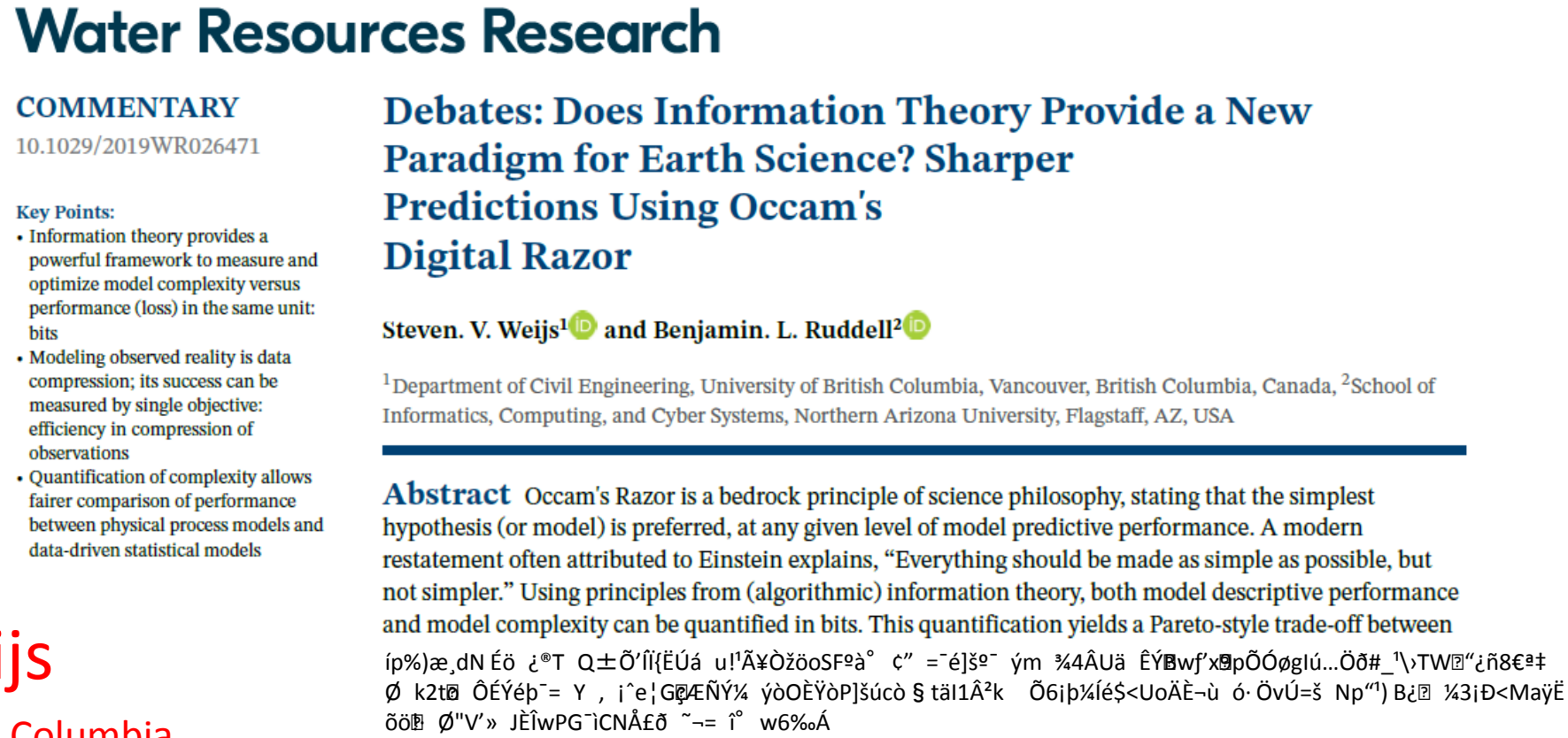
Algorithmic information theory (AIT)

Occam's Digital Razor zips the universe



Water Resources Research

COMMENTARY

10.1029/2019WR026471

Key Points:

- Information theory provides a powerful framework to measure and optimize model complexity versus performance (loss) in the same unit: bits
- Modeling observed reality is data compression; its success can be measured by single objective: efficiency in compression of observations
- Quantification of complexity allows fairer comparison of performance between physical process models and data-driven statistical models

Debates: Does Information Theory Provide a New Paradigm for Earth Science? Sharper Predictions Using Occam's Digital Razor

Steven. V. Weijs[1] and Benjamin. L. Ruddell[2]

[1]Department of Civil Engineering, University of British Columbia, Vancouver, British Columbia, Canada, [2]School of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, AZ, USA

Abstract Occam's Razor is a bedrock principle of science philosophy, stating that the simplest hypothesis (or model) is preferred, at any given level of model predictive performance. A modern restatement often attributed to Einstein explains, "Everything should be made as simple as possible, but not simpler." Using principles from (algorithmic) information theory, both model descriptive performance and model complexity can be quantified in bits. This quantification yields a Pareto-style trade-off between

íp%)æ¸dN Éö ¿®T Q±Õ'ÍÌ{ËÚá u!¹Ã¥ÒžöoSFºà° ¢" =¯é]šº¯ ým ¾4ÂUä ÊÝ₪wƒ'x₪pÕÓøgIú…Öð#_¹\›TW"¿ñ8€ª‡ Ø k2t₪ ÔÉÝéþ¯= Y , ¡^e¦G₪ÆÑÝ¼ ýòOÈŸòP]šúcò § täI1Â²k Õ6¡þ¼Íé$<UoÄÈ¬ù ó· ÖvÚ=š Np"¹) B¿₪ ¼3¡Ð<MaÿË õö₪ Ø"V'» JÈÎwPG¯ìCNÅ£ð ˜¬= î° w6‰Á

Steven Weijs

University of British Columbia

Vancouver, Canada

2020 Summer School for Information Theory in the Earth Sciences (SITES)

Opening quote WRR debate:

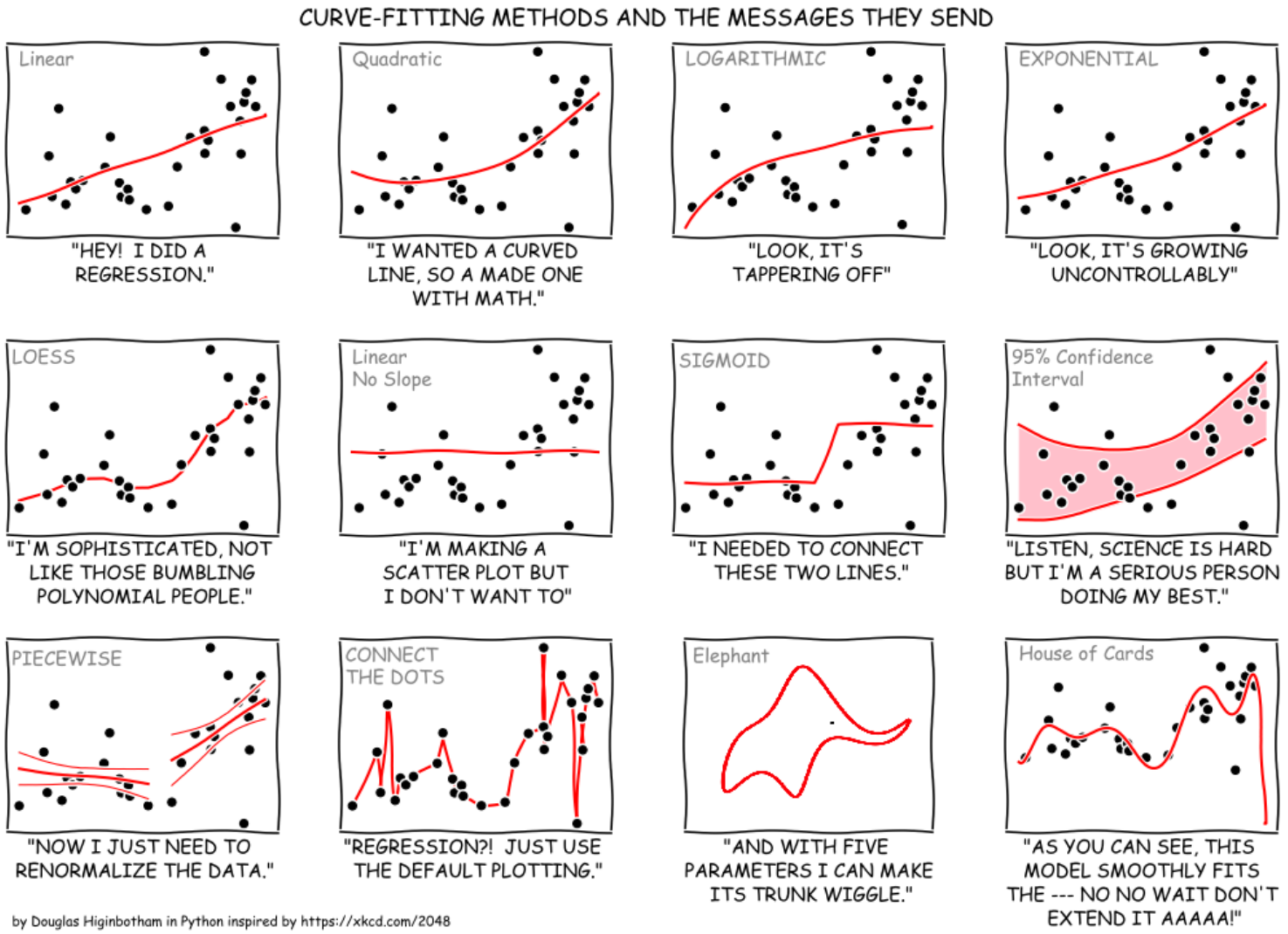
“Suppose we draw a set of points on paper in a totally random manner” ...

“I am saying it is possible to find a geometric line whose notation is constant and uniform, following a certain law, that will pass through all points, and in the same order they were drawn.”...

*“But if that law is **strongly composed**, the thing that conforms to it should be seen as **irregular**”*

Gottfried Wilhelm Leibniz, 1686: Discours de métaphysique V, VI (from French)

What makes model inference different from data fitting?



https://zenodo.org/record/1436555#.Xe6tgdWIY58

Answers

Poll: which need do you think is most fundamental?

1. We want predictive power, not just describe past data
2. We want to explain / understand the mechanics of what is happening
3. We want to know cause and effect, not just relations

The answer:

Simplicity

(Parsimony / Occam’s Razor)

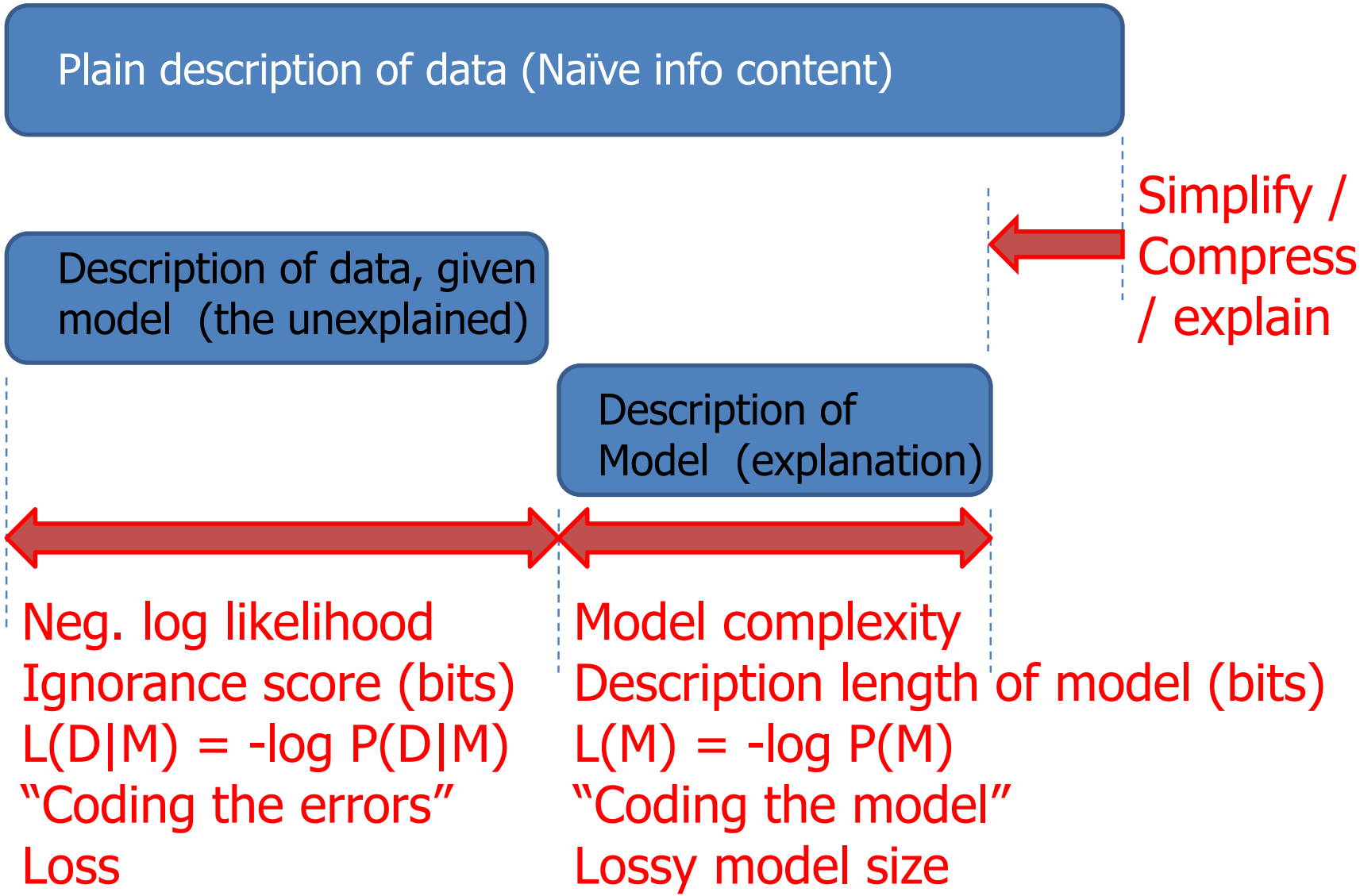
The answer:

Simplicity

(Parsimony / Occam's Razor)

New problem: compare models of different performance and simplicity

Formalizing learning: compression view

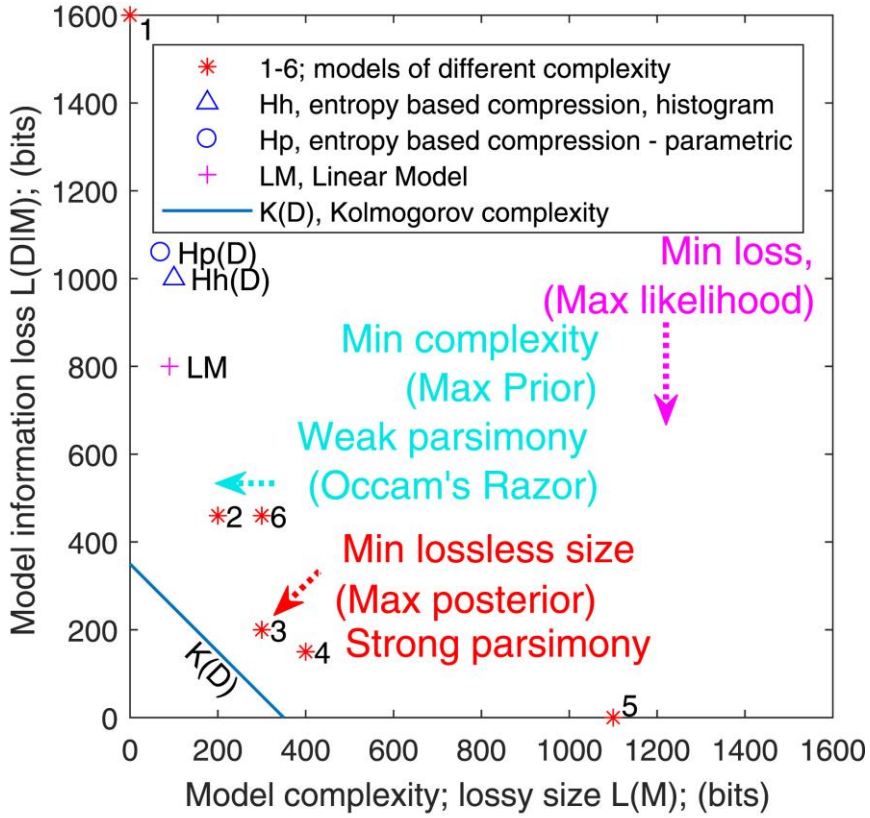


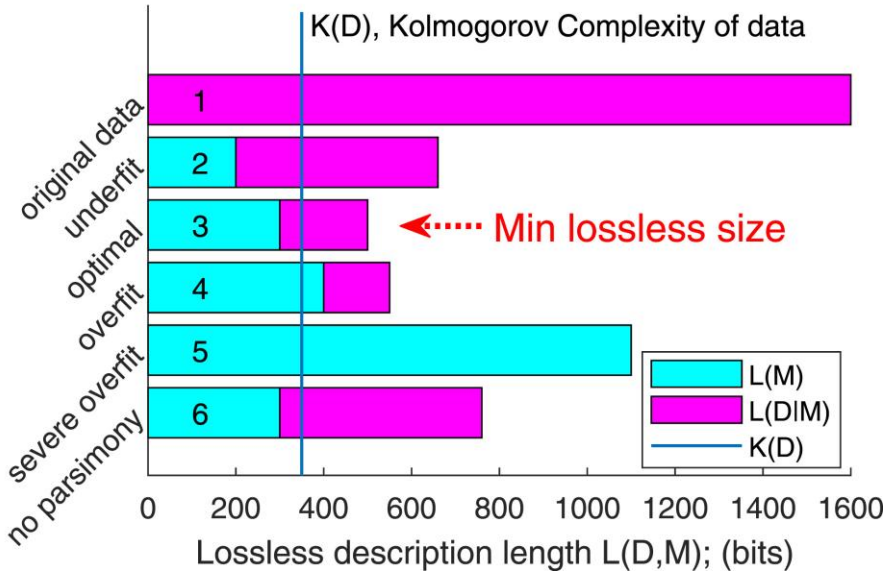
Occam's Digital Razor

- 2D → 1D
- Compression cannot be gamed

(if you are game:
check out: Hutter Prize)

Weijs & Ruddell, WRR, 2020





Why go digital?

Occam's Razor

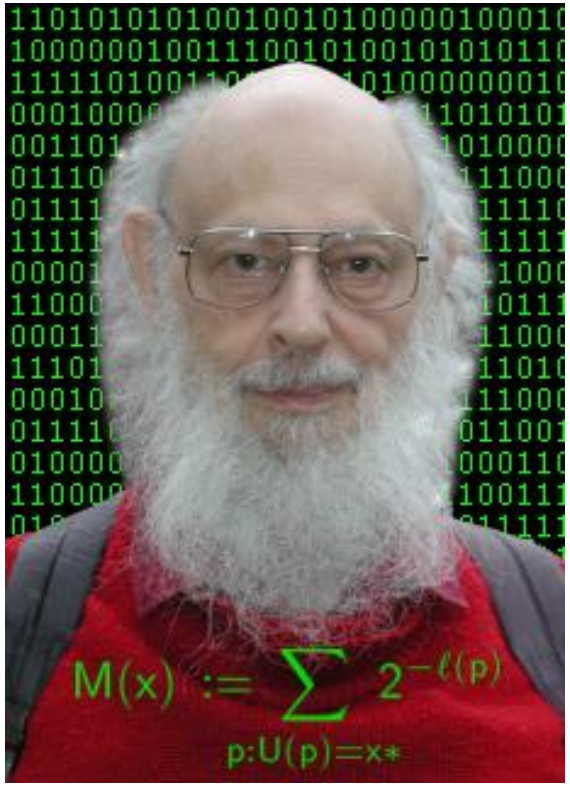
- Compare models of equal performance
- Compare models
- No units

Occam's digital razor

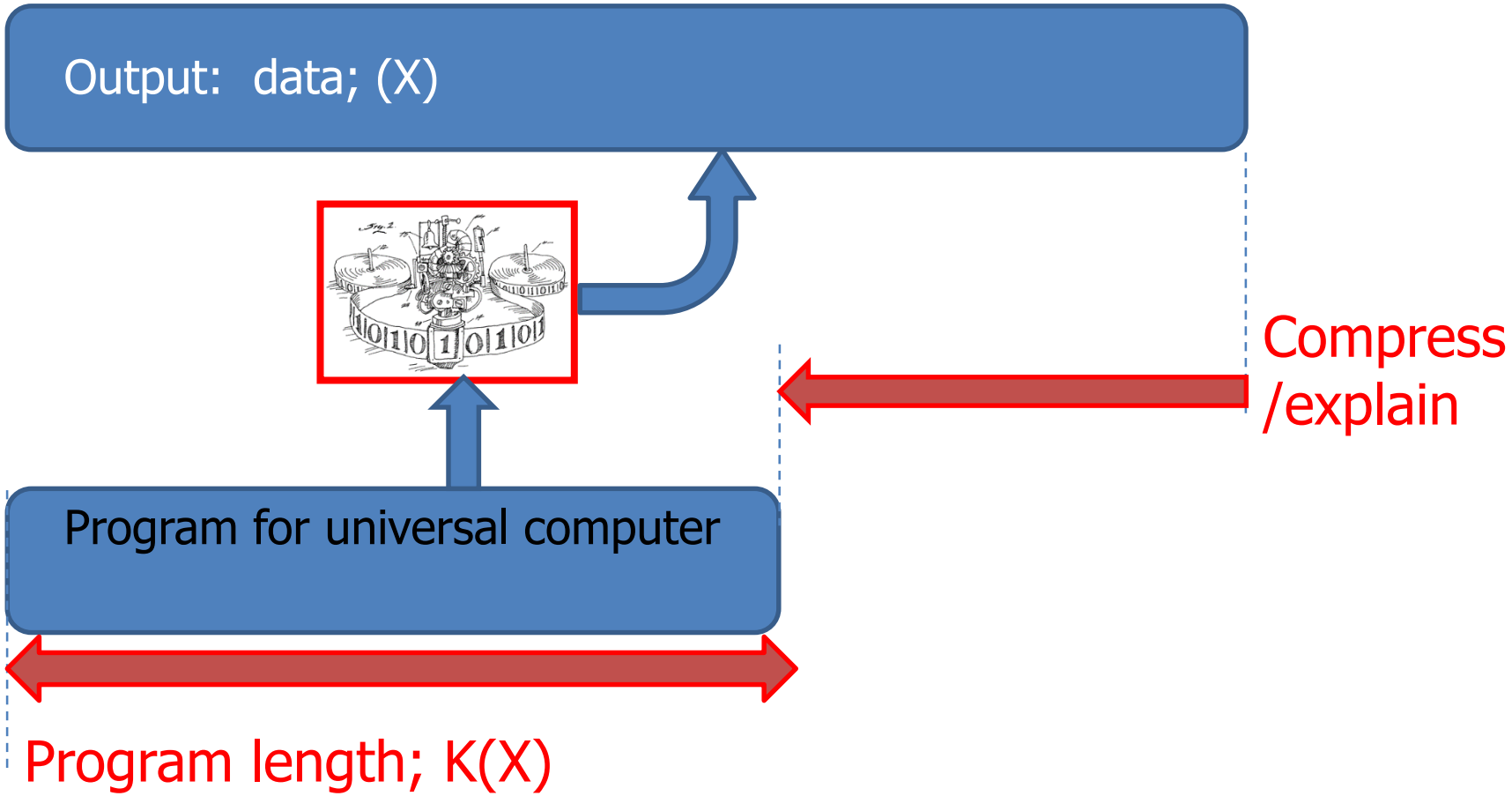
- Compare all models
- Compare models and original data
- All in bits
- Connection to prior probability

Algorithmic Information Theory (AIT)

independently developed by Kolmogorov(1968), Solomonoff (1964) and Chaitin (1966)



AIT view: searching shortest program



Note! No distinction between model (M) and data (D|M)

Key difference IT and AIT

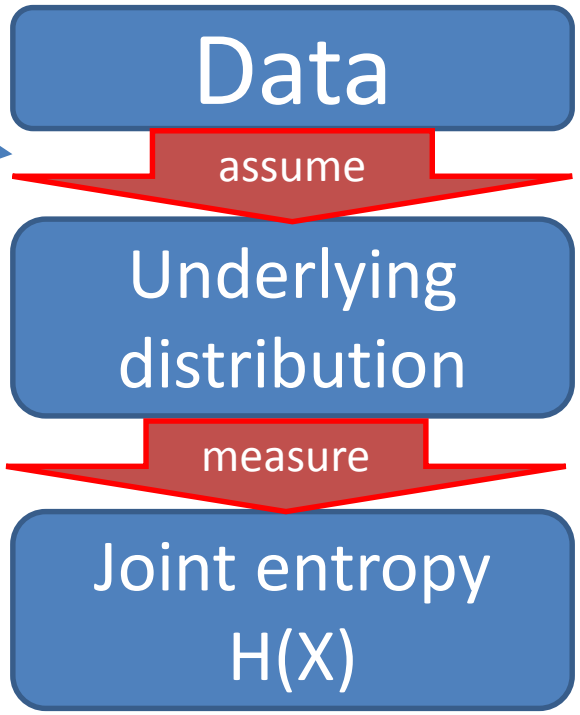
Information Theory

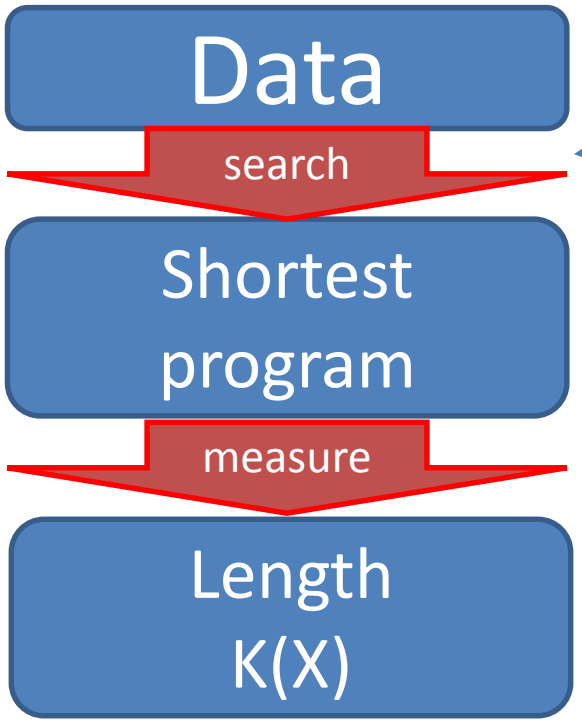
- Information content in bits
- Calculated from estimated distribution of data

Algorithmic Info-theory

- Information content in bits
- Calculated from single instance of data

Implicit complexity control, Limited manual hypotheses





Explicit complexity control, Enumerate computable models

Assumptions / Issues

- Assumes universe is computable (describable in math)
- Deciding shortest program: Incomputable, need approx.
- If applied to sub-system data: include external knowledge
- If not starting from scratch: prior knowledge

Main Challenge

- Move towards practice (encode prior, computable approximations, fair and practical complexity measures for ESM)

Some points for discussion

- Modeling answers questions about **how** not **why**
- Only difference between description is explanation is simplicity
- Being right for the right reasons = being right more