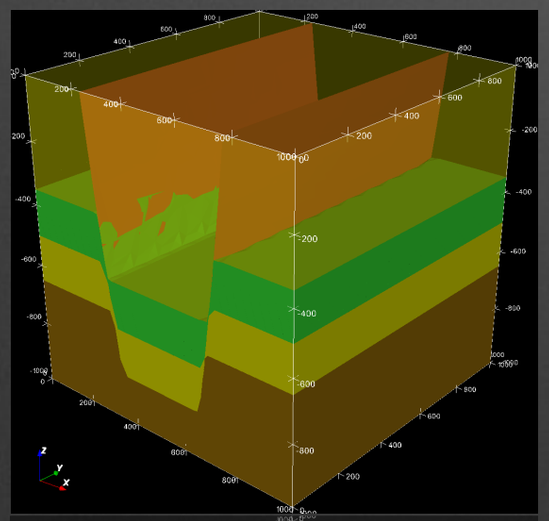
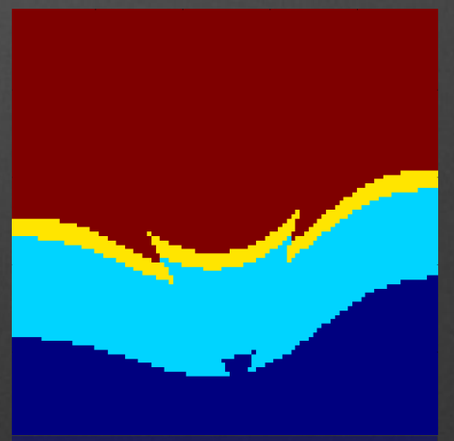
Information Theory and the Analysis of Uncertainties in a Spatial Geological Context

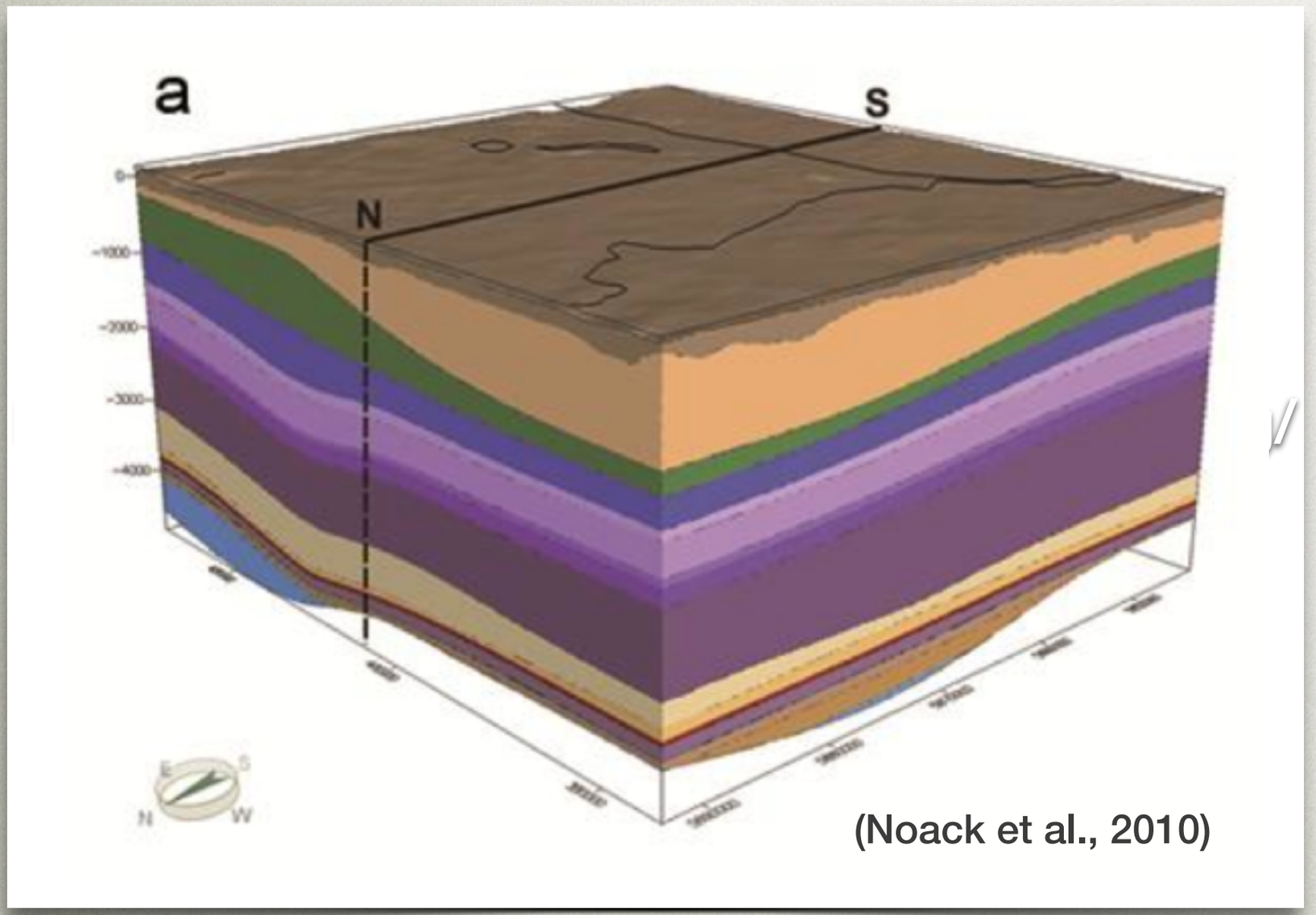
J. Florian Wellmann - RWTH Aachen
Mark Lindsay, Mark Jessell, Jonathan Poh, Klaus Regenauer-Lieb,
Miguel de la Varga Hormazabal, Simin Huang, Wang Hui, Philipp Schendt,
Johannes Aichele





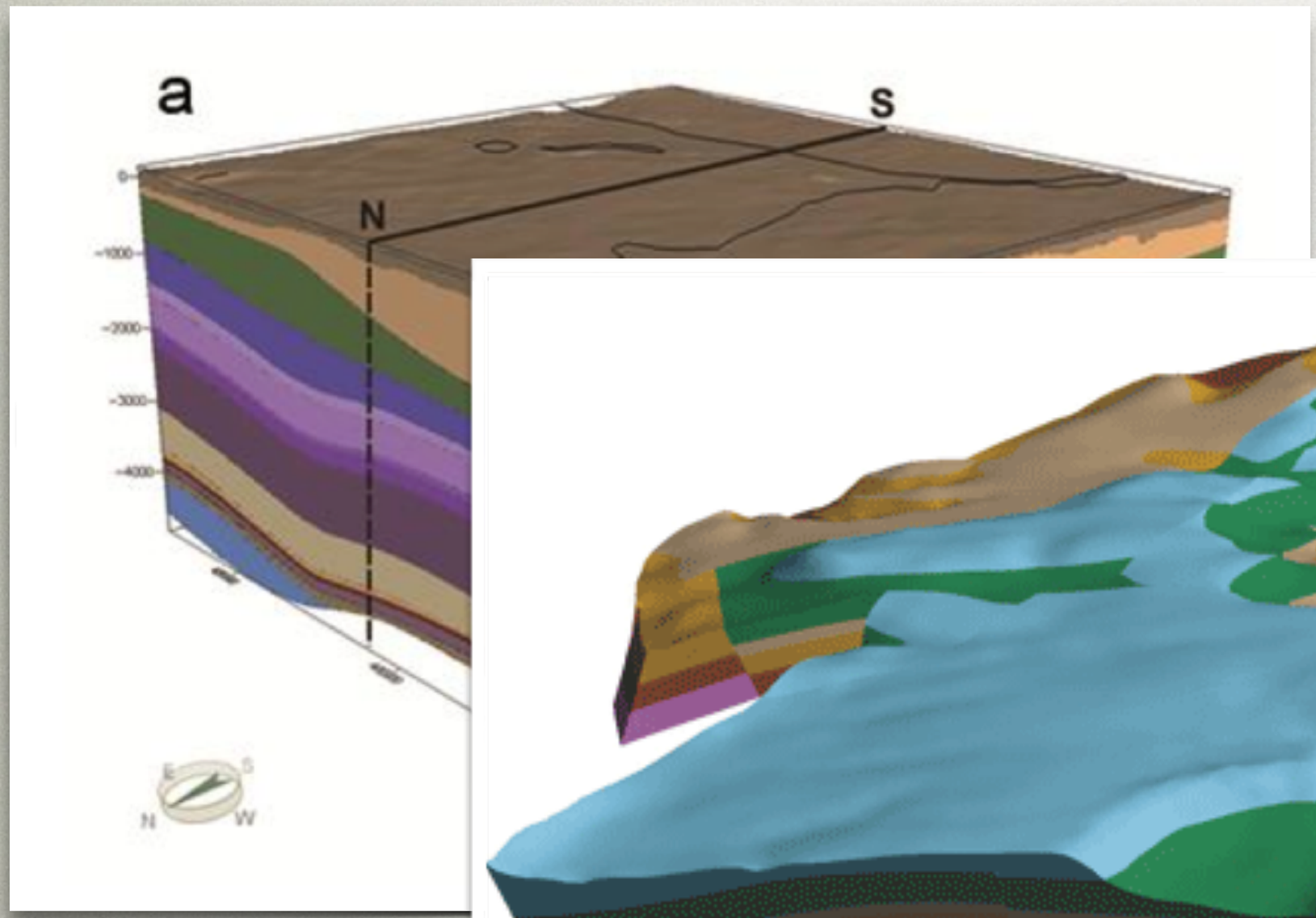
Information Theory and the Earth Sciences, Schneefernerhaus, 27.4.2016

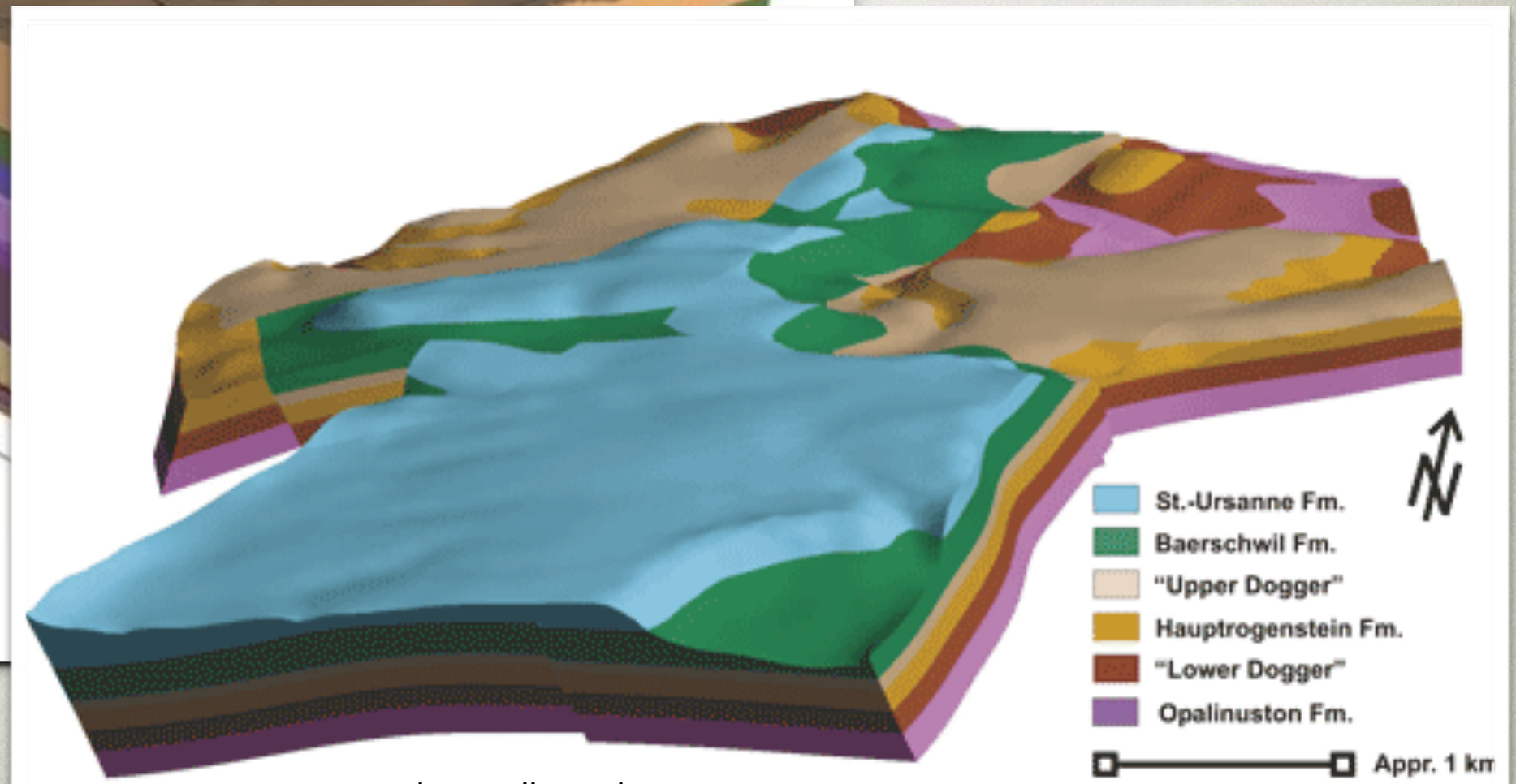
...WHERE OUR WORK (PROBABLY) FITS IN...



(Noack et al., 2010)

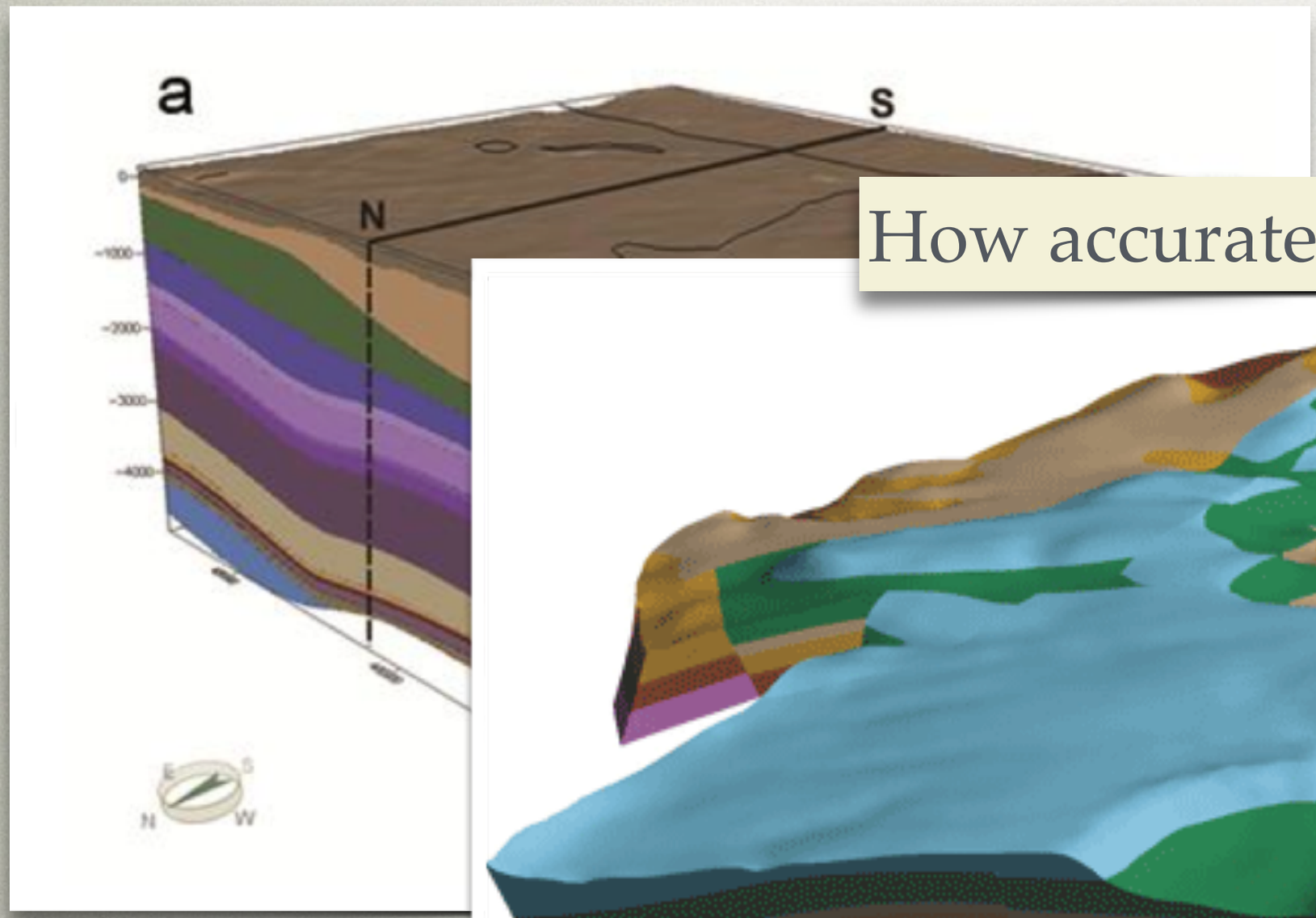
...WHERE OUR WORK (PROBABLY) FITS IN...



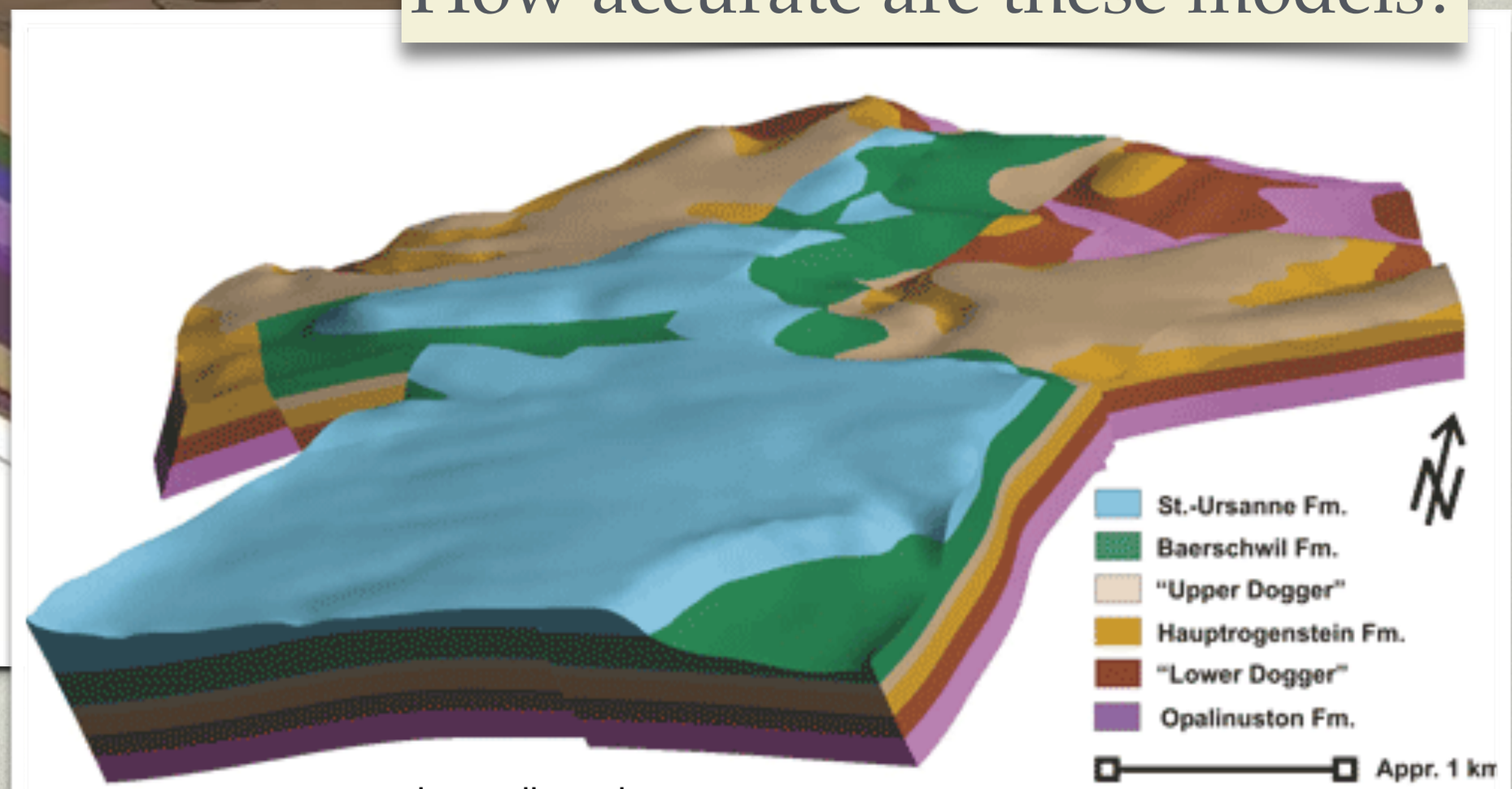


aug.duw.unibas.ch

...WHERE OUR WORK (PROBABLY) FITS IN...



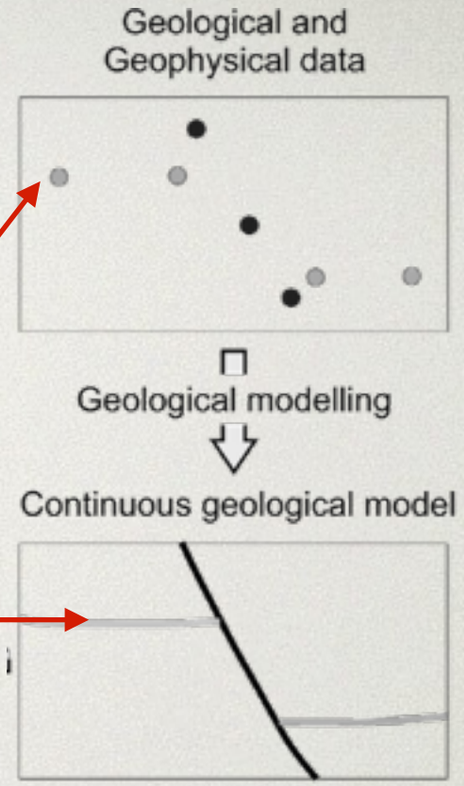
How accurate are these models?



aug.duw.unibas.ch

Conventional workflow

Construction of a continuous (3-D) model from geological data

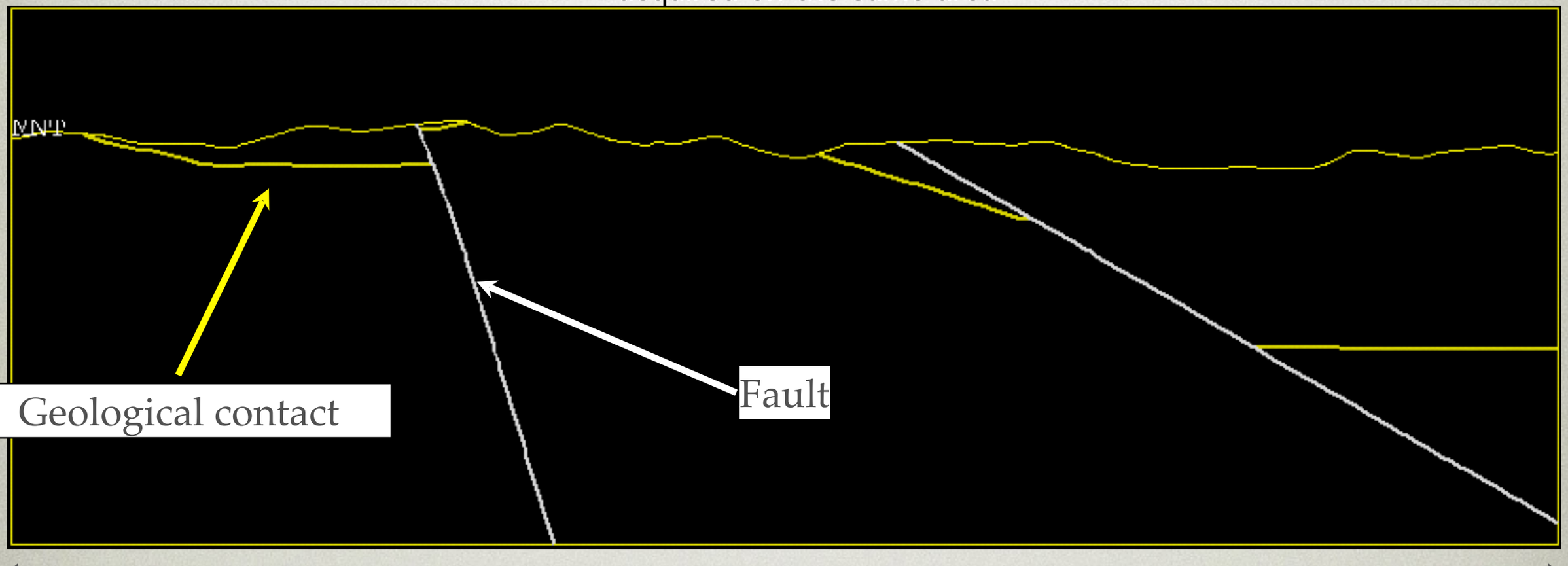


=> Quick look at a real example…

Models comparison

Slide from: Courrioux et al., 34th IGC, Brisbane, 2012

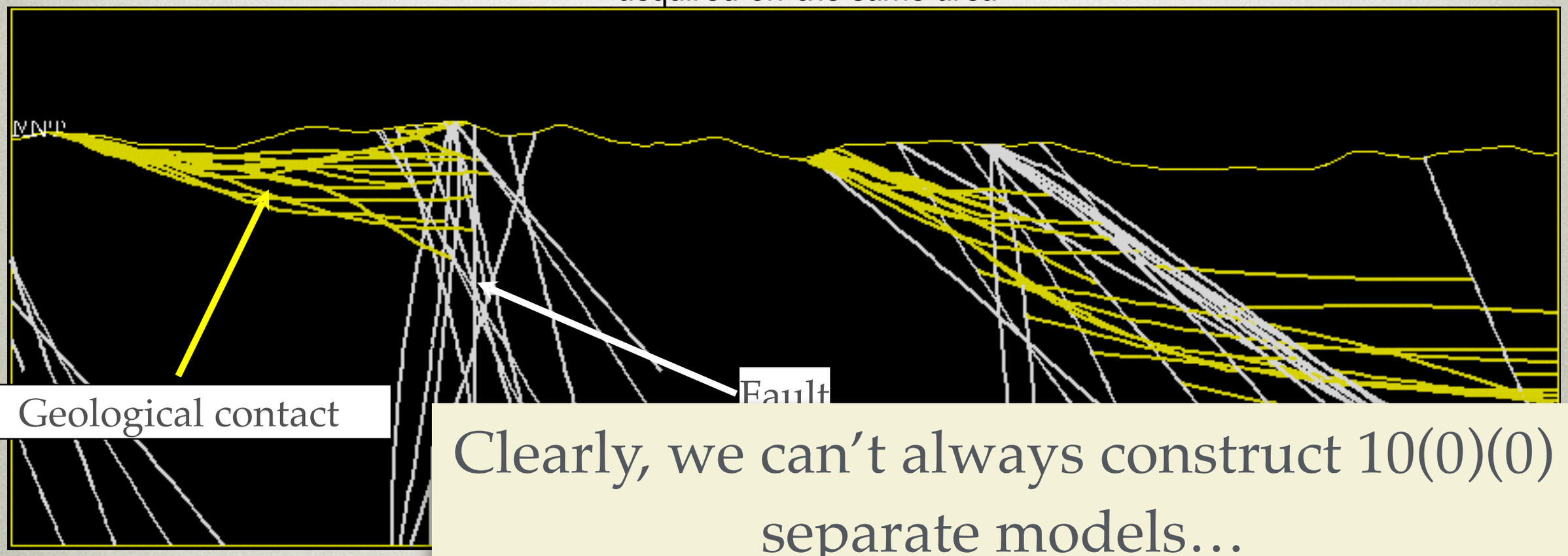
> In this study, we benefit from different realizations of a model from multiple sets of data acquired on the same area



Models comparison

Slide from: Courrioux et al., 34th IGC, Brisbane, 2012

> In this study, we benefit from different realizations of a model from multiple sets of data acquired on the same area



Clearly, we can't always construct 10(0)(0) separate models…
…but what about a randomised approach?

Great variability far from outcrops

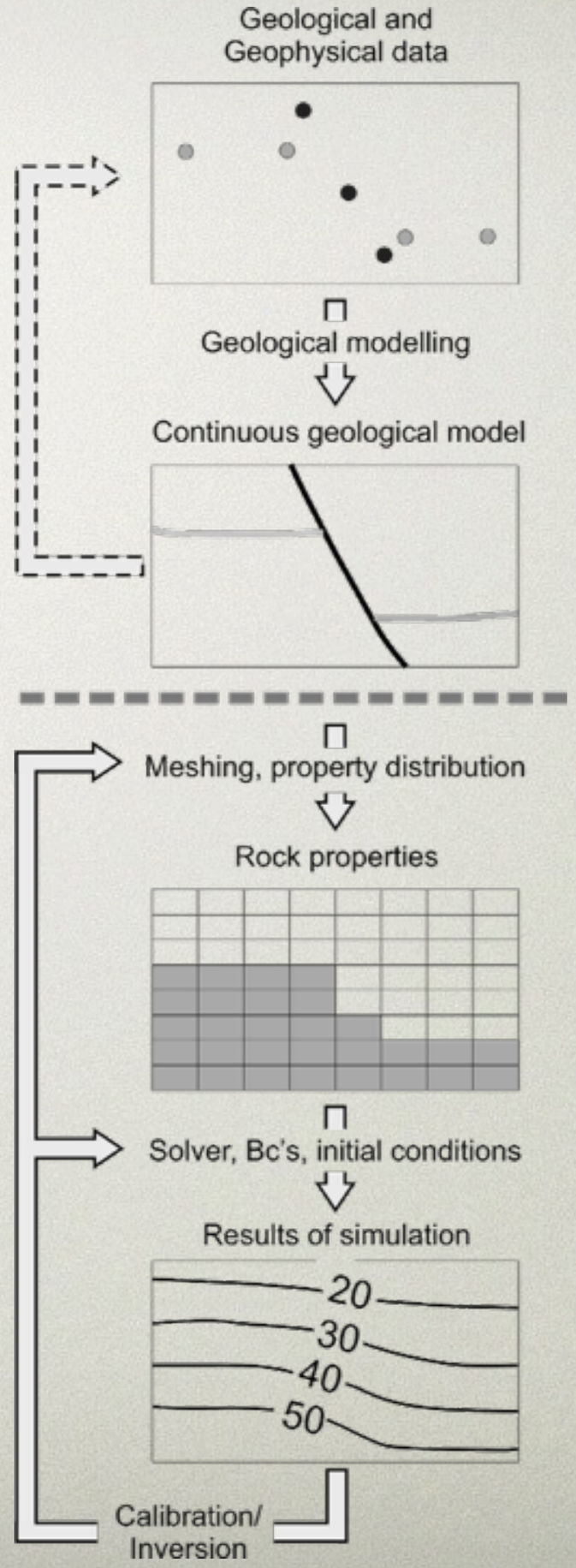
Where Is the reality ? somewhere here ?

Conventional workflow

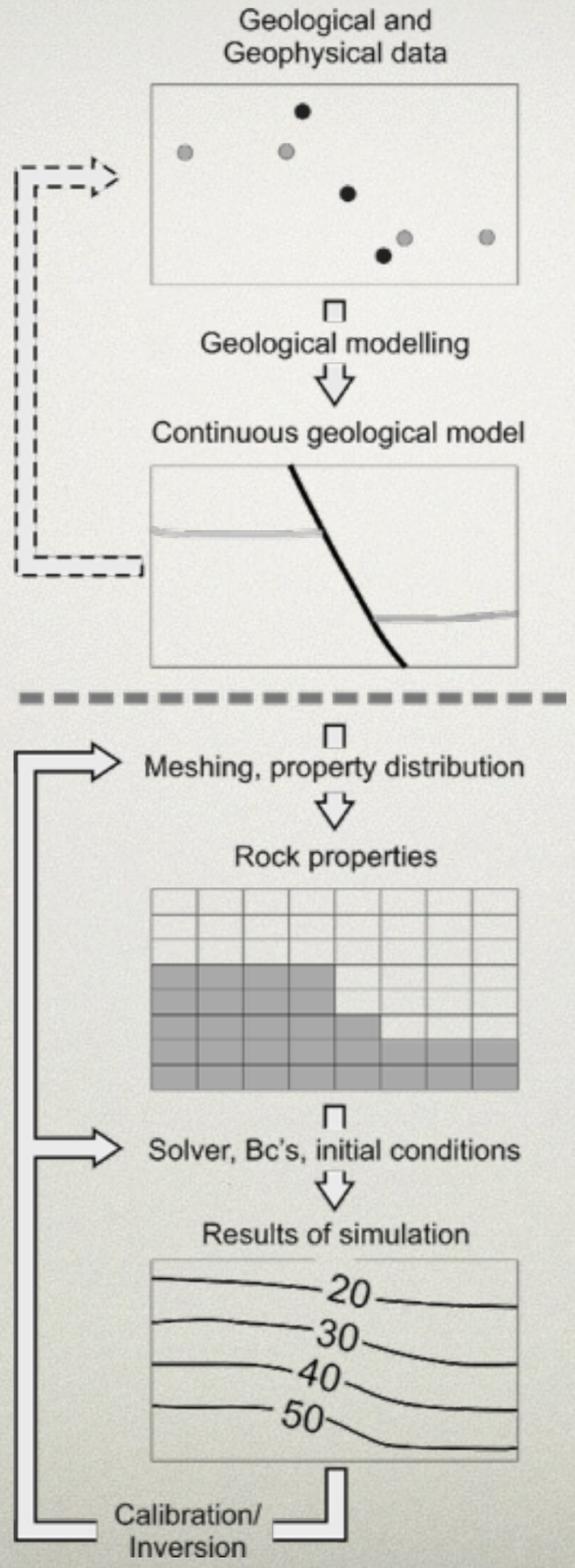
Conventional methods: manual iterations

Mesh generation

=> Limitations for automation!

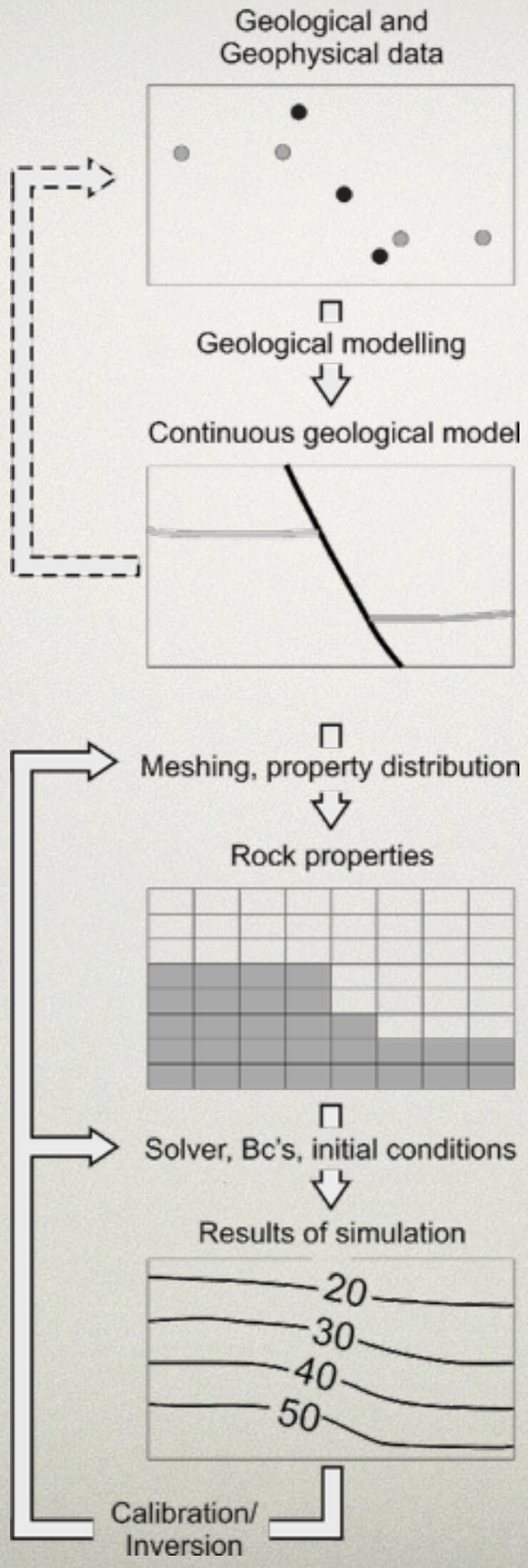


Technical challenge: removing the limiting step between continuous model and mesh



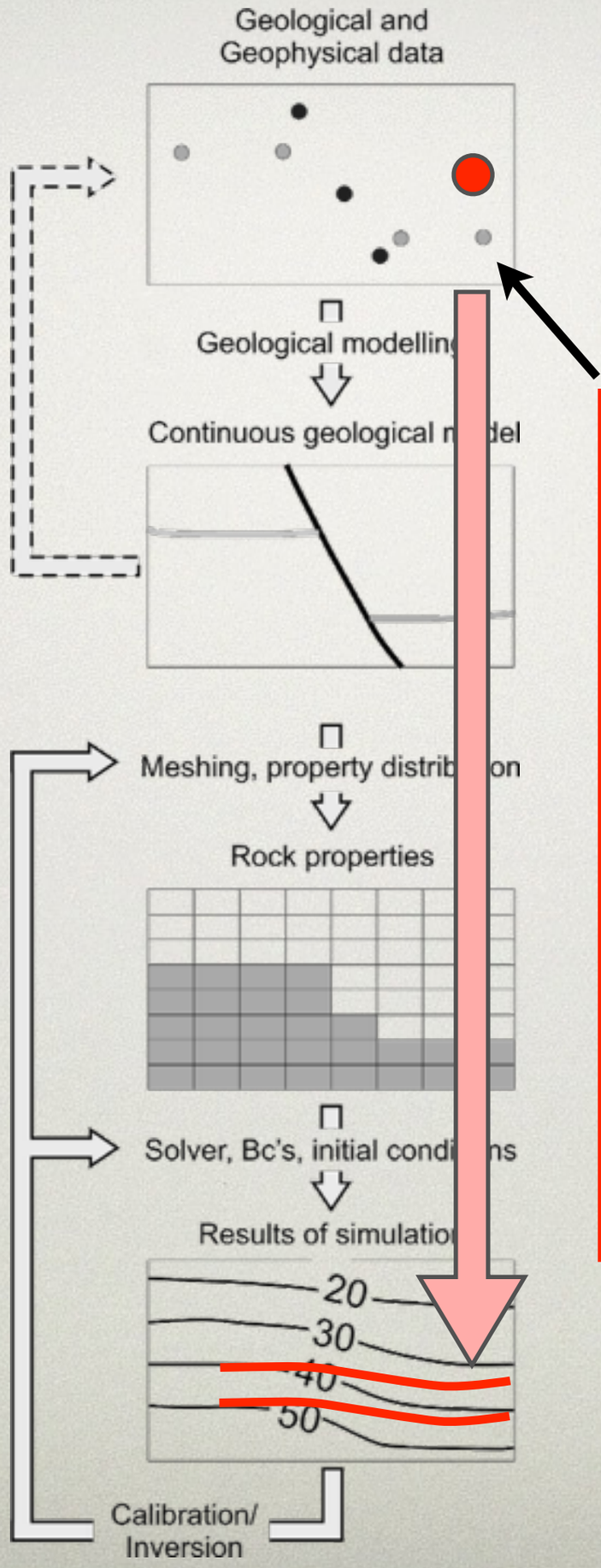
Developed methods to combine implicit modelling and input file generation with PySHEMAT and PyTOUGH (both open source, available on github.com)

(Wellmann et al., 2011, 2013)



Developed methods to combine implicit modelling and input file generation with PySHEMAT and PyTOUGH (both open source, available on github.com)

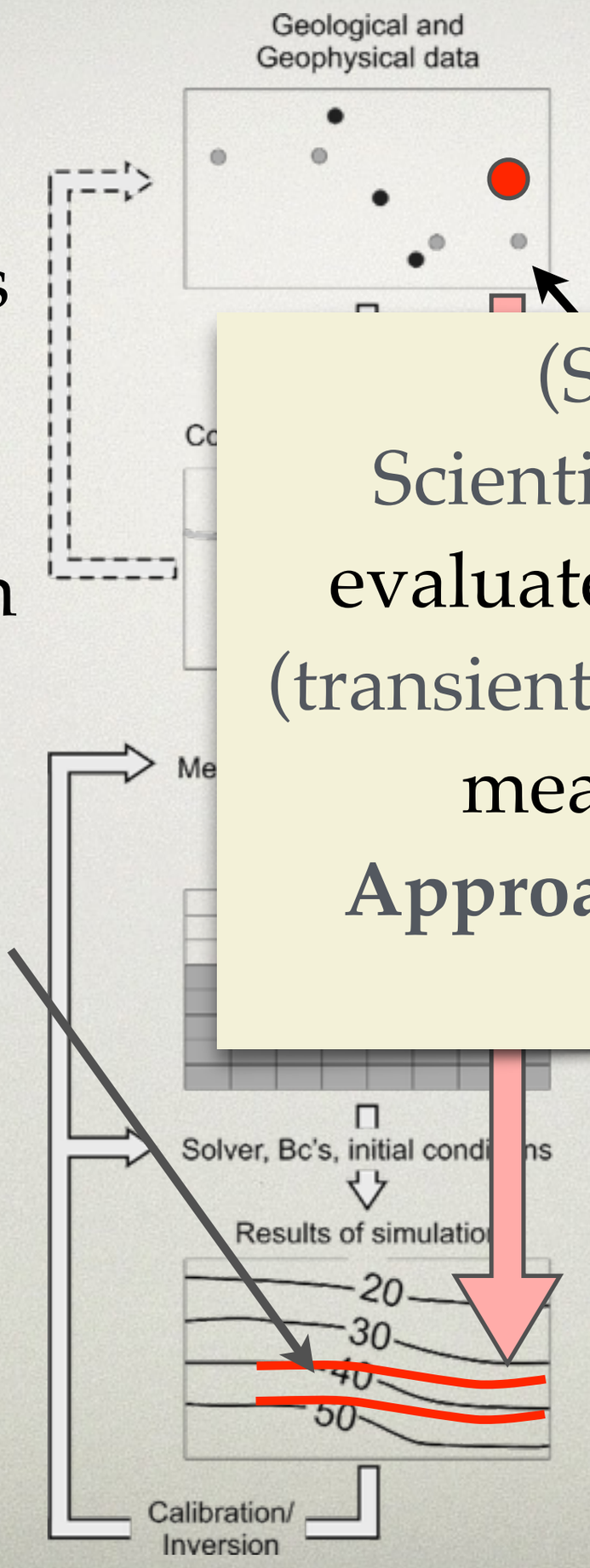
(Wellmann et al., 2011, 2013)



Now possible: change position of geological data point, automatically update effect on simulated flow fields

Developed methods to combine implicit modelling and input file generation with PySHEMAT and PyTOUGH (both open source, available on github.com)

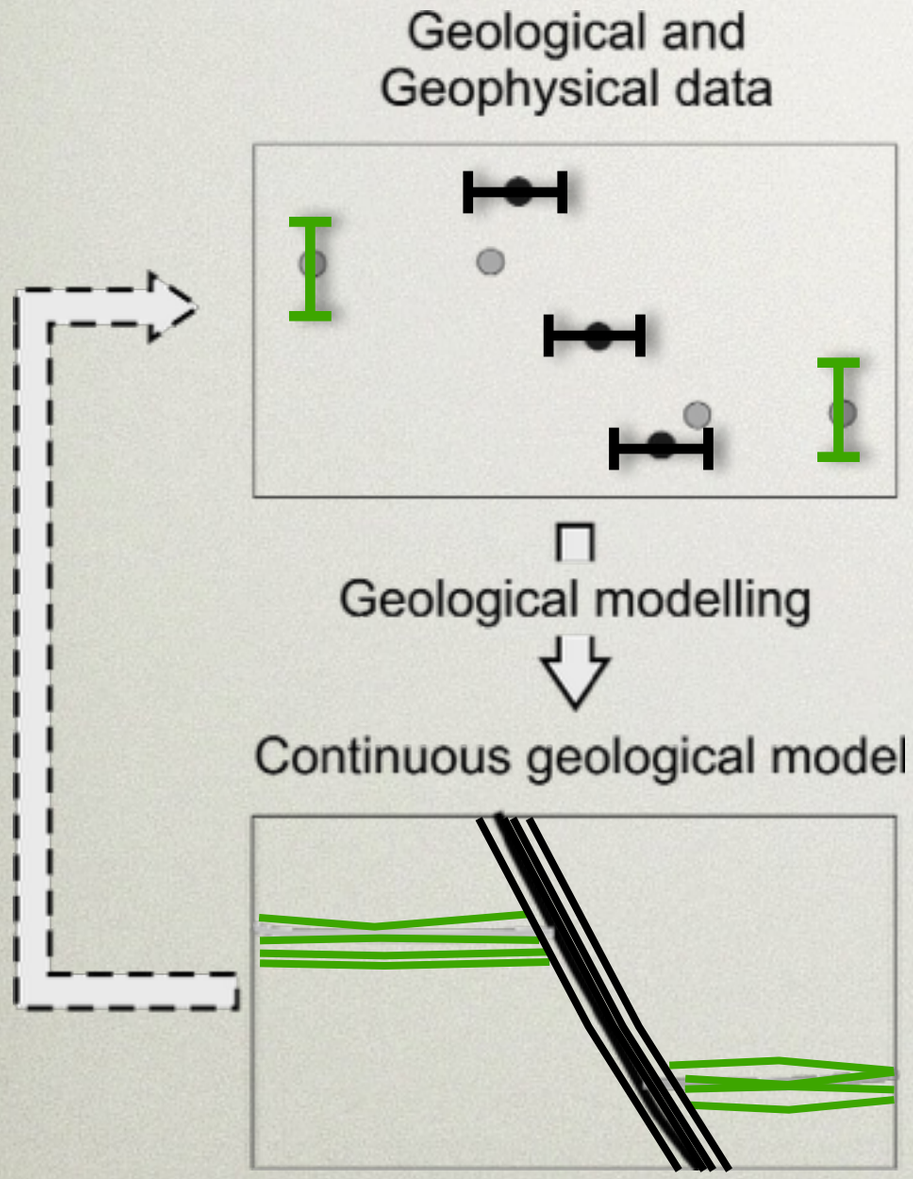
(Wellmann et al., 2011, 2013)



(Small) Side track:
Scientific challenge: How to evaluate output from multiple (transient) simulation results with meaningful measures?
Approach: (thermal) entropy production

simulated flow fields

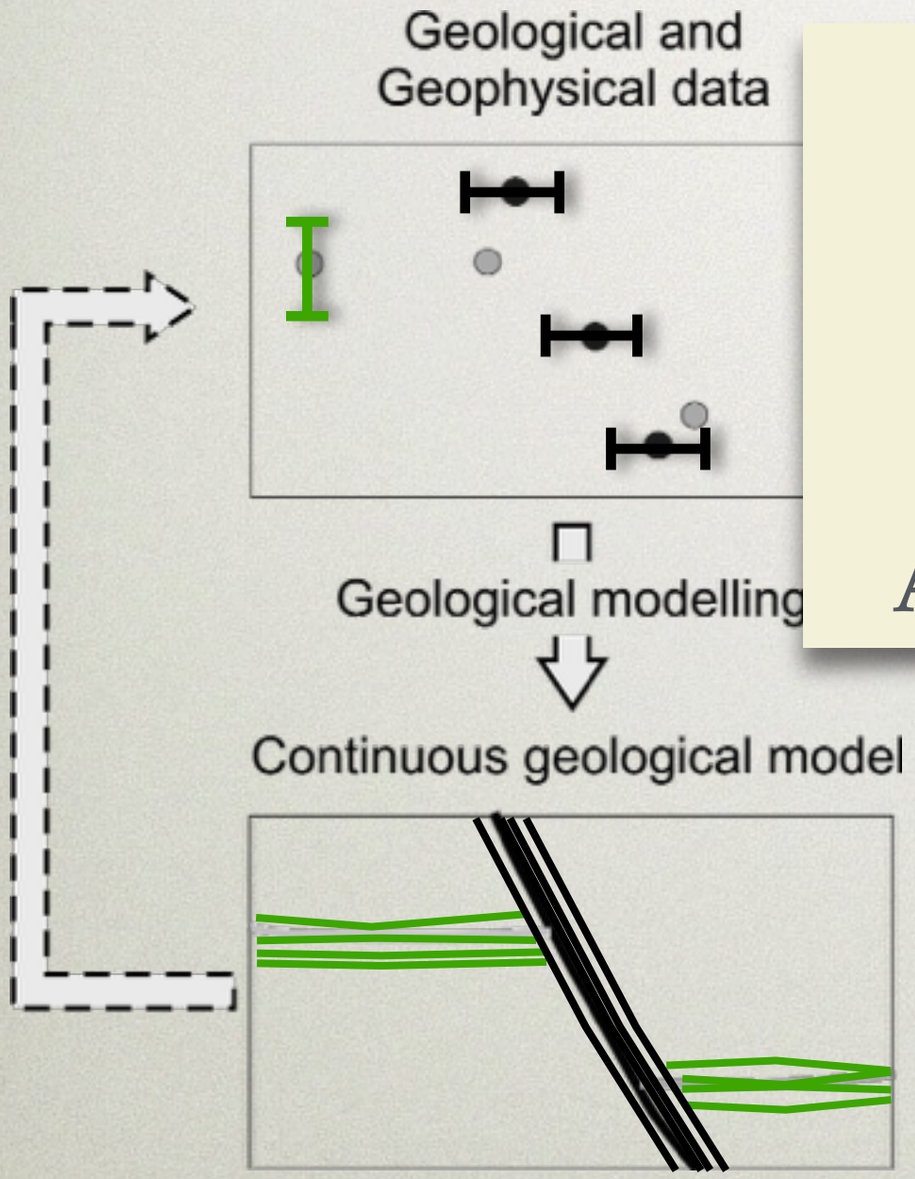
Technical Challenge: Geological Ensemble Modelling



Consider uncertainties in structural data

Automatically generate multiple probable geological models with stochastic approach

Technical Challenge: Geological Ensemble Modelling

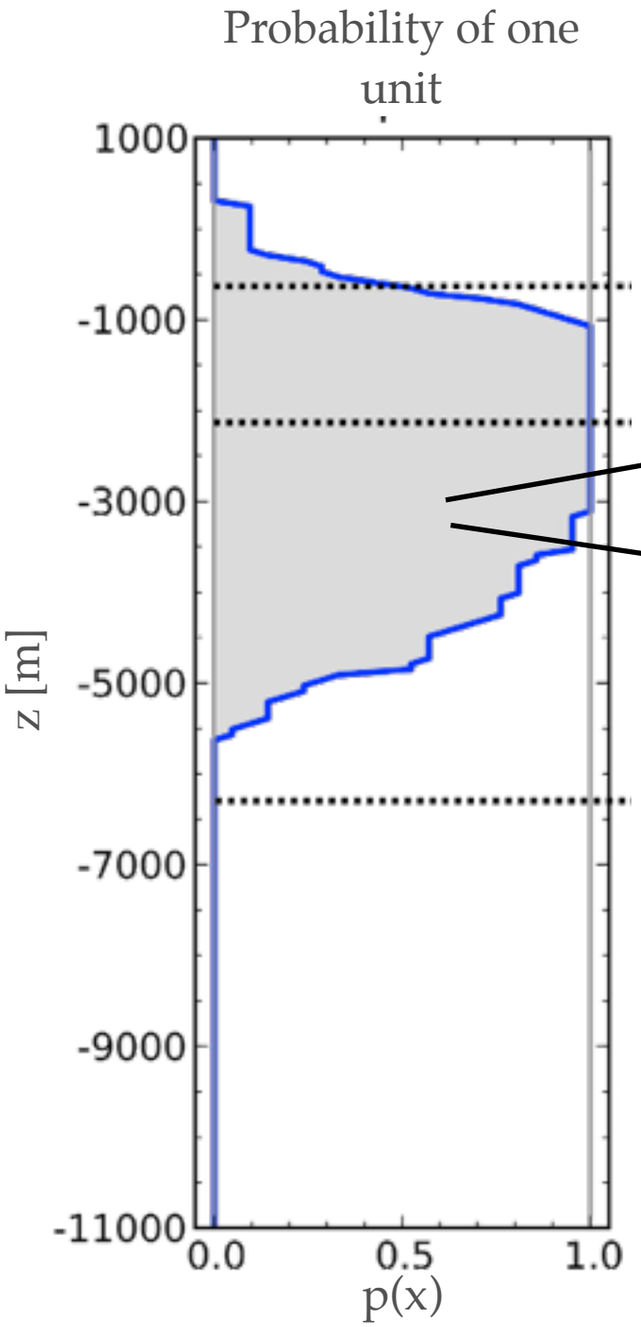


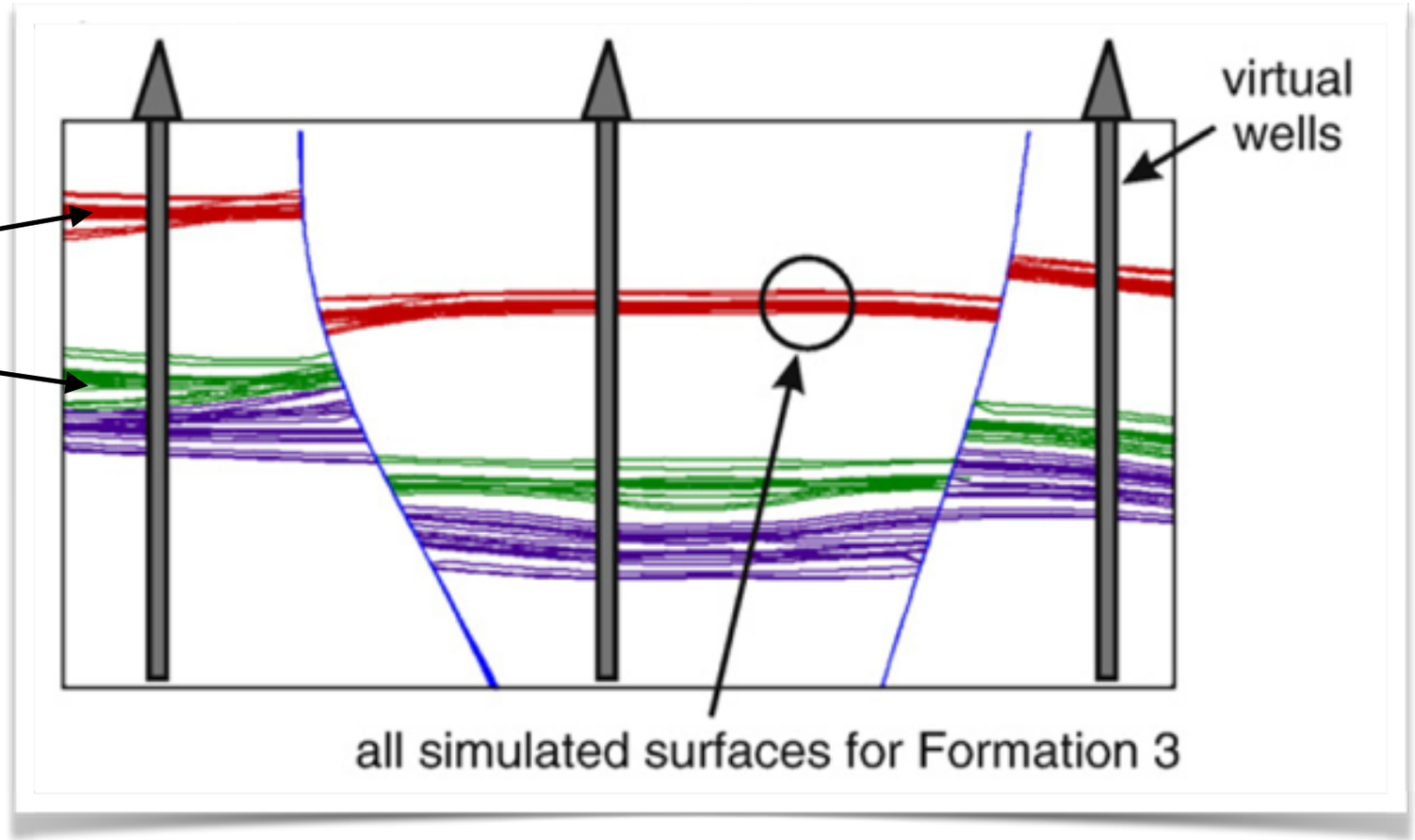
Scientific challenge:
how to analyse and visualise multiple geological modelling results?

Approach: Information entropy

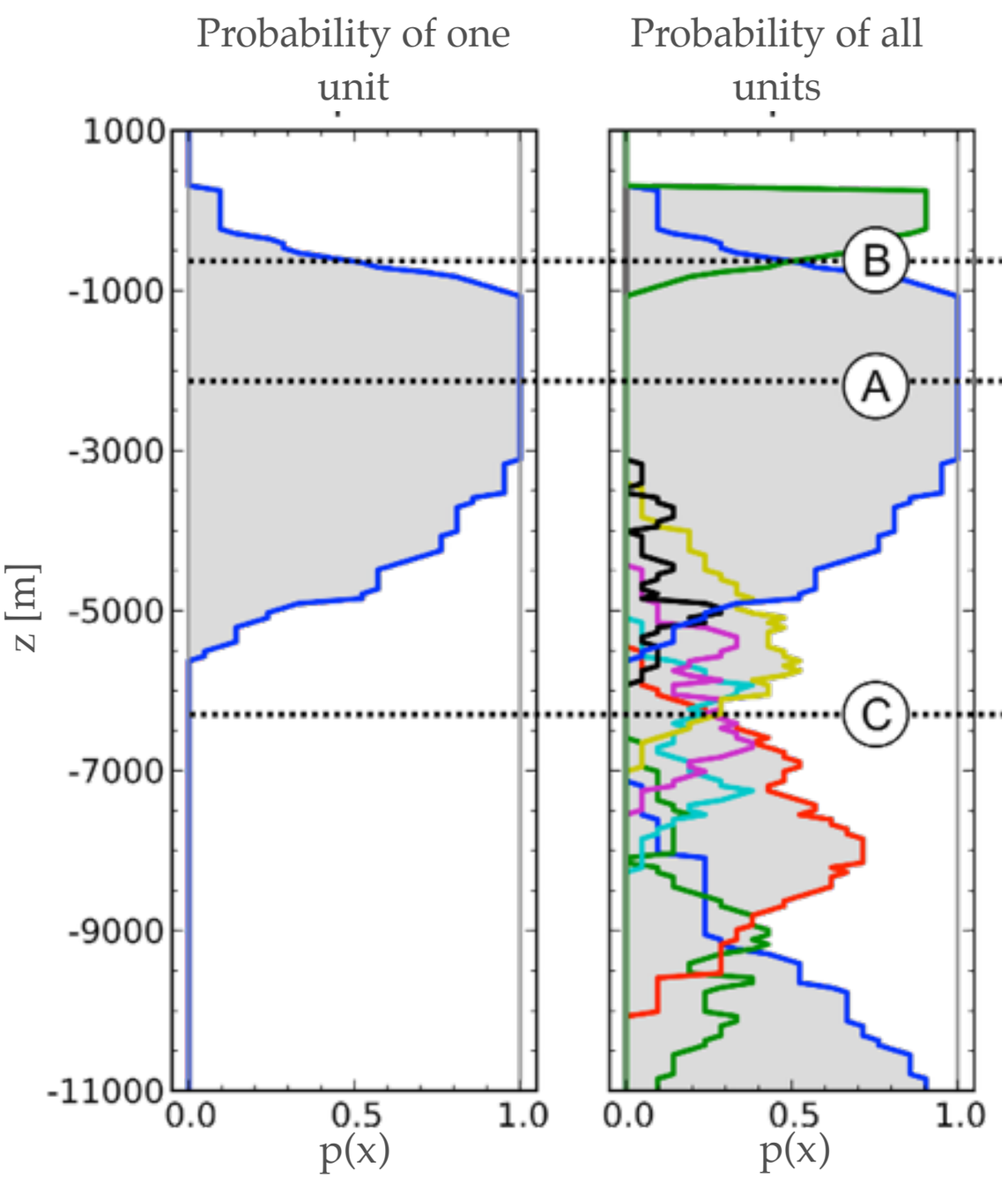
Automatically generate multiple probable geological models with stochastic approach

Quantitative interpretation

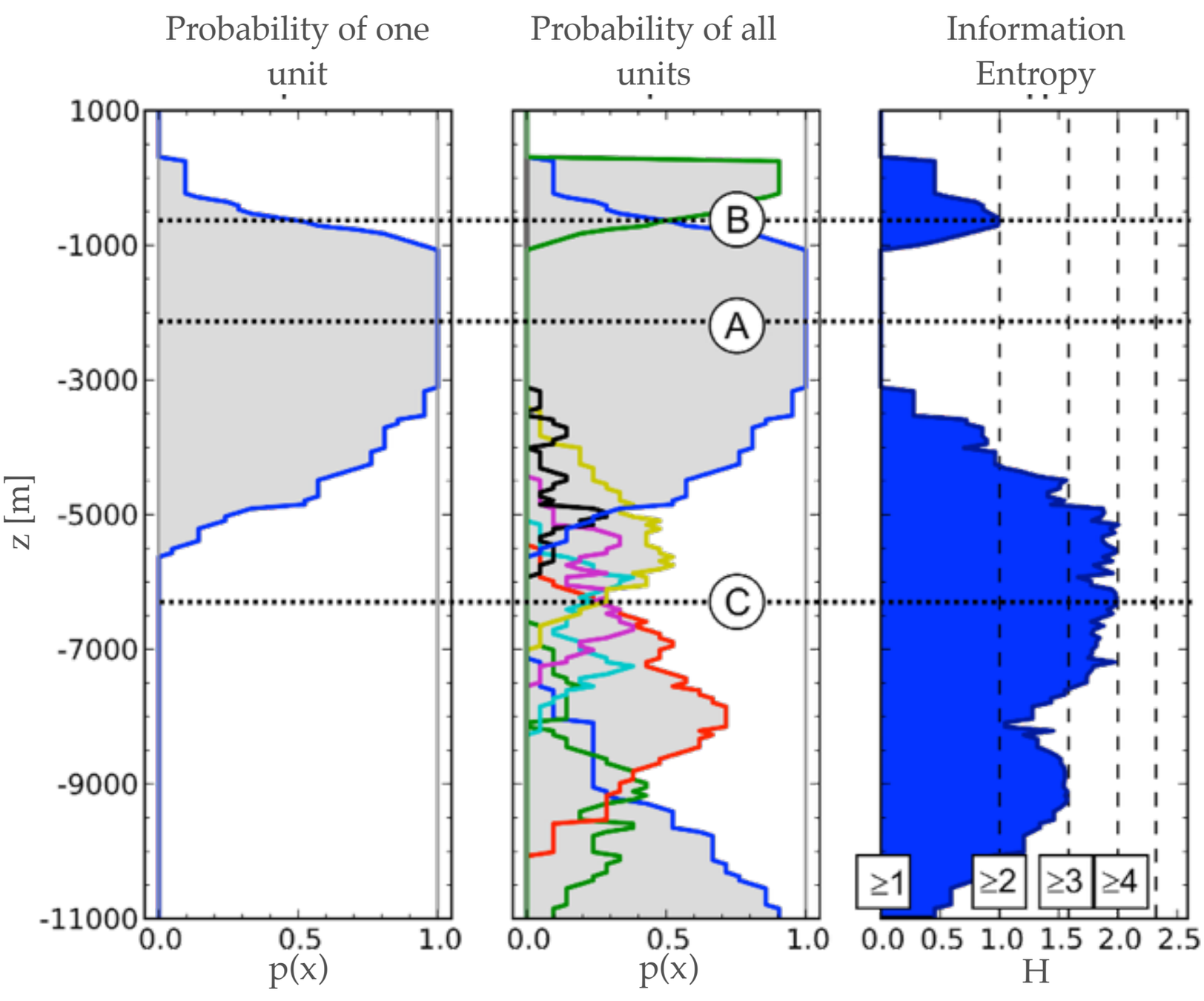




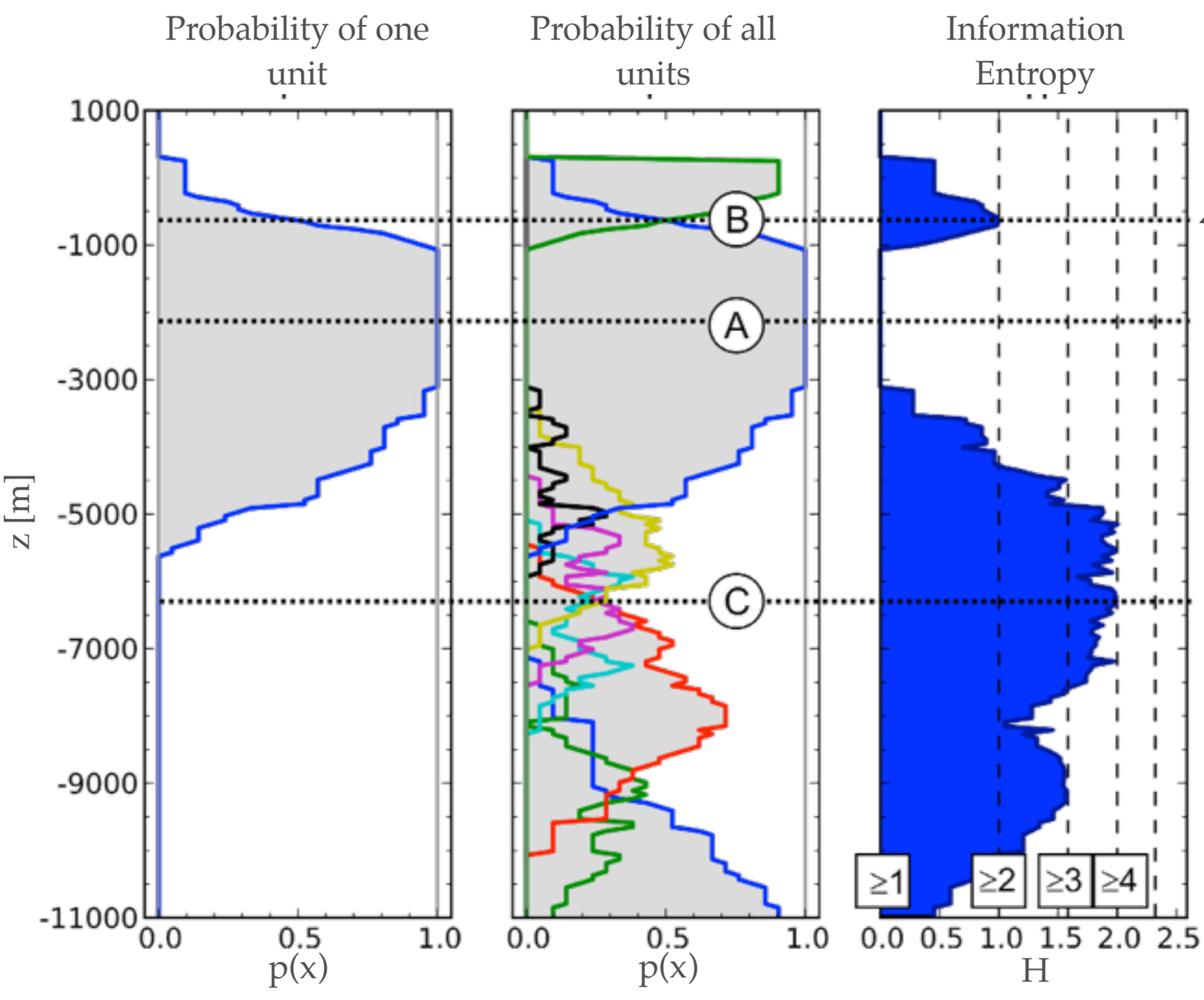
Quantitative interpretation



Quantitative interpretation



Quantitative interpretation

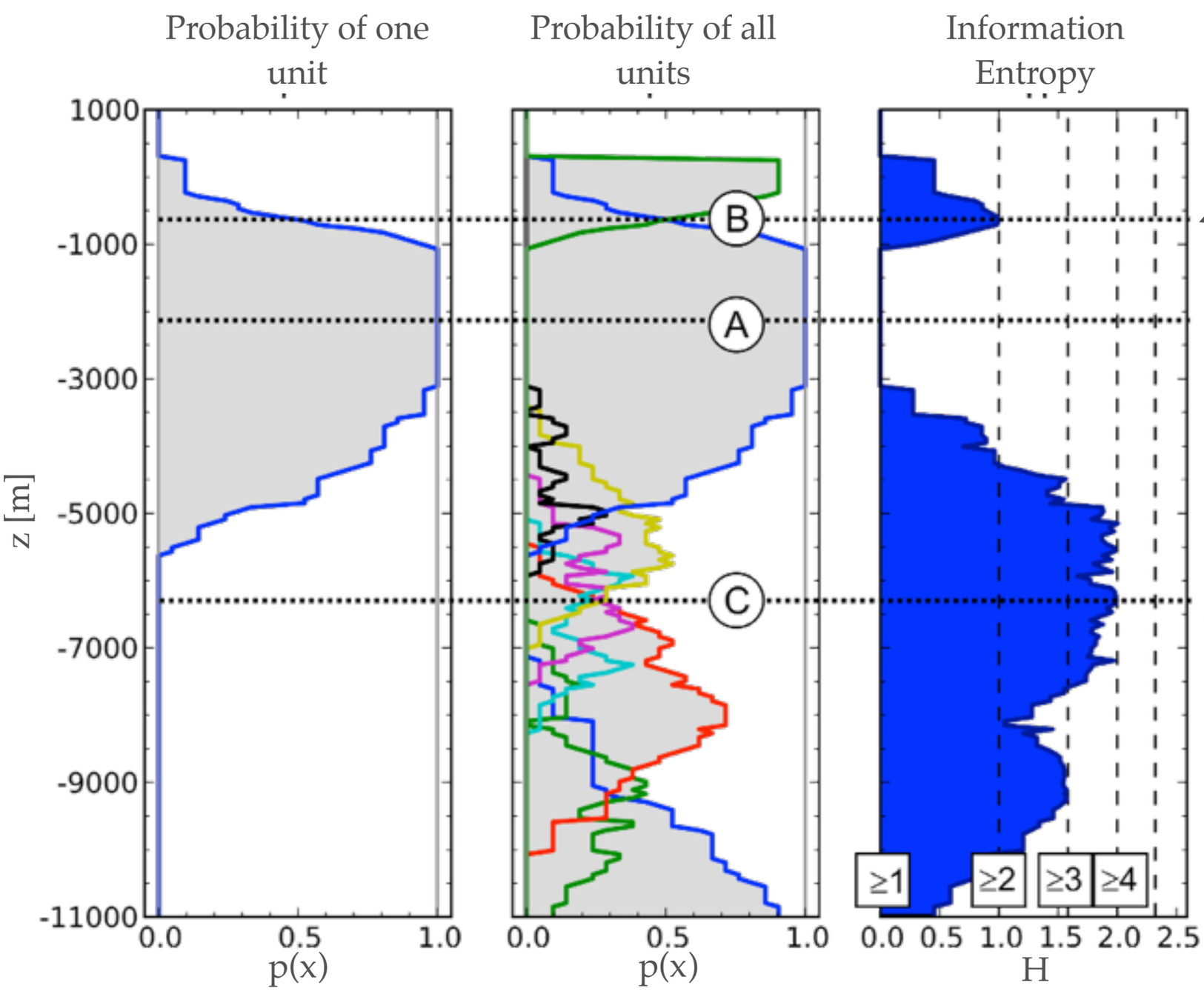


Two outcomes, equally probable ("coin case")

No uncertainty

Generally: more outcomes and more uniform, higher entropy. Max H for n outcomes: $\log_2(n)$

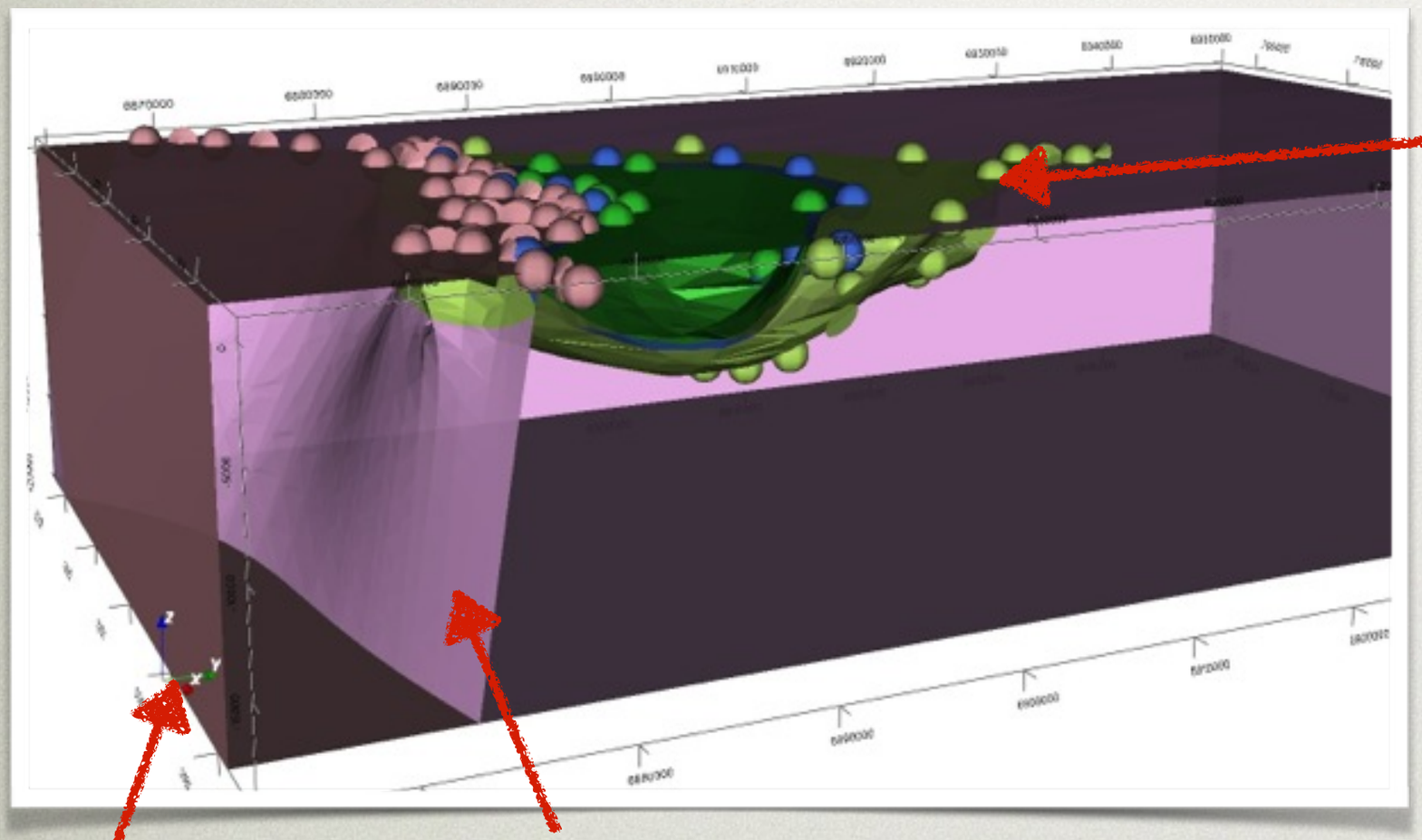
Quantitative interpretation



n	log
2	1
3	1,58
4	2
5	2,43
6	2,58

Generally: more outcomes and more uniform, higher entropy. Max H for n outcomes: $\log_2(n)$

Example: Greenstone Belt, Western Australia

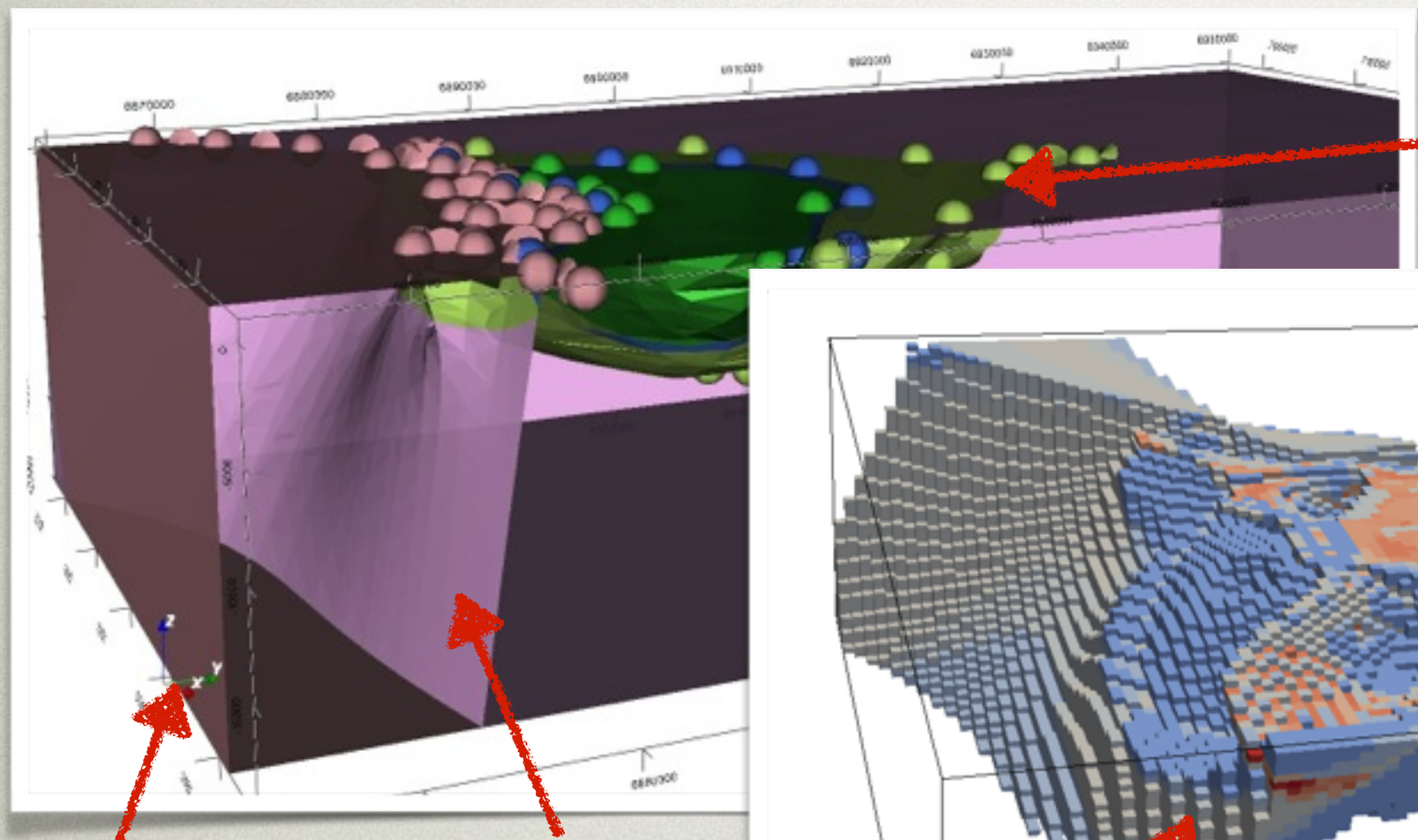


Geological data points

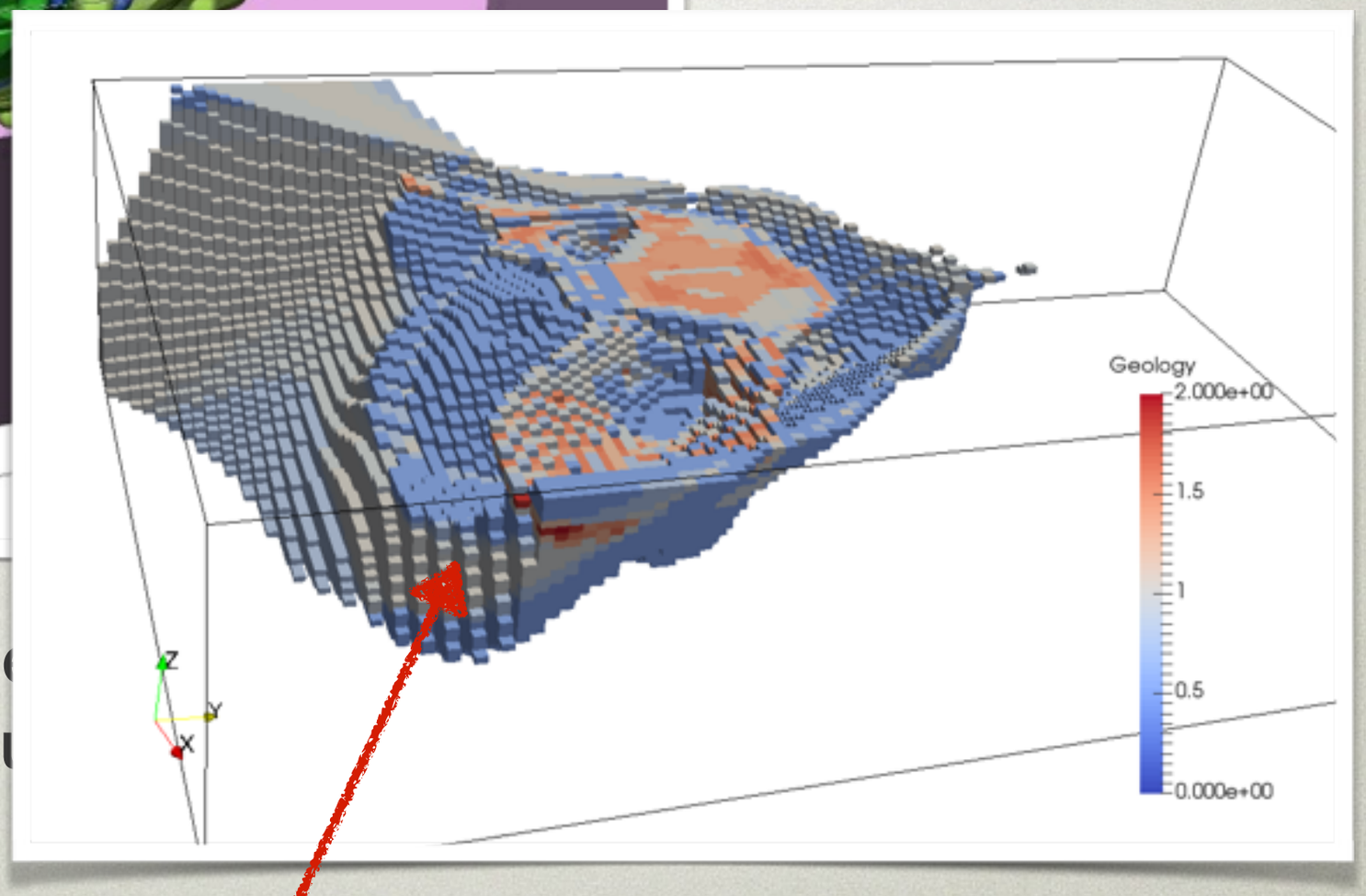
Relevant geological structures

Model scale:
50 x 80 x 20 km

Example: Greenstone Belt, Western Australia



Geological data points

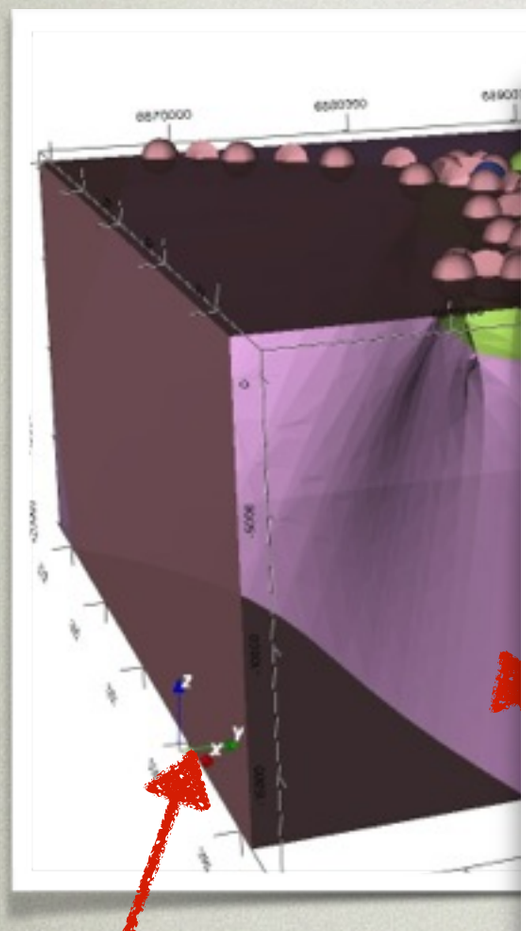


Relevant g
struct

Model scale:
50 x 80 x 20 km

Information entropy in each cell (transparent: 0)

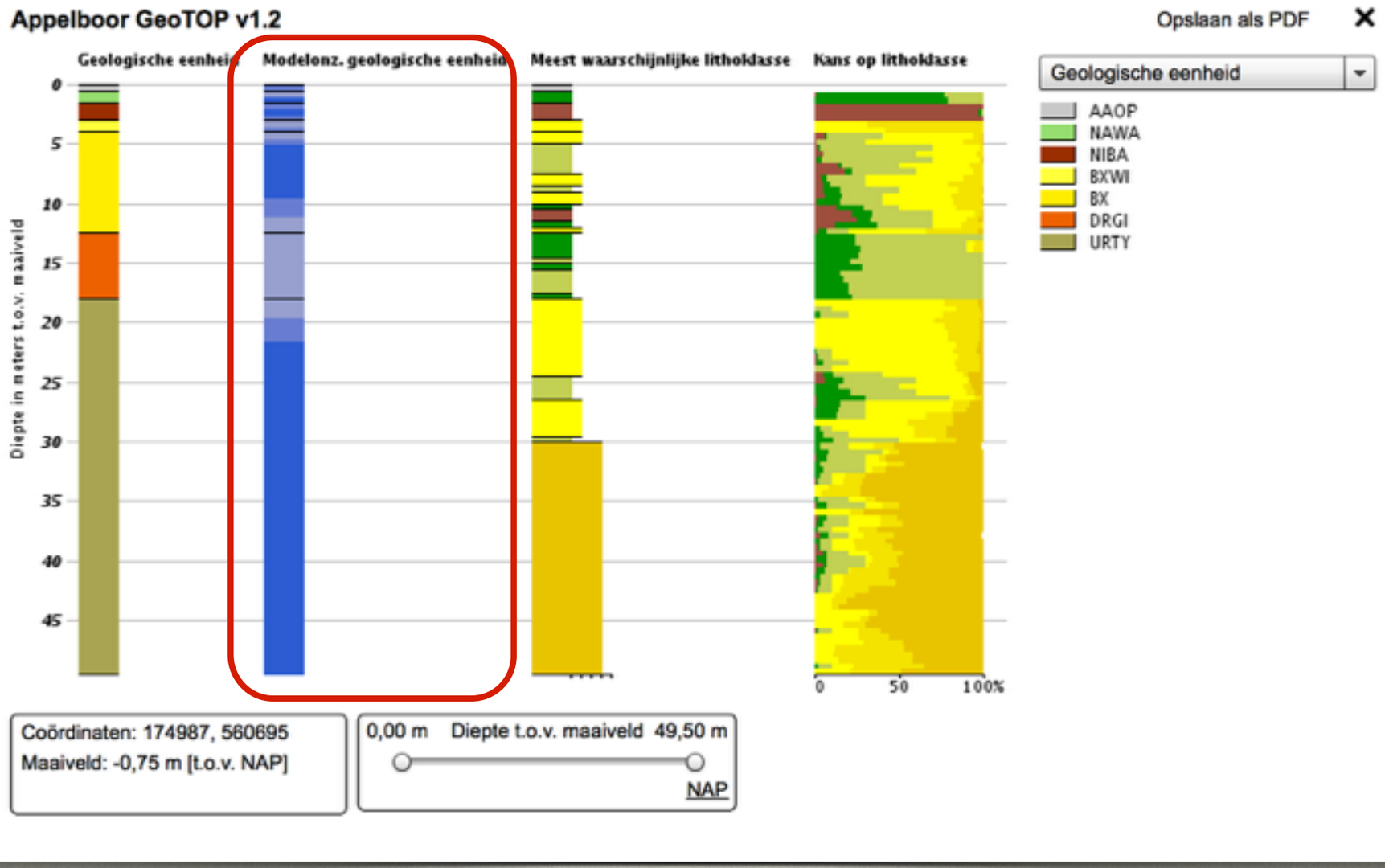
Example: Greenstone Belt, Western Australia



Re

Model scale:
50 x 80 x 20 km

Also used in very similar context in TNO map
(www.dinoloket.nl)

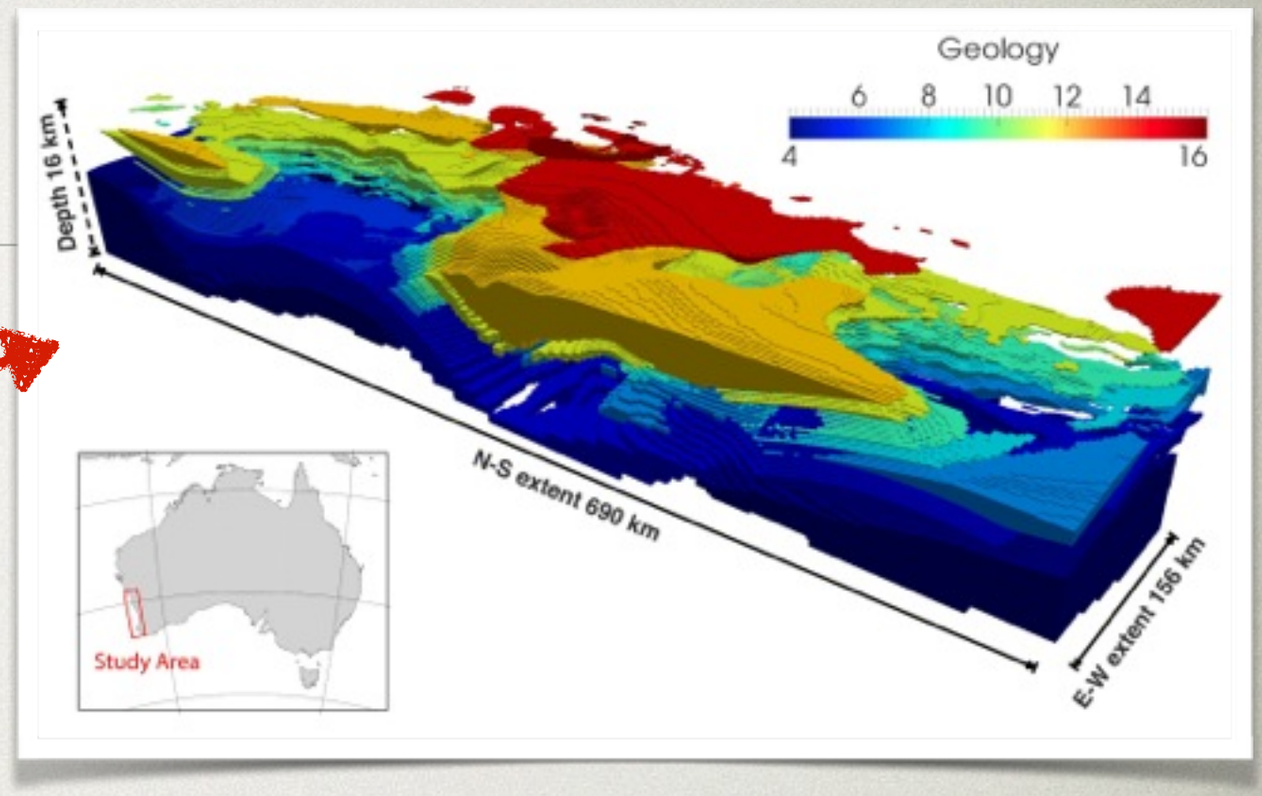


Where next?

- Link to process simulations (hydrogeology, geothermal systems, …)

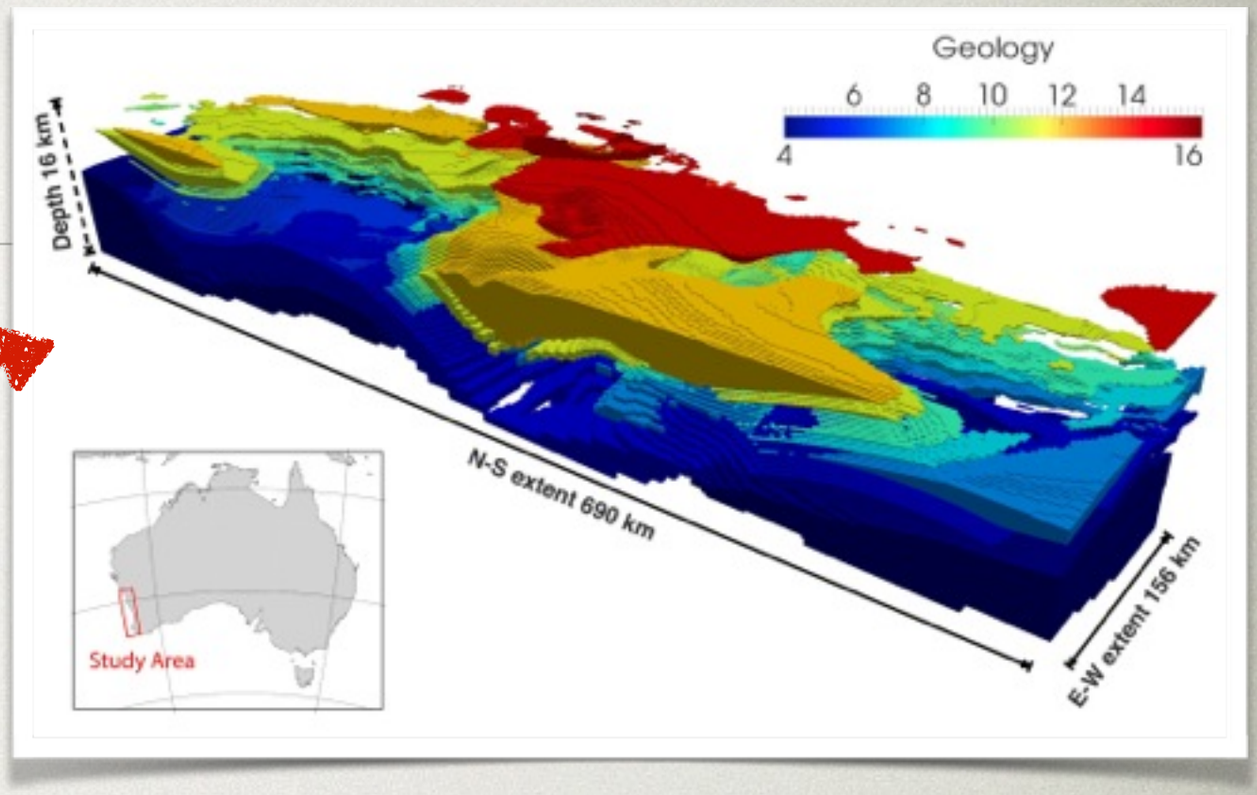
- How to constrain the model with more observations and additional data?

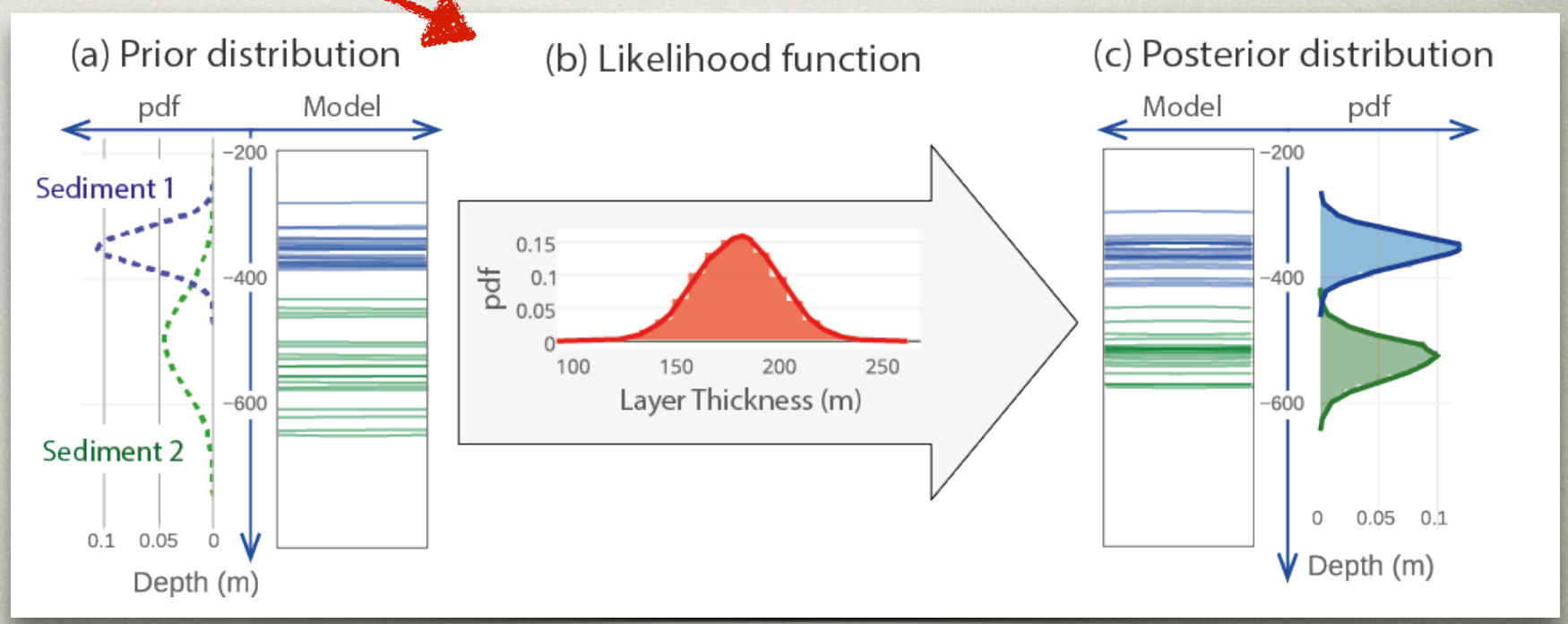
- How to estimate a possible uncertainty reduction?



Where next?

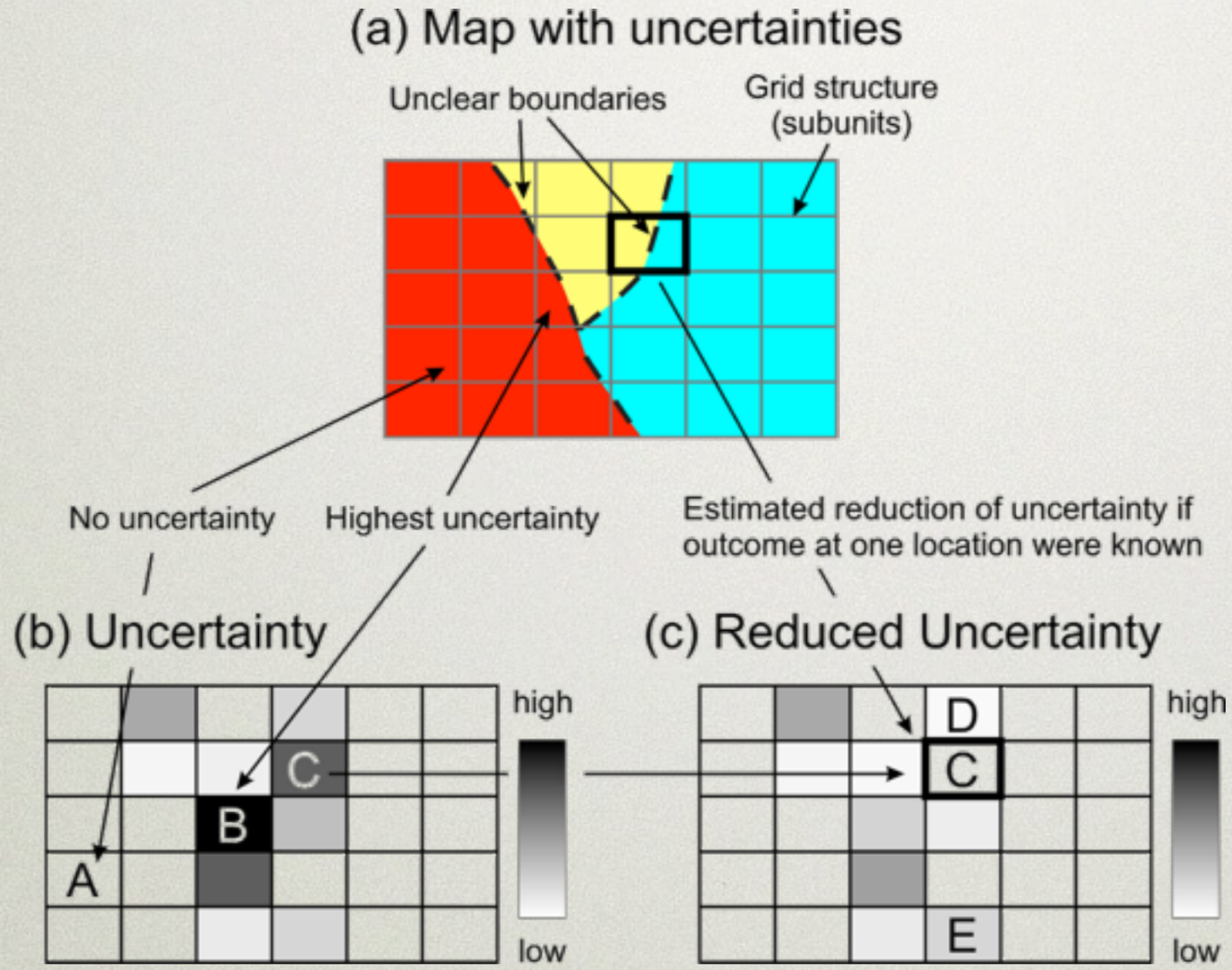
- Link to process simulations (hydrogeology, geothermal systems, …)
- How to constrain the model with more observations and additional data?
- How to estimate a possible uncertainty reduction?



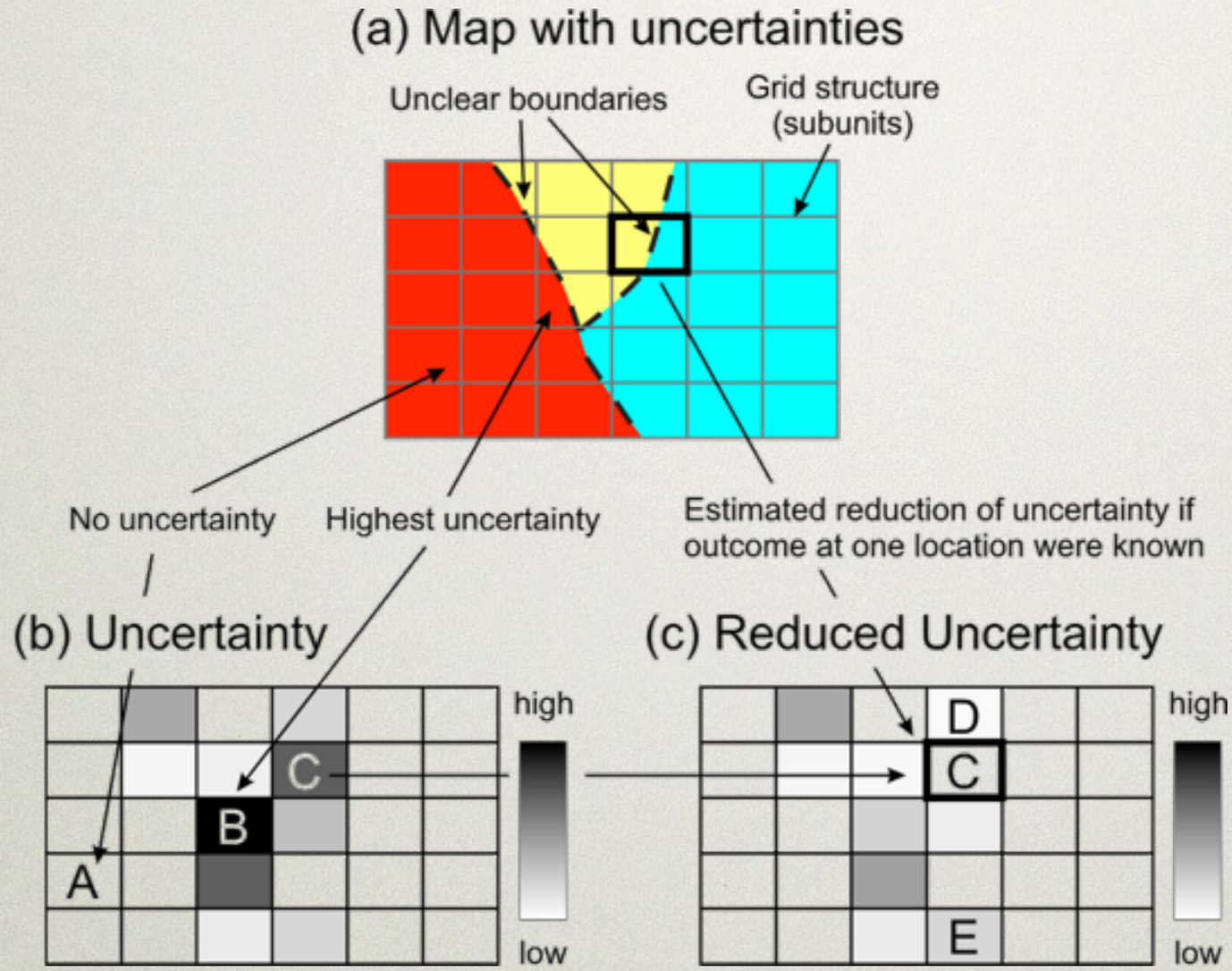


(aka Miguel's "Dirty Work")

Mutual Information and Conditional Entropy



Mutual Information and Conditional Entropy

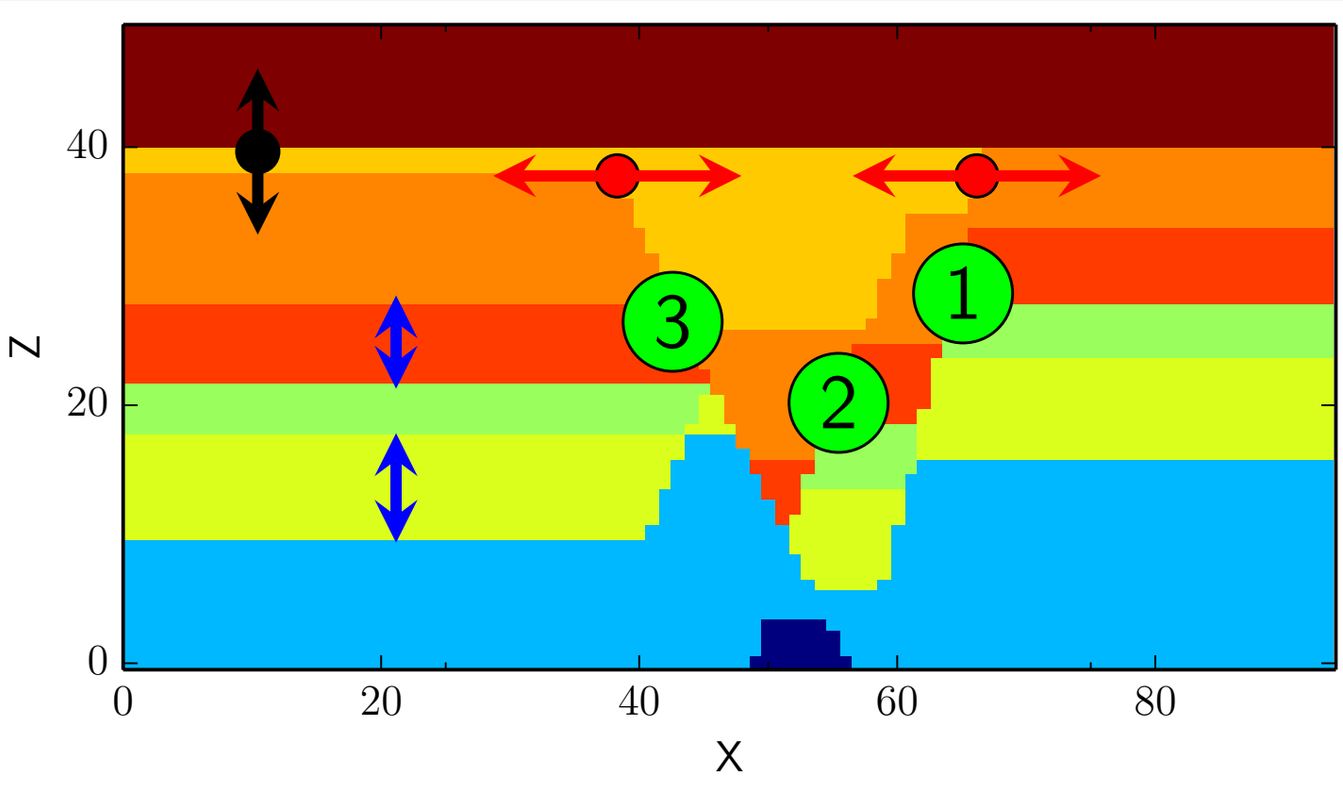


Multivariate Conditional Entropy

$$H(X_m|X_1, X_2, \ldots, X_n) = H(X_1, X_2, \ldots, X_n, X_m) - H(X_1, X_2, \ldots, X_n)$$

Slice in E-W direction and considered uncertainties

The **parameterisation** of the geological events **contains uncertainties**, and we consider here as uncertain:

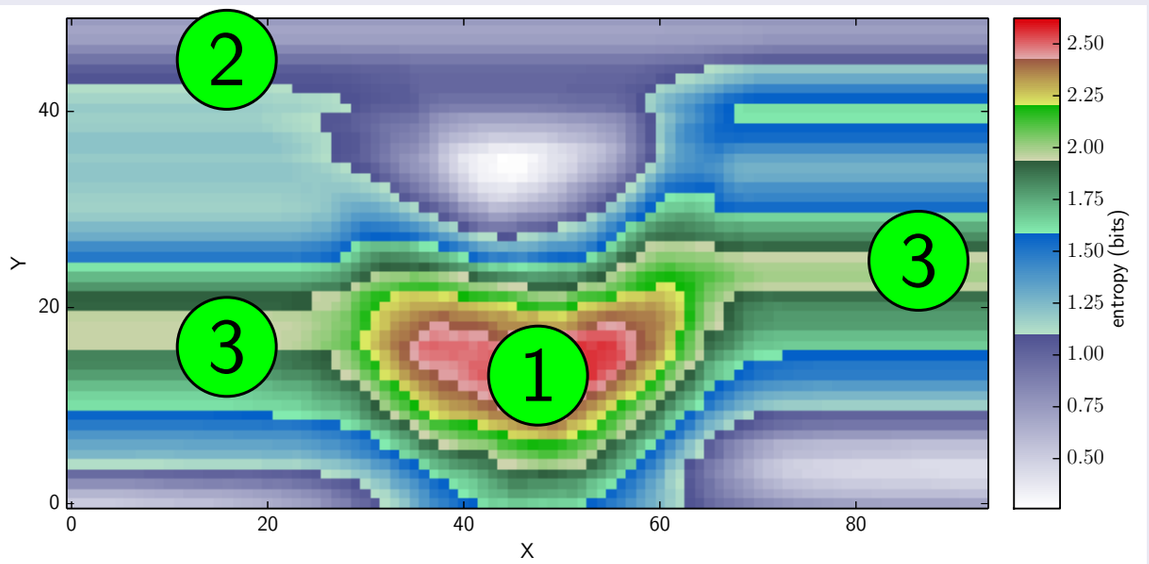


Parameters of geological history:

- Fault **positions** and dip **angle** (•)
- **Age relationship** (order) of faults (•)
- Unit **thickness** (•)
- **Position** of unconformity (•)

Analysis of information entropy

Visualisation of information entropy



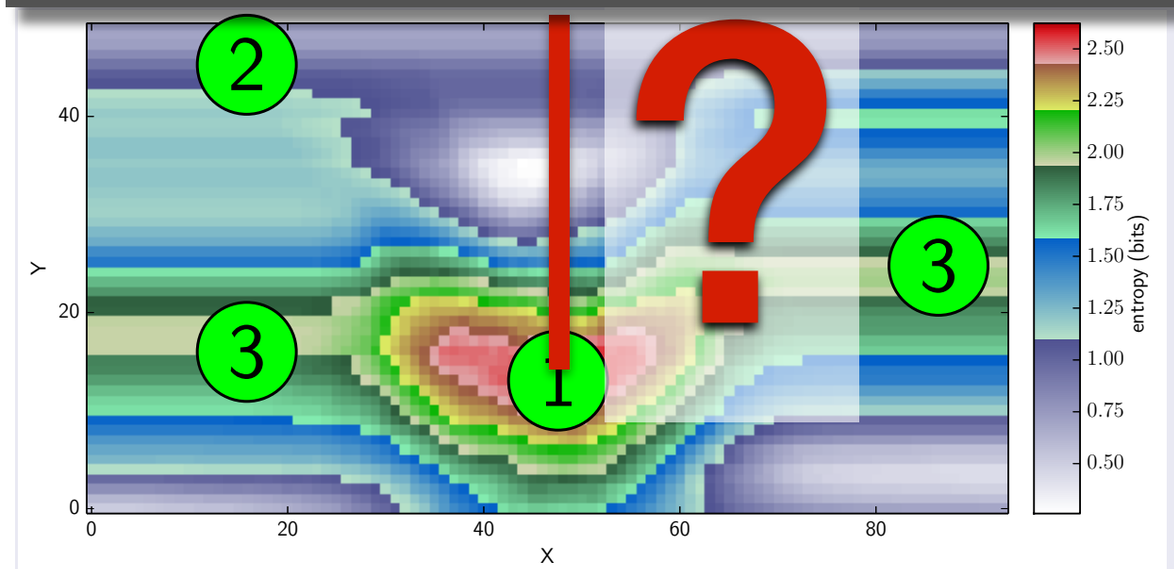
1. Uncertainties are highest in the deep parts of the basin;
2. At shallow depth, only uncertainty due to depth of unconformity;
3. In shoulders uncertainty due to stratigraphic layer thickness.

Entropy is calculated for each cell based on estimated unit probabilities with Shannon's equation:

$$H(X) = -\sum_{i=1}^{n} p_i(X) \log_2 p_i(X)$$

Analysis of information entropy

Assume: we would like to reduce uncertainties about layer position in basin: where to gather information?



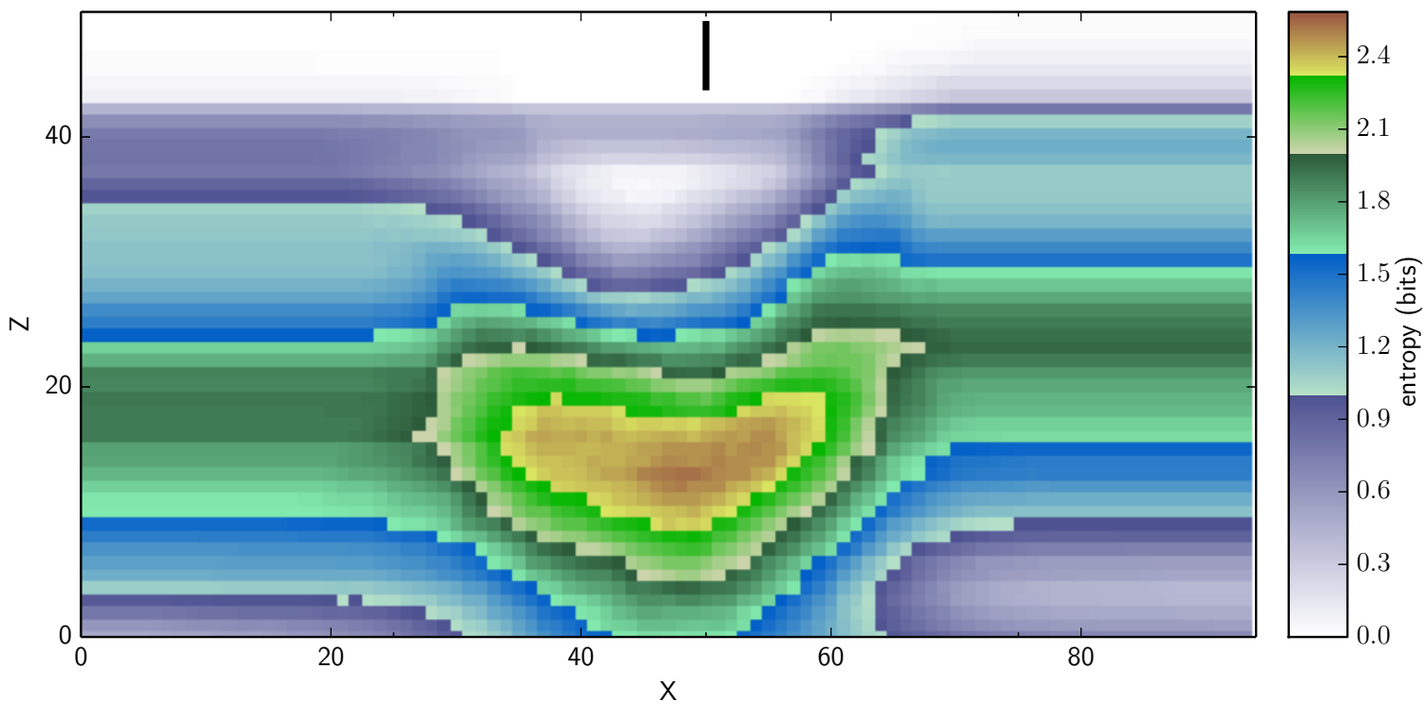
1. Uncertainties are highest in the deep parts of the basin;
2. At shallow depth, only uncertainty due to depth of unconformity;
3. In shoulders uncertainty due to stratigraphic layer thickness.

Entropy is calculated for each cell based on estimated unit probabilities with Shannon's equation:

$$H(X) = -\sum_{i=1}^{n} p_i(X) \log_2 p_i(X)$$

Uncertainty reduction with additional information

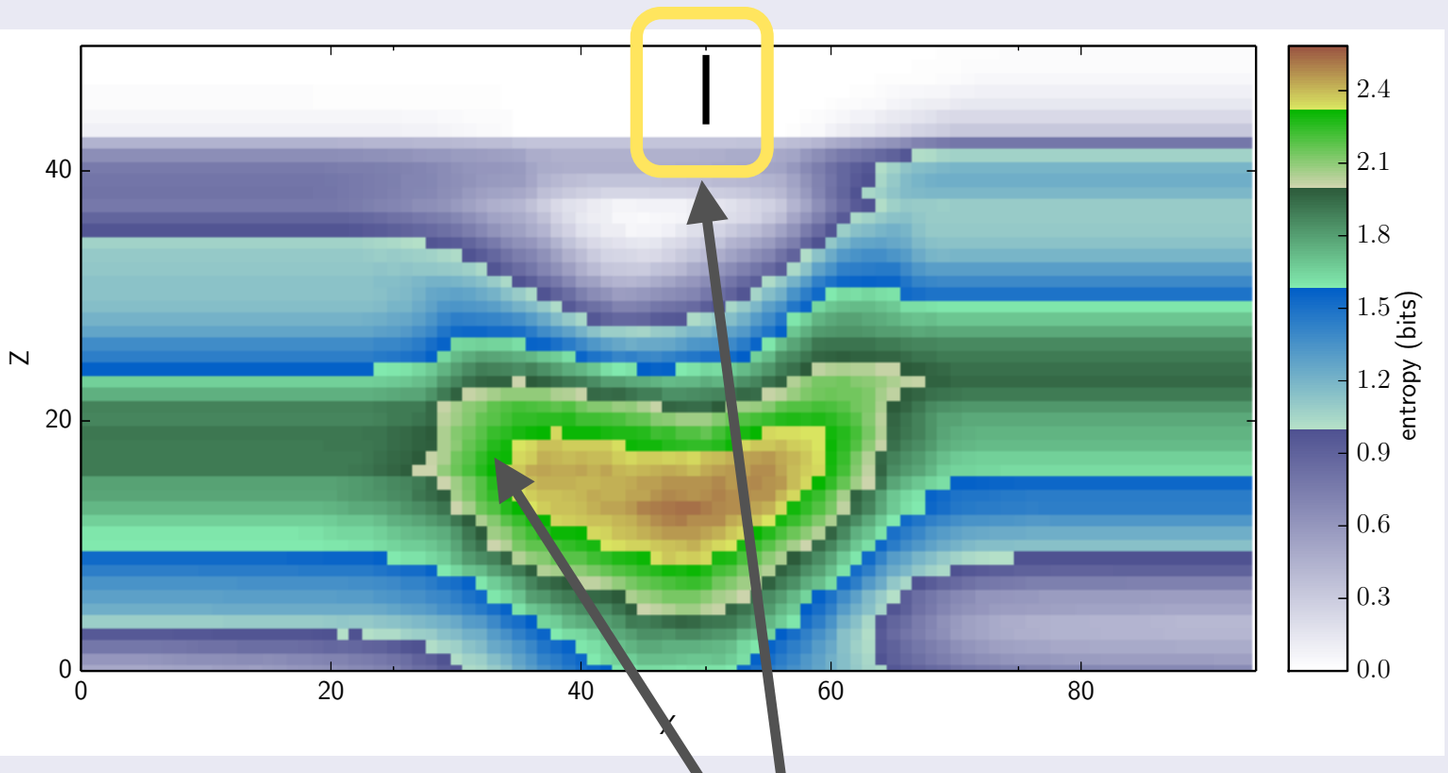
Gathering subsequent information at one location ("drilling"):



Conditional entropy of each cell given information at subsequent locations along a line ("drillhole"): **uncertainty** in the model is **reduced with new knowledge**.

Uncertainty reduction with additional information

Gathering subsequent information at one location ("drilling"):

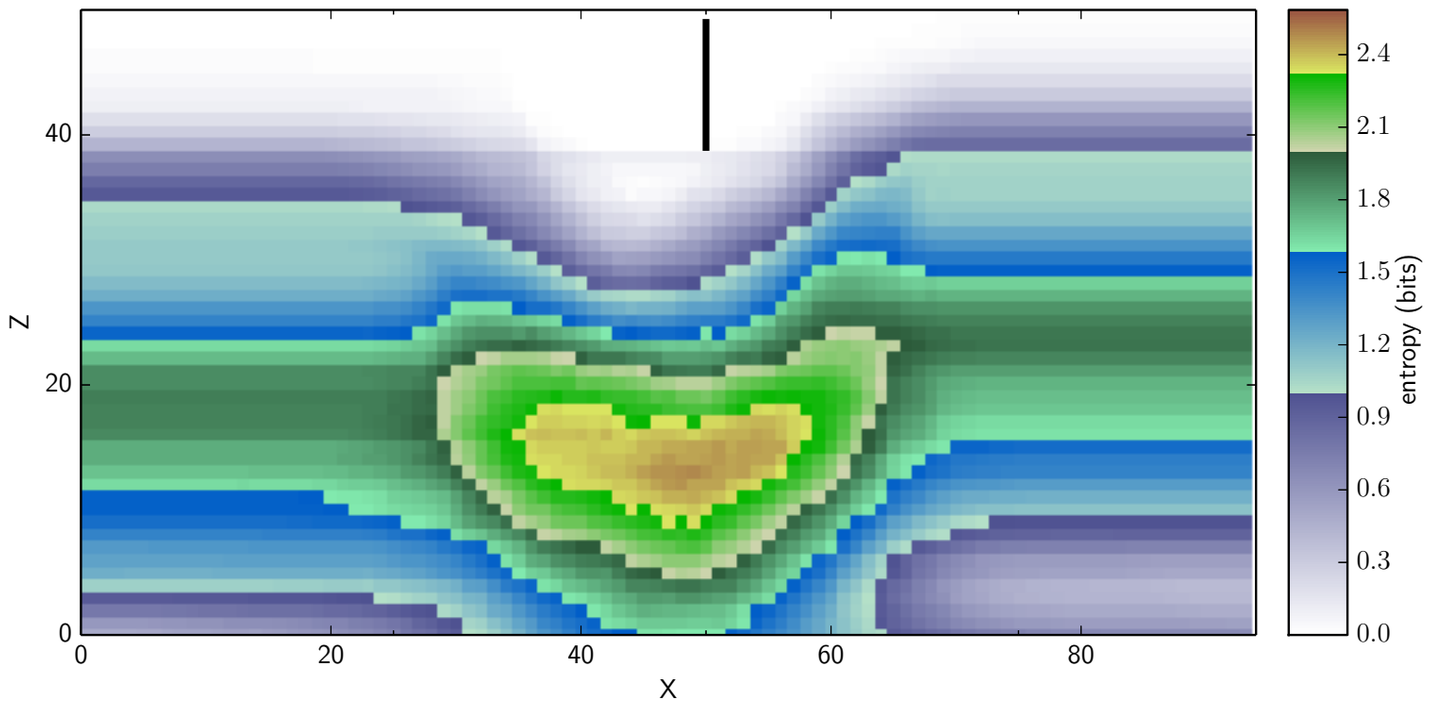


Conditional entropy of each cell given information at subsequent locations along a line ("drillhole"): **uncertainty** in the model is **reduced with new knowledge**.

Multivariate conditional entropy:
remaining uncertainty at each location,
given information of all points on line

Uncertainty reduction with additional information

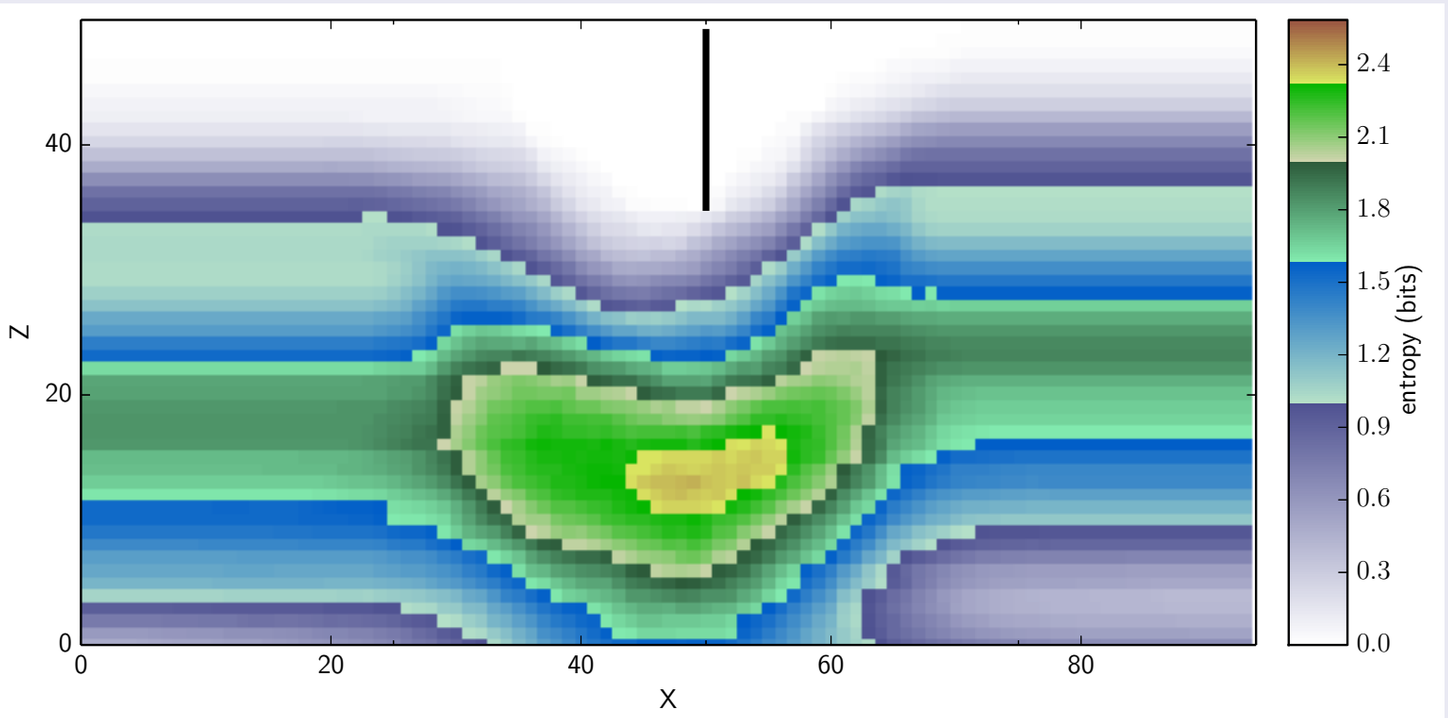
Gathering subsequent information at one location ("drilling"):



Conditional entropy of each cell given information at subsequent locations along a line ("drillhole"): **uncertainty** in the model is **reduced with new knowledge**.

Uncertainty reduction with additional information

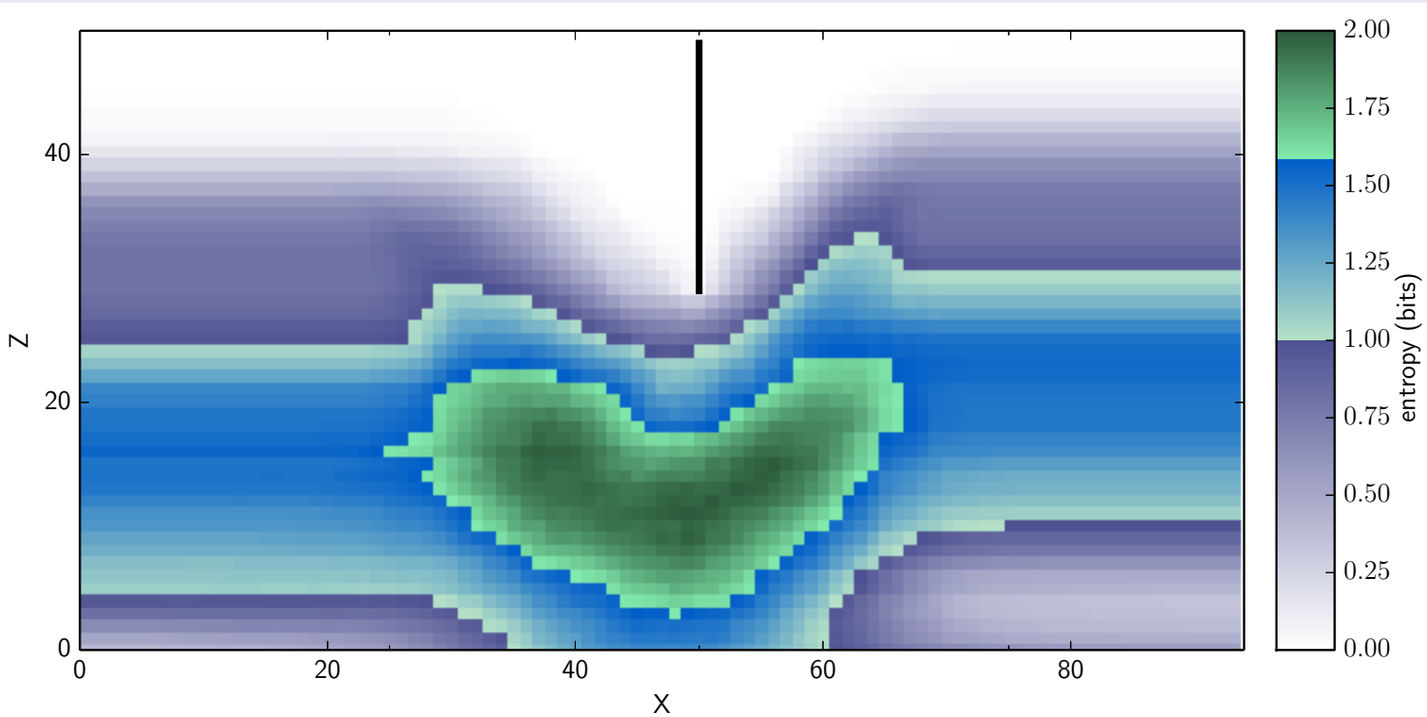
Gathering subsequent information at one location ("drilling"):



Conditional entropy of each cell given information at subsequent locations along a line ("drillhole"): **uncertainty** in the model is **reduced with new knowledge**.

Uncertainty reduction with additional information

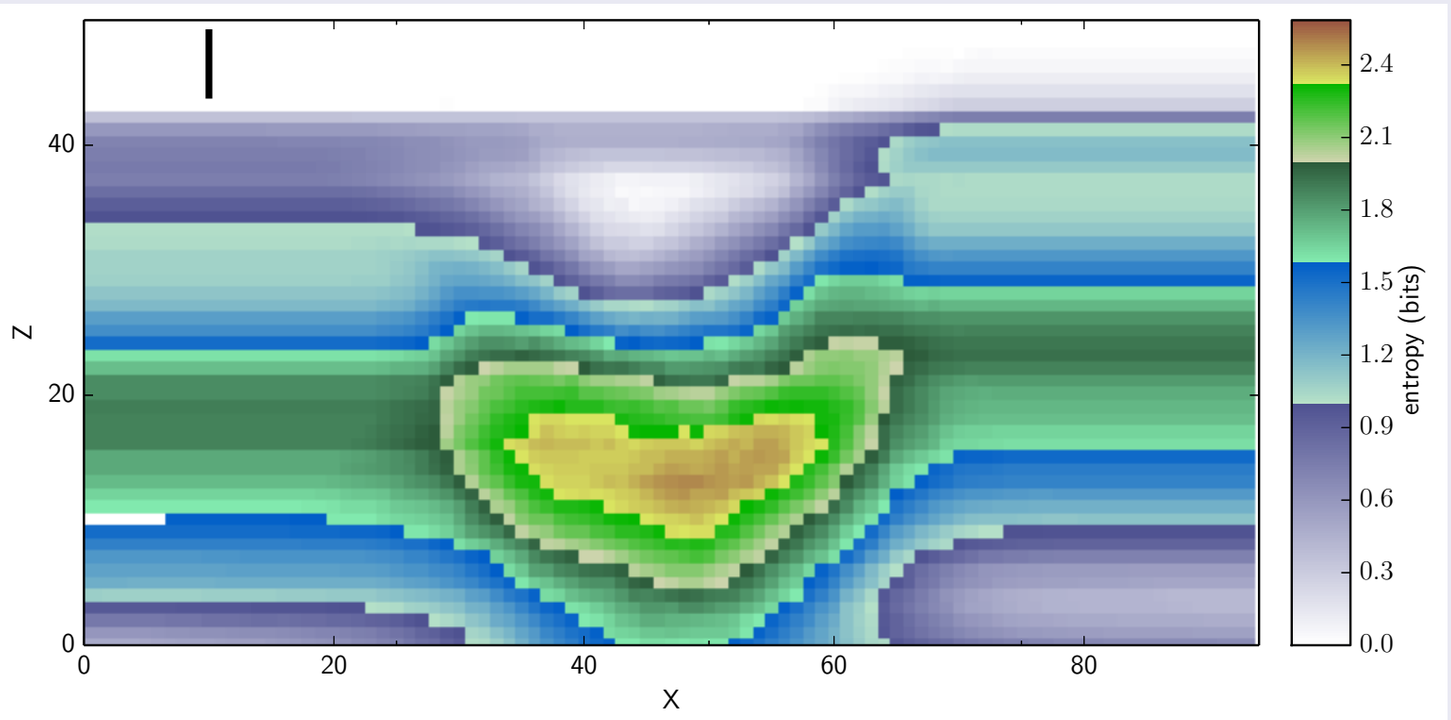
Gathering subsequent information at one location ("drilling"):



Conditional entropy of each cell given information at subsequent locations along a line ("drillhole"): **uncertainty** in the model is **reduced with new knowledge**.

Uncertainty reduction with additional information

Gathering subsequent information at one location ("drilling"):

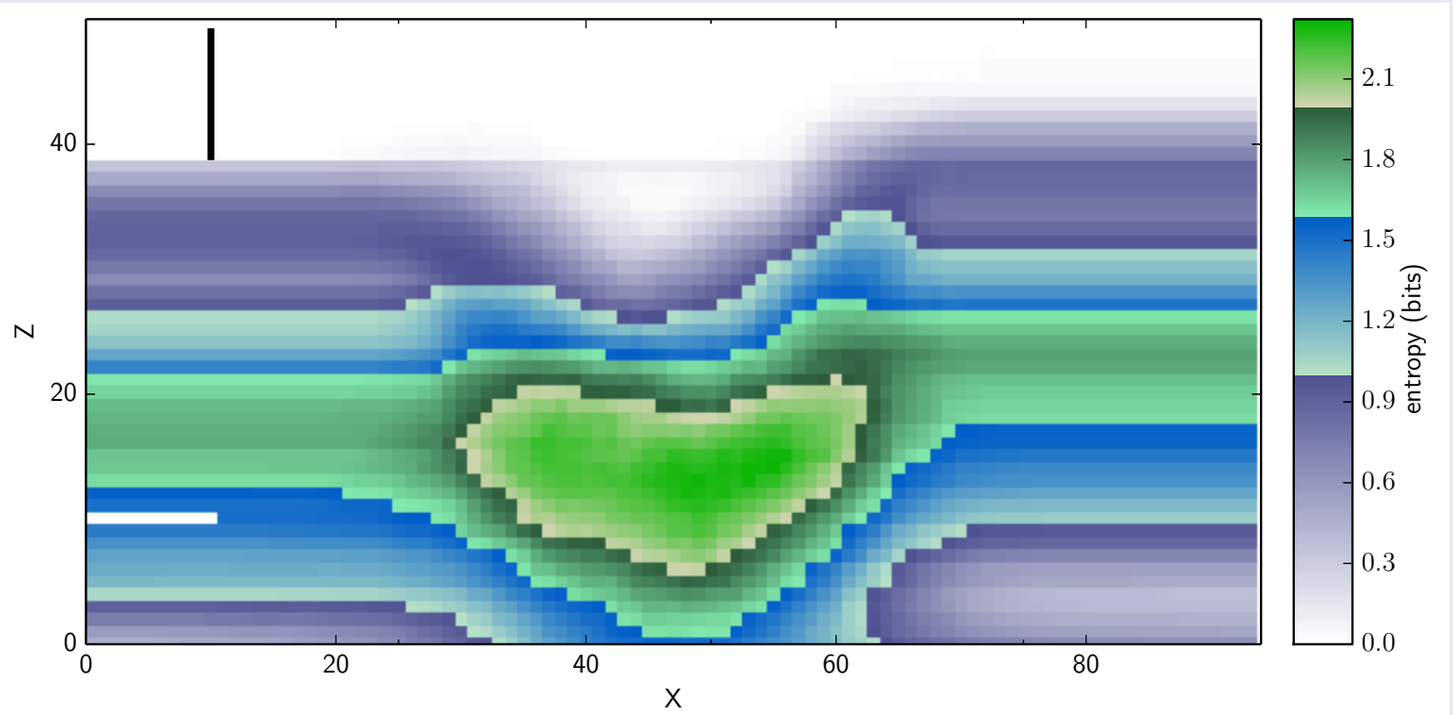


Conditional entropy of each cell given information at subsequent locations along a line ("drillhole"): **uncertainty** in the model is **reduced with new knowledge**.

Note the lateral reduction due to underlying structure.

Uncertainty reduction with additional information

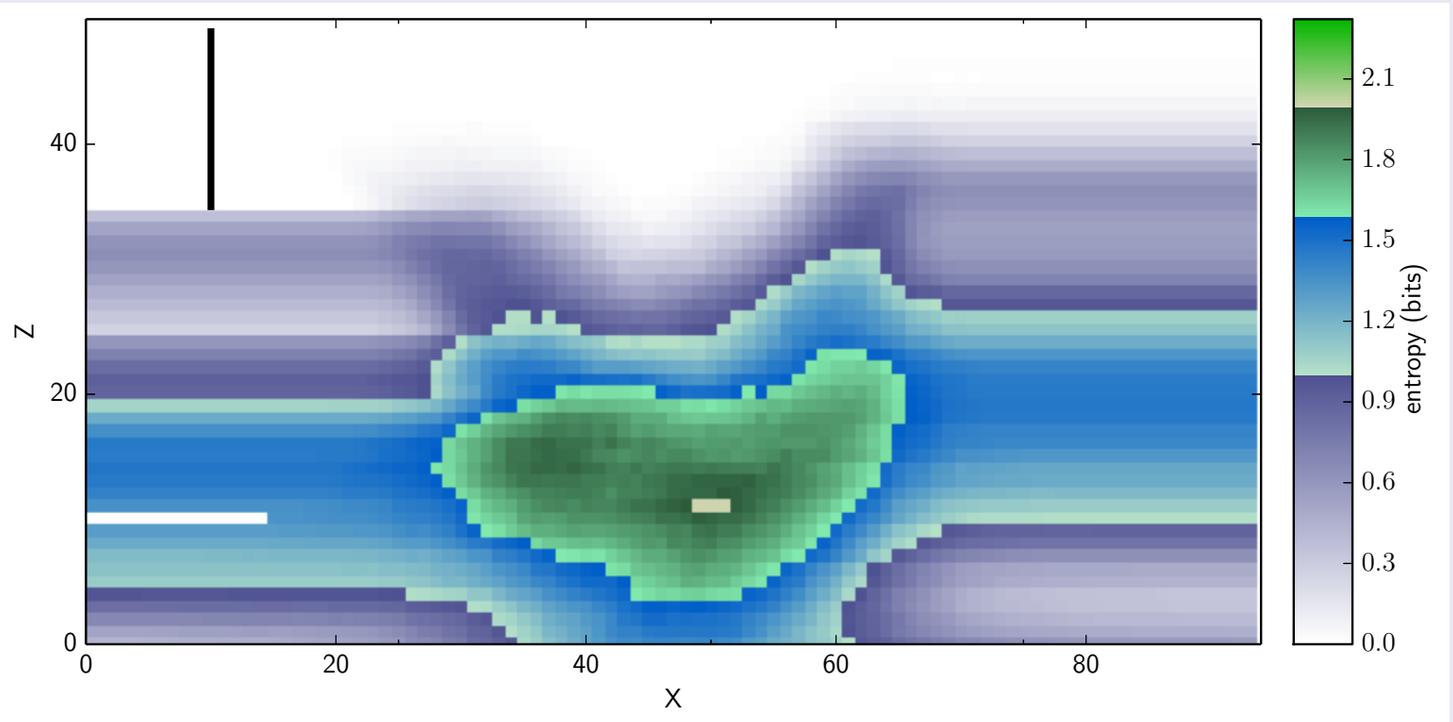
Gathering subsequent information at one location ("drilling"):



Conditional entropy of each cell given information at subsequent locations along a line ("drillhole"):
uncertainty in the model is **reduced with new knowledge**.
Note the lateral reduction due to underlying structure.

Uncertainty reduction with additional information

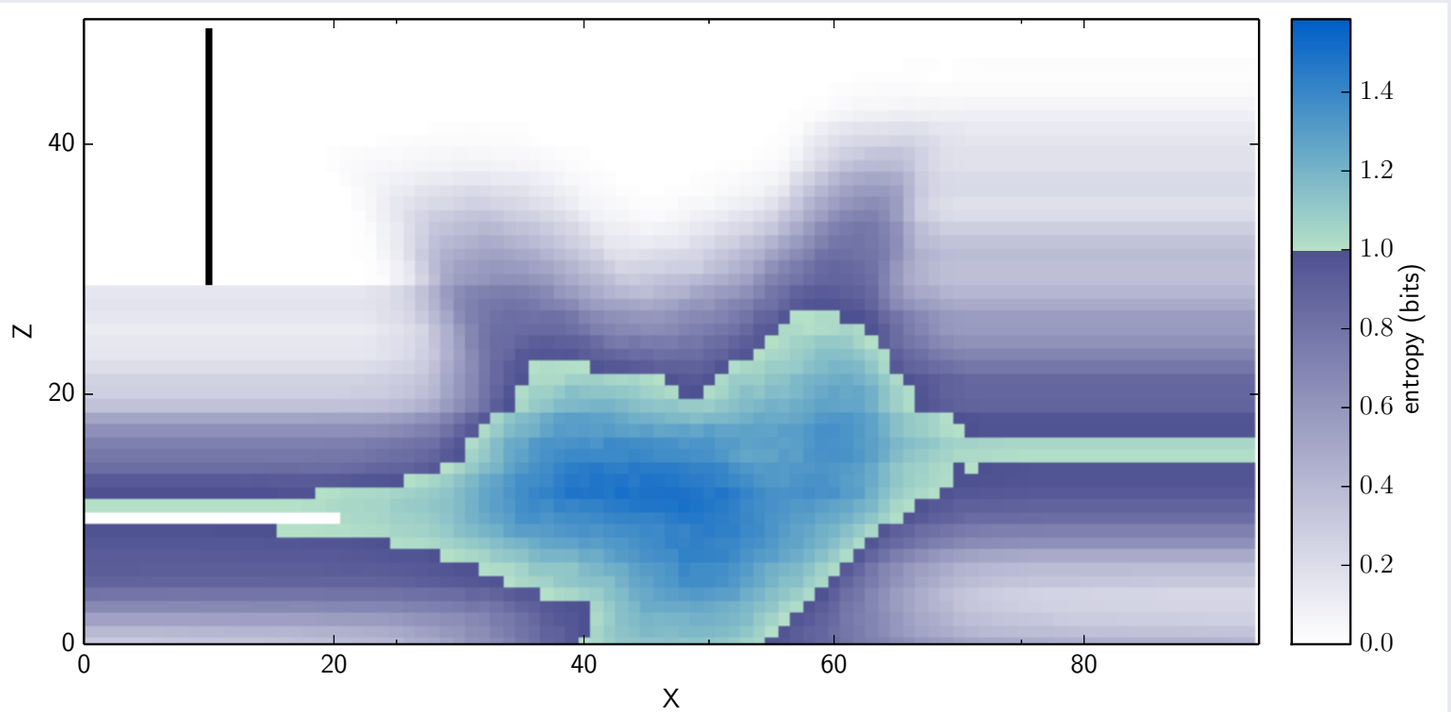
Gathering subsequent information at one location ("drilling"):



Conditional entropy of each cell given information at subsequent locations along a line ("drillhole"):
uncertainty in the model is **reduced with new knowledge**.
Note the lateral reduction due to underlying structure.

Uncertainty reduction with additional information

Gathering subsequent information at one location ("drilling"):

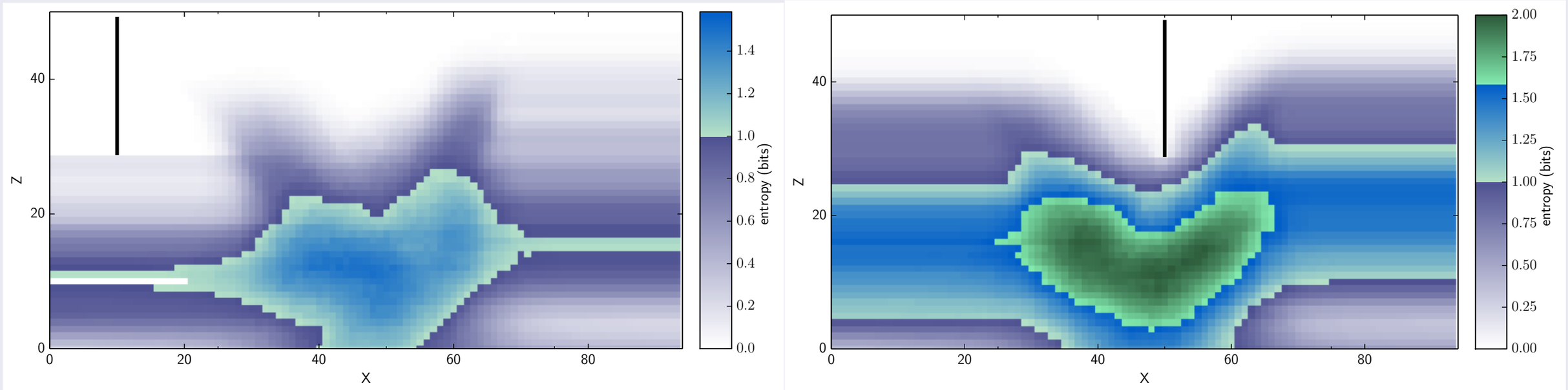


Conditional entropy of each cell given information at subsequent locations along a line ("drillhole"):
uncertainty in the model is **reduced with new knowledge**.
Note the lateral reduction due to underlying structure.

Comparison of "drillhole" positions

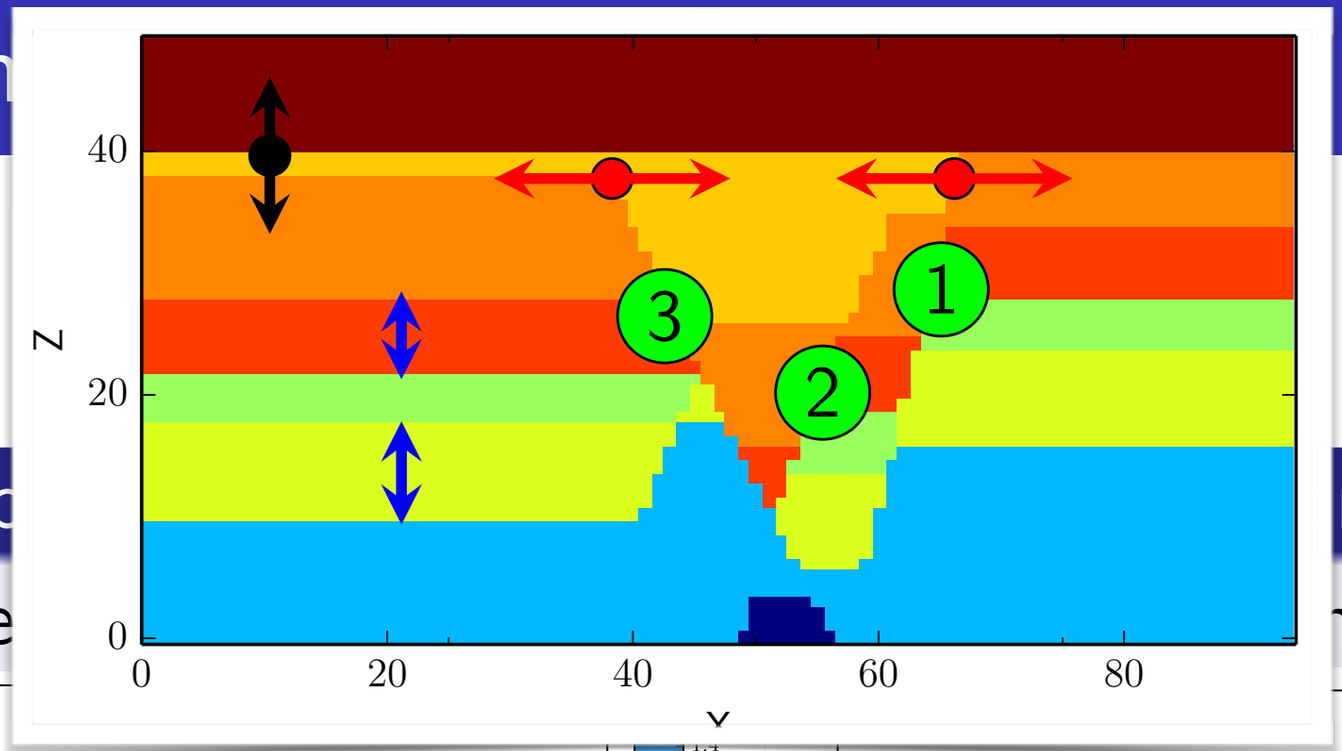
Comparison of remaining uncertainty for different drillhole positions

The difference is **clearly visible** when we compare both results:



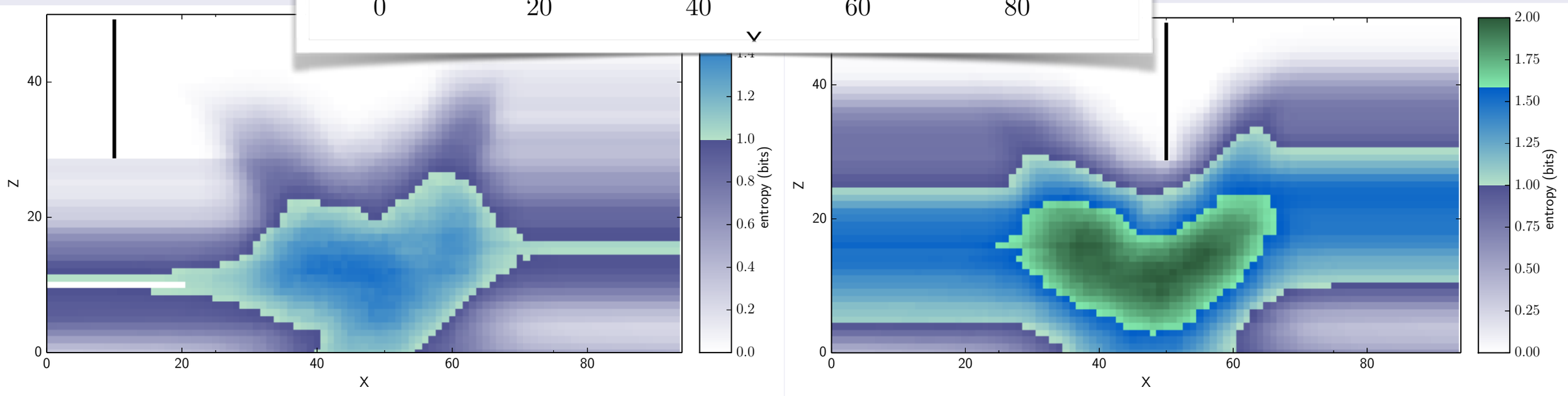
This analysis can give us an insight **where additional information** can be **expected to reduce uncertainties**.

Comparison



Comparison c rillhole positions

The difference n results:



This analysis can give us an insight **where additional information** can be **expected to reduce uncertainties**.

More Ideas after this workshop...

- Examine "structure" of the joint probability tables (and the relationship to the underlying model structure)
- Deeper exploration of information correlations (joint entropy for entire model?) -> how much "information" does the entire model contain?
- AIT: test model compression for several synthetic models?
- Apply MaxEnt to model inference problem

Summary

- Methods from information theory:
  - Information entropy for analysis of uncertainties
  - Conditional entropy and mutual information for analysis of correlation and uncertainty reduction
- Reliability filters for geological models (“sanity checks”) to consider additional geological knowledge -> geological modelling as an inference problem
- Methods enable us to learn about geological parameter correlations and information correlation in the subsurface