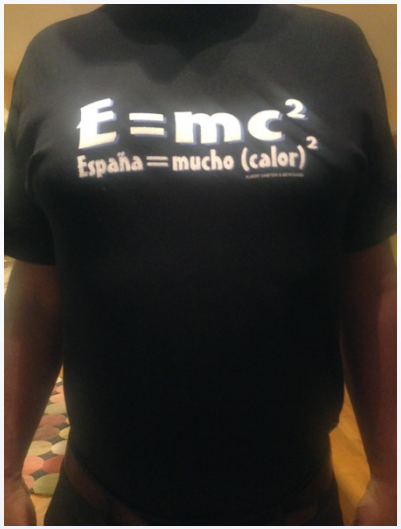
BIG Data for BIG Floods:

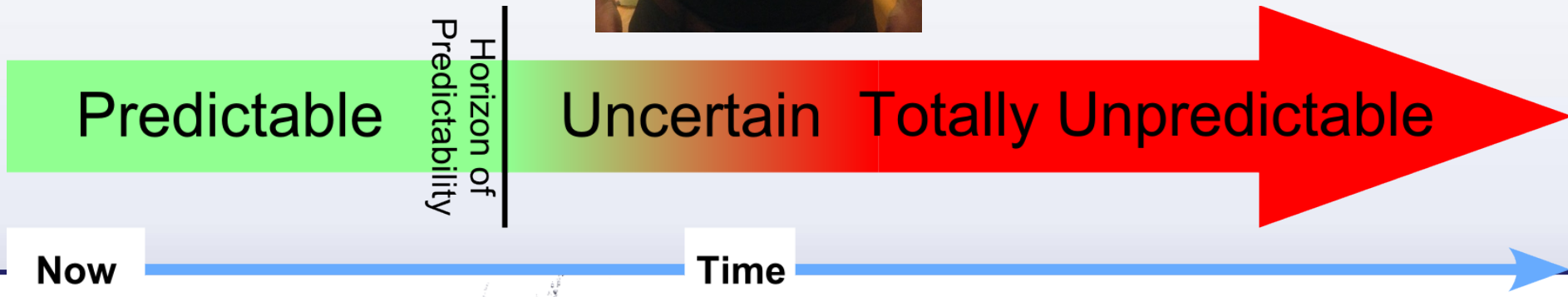
Can we BREAK the Predictability Limits?

Shafiqul (@ShafikIslam) Islam

School of Engineering & Fletcher School of Law & Diplomacy

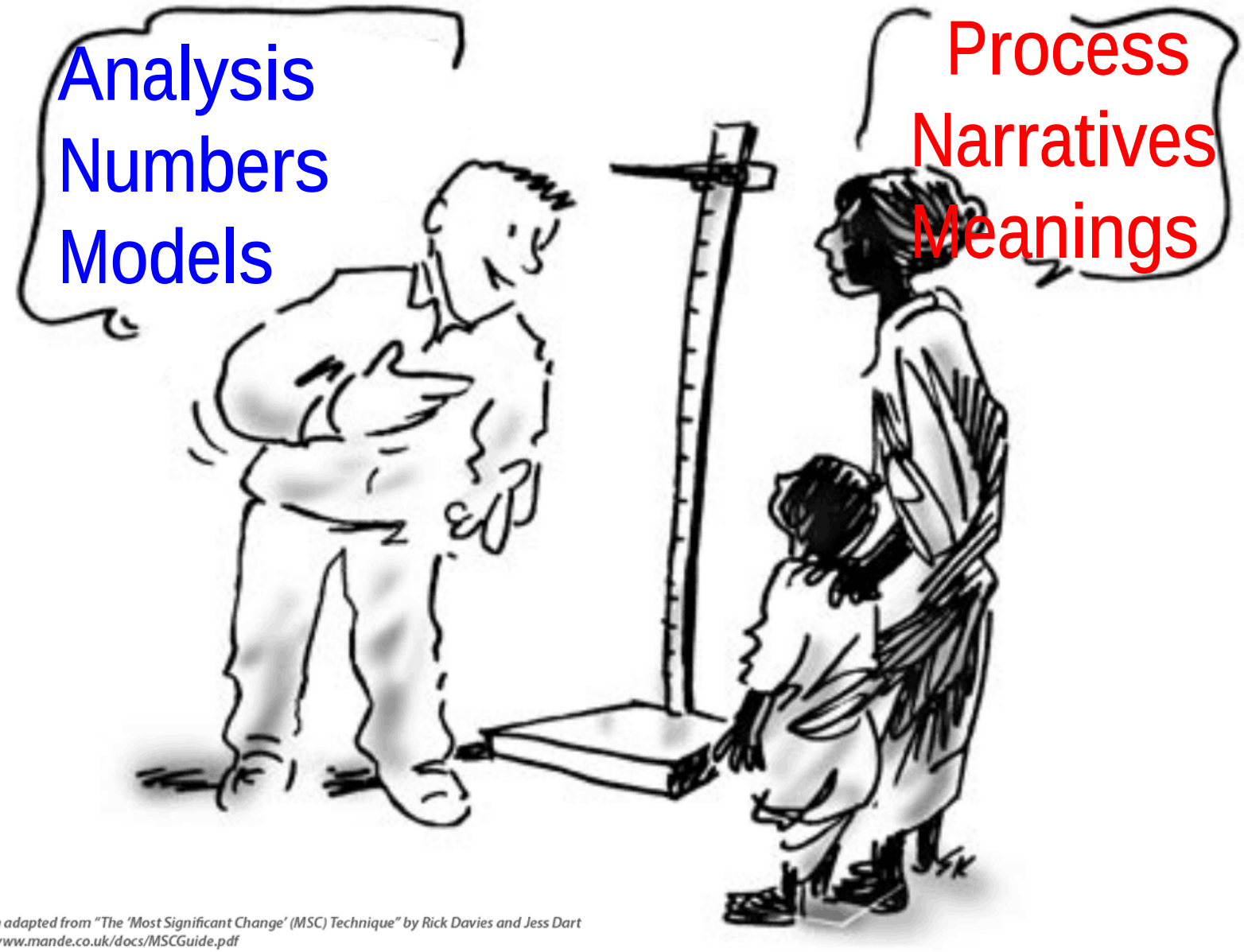
MN Workshop on Understanding Climate Change from Data 04 August 2014





WATER DIPLOMACY

Understanding Climate Change from Data



cartoon adapted from "The 'Most Significant Change' (MSC) Technique" by Rick Davies and Jess Dart
http://www.mande.co.uk/docs/MSCGuide.pdf

Predictability of Weather: Where is the Disconnect?

- *"Hitherto the long range forecaster has been denied a seat in the banquet hall of science; ...the general scientist has denounced him; the professional weather man has treated him with supercilious scorn"* (Ricard 1911).

- This is a snapshot of ideas and challenges related to long range weather forecasting 104 years ago when computers did not exist; and science of prediction was limited to reading the weather maps.

- Then, came the era of quantitative weather prediction with the vision of three brilliant scientists – Bjerknes (1862 -1951), Richardson (1881-1953), and von Neumann (1903-1957) – and the problem of weather forecasting was stated mathematically as an initial value problem by pulling together the Newtonian dynamical equations, the conservation of mass, the laws of classical thermodynamics and interactions.

- The system of equations described by Bjerknes represented the complexity of atmospheric processes in a highly simplified form; yet, the system was analytically intractable (Tribbia and Anthes 1987).

A Fundamental Question in Predictability

A Dynamical System is :

TYPE 1 – characterized by an infinite range of predictability

TYPE 2 – the range of predictability is *finite*, but can be *increased indefinitely* by *decreasing the size of the initial error*

TYPE 3 – the range of predictability is *finite and intrinsically limited*

Does the Weather or Climate Constitute a Type 2 or a Type 3 System?

Evolution of One-Day Forecast Error, Lorenz Error Growth, and Forecast Skill for ECMWF Model

(500 hPa NH Winter)

	1982	1987	1992	1997	2002
"Initial error (1-day forecast error) (m)	20	15	14	14	8
Doubling time (days)	1.9	1.6	1.5	1.5	1.2
Forecast skill (day 5 ACC)	0.65	0.72	0.75	0.78	0.84

Predictability: Where is the Disconnect?

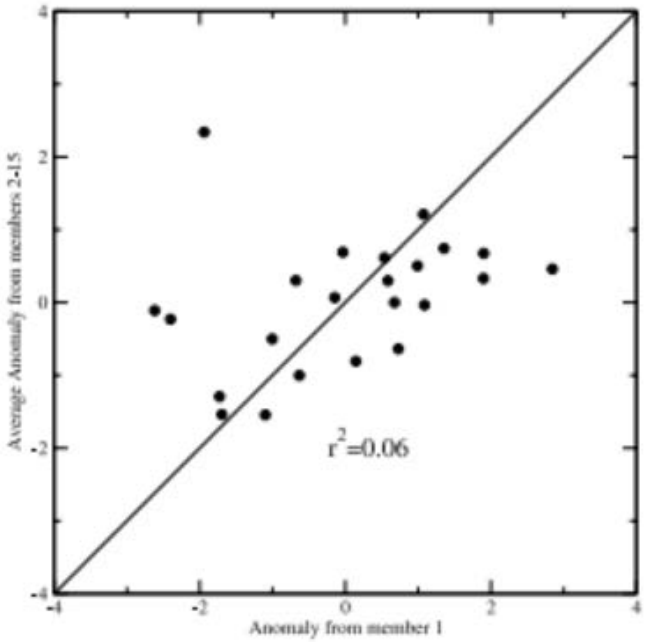


Figure 1. Scatter plot for the idealized predictability analysis, showing the degree to which the Climate Forecast System can "predict itself" at a central U.S. grid cell (37.5°N, 97.5°W). The lead-time for this prediction is 16 days, and the anomaly is from a 32-day running mean. The x-axis represents the precipitation anomaly generated by the first members of all May forecasts, and the y-axis represents the forecasted precipitation anomaly averaged over the remaining fourteen ensemble members from the same ensemble set. Twenty-four points are from the twenty-four years (1981–2004). The solid line is the 1:1 line.

Luo & Wood 2006

Lorenz (1963) discovered that even with a perfect model and almost perfect initial conditions the forecast loses all skill in a finite time interval: “A butterfly in Brazil can change the forecast in Texas after one or two weeks”.

Definition of Deterministic Chaos
(Lorenz, March 2006, 89)

WHEN THE PRESENT DETERMINES

THE FUTURE

BUT

THE APPROXIMATE PRESENT DOES NOT

APPROXIMATELY DETERMINE THE FUTURE

Chaos: Where is the Bottleneck for Flood Forecasting?

- Chaos: **Sensitivity to initial conditions** – causing large divergence in the prediction. But this divergence is not infinite, it oscillates within bounds.
- Discovered by Ed Lorenz in Weather Modeling in 1963

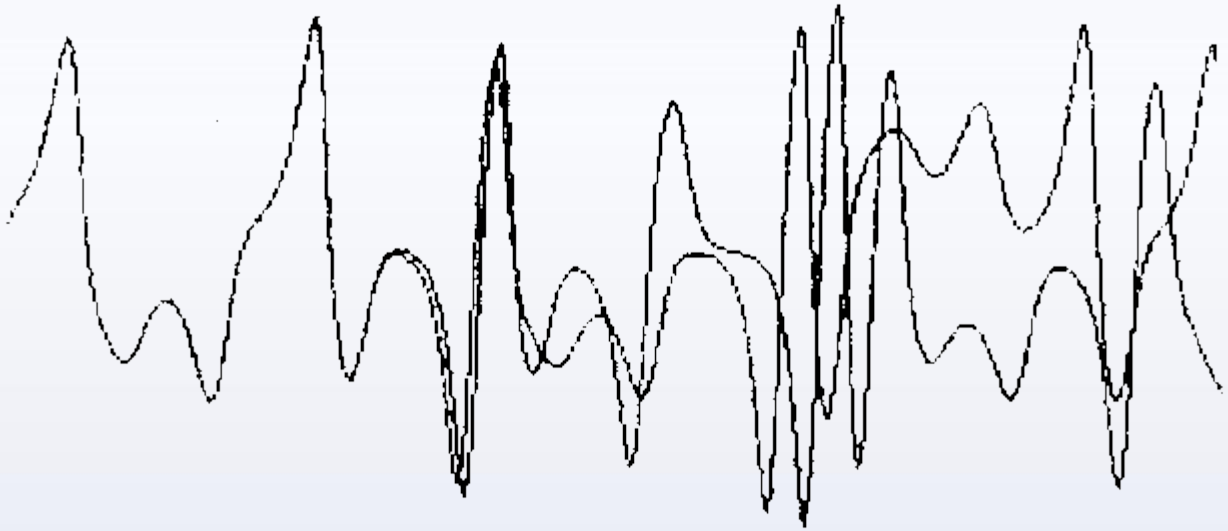
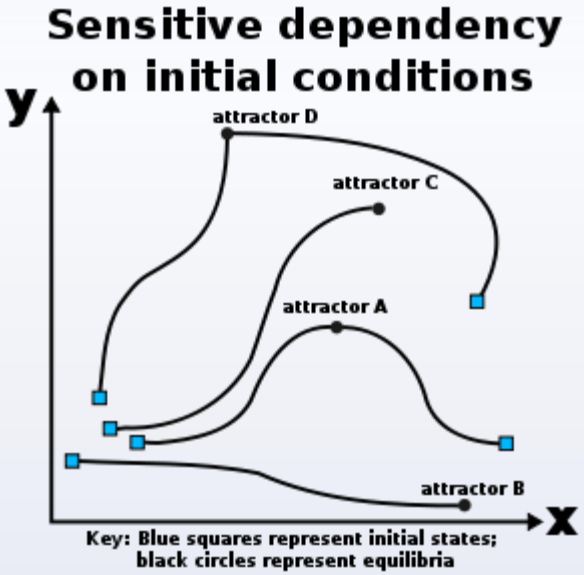


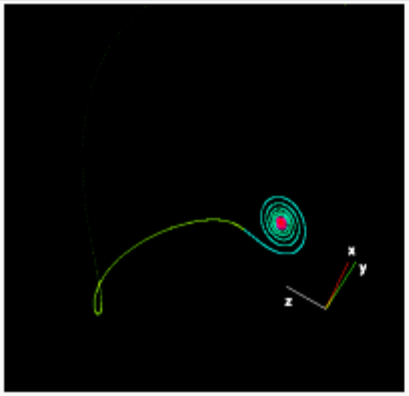
Figure 1: Lorenz's experiment: the difference between the start of these curves is only .000127. (Ian Stewart, Does God Play Dice? The Mathematics of Chaos, pg. 141)



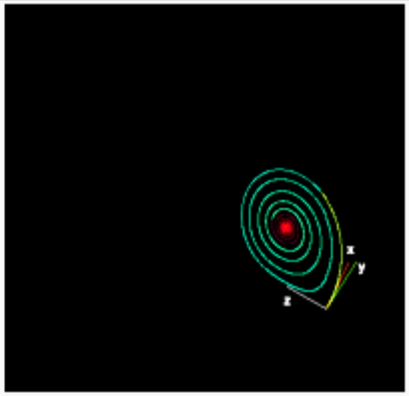
Strange Attractor: Lorenz 3-Equations of Convection

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \sigma(y - x),$$
$$\frac{\mathrm{d}y}{\mathrm{d}t} = x(\rho - z) - y,$$
$$\frac{\mathrm{d}z}{\mathrm{d}t} = xy - \beta z.$$

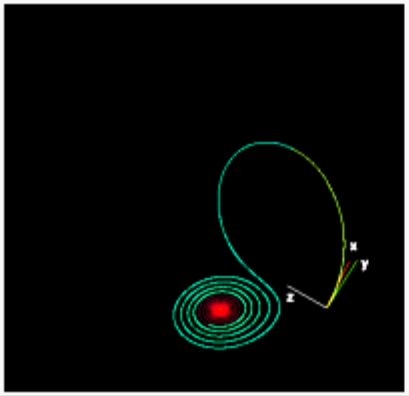
Example solutions of the Lorenz system for different values of ρ



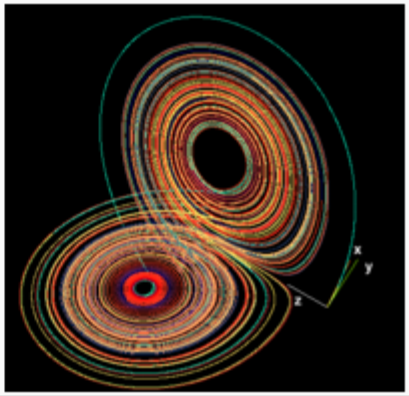
ρ=14, σ=10, β=8/3 (Enlarge)



ρ=13, σ=10, β=8/3 (Enlarge)



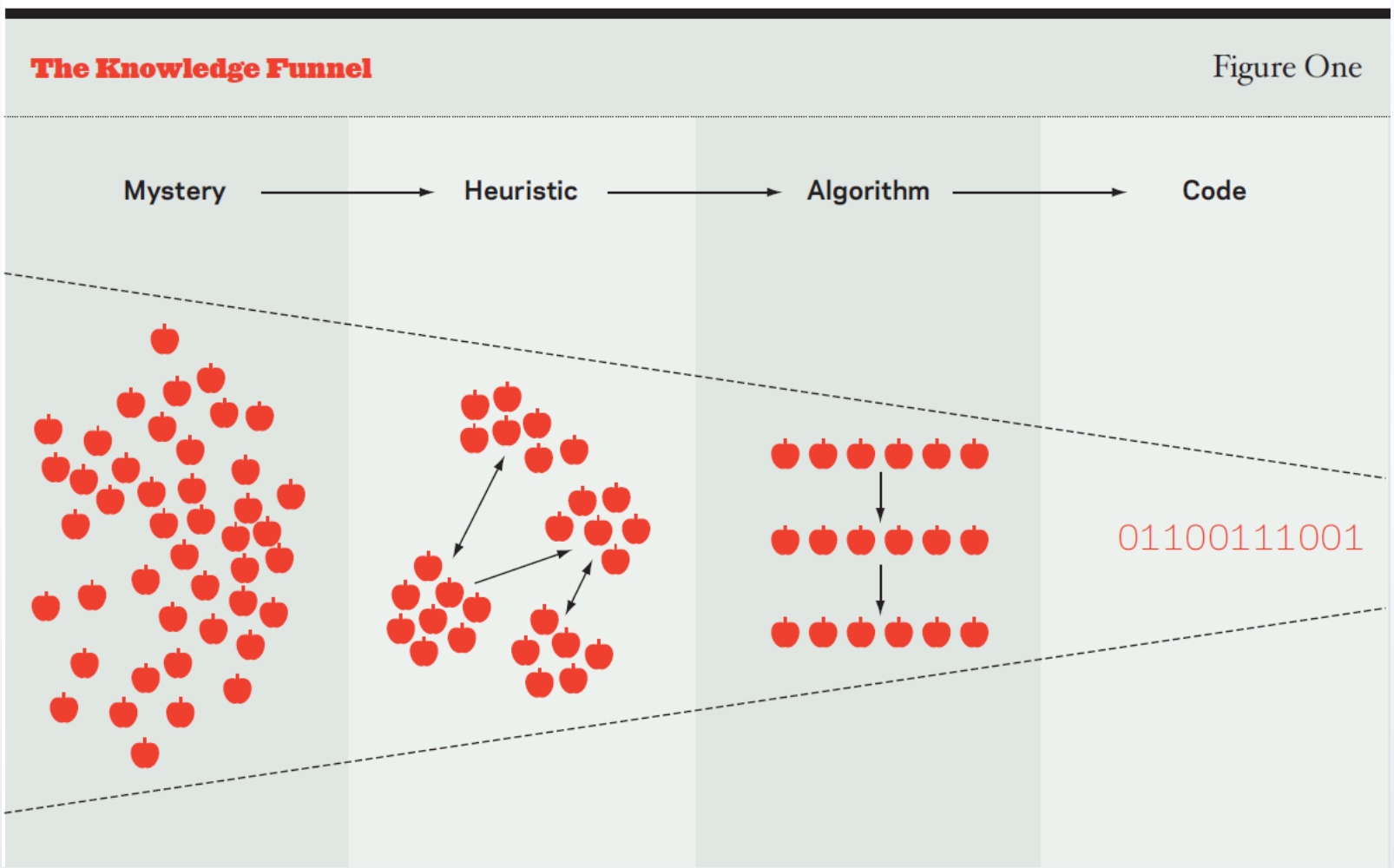
ρ=15, σ=10, β=8/3 (Enlarge)



ρ=28, σ=10, β=8/3 (Enlarge)

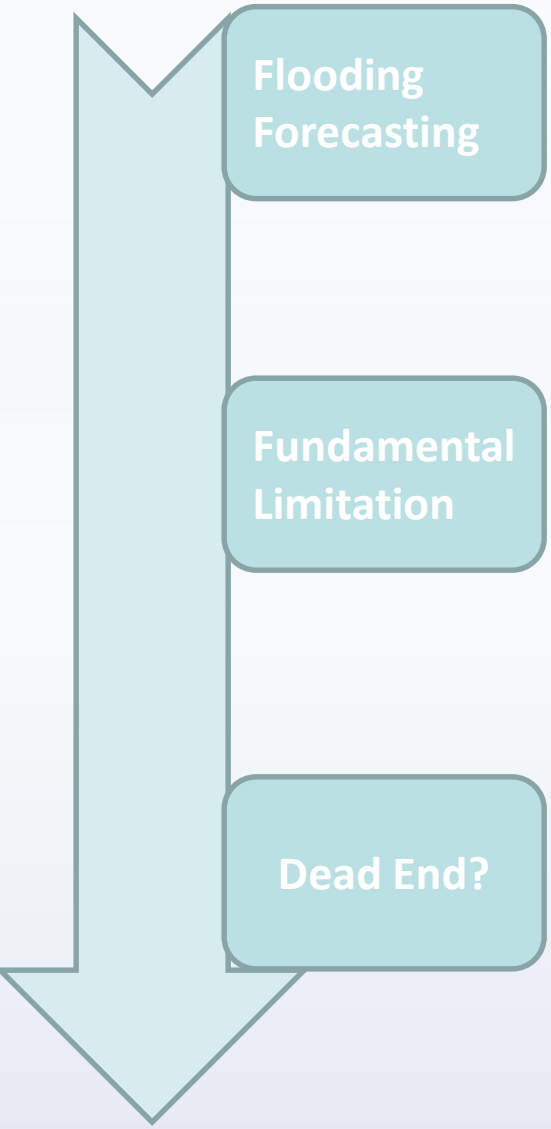
For small values of ρ, the system is stable and evolves to one of two fixed point attractors. When ρ is larger than 24.74, the fixed points become repulsors and the trajectory is repelled by them in a very complex way.

Mystery to Code: Where is the Bottleneck for Flood Forecasting?



Martin 2009

Mystery to Code: Where is the Bottleneck for Flood Forecasting?



- **Existing operational flood forecasting systems usually rely on precipitation inputs.**

- **Fundamental limitation on the predictability of Precipitation is Sensitive Dependence on Initial Conditions**

Predictability VS Predictive Ability

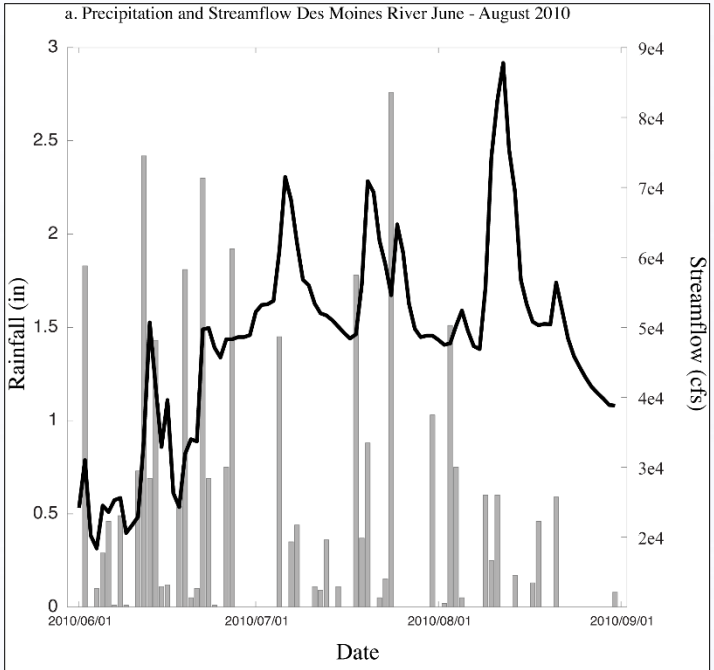


Figure 1. Precipitation (grey bars; inches; left axis) and Des Moines River discharge (dark line; cfs; right axis) for June to August 2010.

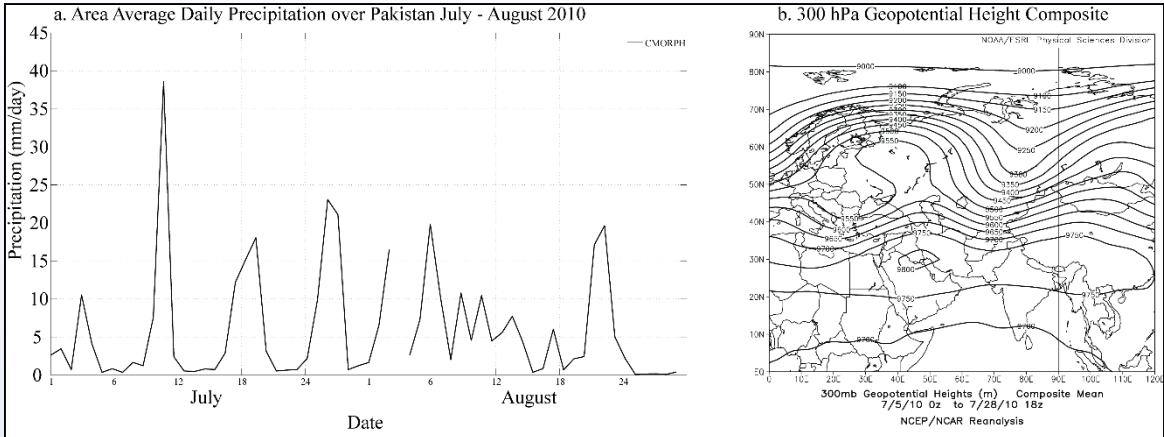
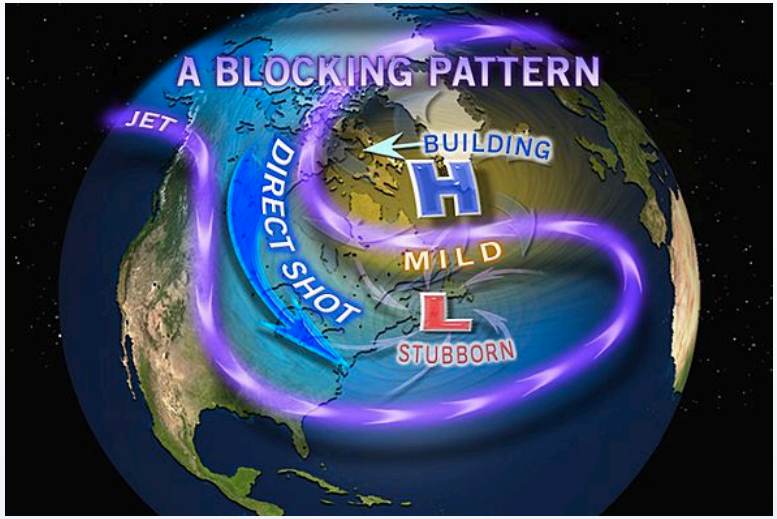


Figure 2. Panel a. is the daily average precipitation from CMORPH averaged over Pakistan, July and August 2010. Panel b. is the composite of the 300hPa geopotential height from July 5 to July 28, 2010 taken from NCEP/NCAR Reanalysis.

- **Atmospheric regimes are usually global events.**
  - **Covering large geographical areas**
  - **Lasting up to months**



"A major contributing factor to the major winter storms" **in the U.S in Spring, 2013 is** ***" blocking in the atmosphere".***
(http://rapidrestoration.co)

- **Number of features contributing to the regimes, from the Cartesian product between the spatial and temporal domains, is enormous.**

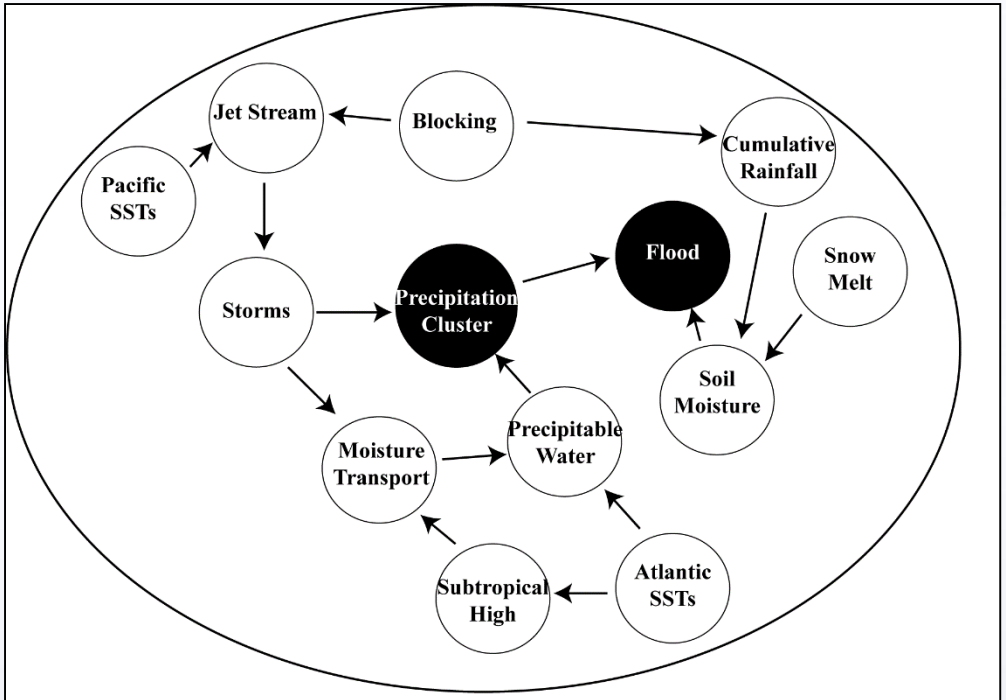
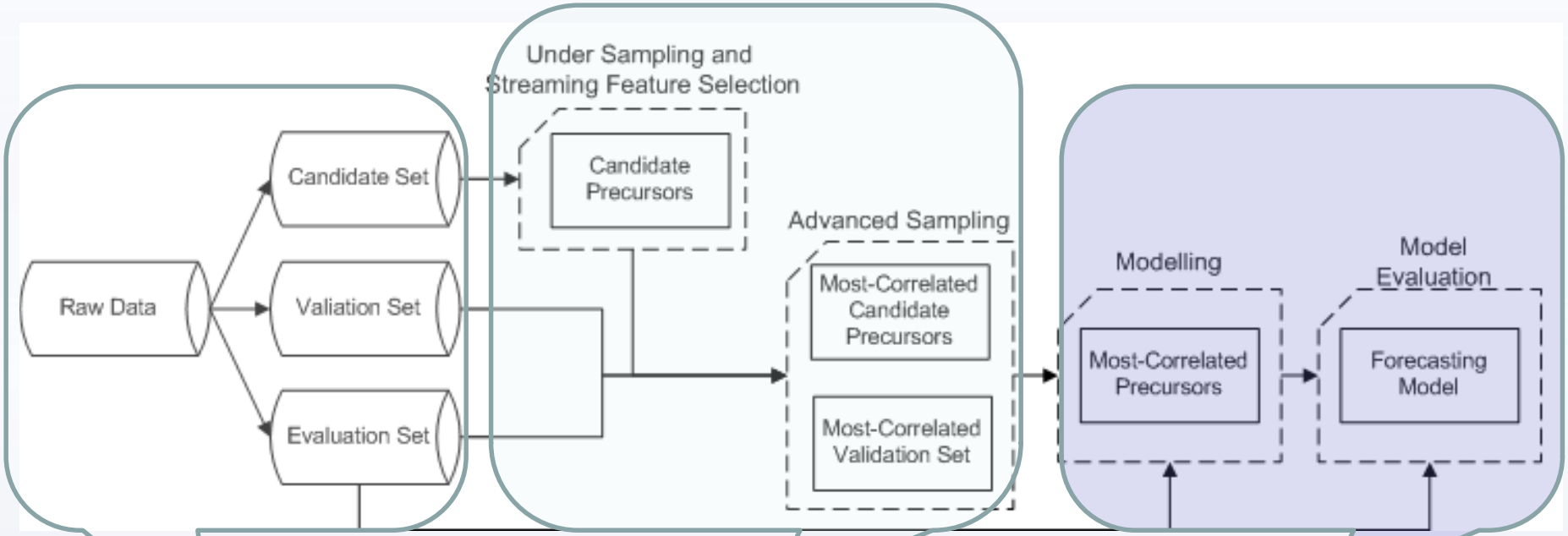


Figure 4. Hypothetical Bayesian network for the problem of flood prediction.

Mystery to Code: Think Differently and Model Differently



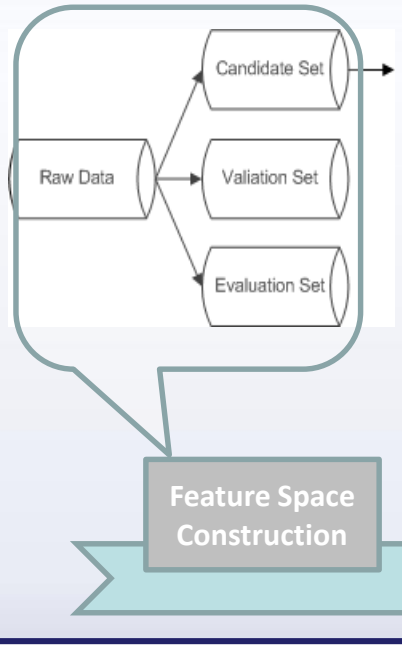
Feature Space Construction

Identifying precursors through Online Stream Feature Selection and Hierarchical Re-sampling

Building and evaluating the Forecasting Model

WATER DIPLOMACY

- **The precursors we are looking for are meteorological variables with certain spatial and temporal information.**



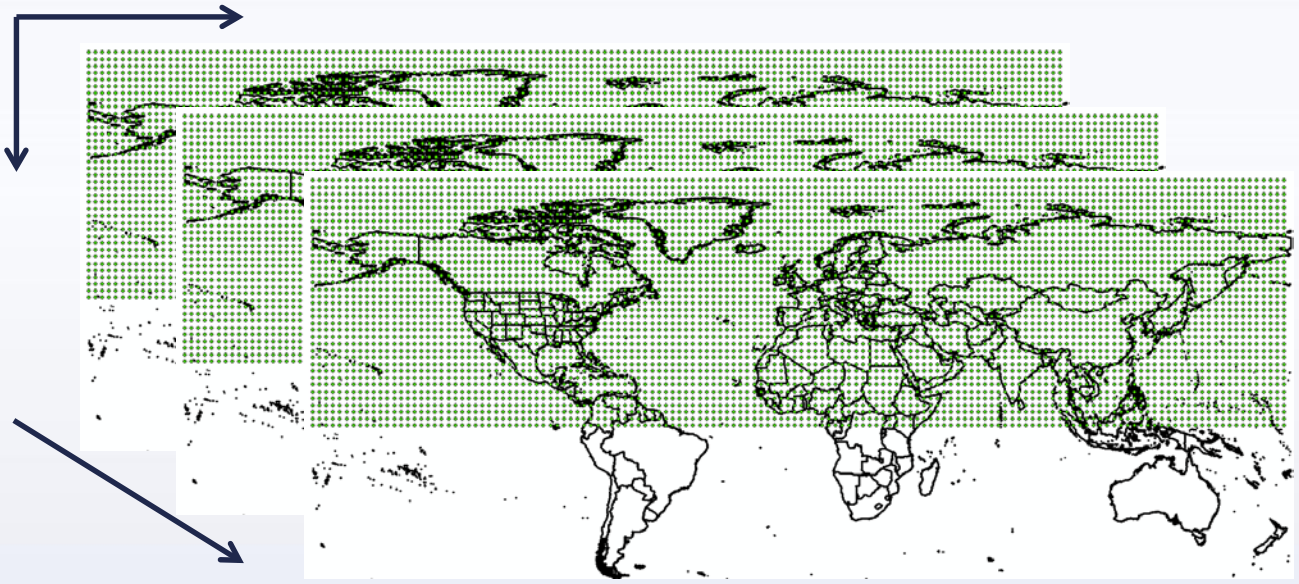
Meteorological Variables	
Z300	300hPa Geopotential Height
Z500	500hPa Geopotential Height
Z1000	1000hPa Geopotential Height
U300	300hPa Zonal Wind
V300	300hPa Meridional Wind
U850	850hPa Zonal Wind
V850	850hPa Meridional Wind
T850	850hPa Temperature
PW	Precipitable water

Mystery to Code: Feature Space Reconstruction

- Build a feature space with the **spatial and temporal** information of the predictors to achieve a comprehensive coverage of the potential precursors

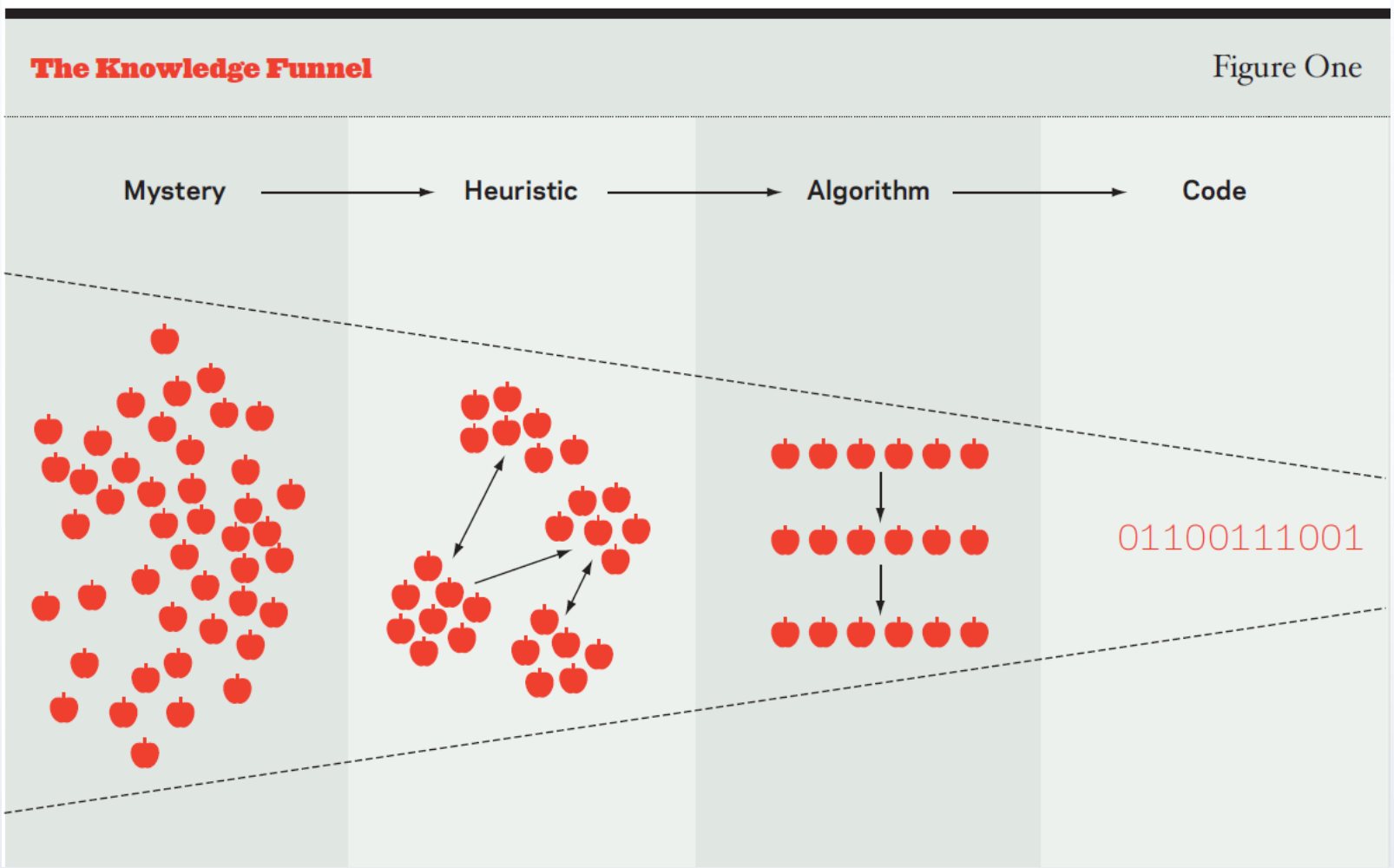
Spatial :
Sampling from 5,328 locations evenly distributed between the equator and the North pole.

Temporal :10 days



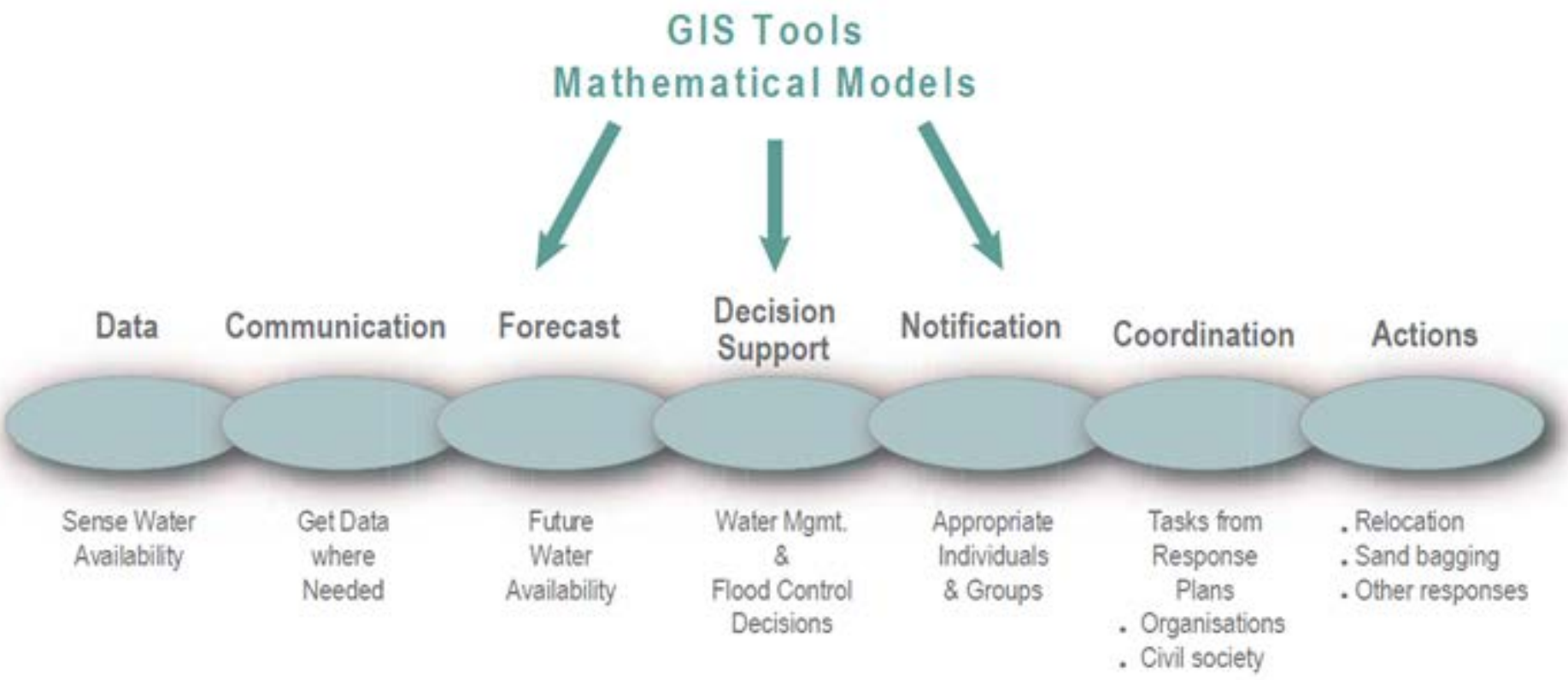
$$9\,Variabls \times 5328\,locations \times 10\,days = \mathbf{479,520\,features}$$

Mystery to Code: How to “discover” appropriate Heuristics?



Martin 2009

Flood and Water Resources Management: a Critical Chain of Events and Actions



Processes, People, and Politics are LINKED

Neither Numbers Nor Narratives:

Neither Objectivists Nor Interpretivists

Principled Pragmatist ?

WATER DIPLOMACY