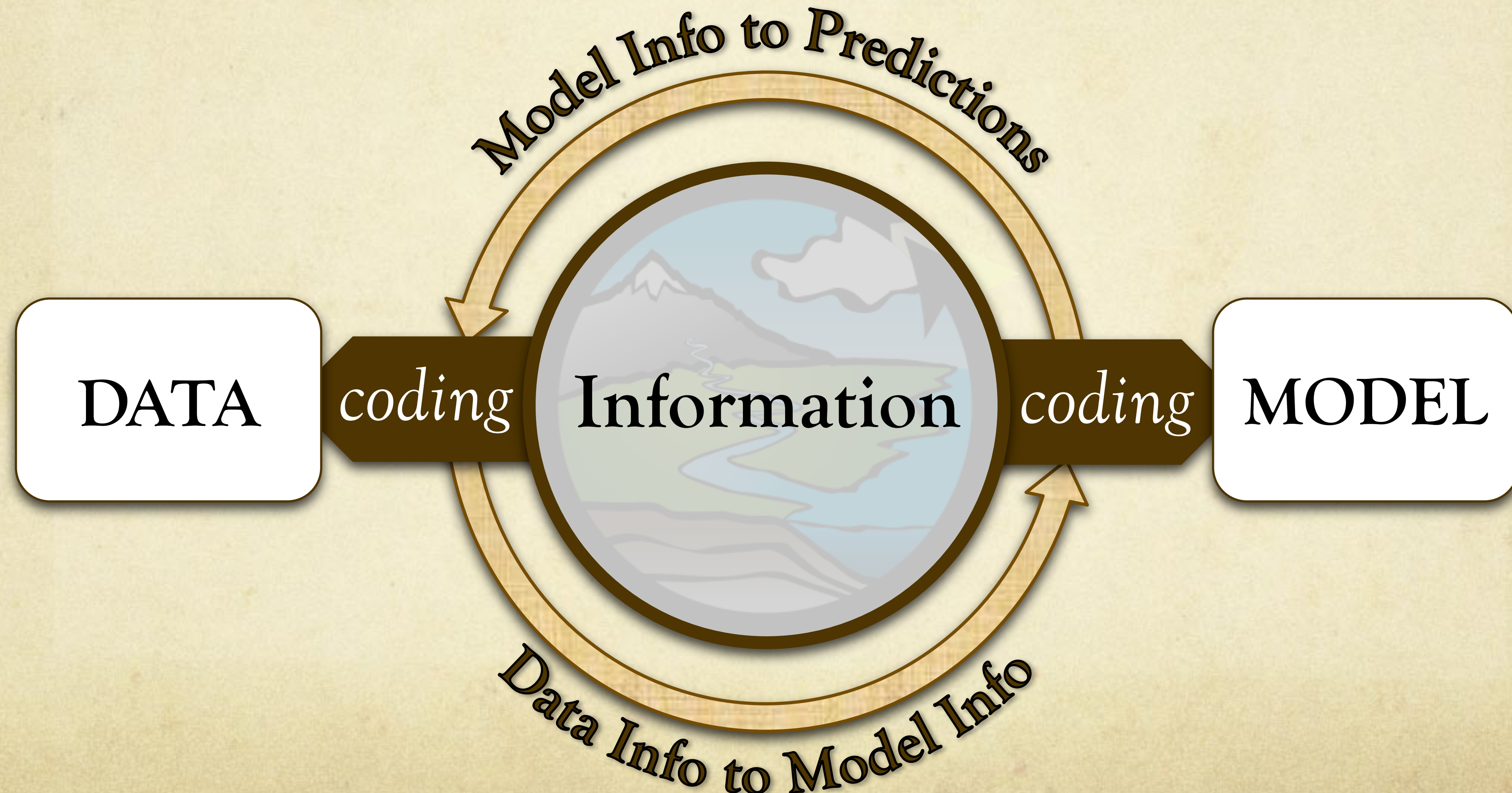
Towards Improved Model Structural Inference: Application to Hydrologic Modeling

To Conclude

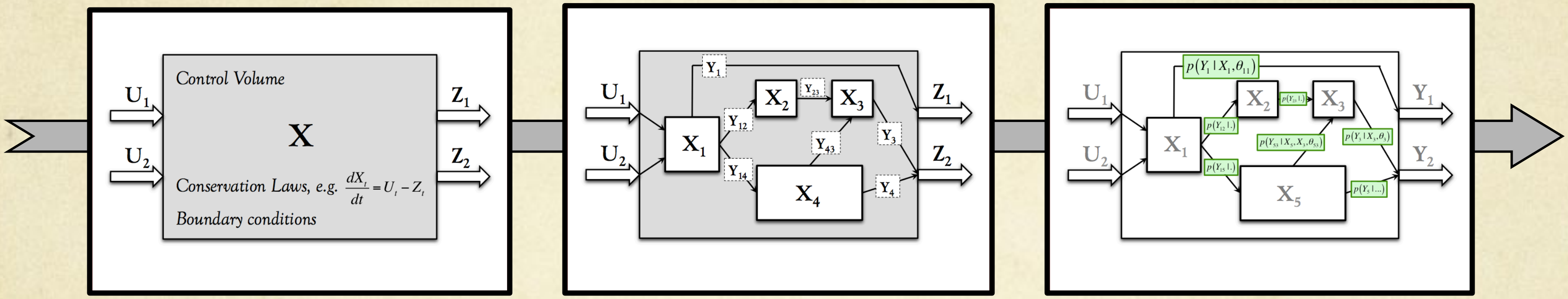


"Information" comes from 3 Sources

Data, Models, Inference



Its Useful to Consider How Info is Coded in Models



1. **System Diagram** *(Conservation Law Hypothesis)*
2. **Sub-system Architecture** *(Process Model Hypothesis)*
3. **Parameterization** *(Process Equations Hypothesis)*
4. *Computational Implementation*
5. *Inference*

Systems Theory can Help Maximize *Effectiveness* of "Learning"

(about dynamical systems)

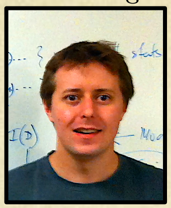
1. *How to Explicitly Decode Info in Data (Diagnostics)*

2. *How Info is encoded in Models (Hypotheses/Assumptions)*

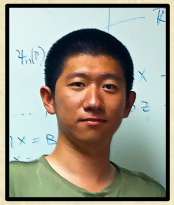
3. *How to Learn from the Model-Data encounter (Inference)*

Beth Jackson

Grey Nearing



Gong Wei



Gab Abramowitz

Martyn Clark

Jasper Vrugt

Uwe Ehret

Steven Weijs

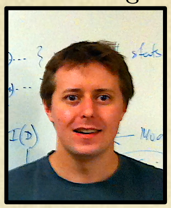
Shervan Gharari

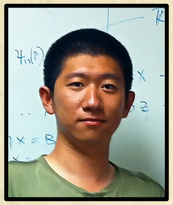
Some of the people who help to keep me somewhat close to "reality"

Anyone who make a contribution to any field of endeavor, and stays in that field long enough ..... becomes an obstruction to progress

Jones Conservation Law (Blosch 2003)

Beth Jackson


Grey Nearing


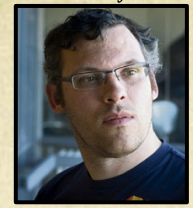
Gong Wei


Gab Abramowitz


Martyn Clark


Jasper Vrugt


Uwe Ehret


Steven Weijs


Shervan Gharari


Some of the people who help to keep me somewhat close to "reality"

THANK YOU

Anyone who make a contribution to any field of endeavor, and stays in that field long enough ….. becomes an obstruction to progress

Jones Conservation Law (Blosch 2003)