Australia's National Science Agency

Spatial error constraints reduce overfitting for potential field geophysical inversion

Mark Lindsay

CSIRO Mineral Resources, DARE ITTC, SES @ UWA

mark.lindsay@csiro.au

Vitaliy Ogarko (UWA), Jérémie Giraud (Uni Lorraine, UWA)

Mosayeb Khademi (UniSA, GA)

Minerals 4D

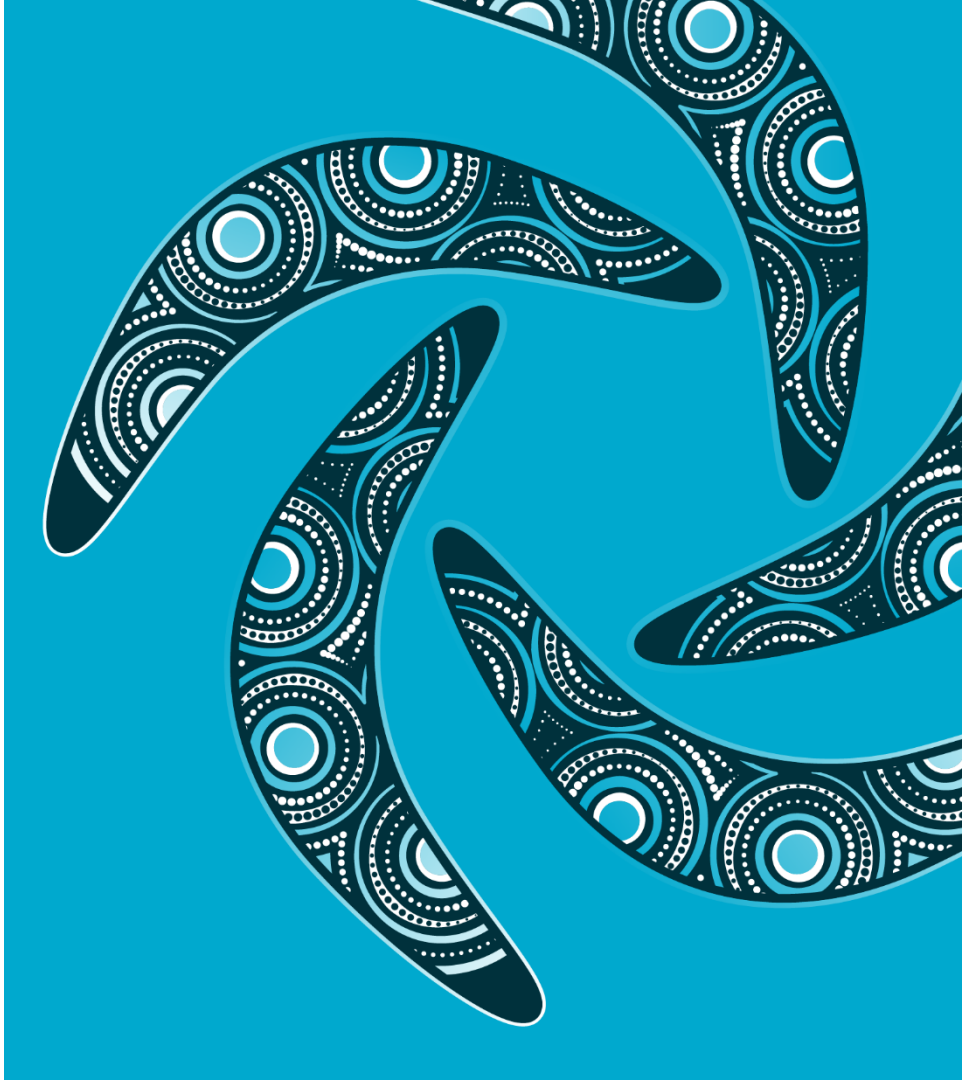
DARE ARC Training Centre in Data Analytics for Resources & Environments

Loop

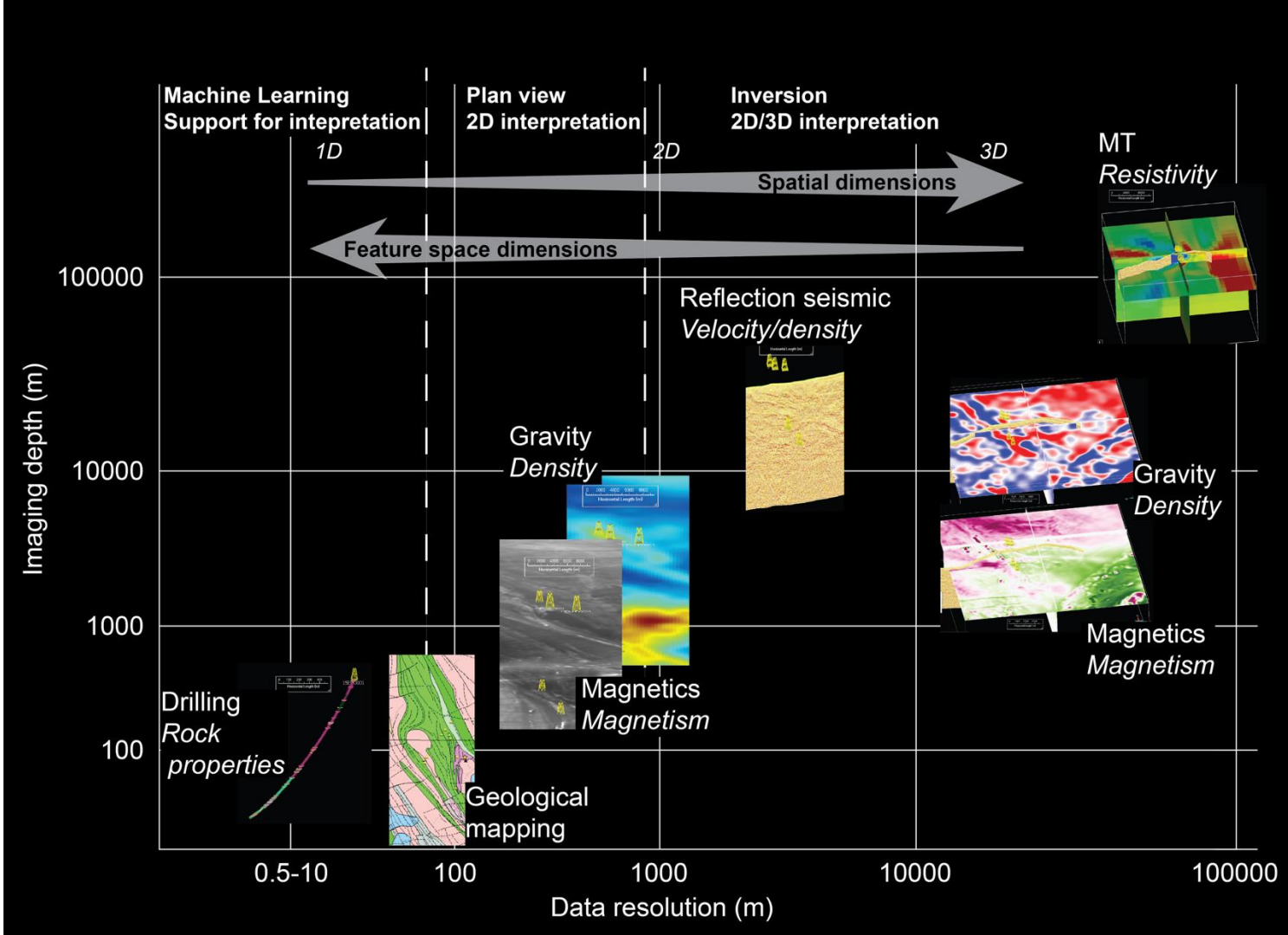
3D Geological Modelling Conference

Fremantle, 8 & 9 April 2025

I would like to begin by acknowledging the lands of the Whadjuk people of the Nyoongar nation, and pay our respect to their Elders past and present.



Scale, Data, Dimensions



Uncertainty in geological modelling

- Integration of different data

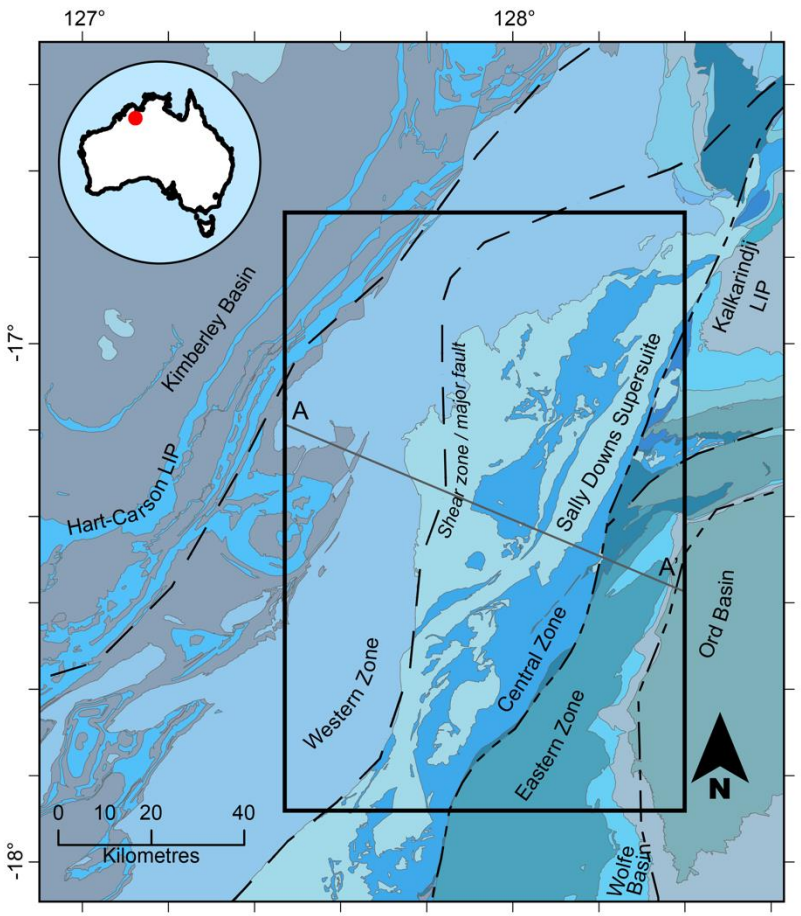
➔Inconsistencies

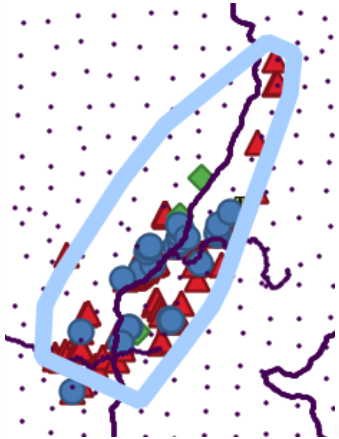
➔Exclusion of data

➔Inclusion of interpretations

- Models are variable
- Variability leads to uncertainty

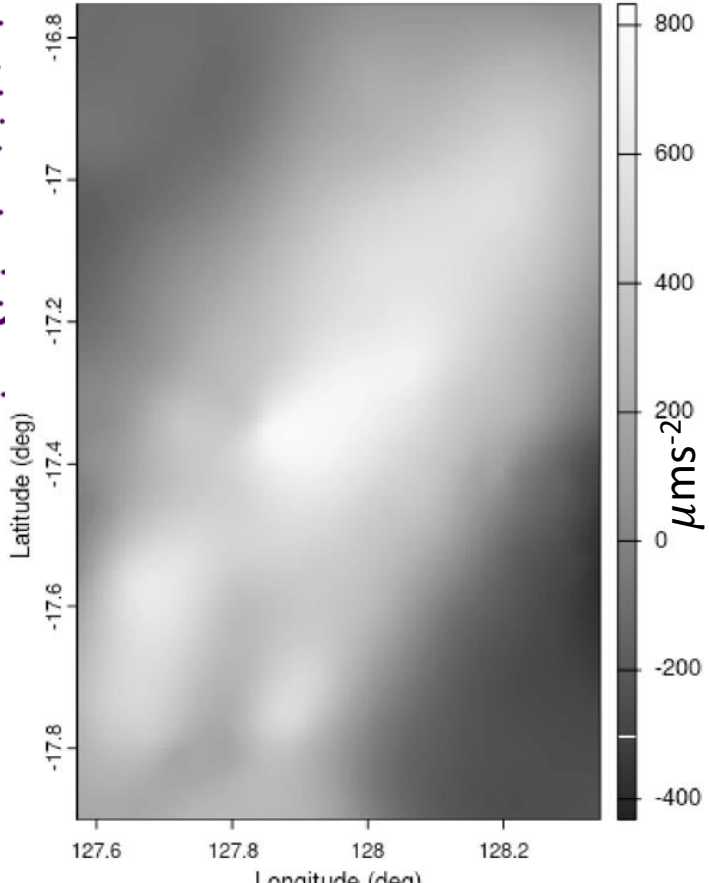
Lindsay et al 2025
In review





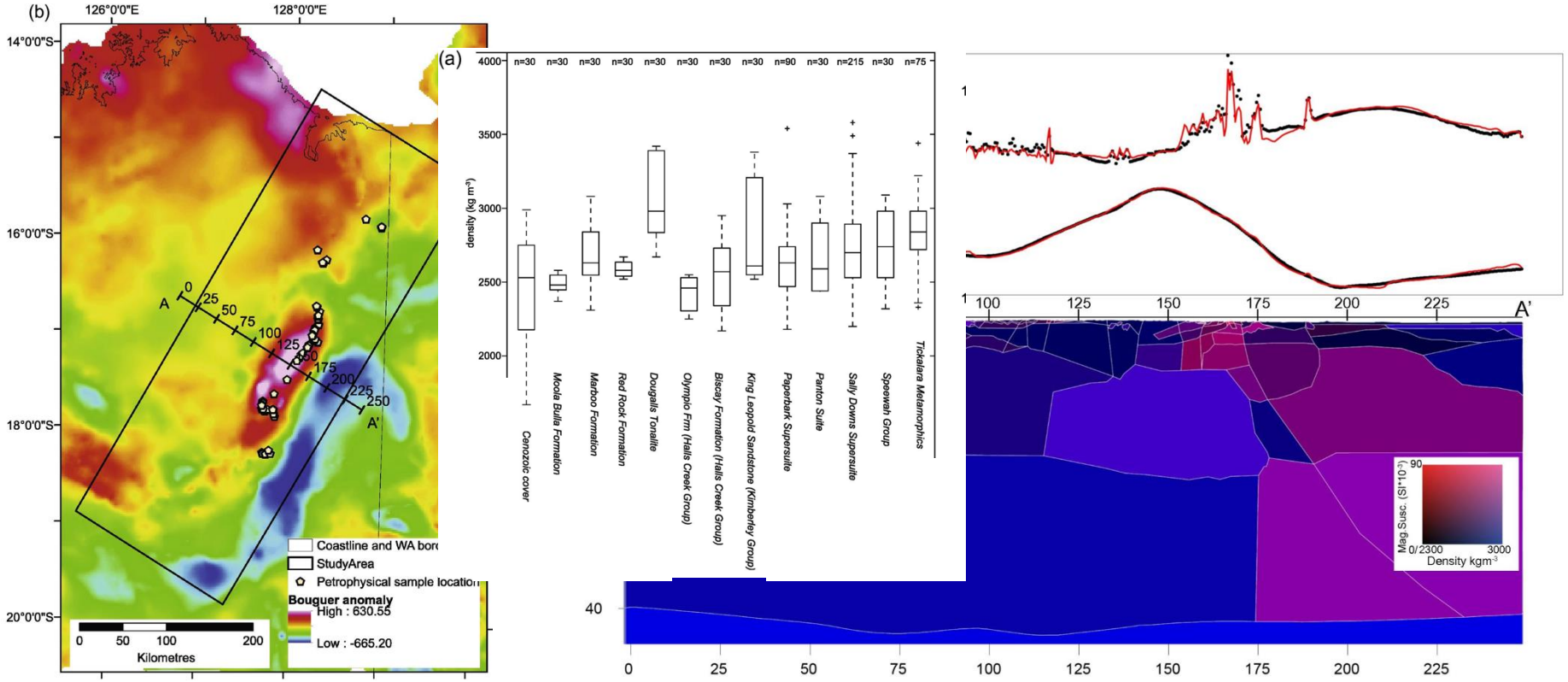
Co
V
Ni
REE

Bouguer gravity anomaly

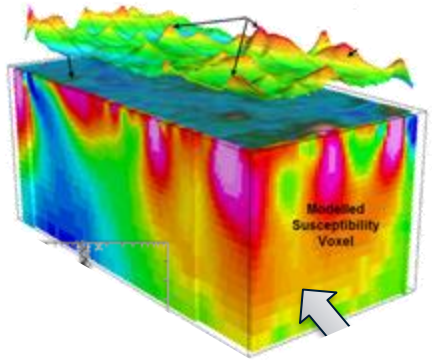


Lindsay et al 2016

Forward modelling



Geophysical inversion

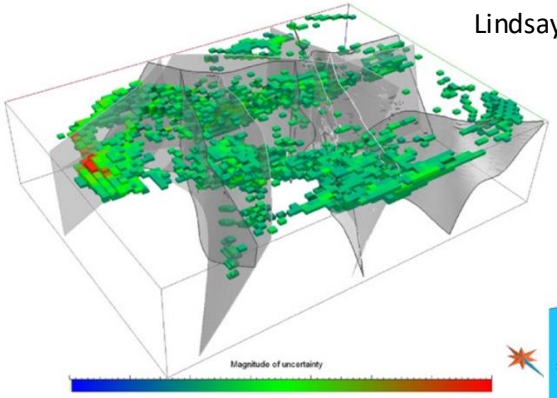


Thanks to Jeremie Giraud

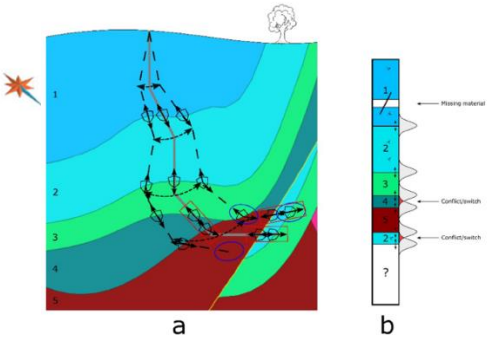
Previous work

- Geological model uncertainty
  - Generating ensembles varying orientation measurements
  - Improved statistical treatment
  - Improved perturbation schemes
  - Topological variations

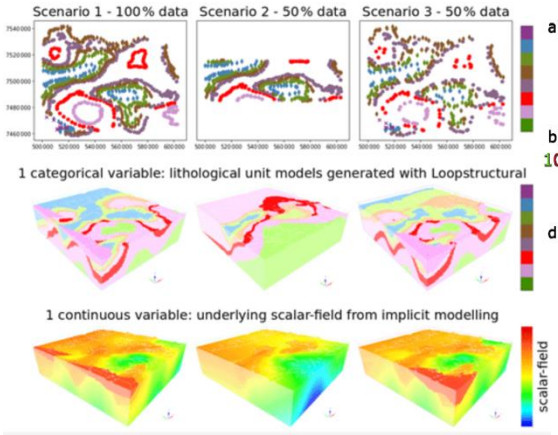
Lindsay et al. 2012

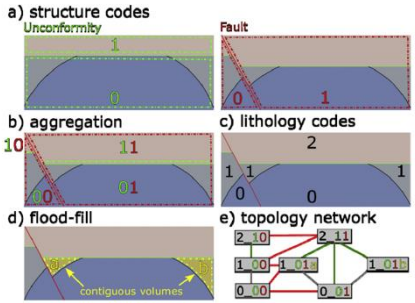


Pakyuz-Charrier et al. 2018



Pirot et al. 2022

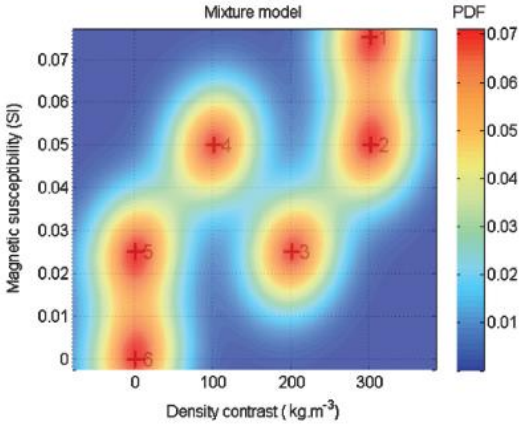




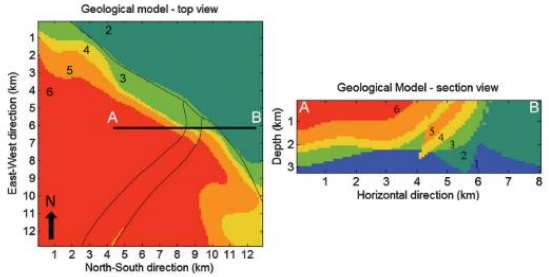
Thiele et al. 2016

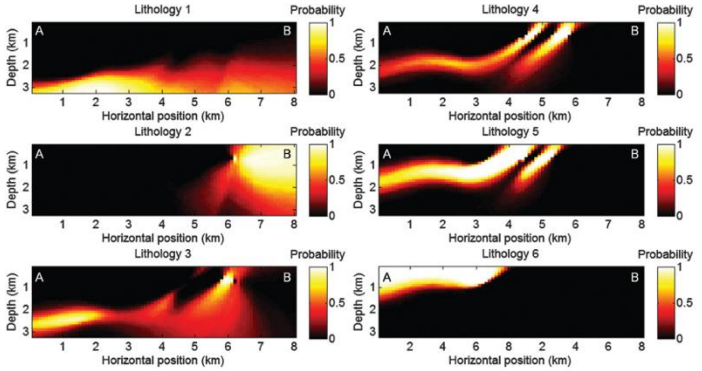
Previous work

- Geophysical inversion
  - Petrophysical uncertainty
  - Geological model uncertainty
  - Geological plausibility

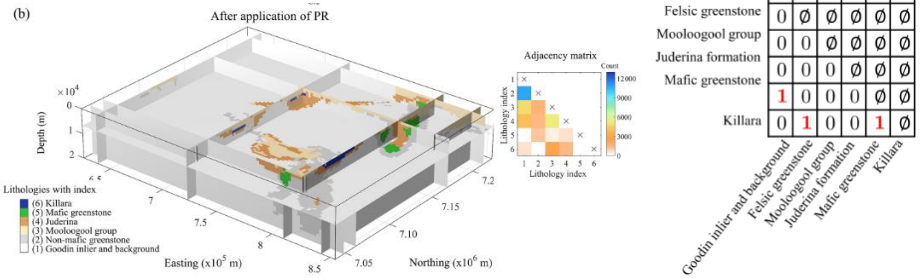


Giraud et al. 2017





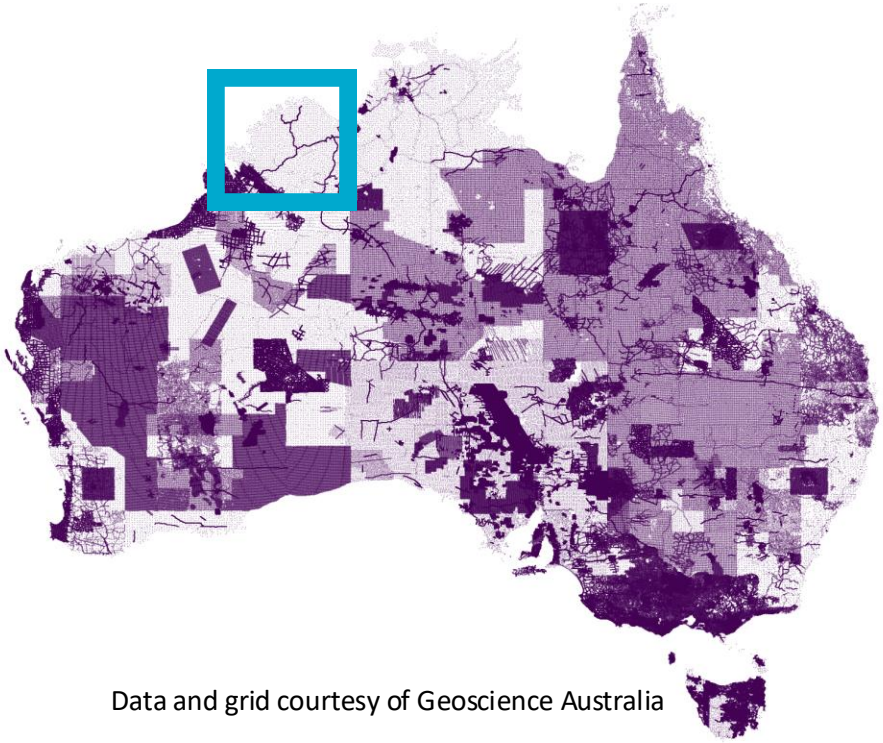
Giraud et al. 2020



	Goodin inlier and background	Felsic greenstone	Mooloogool group	Juderina formation	Mafic greenstone	Killara
Goodin inlier and background	Ø	Ø	Ø	Ø	Ø	Ø
Felsic greenstone	0	Ø	Ø	Ø	Ø	Ø
Mooloogool group	0	0	Ø	Ø	Ø	Ø
Juderina formation	0	0	0	Ø	Ø	Ø
Mafic greenstone	1	0	0	0	Ø	Ø
Killara	0	1	0	0	1	Ø

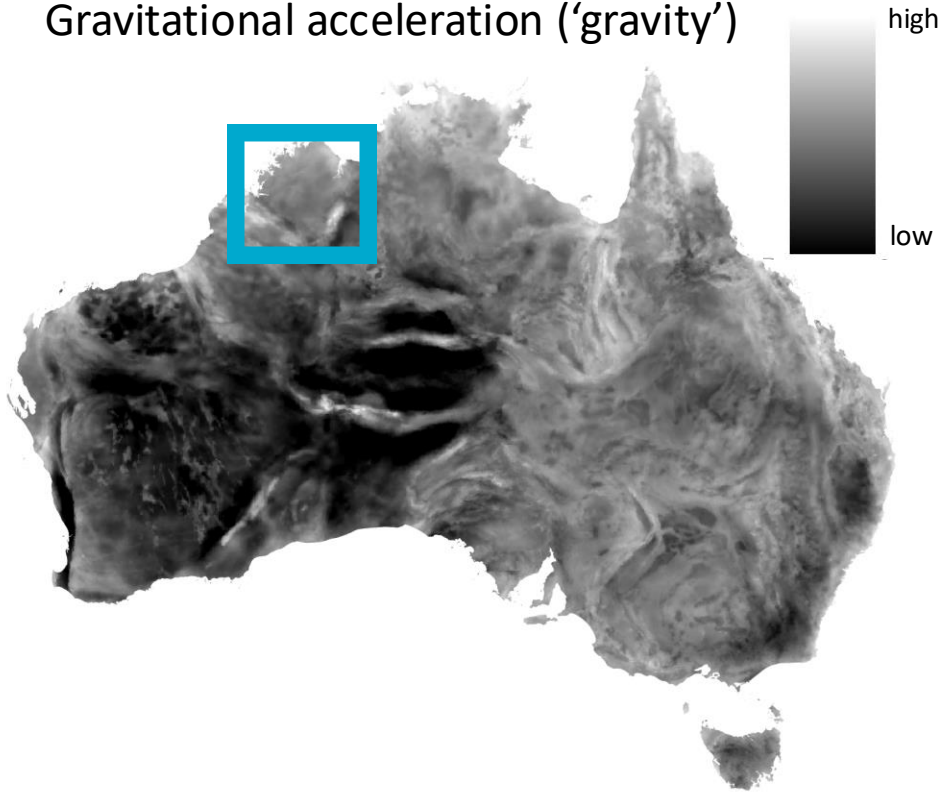
An Australian gravity model and its data

~1.6 M data (continental scale compilation)



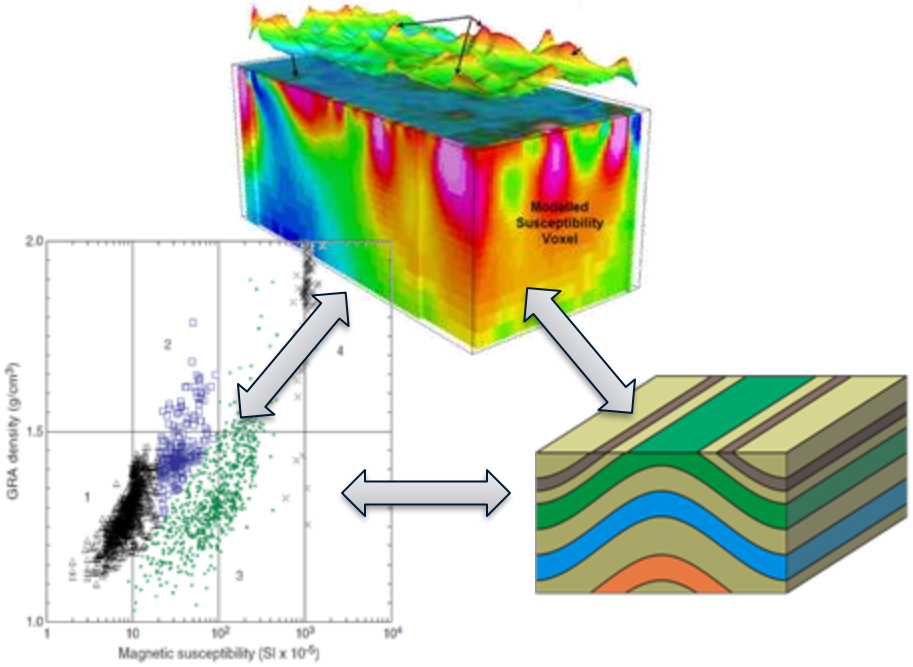
Data and grid courtesy of Geoscience Australia

Gravitational acceleration (‘gravity’)



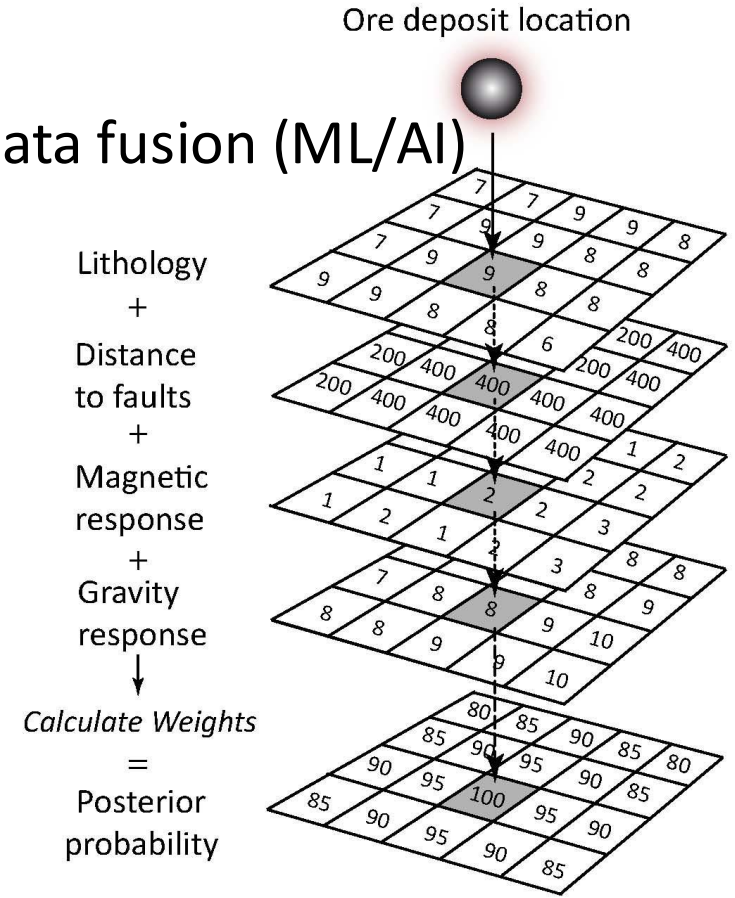
Why is this important?

- Inversion



Thanks to Jeremie Giraud

- Data fusion (ML/AI)



Lindsay et al 2014

Modelling

- INLA
- Integrated Nested Laplacian Approximation
- Rue et al 2009, r-inla.org
- Performs approximate Bayesian inference in latent Gaussian models
- Generalised linear mixed models, **spatial**, spatio-temporal models
- SPDE (Stochastic Partial Differential Equation)

From Cressie et al 2021 using SST (Sea Surface Temp data)

Table 1 Comparison of spatial predictions of SST residuals in the Brazil-Malvinas confluence zone using the five R packages, `FRK`, `INLA`, `LatticeKrig`, `mgcv`, and `gstat`, and using a C++ implementation of MRA provided by Huang et al. (2019)

Method	RMSPE	COV90	IS90	CRPS	Run time (minutes)
`FRK`	0.46	0.88	2.21	0.23	8.10
`INLA`	0.46	0.93	2.18	0.24	0.86
`LatticeKrig`	0.45	0.91	2.09	0.23	9.63
`mgcv`	0.45	0.90	2.12	0.23	5.73
MRA	0.44	0.92	2.06	0.23	NA[a]
`gstat`	0.44	0.91	2.07	0.22	7.56

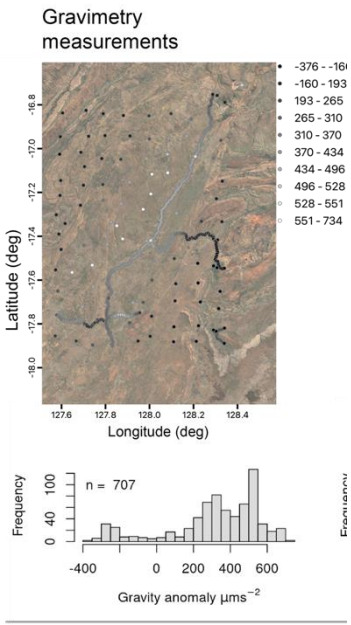
FRK v2

(not used in this presentation)

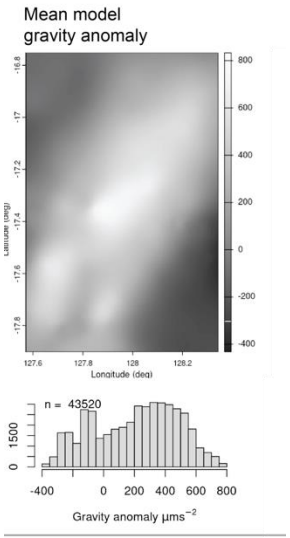
- Zammit-Mangion et al 2021
- Sainsbury-Dale et al 2024

Lindsay et al 2025
In review

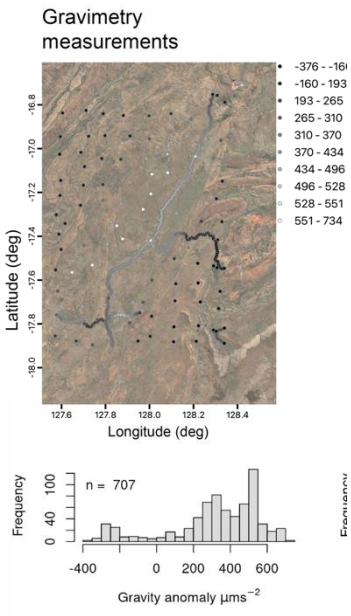
- Pre-processing
- High pass filter (remove regional)



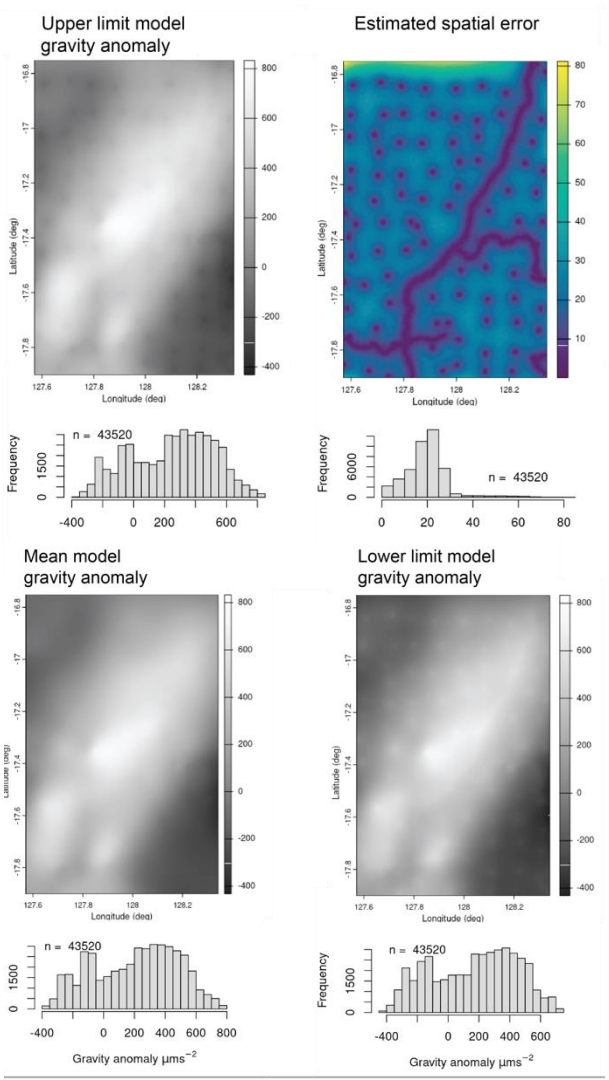
Geostatistical interpolation



- Pre-processing
- High pass filter (remove regional)



Geostatistical interpolation



Lindsay et al 2025
In review

lower limit = 10th %ile
upper limit = 90th %ile

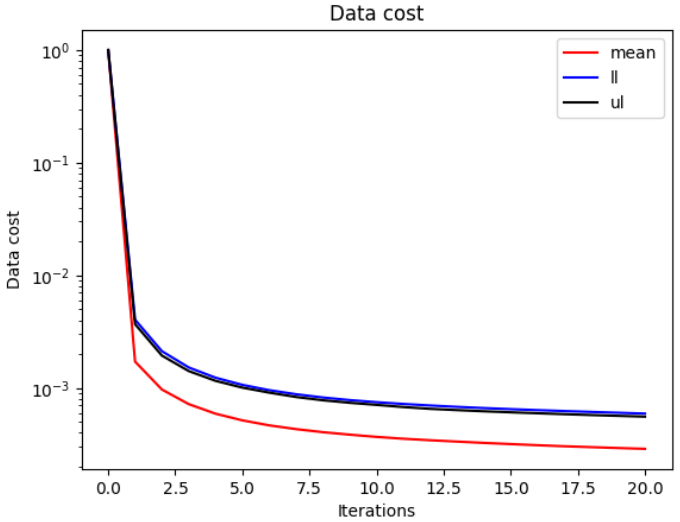
Inversion run

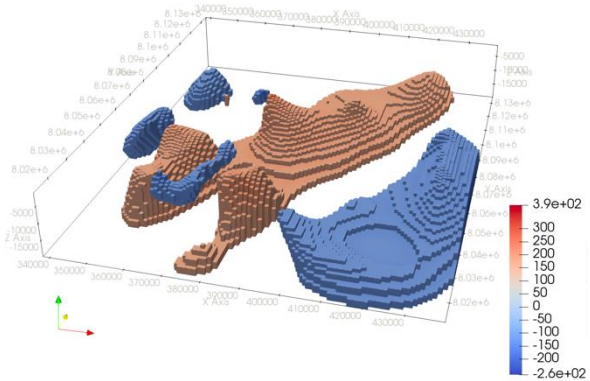
- 20 iterations
- Model grid:
  - 102 x 147 x 45
- Gravity data
  - 10880
- Depth weighted = Yes
- **Spatial error = No**

Tomofast-x

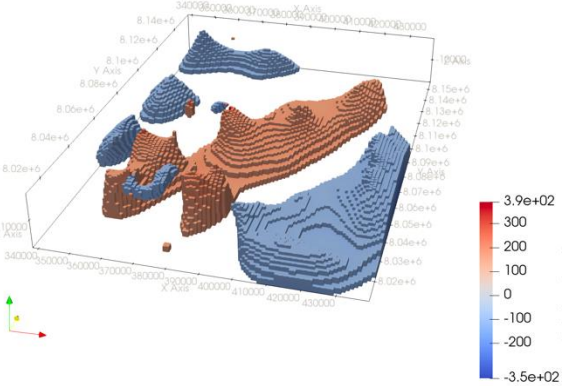
Open-source inversion code

Reference density = 2670 kg/m^3
Density values threshold
> 150 kg/m^3
< 105 kg/m^3

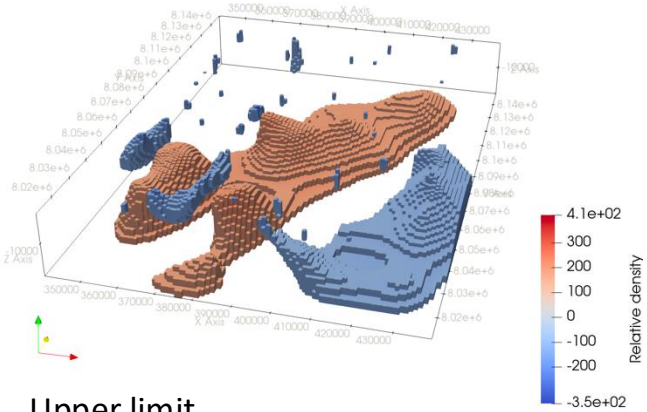




Mean



Lower limit



Upper limit

An input representing coverage

Lindsay et al 2025
In review

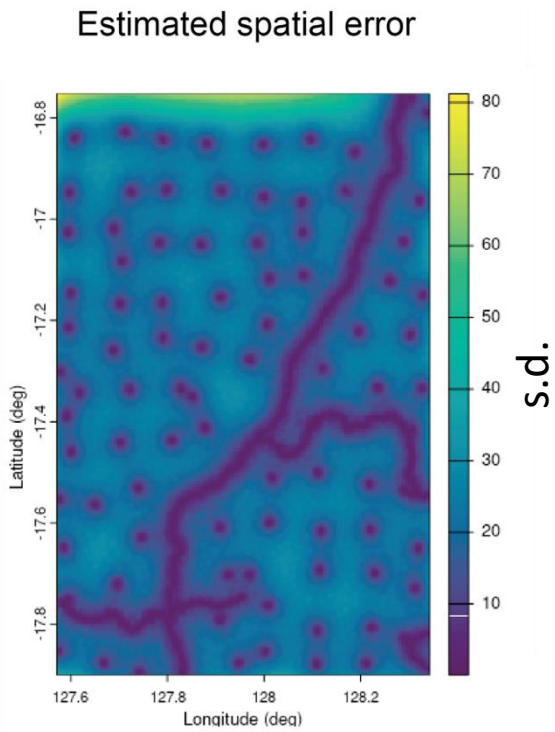
- Can be input to Tomofast
- Locations of low and high data density
- Used to weight regions during inversion according to data availability

$$\Phi_d = \|\boldsymbol{W}_d(\boldsymbol{d}^{calc} - \boldsymbol{d}^{obs})\|_2^2,$$

Diagonal weighting matrix

$$W_d = 1/\sigma^2$$

Data distance related uncertainty



Inversion run

- 20 iterations
- Model grid:
  - 102 x 147 x 45
- Gravity data
  - 10880
- Depth weighted = Yes
- **Grid covariance = Yes**

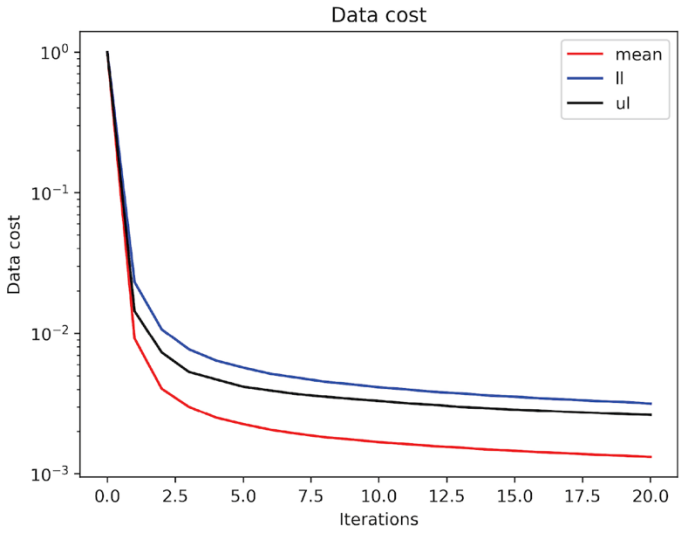
Tomofast-x

Open-source inversion code

Reference density = 2670 kg/m^3

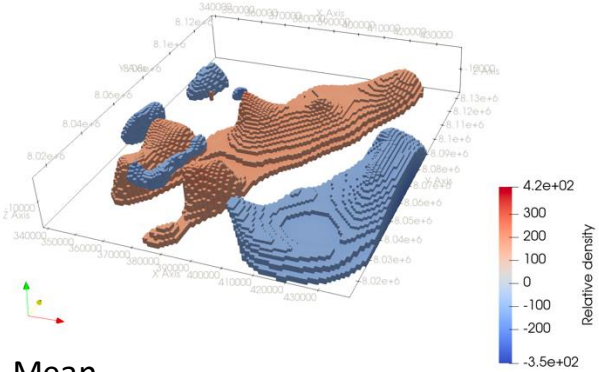
Density values threshold

- > 150 kg/m^3
- < 105 kg/m^3

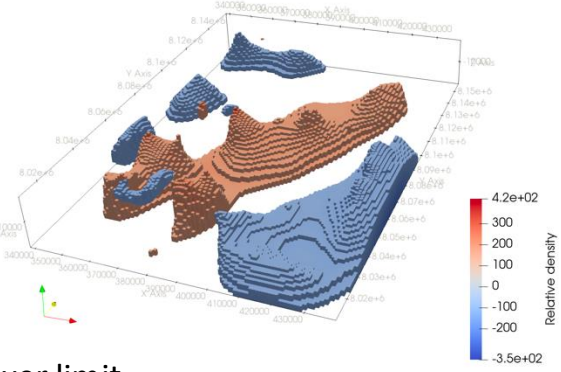


lower limit = 10th %ile

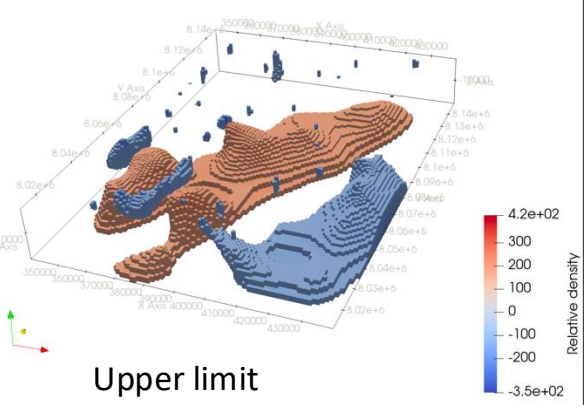
upper limit = 90th %ile



Mean



Lower limit



Upper limit

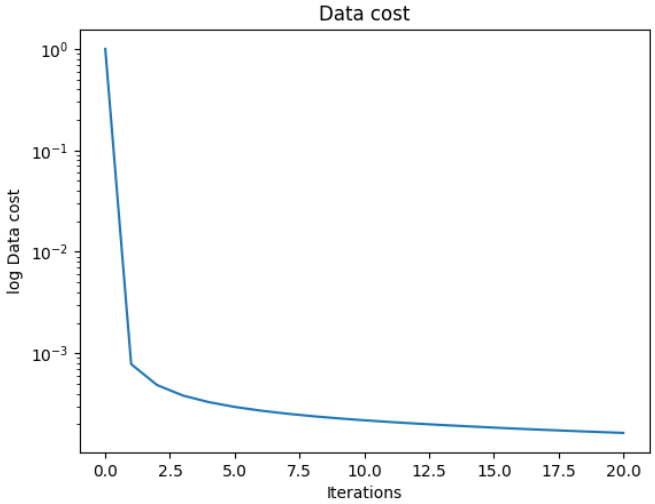
Station points only

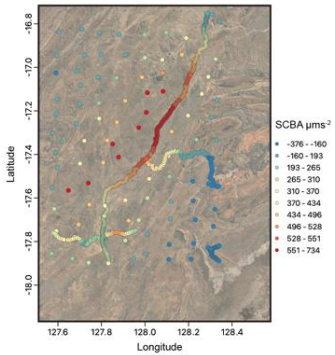
- 20 iterations
- Model grid:
  - 101 x 147 x 45
- Gravity data
  - 707
- Depth weighted = Yes
- Grid covariance matrix = No

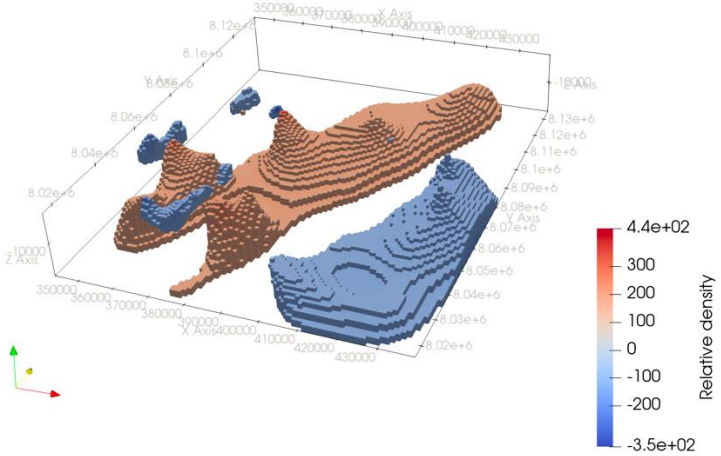
Reference density = 2670 kg/m^3

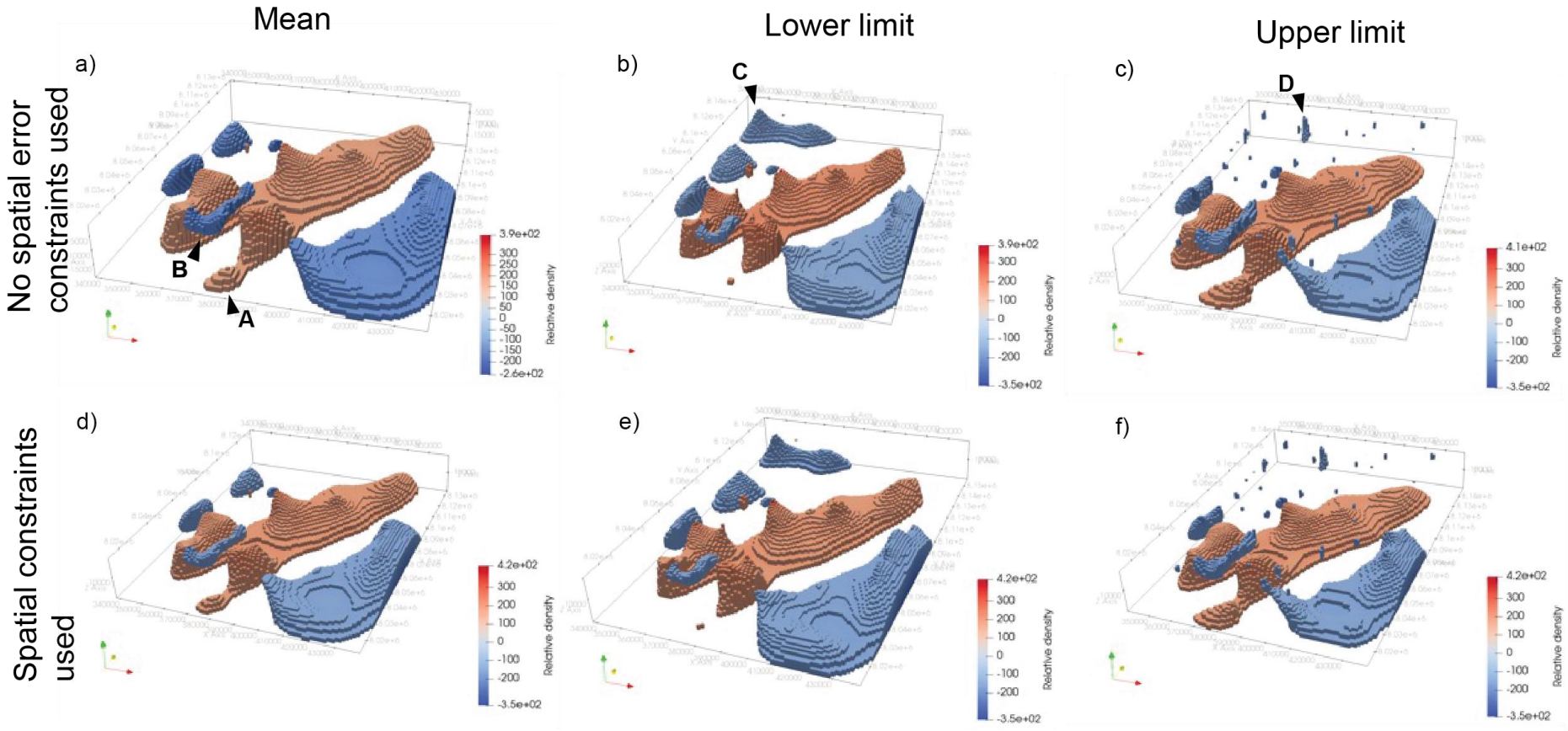
Density values threshold

- > 150 kg/m^3
- < 105 kg/m^3





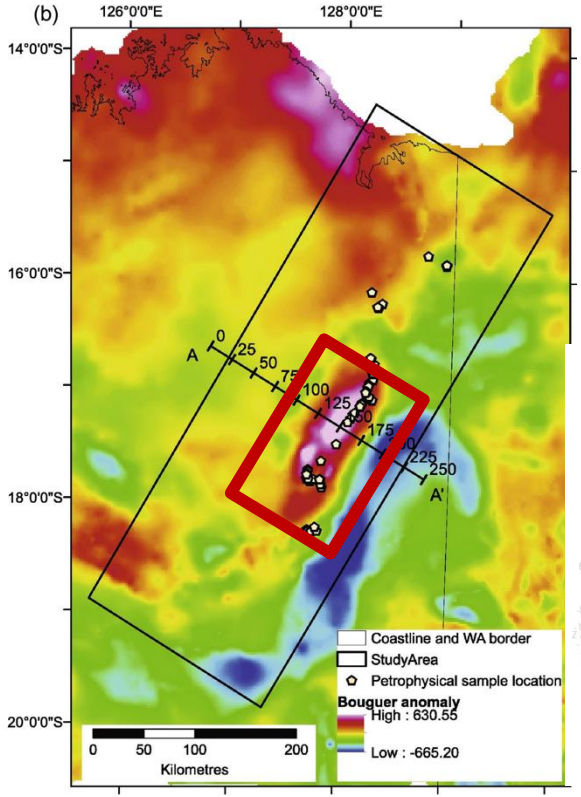


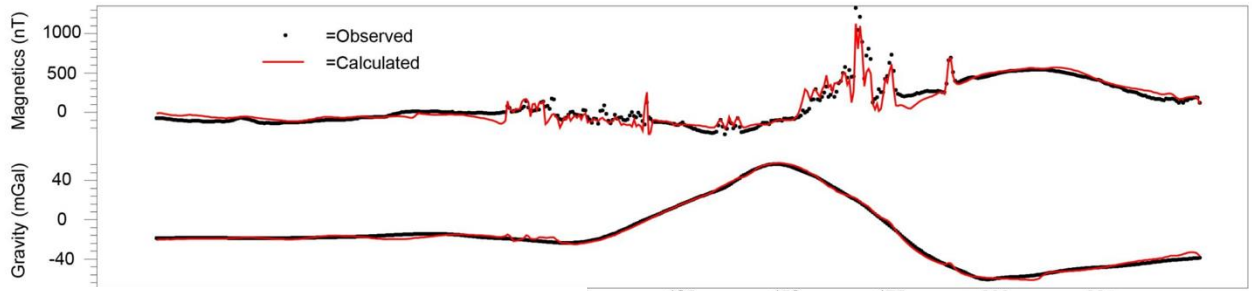


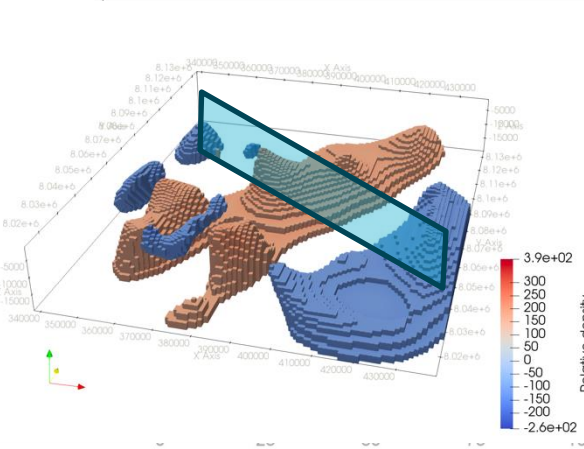
Lindsay et al 2025
In review

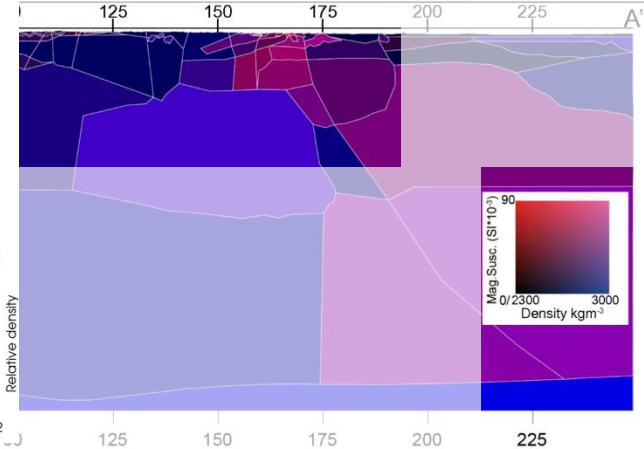
Lindsay et al 2016

Forward modelling



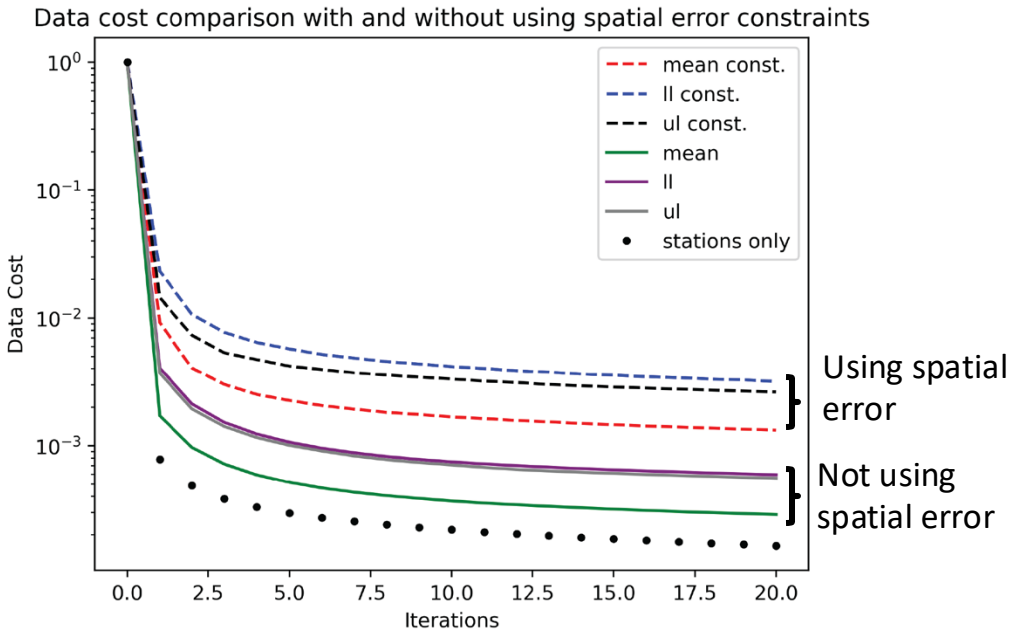


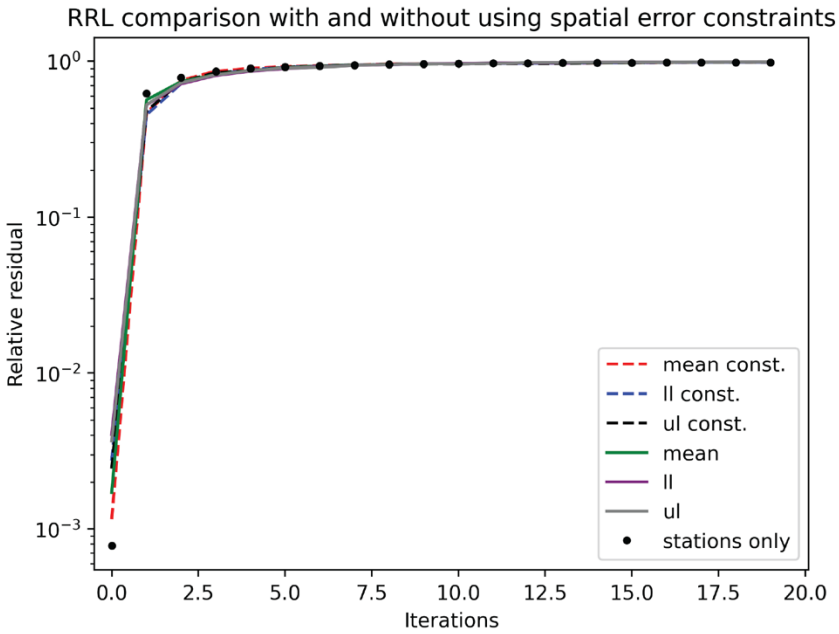




Lindsay et al 2025
In review

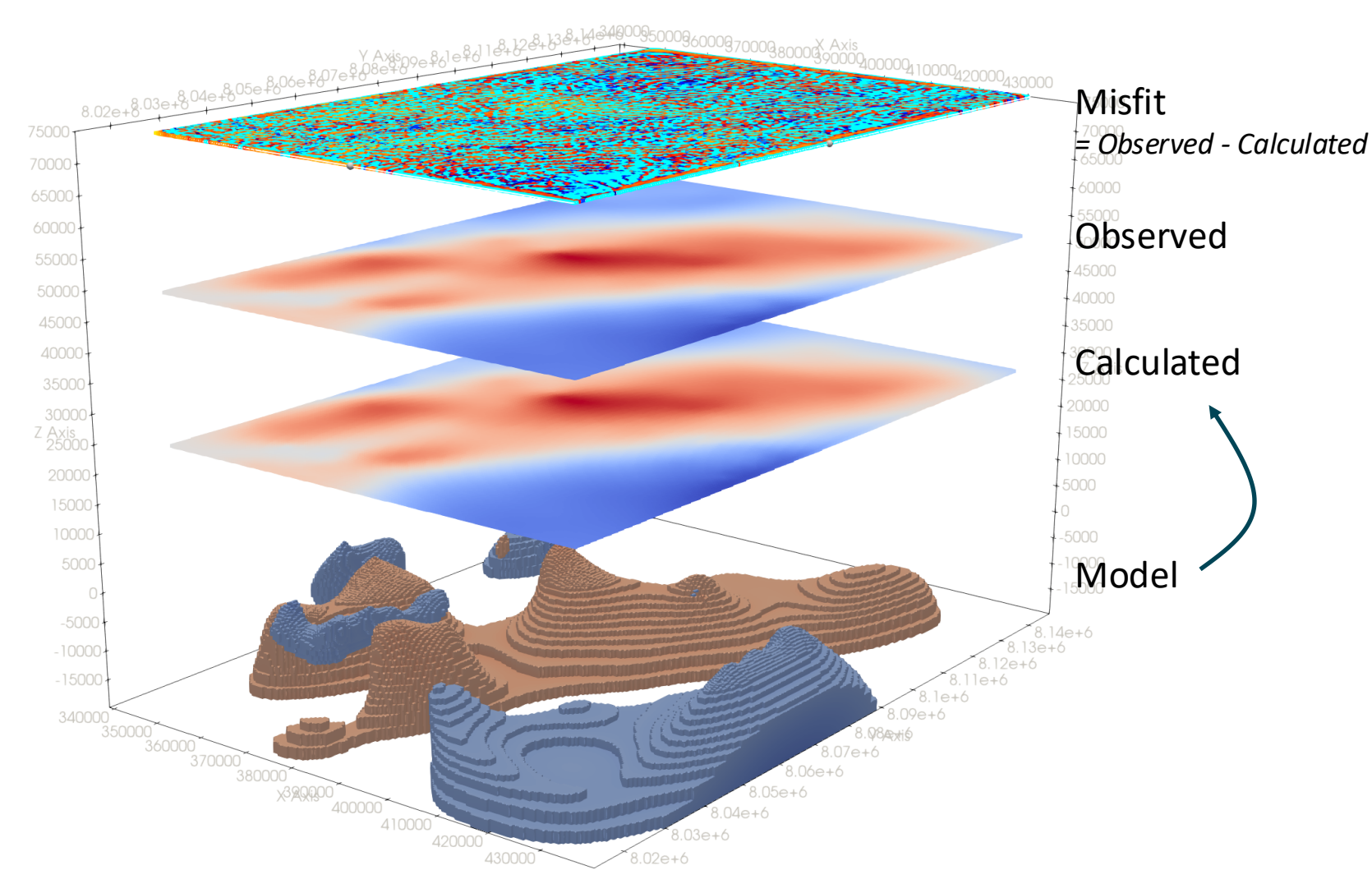
Comparison data cost with/without covar

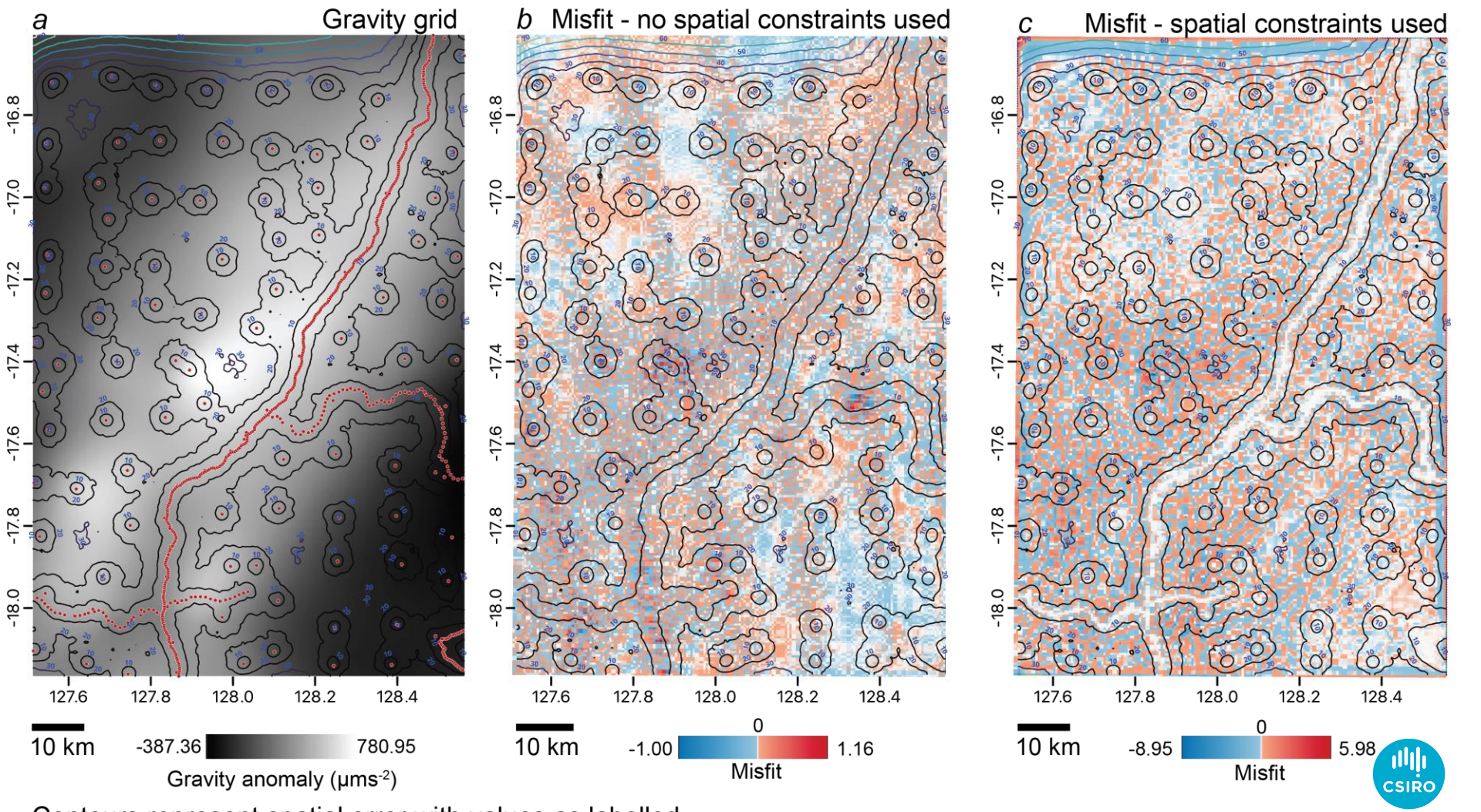




Data cost: normalised version of RMSE
Low values are desired

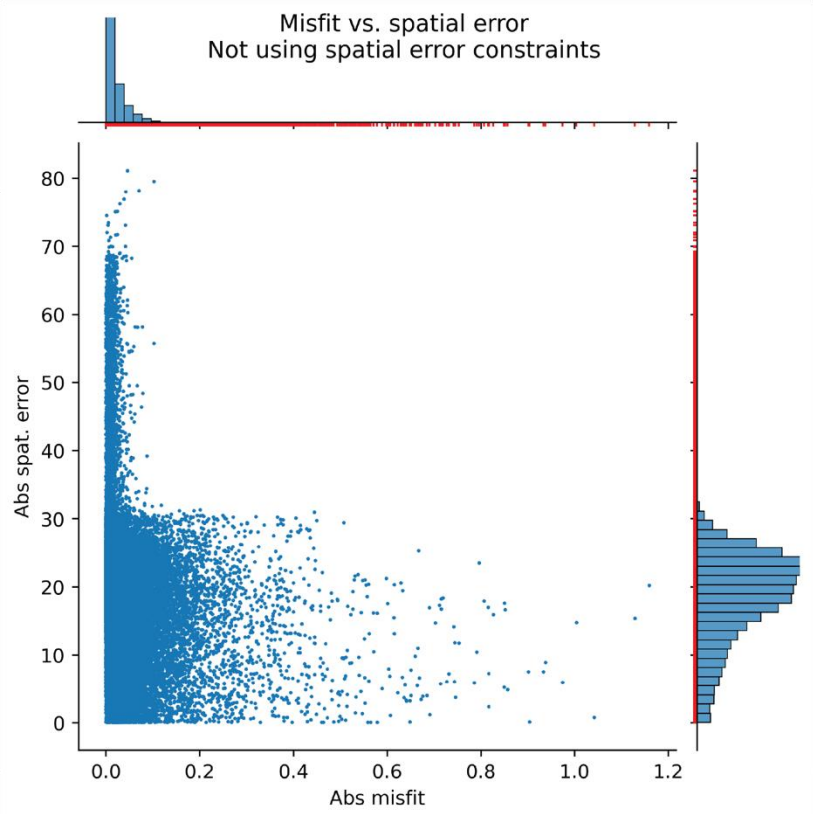
Relative residual: measure from the solver representing model variance. High values are good. Convergence is achieved when RR -> 1.0

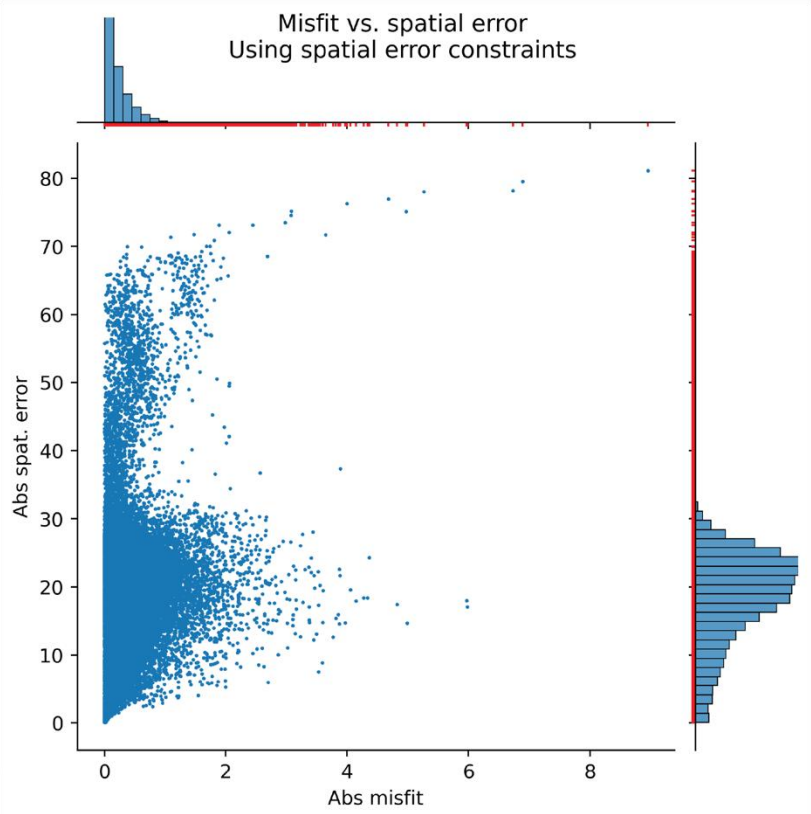

Misfit
= Observed - Calculated
Observed
Calculated
Model
X Axis
Y Axis
Z Axis



Contours represent spatial error with values as labelled.

Misfit vs Spatial Error



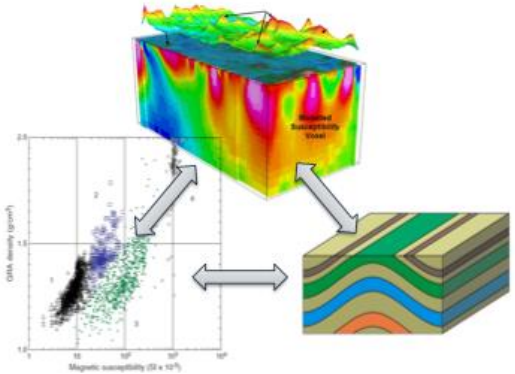


Thoughts

- Magnetic data
  - Tends to be closer spaced
  - Effects of spatial error constraints lesser (?)
  - Near-surface noise fitting more potentially more likely
- L1/L2 regularisation (use this already)
- Cross-validation (withholding n% of the data)
- Bayesian methods
  - Using priors
  - Using Bayesian Information Criterion to assess
- Measures of overfitting (e.g. Durbin-Watson statistic)
  - Autocorrelation
- Applications to prospectivity analysis and machine learning

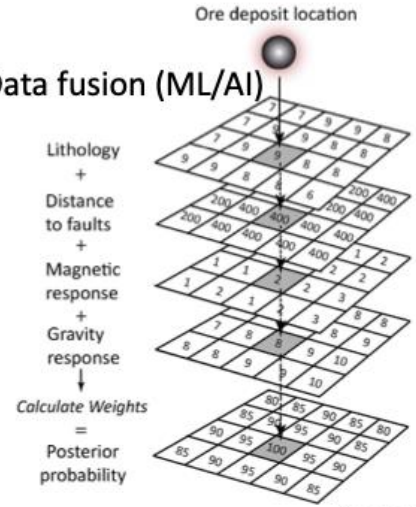
Why is this important?

- Inversion



Thanks to Jeremie Giraud

- Data fusion (ML/AI)



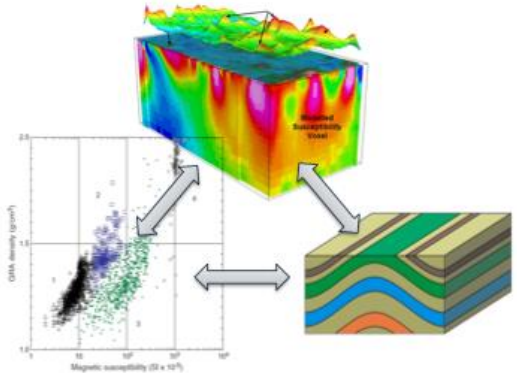
Lindsay et al 2014

Conclusions

- Near surface variations
- Spatial error reduces overfitting
- Benefits near-surface noise fitting
- Using different percentiles has no real impact

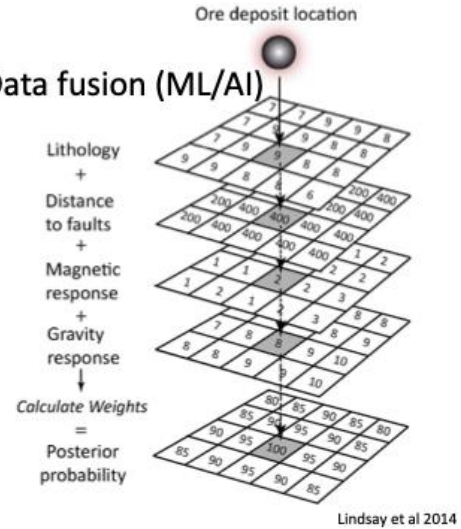
Why is this important?

- Inversion

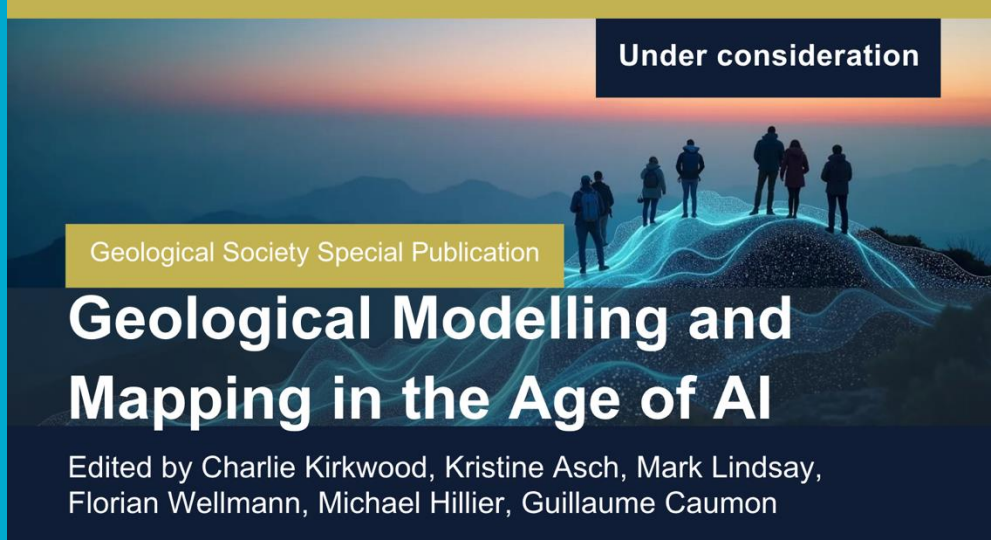


Thanks to Jeremie Giraud

- Data fusion (ML/AI)



Lindsay et al 2014



Thank you for your attention

mark.lindsay@csiro.au

0407 3827504

Minerals 4D

DARE ARC Training Centre in Data Analytics for Resources & Environments

CSIRO