Feeding Machine Learning Algorithms with Massive Model Suites

Leo Portes[1,2], Mark Jessell[1,2,3], Mark Lindsay[2,4], Guillaume Pirot[1,2,3], Michel Nzikou[1,3], Ed Cripps[1,2]

[1]UWA, [2]ITTC DARE, [3]MinEx CRC, [4]CSIRO

MinEx CRC

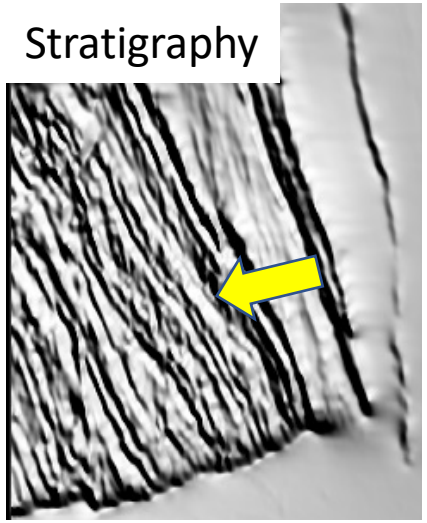
Loop

DARE
ARC Training Centre in
Data Analytics for Resources & Environments

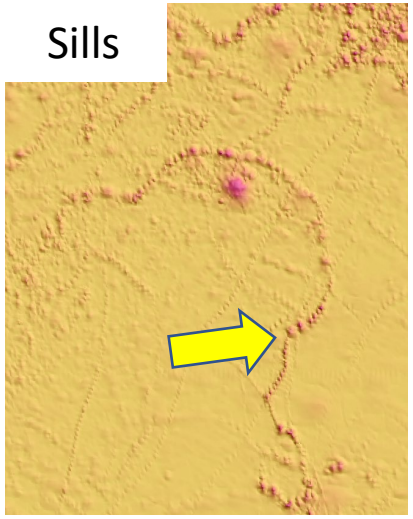
THE UNIVERSITY OF WESTERN AUSTRALIA

The challenge: Extracting geological meaning from geophysical data, e.g. Single or multiple linear features

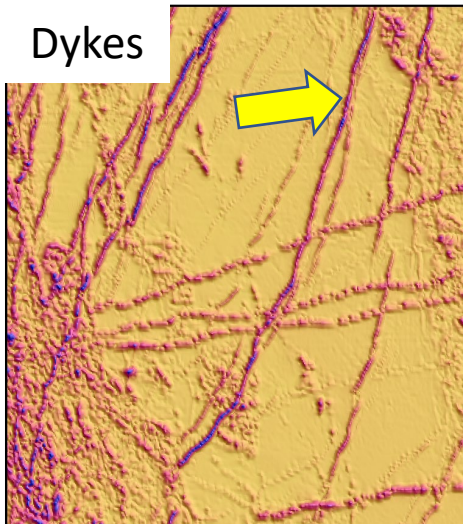
Stratigraphy



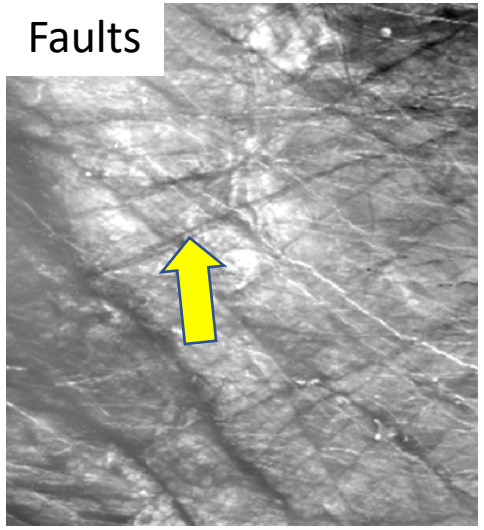
Sills



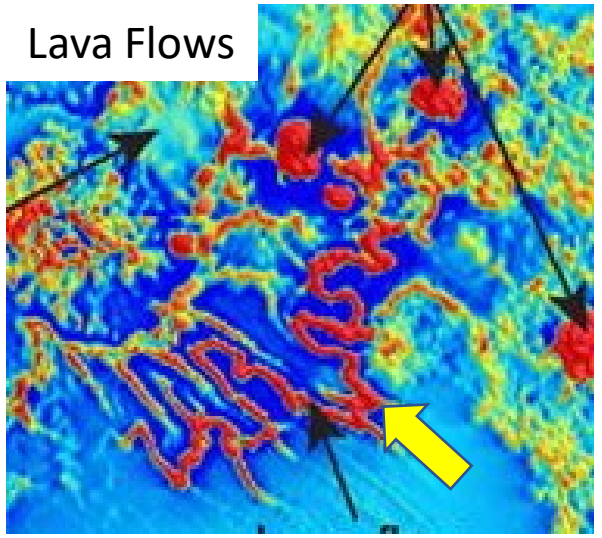
Dykes



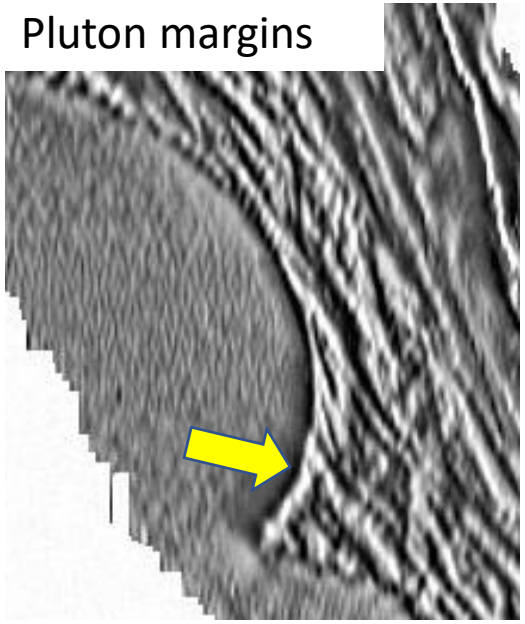
Faults



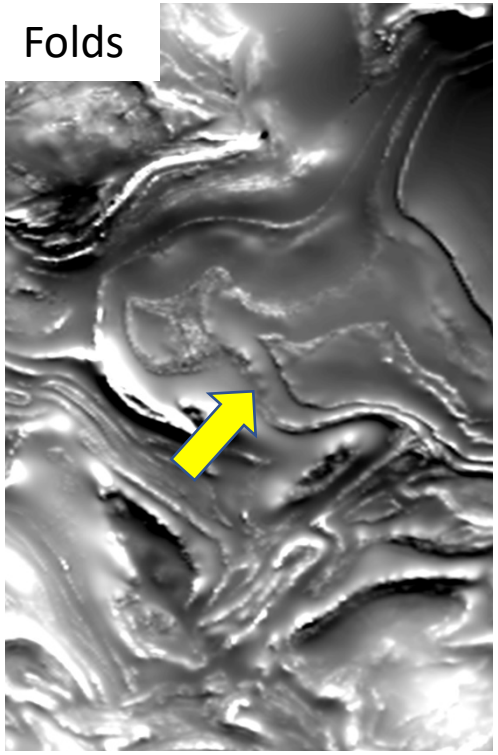
Lava Flows



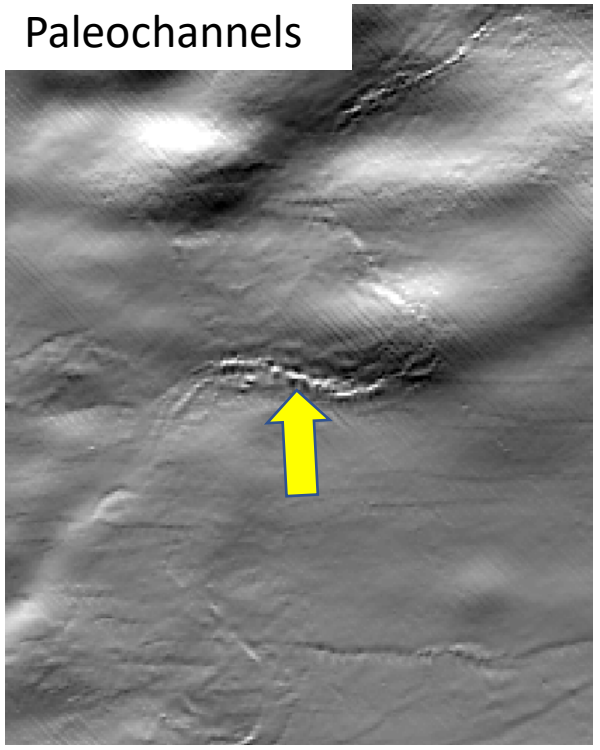
Pluton margins



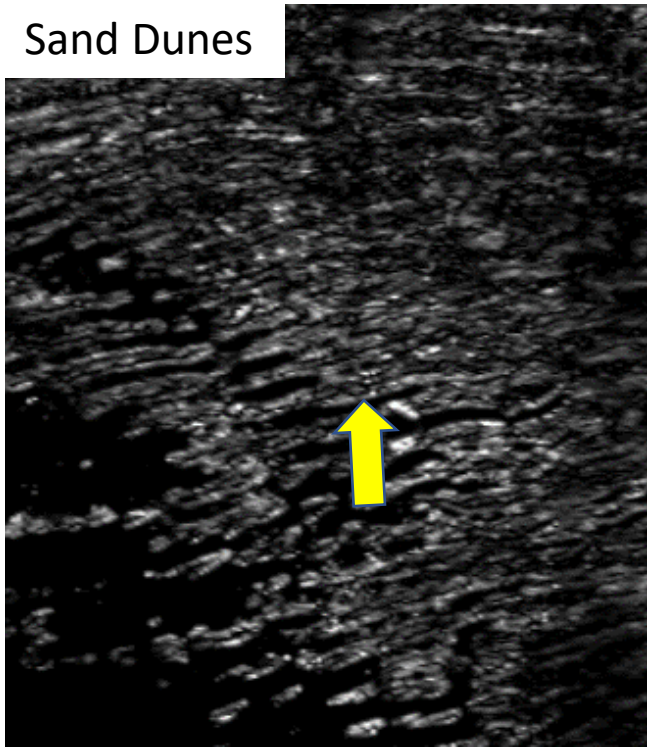
Folds



Paleochannels

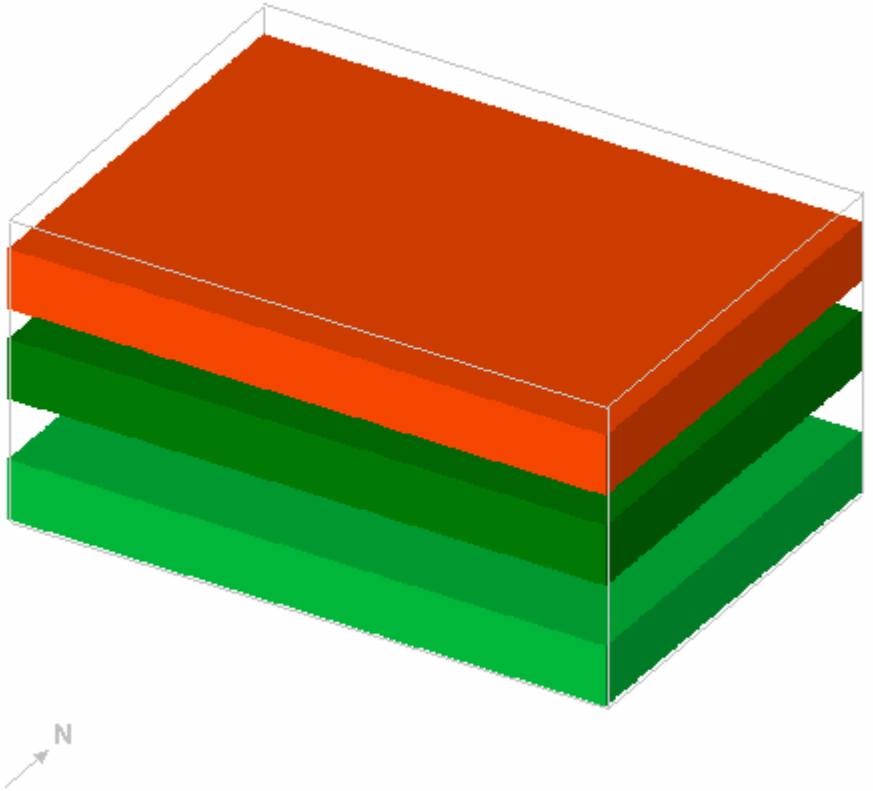


Sand Dunes



Noddy

- Semantic/kinematic/implicit modelling engine
- Calculates 3D models based on an idealised description of the kinematics of a sequence of geological events
- Density, magnetic susceptibility (K_{ij}) and remanence properties assigned to each unit so we can calculate gravity and magnetic responses

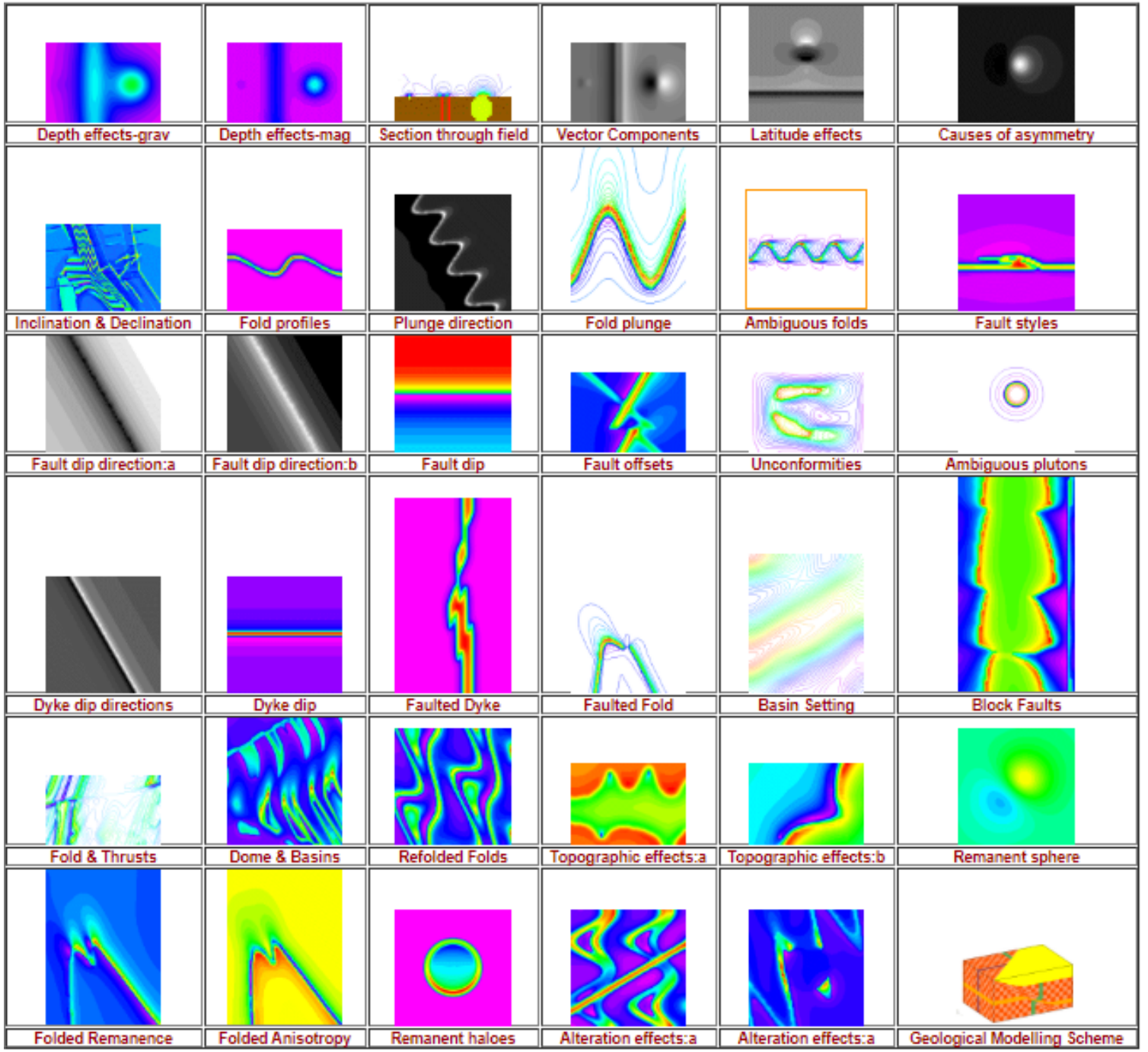


Jessell, 1981

Atlas of Structural Geophysics

http://tectonique.net/asg

Human learning dataset



Farrell et al., 1996
Genetic Algorithm sampling of Noddy model space for geophysical inversion

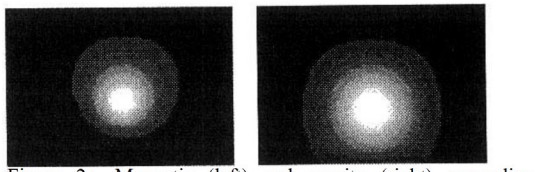


Figure 2a. Magnetic (left) and gravity (right) anomalies of the target spherical plug.

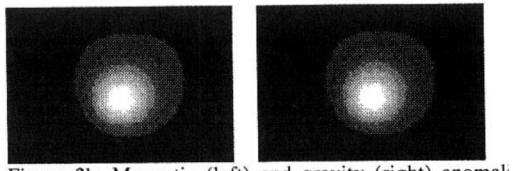


Figure 2b. Magnetic (left) and gravity (right) anomal of a model spherical plug with correlation coefficie 0.992.

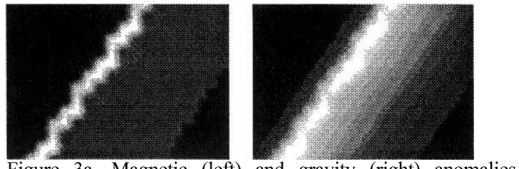


Figure 3a. Magnetic (left) and gravity (right) anomalies of the target dike.

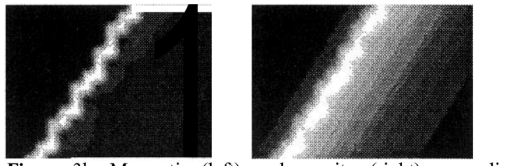
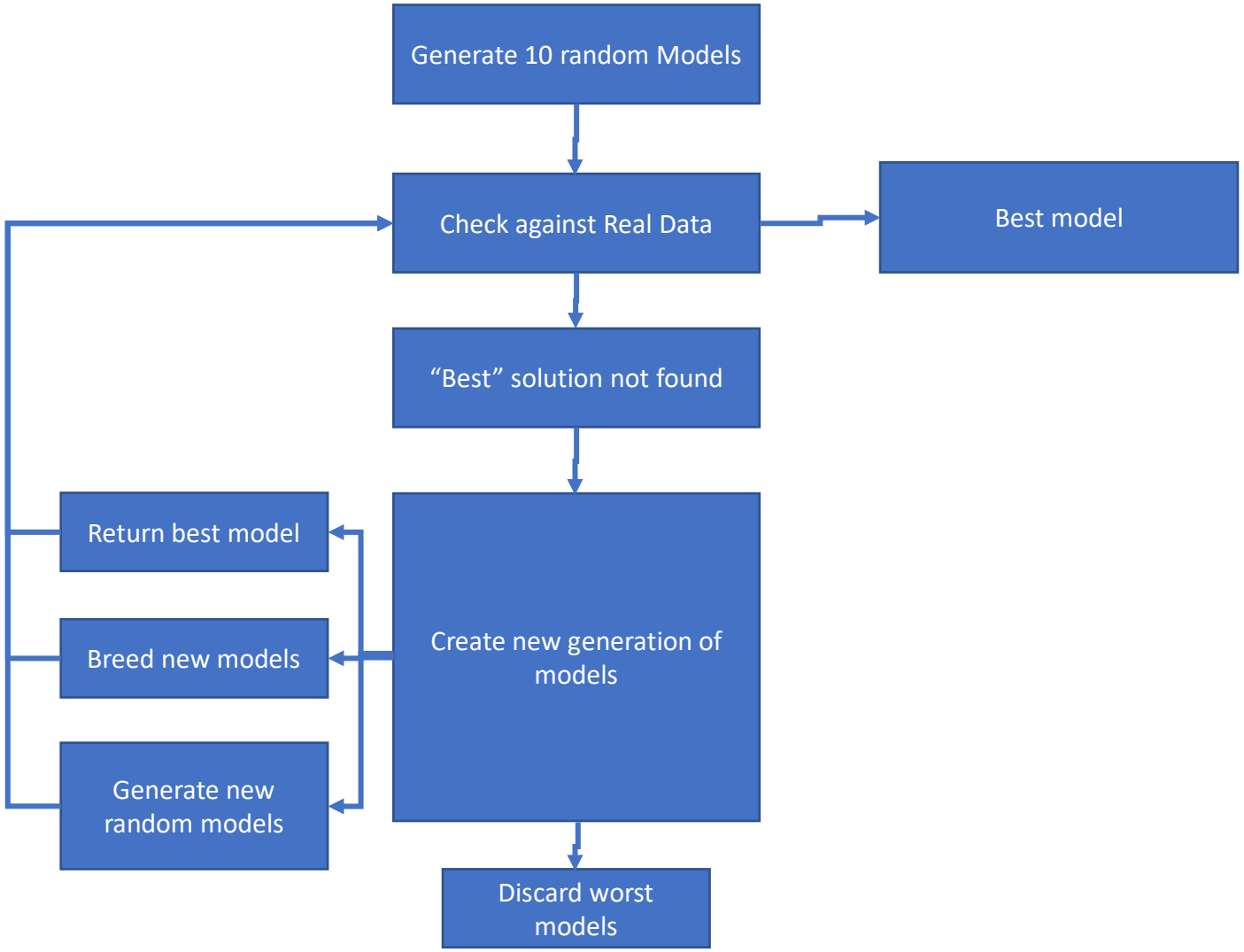


Figure 3b. Magnetic (left) and gravity (right) anomalies of a model dike with correlation coefficient 0.82.



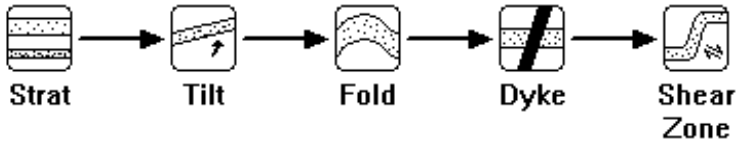
How to make a random model…

Base Stratigraphy + Tilt + 3 random events with random spatial & orientation parameters

= 343 event orders (e.g. STRAT-TILT-FOLD-DYKE-SHEA)
~ 60-70 independent variables per history

Lots of arbitrary decisions:

- Size of model (4k x 4k x 4k)
- Size of voxels (20m x 20m x 20m)
- Number of random events (3)
- Start with Strat & Tilt event
- Range and Probability Density Function of parameters for each event
  - *What range of wavelengths do folds form in nature?*
  - *How wide are mafic dykes usually, Felsic dykes? Petrophysically coherent stratigraphic units?*



Allowable events:

- Fold
- Fault
- Unconformity
- Shear Zone
- Dyke
- Plug
- Tilt

Each model and associated geophysics takes ~5-10 seconds to run on a multi-core laptop

200,000 cpu hours would allows us to make 72M models

Wellmann et al. 2016 *pynoddy*

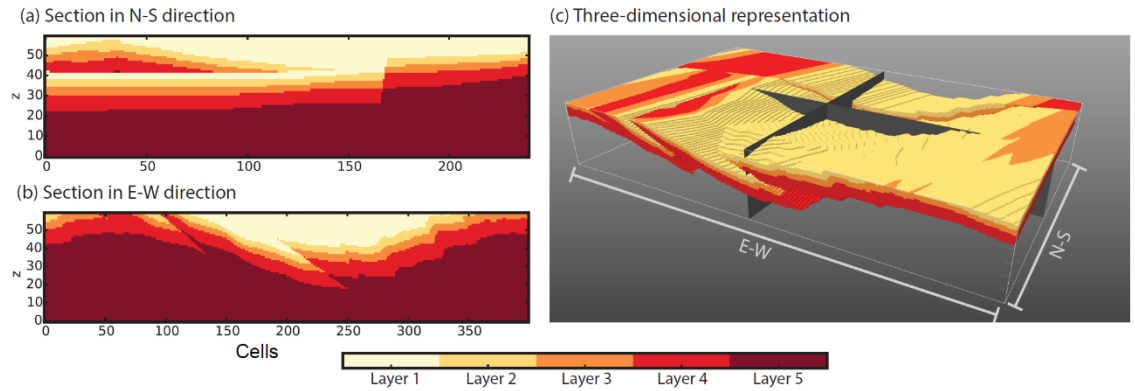


Figure 4. Sections through the fold and thrust belt model in **(a)** north–south direction, and **(b)** east–west direction (vertical exaggeration of 1.5) through the centre of the model. **(c)** Three-dimensional representation for the central three layers of the fold and thrust belt model. The grey surfaces correspond to the location of the sections in the figure above.

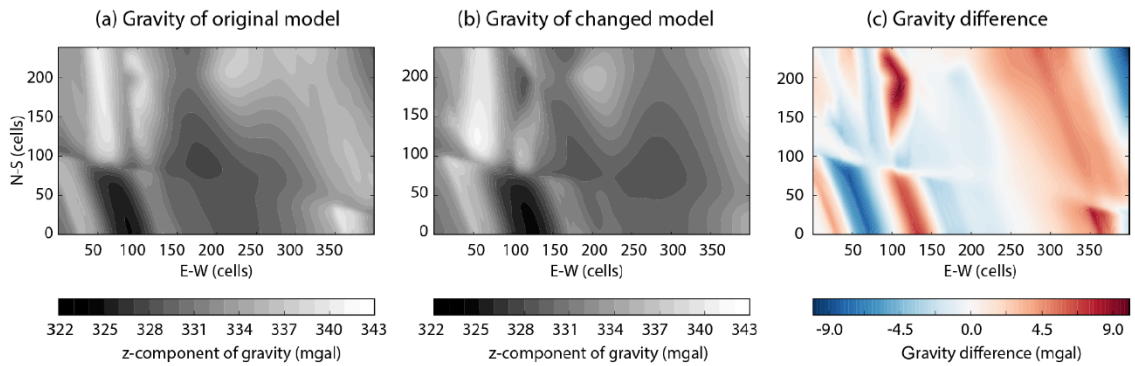


Figure 5. Evaluation of the effect of a wavelength change in a late folding event on the forward calculated gravity field: **(a)** gravity field of original model, **(b)** gravity field of model with changed event parameters, and **(c)** difference plot of gravity fields.

Thiele et al 2016. topological uncertainty

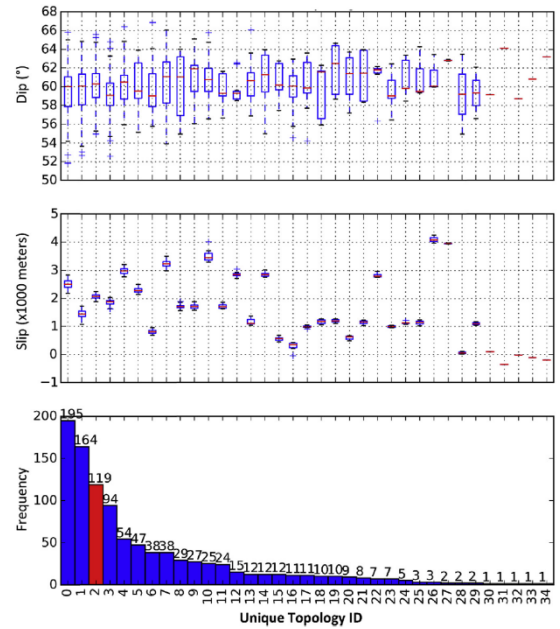


Fig. 4. Boxplots showing the range of parameter space occupied by models with different topological configurations. The red bar in each boxplot represents the median value of each parameter, the box is the interquartile range and the whiskers are the total range. The bottom plot shows frequency of each topology class. The topology class to which the original model belongs is shaded red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

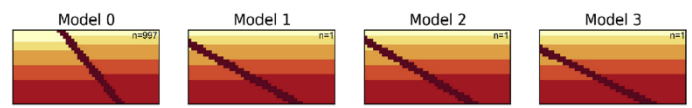


Fig. 5. The four different topological configuration observed in the *Dyke_d* model suite and the number of times (n) that they were observed. Note that all models would be identical if the boundaries were enlarged.

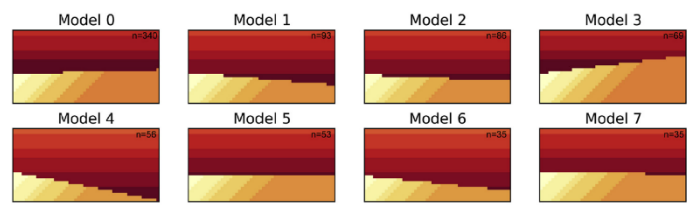


Fig. 6. Examples of the eight most frequently observed topologies resulting from model *unconf_de* and the number of times (n) they were observed. Note how the variations in unconformity dip changes the thickness of the unit immediately above the unconformity, resulting in many different contact relationships, and hence different topologies.

Karpatne et al. (2017) summarise the principal challenges to applying Machine Learning to the geosciences:

1. **Spatio-temporal structure**
2. **High dimensionality**
3. **Paucity of ground truth**
4. Objects with amorphous boundaries
5. Heterogeneity in space and time
6. Interest in rare phenomena
7. Multi-resolution data
8. Noise, incompleteness and uncertainty in data

Karpatne et al. (2017) Machine Learning for the Geosciences: Challenges and Opportunities. IEEE Transactions on Knowledge and Data Engineering 10.1109/TKDE.2018.2861006

As reported by CNN, there are estimated to be 6 Billion images of cats indexed by Google*

*reliable source


National Geographic
Domestic cat


Encyclopedia Britannica
Cat | Breeds & Facts | Britannica


Wikipedia
Cat - Wikipedia


The Atlantic
How Humans Created Cats - The Atlantic


Wikipedia
Tabby cat - Wikipedia


NBC News
and an invasive alien species


Wikipedia
Kitten - Wikipedia


BBC
Why do we think cats are unfriendly ...


Encyclopedia Britannica
Cat | Breeds & Facts | Brit...


Good Housekeeping
25 Best Cat Instagram Captions - Short ...


The Atlantic
How Cats Used Humans to Conquer the ...


The Conversation
Australian shelters and pounds kill 50 ...


Feliway
How Much Space Does a Cat Need ...


The Guardian
cats' right to roam? | Wil...

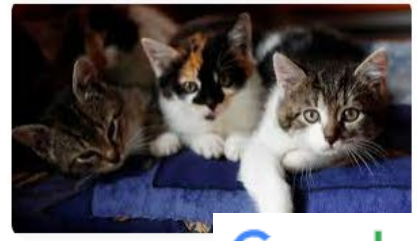

Reader's Digest
Where Do Cats Like to Be Petted — Ho...


Natural History Museum
Fantastic felines: a world of wild cats ...


DC News Now | Washington, DC
Cats classified as 'invasive alien ...


National Geographic Kids
Moment of Meow: Goofy Cats


The Guardian
The inner lives of cats: what our ...


The New York Times
Your Cat Might Not Be I

Google

Google Image search has ~700 images of "geological folds"

Google has 0 images of “geological folds in magnetic data”

Atlas of
Structural Geophysics II

contact
Cover

Table of Contents | Help | Image Index | Previous | Next

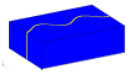
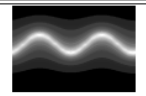
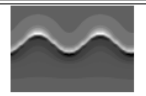
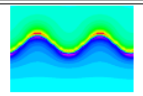
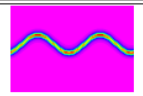
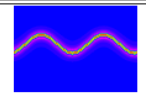
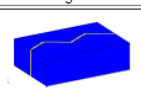
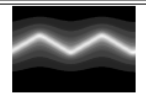
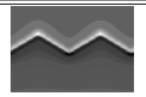
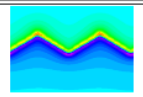
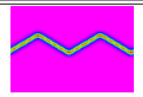
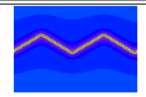
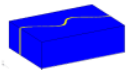
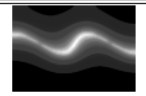
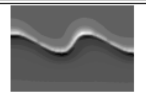
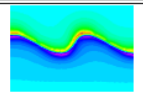
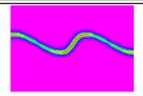
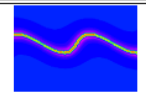
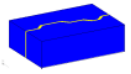
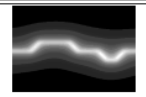
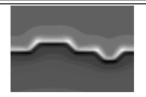
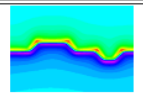
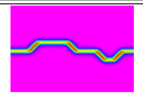
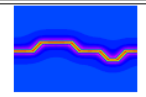
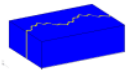
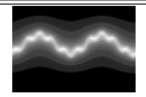
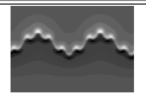
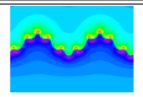
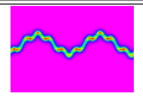
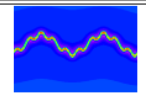
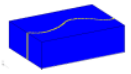
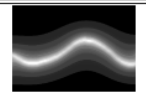
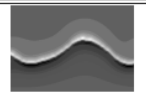
2.1.1 Variation in fold profile

This sequence shows the affect of varying the fold profile geometry for a 200 m thick layer.

All block diagrams are viewed from SW.

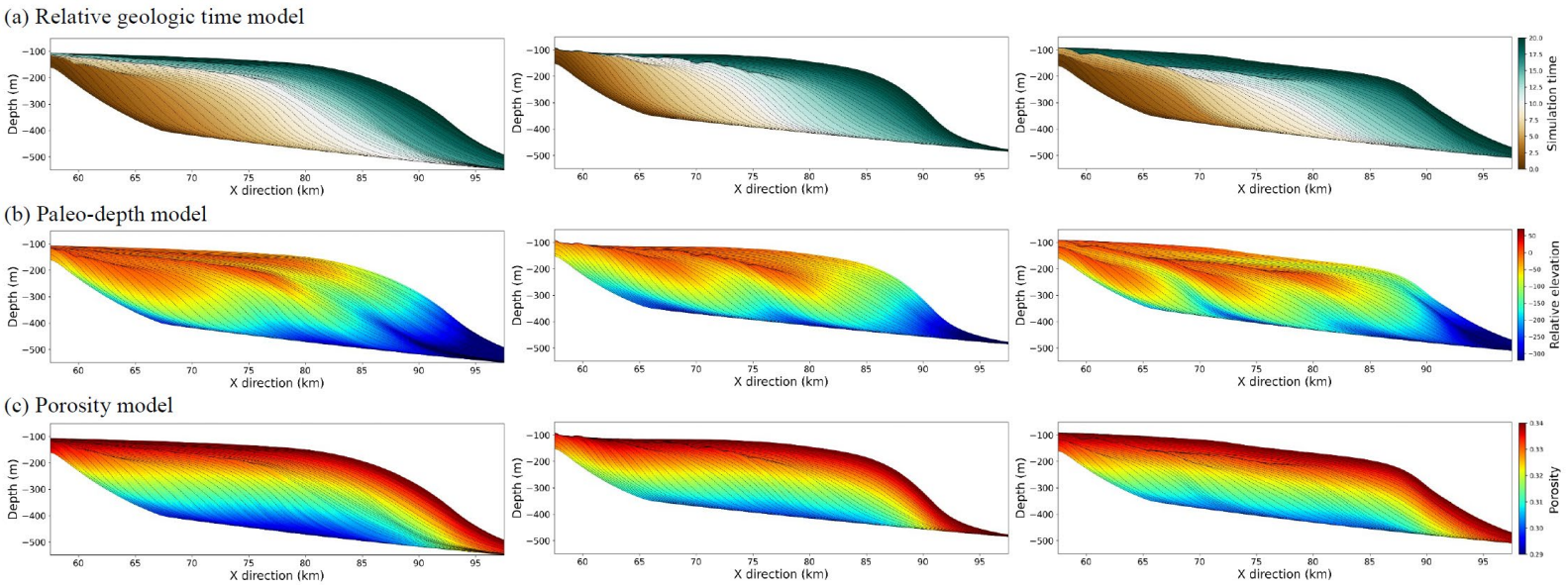
Click on the images to launch **Noddy**. A **legend** is provided at the end of this page.

tectonique.net/asg

Link	Block	Gravity	Magnetics	Magnetics	Magnetics at Pole	Magnetics at Pole 1VD
Sinusoidal Folds 	 LOAD JPEG IMAGE	 LOAD JPEG IMAGE	 LOAD JPEG IMAGE	 LOAD JPEG IMAGE	 LOAD JPEG IMAGE	 LOAD JPEG IMAGE
This model was generated using a similar fold style, so there is some hinge thickening relative to the limbs.						
Chevron Folds 	 LOAD JPEG IMAGE	 LOAD JPEG IMAGE	 LOAD JPEG IMAGE	 LOAD JPEG IMAGE	 LOAD JPEG IMAGE	 LOAD JPEG IMAGE
Asymmetric Folds 	 LOAD JPEG IMAGE	 LOAD JPEG IMAGE	 LOAD JPEG IMAGE	 LOAD JPEG IMAGE	 LOAD JPEG IMAGE	 LOAD JPEG IMAGE
The west limbs of this fold are structurally thickened.						
Box Folds 	 LOAD JPEG IMAGE	 LOAD JPEG IMAGE	 LOAD JPEG IMAGE	 LOAD JPEG IMAGE	 LOAD JPEG IMAGE	 LOAD JPEG IMAGE
These folds are actually formed as a result of conjugate kink sets, and results in the thickening of the limbs relative to the hinges.						
Anticlinoriums & Synclinoriums 	 LOAD JPEG IMAGE	 LOAD JPEG IMAGE	 LOAD JPEG IMAGE	 LOAD JPEG IMAGE	 LOAD JPEG IMAGE	 LOAD JPEG IMAGE
Buckle Fold 	 LOAD JPEG IMAGE	 LOAD JPEG IMAGE	 LOAD JPEG IMAGE	LOAD JPEG IMAGE	LOAD JPEG IMAGE	LOAD JPEG IMAGE
This fold has uniform limb and hinge thickness, but still shows anomaly peaks at the hinges.						

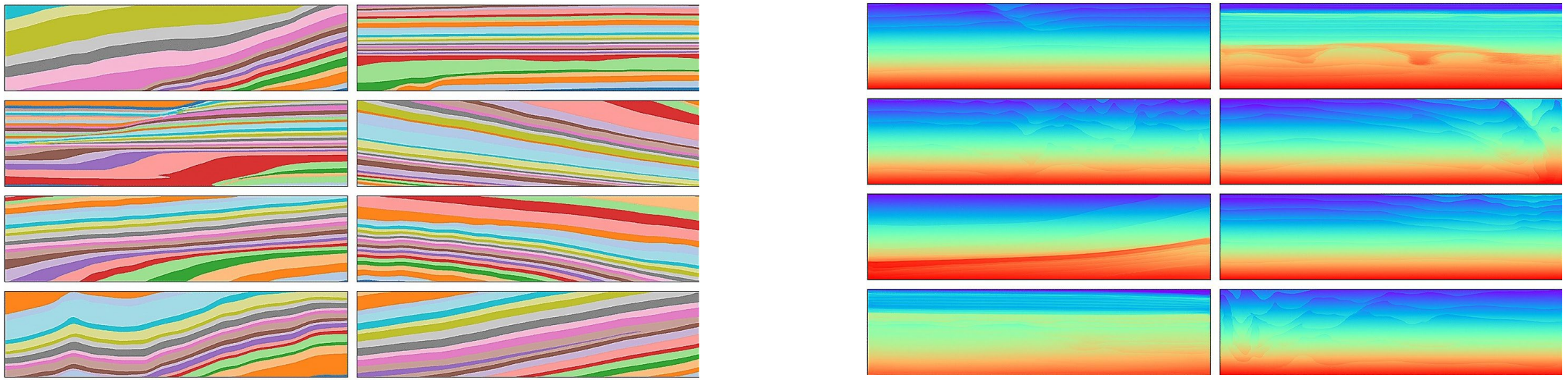
ClinoformNet-1.0: stratigraphic forward modeling and deep learning for seismic clinoform delineation

Hui Gao et al., Geoscientific Model Development, 2022-245



Geophysical model generation with generative adversarial networks

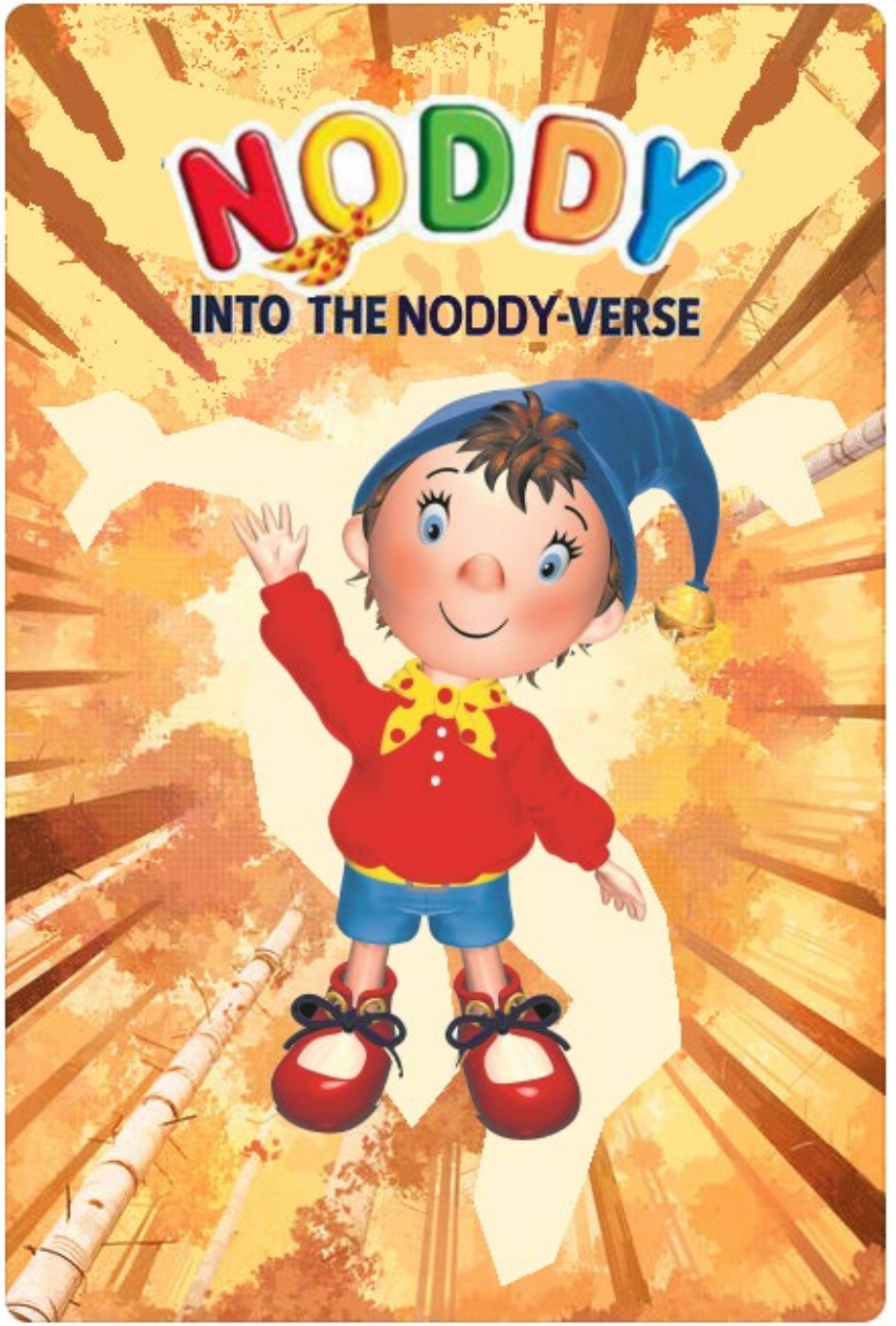
Puzyrev et al., Geoscience Letters 32 2022



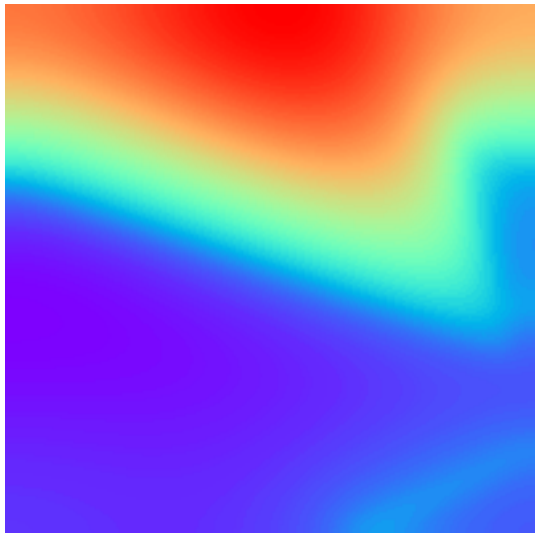
Another Solution for more complex 3D geology?

We are using Noddy to generate very large training sets for 3D geology and resulting gravity and magnetic datasets as inputs to ML systems and as tests of geophysical inversion schemes

1. Spatio-temporal structure
2. High dimensionality
3. Paucity of ground truth

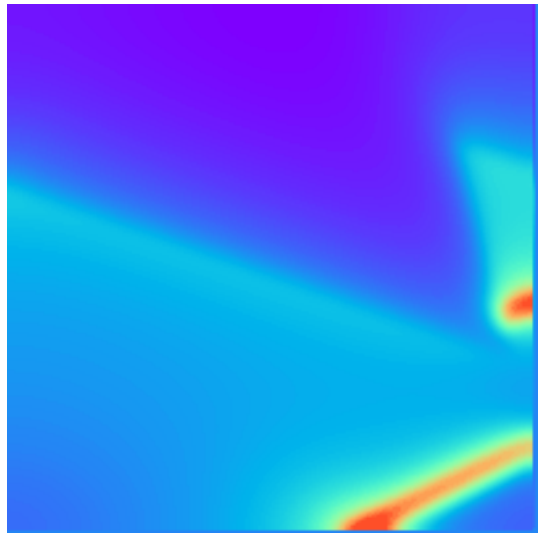


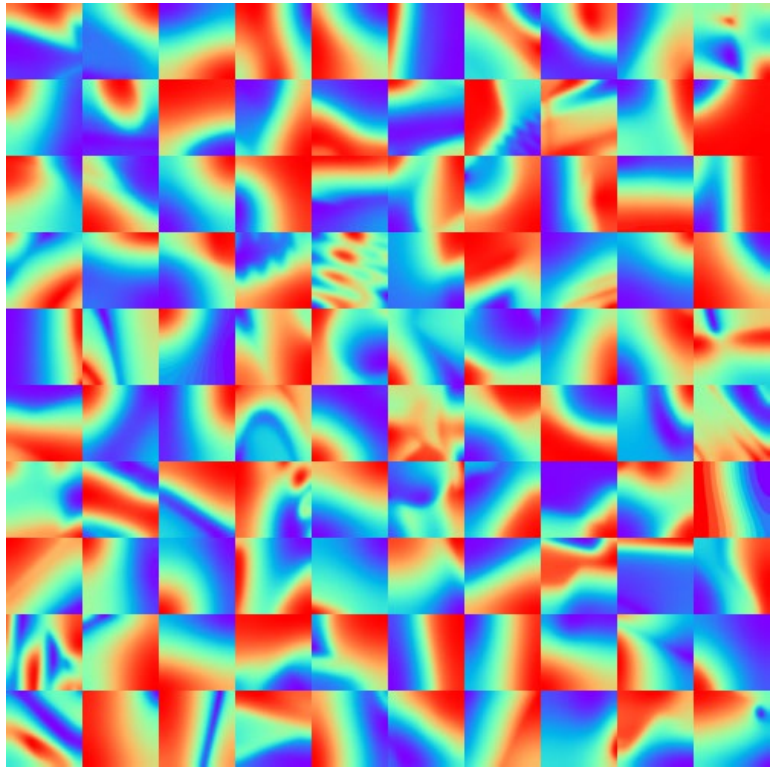
Example models



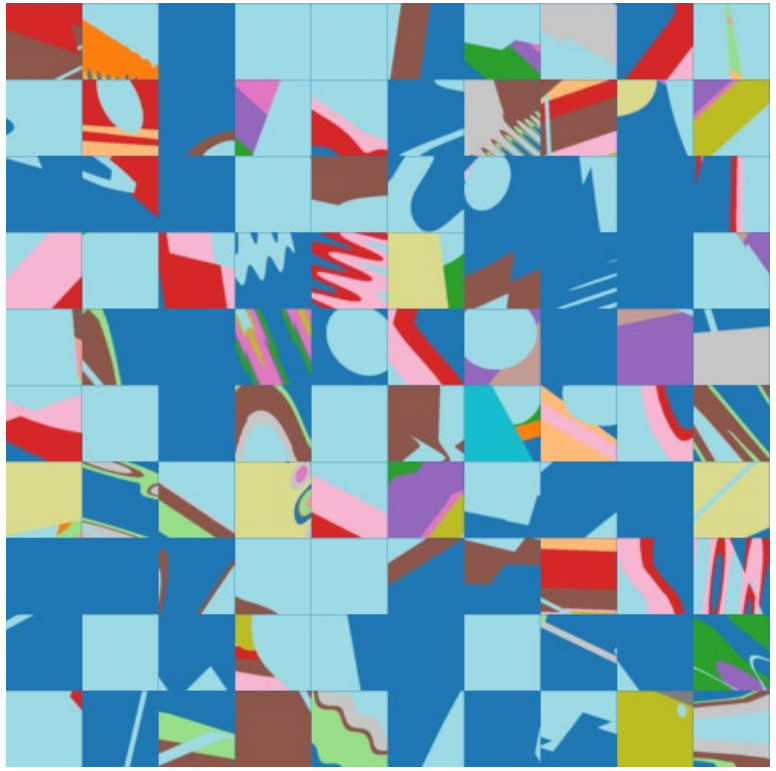
100 randomly generated models randomly selected from database

4kx4kx4k
20m voxels

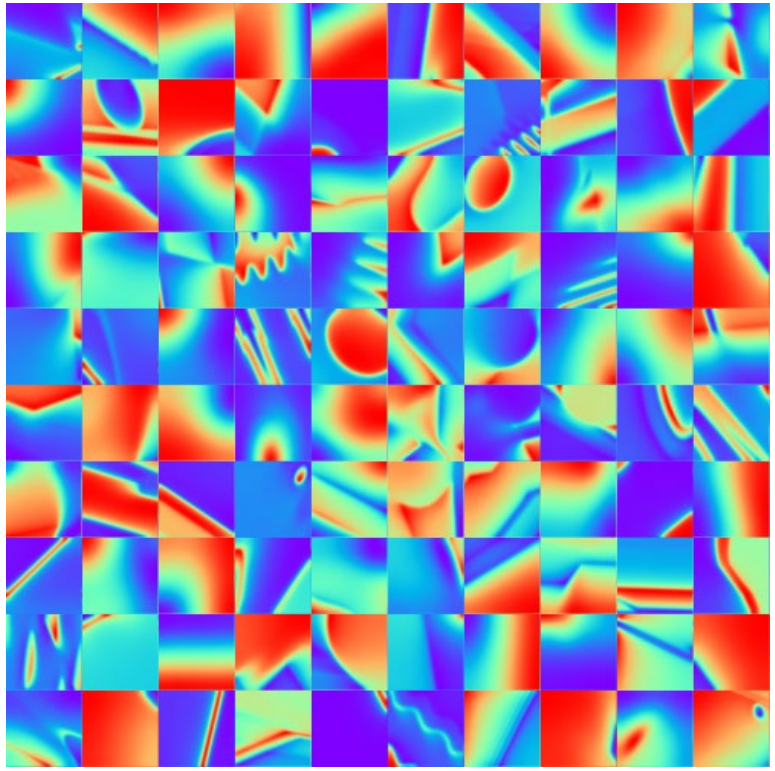




grav



Top surface of model



mag

Goal: predict the parameters of the geological structure (eg, dip direction and the dip angle of the fold axial surface)

Magnetic field 2D images
Noddy dataset, ~50k samples

set	train	val	test
samples	31068	10187	8454

Classifier
Inception-v3
(very deep) CNN

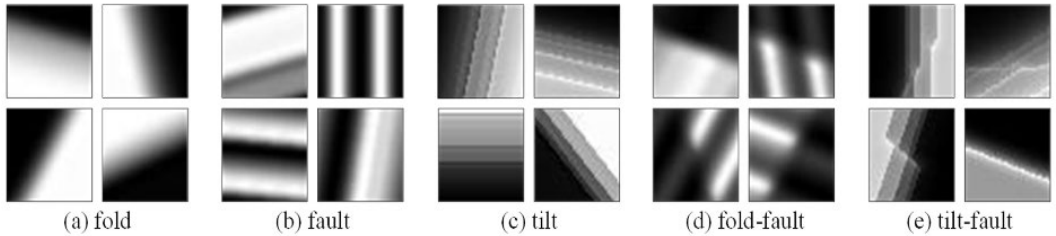


Fig. 1. Samples of the magnetic dataset generated by Noddy. These images are normalized before training.

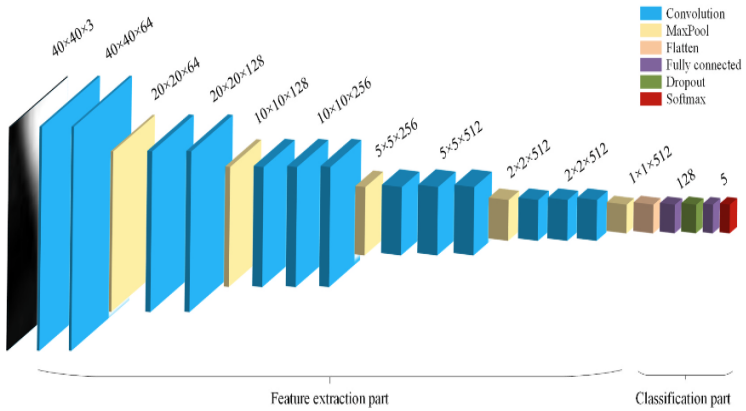
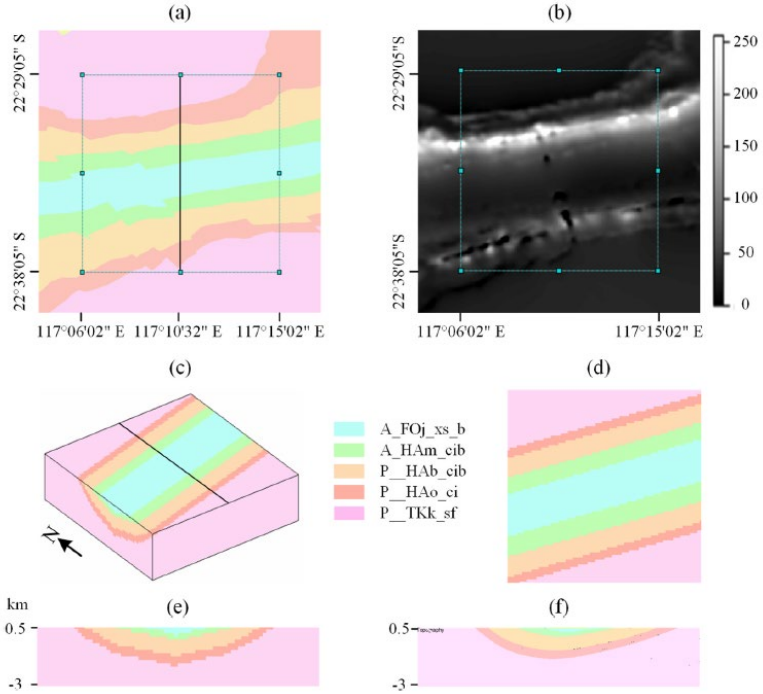


Fig. 3. Network structure of CNN classification model. The feature extraction layer here takes VGG16 as an example. When training with Inception-v3 and ResNet50, the corresponding feature extraction layer is replaced accordingly.

~1 day training

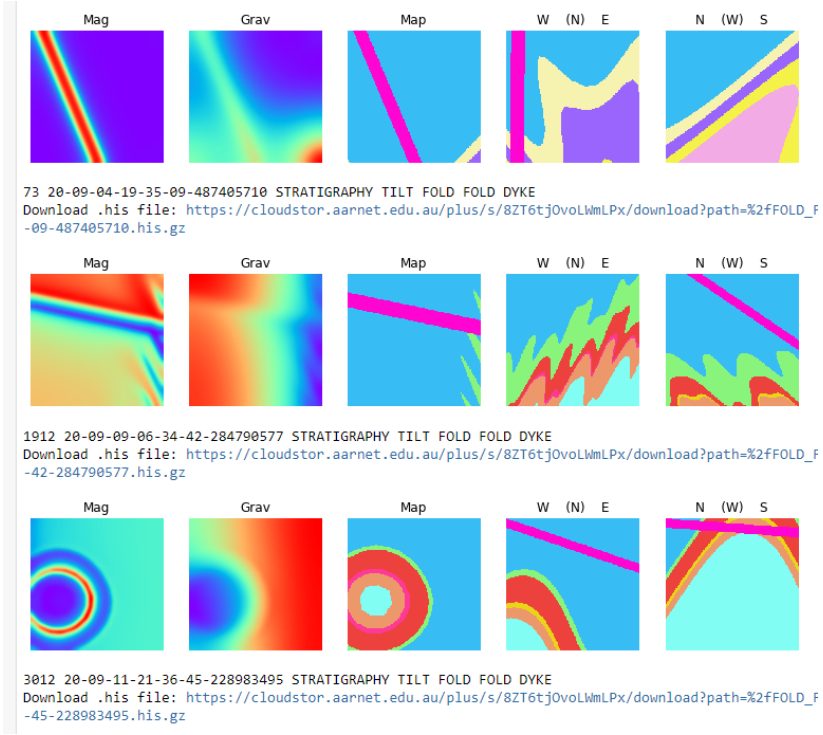


3D geological structure inversion from Noddy-generated magnetic data using deep learning methods, Guo et al., Computers & Geosciences 2021, 104701

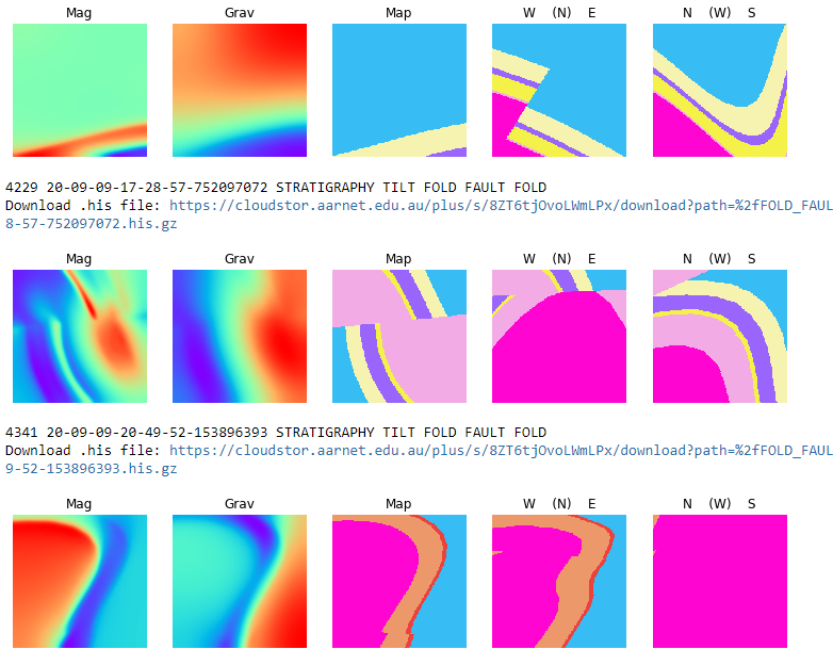
A new Challenge

Open Access 1M labelled model suite with more diverse geological histories including diverse stratigraphies available here:

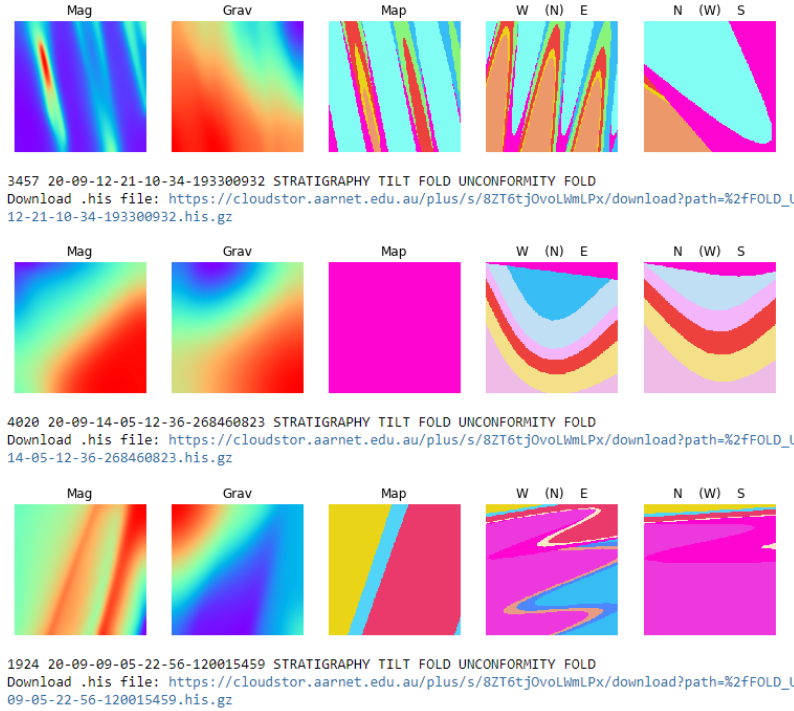
github.com/Loop3D/noddyverse



sampling from 4036 models matching
['FOLD', 'FOLD', 'DYKE']
event sequence



sampling from 7994 models matching
['FOLD', 'FAULT', 'FOLD']
event sequence



sampling from 4094 models matching
['FOLD', 'UNCONFORMITY', 'FOLD']
event sequence

cloudstor → aws

Tool to sample random 3D geological models from noddyverse 1M model suite

This notebook randomly samples and displays models from the 1 Million model suite, with the possibility to filter to specific deformation event sequences.

For each model the magnetic response, the gravity response, the top surface, the north facing vertical section (looking from the south) and the west-facing vertical section (looking from the west) of the 4 x 4 x 4 km cube are displayed:


```
8549 20-09-05-16-34-55-585212604 STRATIGRAPHY TILT UNCONFORMITY PLUG FOLD
Download .his file: https://cloudstor.aarnet.edu.au/plus/s/3GhvhCZRAU3fXdo/download?path=%2F&files=20-09-05-16-34-55-585212604.his.gz
```

The models consist of 20 m cubic voxels (200x200x200 voxels per model), and the history file used to define each model is provided as a link for each model. Windows software to read this history file is available at the Noddy Site, and the source code for a command line version of Noddy, as well as this notebook, is available from github The 1M models themselves are stored at https://cloudstor.aarnet.edu.au/plus/s/8ZT6tjOvoLWmLPx as individual files and as 343 tar files, one per history sequence from https://cloudstor.aarnet.edu.au/plus/s/UxnVSkHfnr7chW9. Each model displayed below has a link to the history file compressed with gzip, so you will need to ungzip it before loading it into Noddy.

Usage

To use this notebook modify the parameters in the first cell and then both cells, with all outputs displayed at the bottom of the notebook. The code in noddyverse.py provides an example so users can see how to parse the individual files.

Can we generalise this method to more complex scenarios?

- Use of Machine Learning in the geosciences needs labelled datasets, but the Google user base is too busy looking at cats to help us
- In this study we have shown the potential of non-physics-based 3D modelling to provide large arbitrarily complex time-aware 3D models and their potential-field response
- We can make a million models, we could even make a billion models, but we can’t make the trillions of models right now that would be needed to generalise this approach.

Hand over to Leo…