EMinar: EM/MT data inversion

Colin Farquharson

Memorial University of Newfoundland,
St. John's, NL, Canada

MEMORIAL
UNIVERSITY

Pre-amble

My intention/aims/hopes with this presentation; E/R.

Big problems (in terms of computations; cells, frequencies/times, sources).

Non-linear (sensitivities/Jacobian, iterations).

Disclaimer.

Outline

~~Inversion~~ optimization background:

- why optimization
- data misfit
- minimizing data misfit, non-uniqueness
- measure of model somethingorother
- descent-based, gradient-based optimization (linearization)
- or sampling and selection of collections of models we like

Forward modelling

Sensitivities

Descent/gradient/derivative/linearization-based algorithms

The Conclusion

Future work/thoughts

- different models, different approaches?
- D+ ?

Outline

→ ~~Inversion~~ optimization background:

- why optimization
- data misfit
- minimizing data misfit, non-uniqueness
- measure of model somethingorother
- descent-based, gradient-based optimization (linearization)
- or sampling and selection of collections of models we like

Forward modelling

Sensitivities

Descent/gradient/derivative/linearization-based algorithms

The Conclusion

Future work/thoughts

- different models, different approaches?
- D+ ?

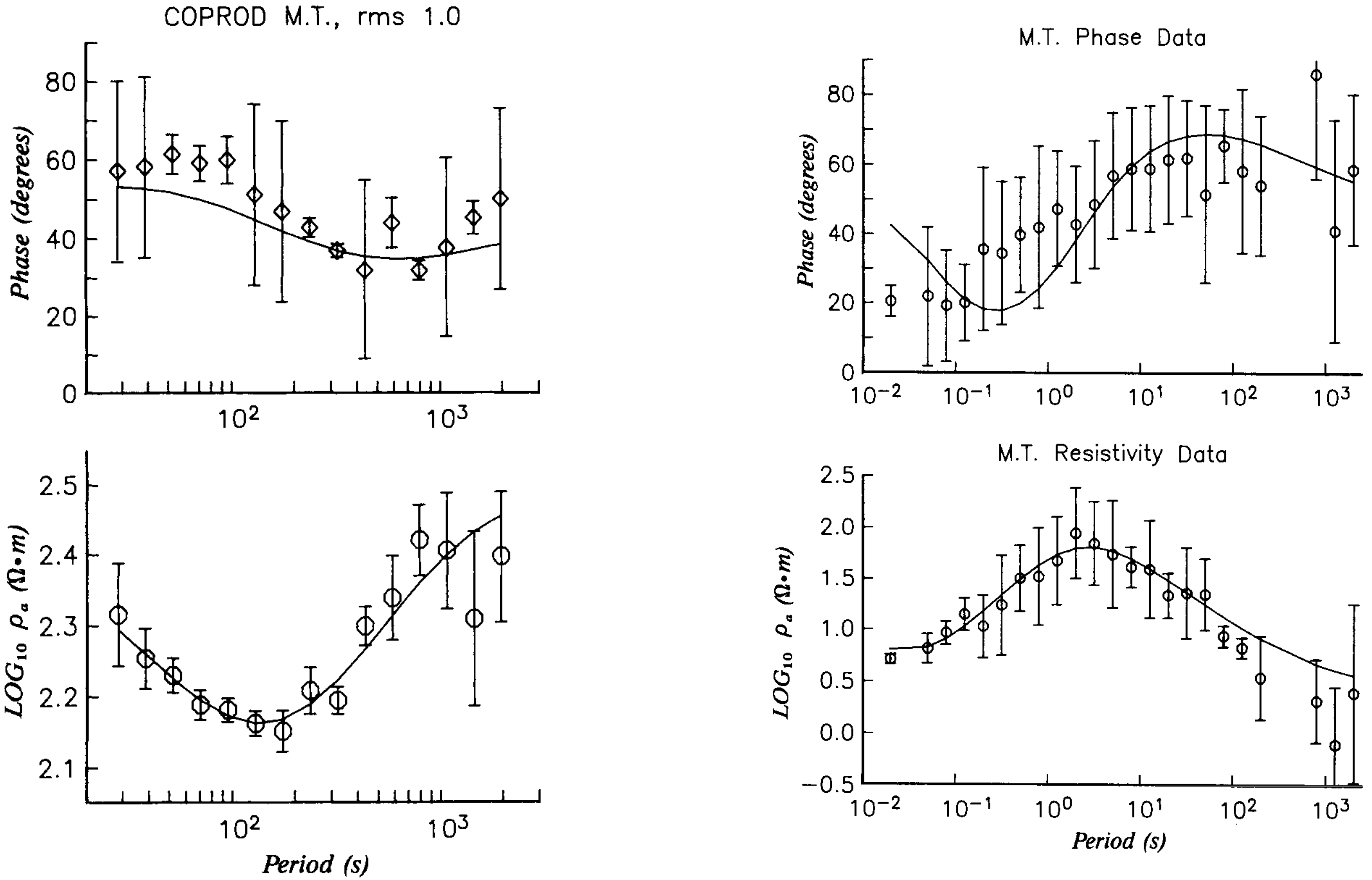
Inversion ... (!)

$$c(z,\omega) = -E(z,\omega)/E'(z,\omega)$$

$$\frac{1}{\sigma(z)} = \frac{1}{\sigma(Z)} + \frac{2\mu_0}{\pi}\int_0^\infty \mathrm{Re}[c^2(z,\omega)]\,d\omega,$$

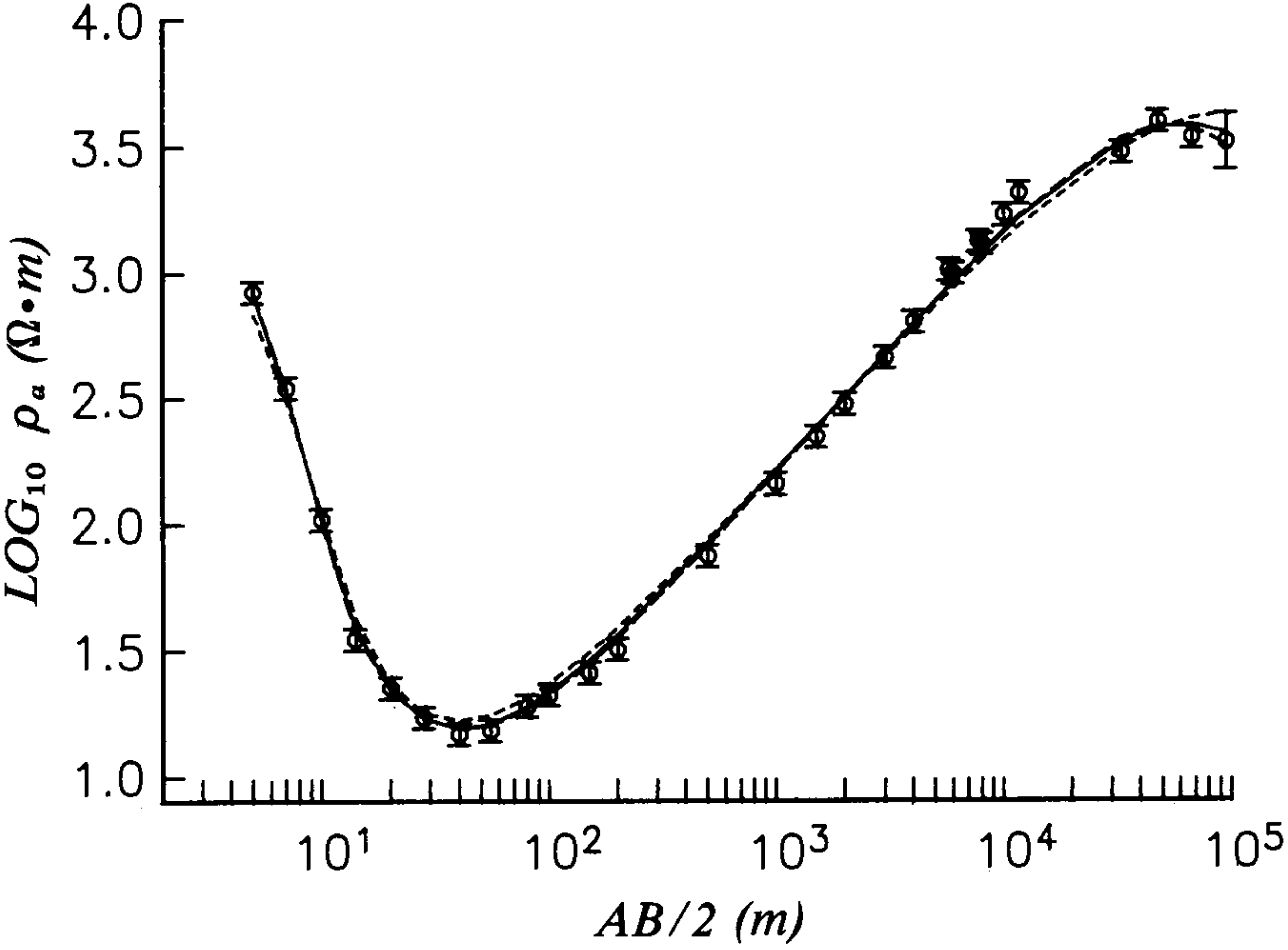
Bailey (1970), Whittall and Oldenburg (1992)

... discrete data, and noise.



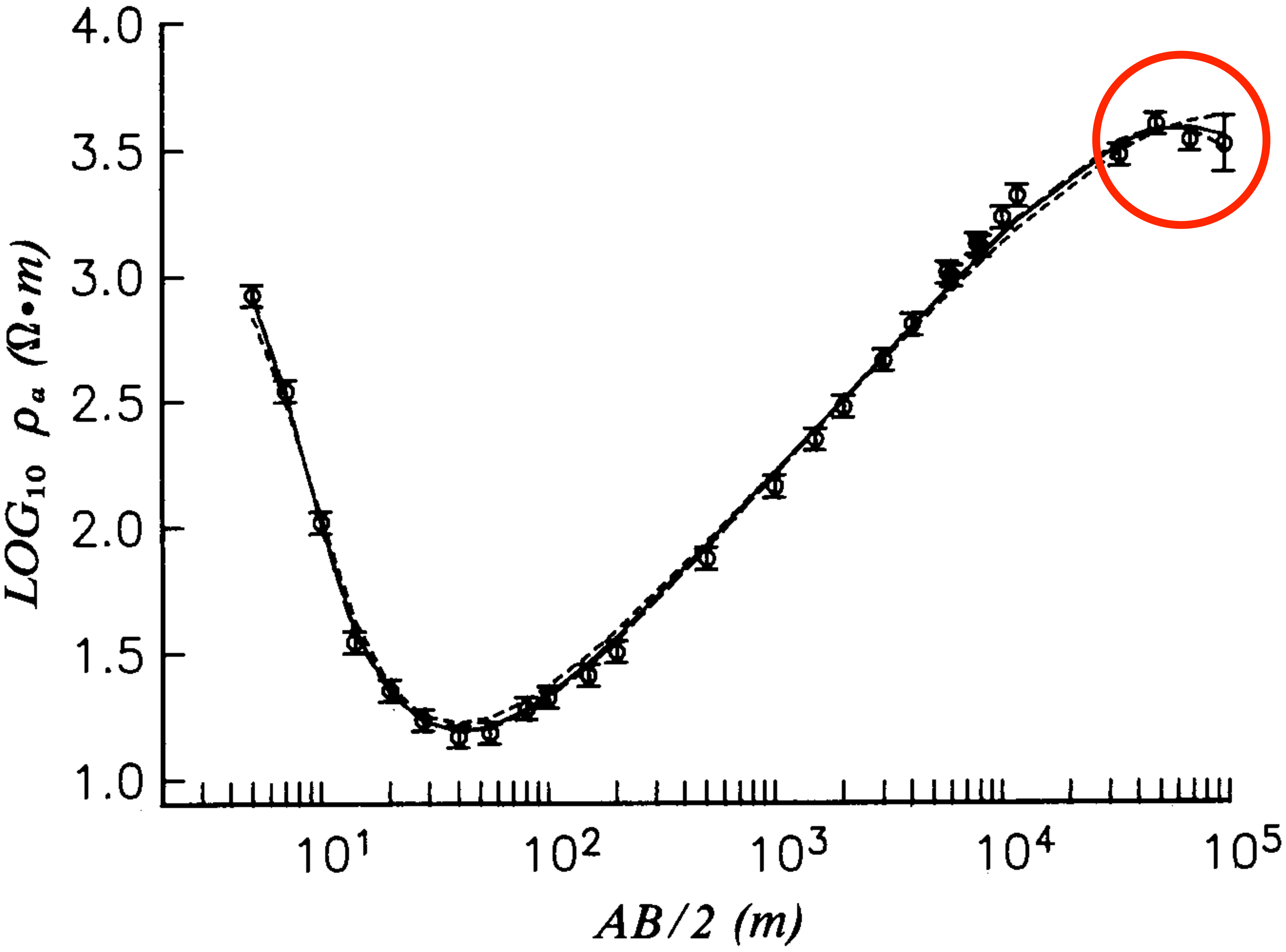
Constable et al. (1987)

Importance of measurement uncertainties.



Constable et al. (1987)

Importance of measurement uncertainties.



Constable et al. (1987)

Quantifying how well the data from our candidate model reproduce the observations.

$$\phi_d = \sum_{i=1}^{M} \left(\frac{d_i^{\text{obs}} - F[\mathbf{m}^*]_i}{\sigma_i} \right)^2$$

Quantifying how well the data from our candidate model reproduce the observations.

$$\|\mathbf{x}\|_p^p = \sum_j |x_j|^p$$

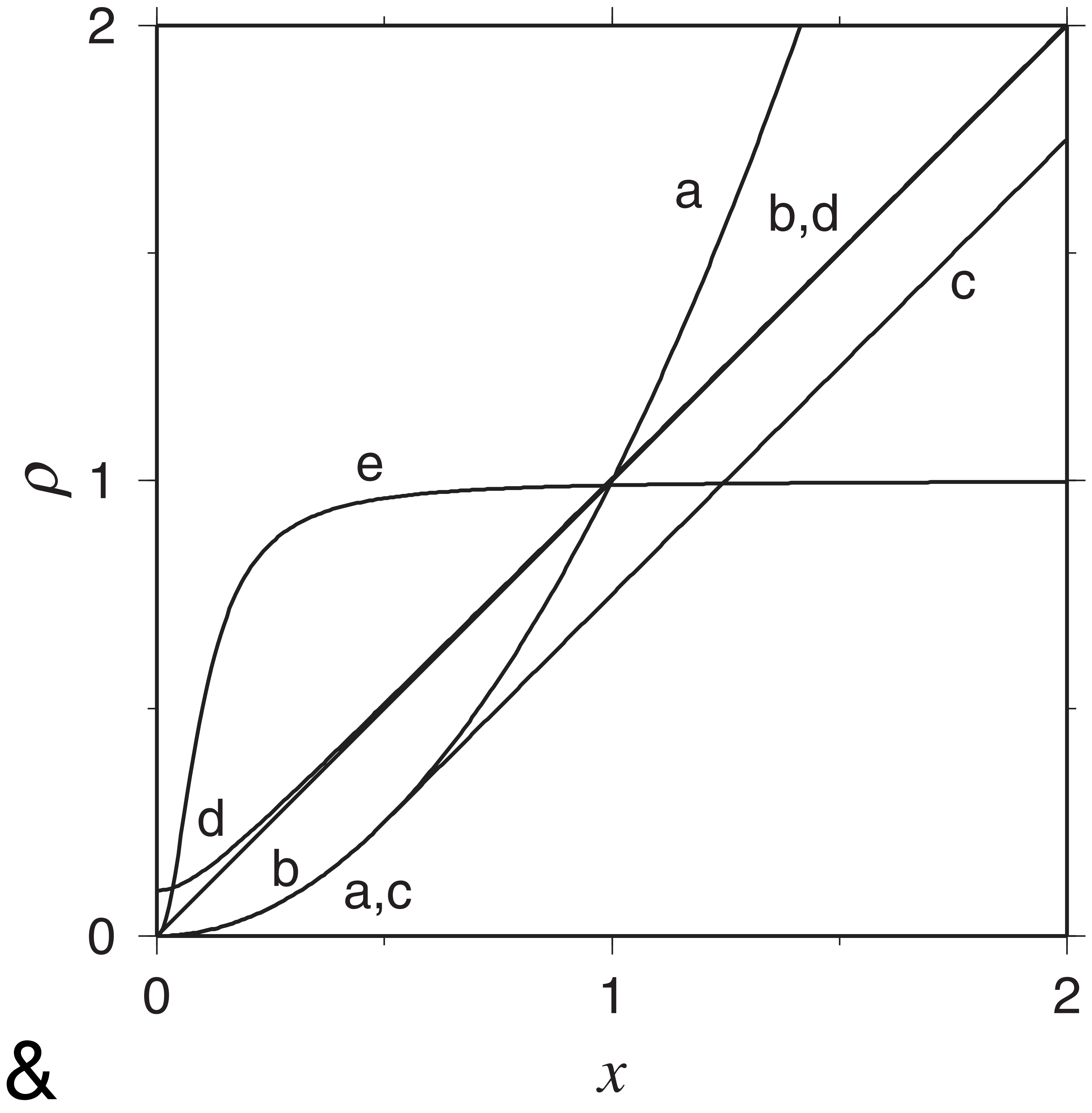
$$\phi(\mathbf{x}) = \sum_j \rho(x_j),$$

$$\rho(x) = \begin{cases} x^2 & |x| \le c \\ 2c|x| - c^2 & |x| > c \end{cases}$$

Huber M-measure

$$\rho(x) = \frac{x^2}{x^2 + \varepsilon^2}$$

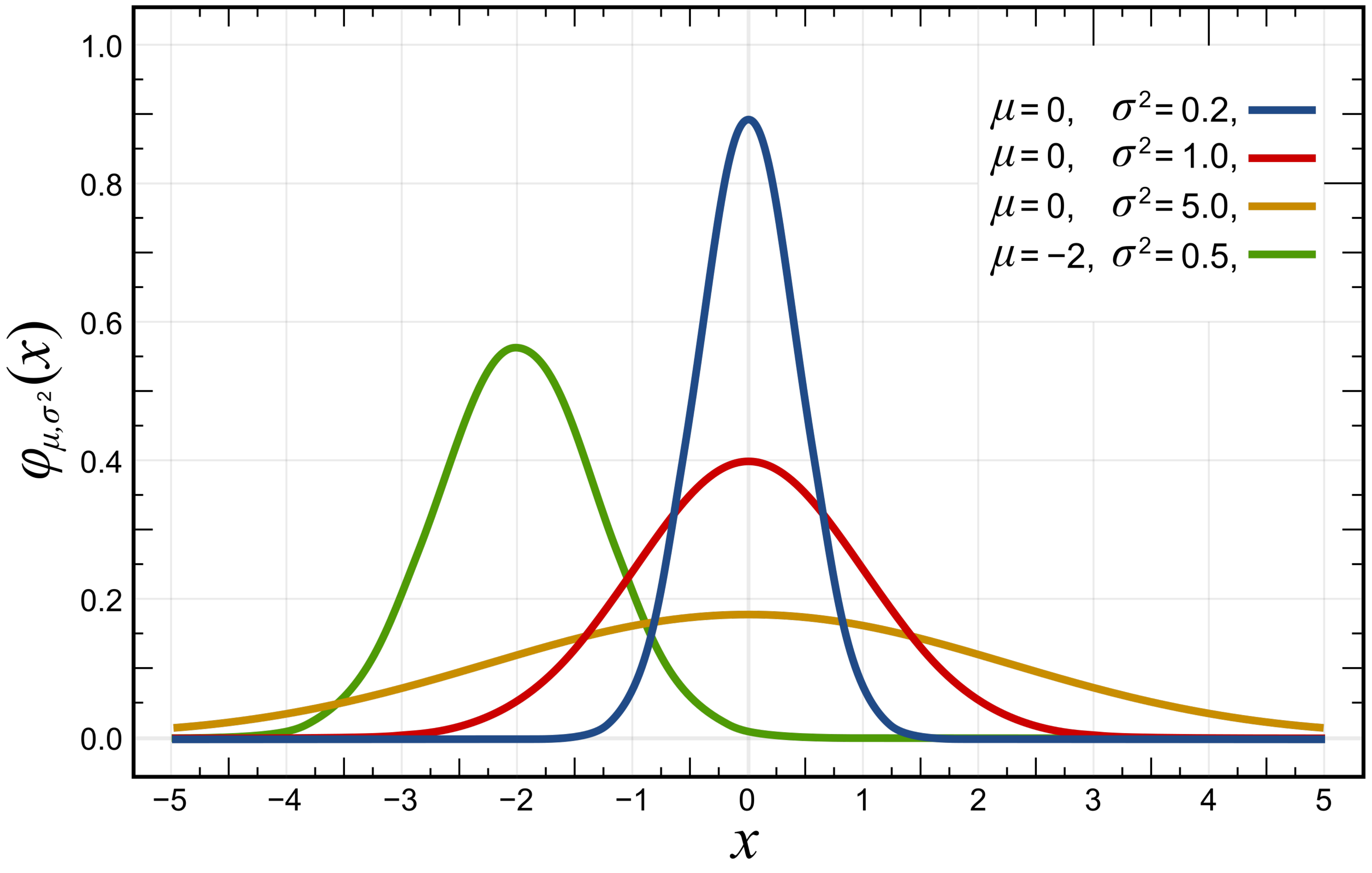
Minimum support (Portniaguine & Zhdanov, 1999; Last & Kubik, 1983)



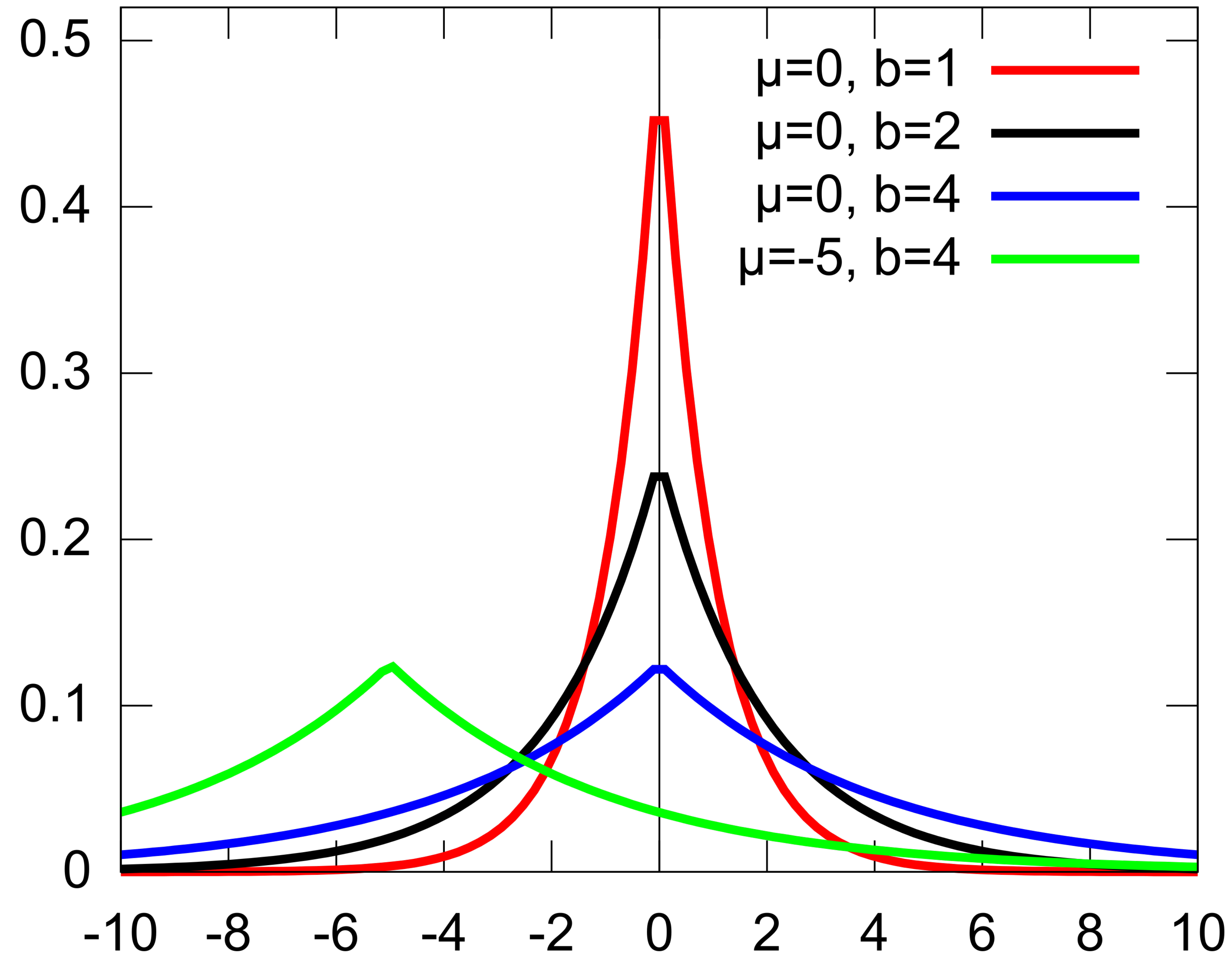
Farquharson (2008)

Quantifying how well the data from our candidate model reproduce the observations.

Gaussian



Laplace



By Inductiveload - self-made, Mathematica, Inkscape, Public Domain, https://commons.wikimedia.org/w/index.php?curid=3817954

By IkamusumeFan - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=34776178

Quantifying how well the data from our candidate model reproduce the observations.

Sometimes, the relation between data and model parameters is functional, $\mathbf{d} = \mathbf{g}(\mathbf{m})$, and in this case the likelihood function is (see equations (1.93)–(1.95))

$$L(\mathbf{m}) = \rho_D(\mathbf{g}(\mathbf{m})) \quad . \qquad (2.3)$$

A couple of examples of such a likelihood function are given as a footnote.[28]

[28]For instance, if d^i_{obs} represents the observed data values and σ^i the estimated mean deviations, assuming double exponentially distributed observational errors gives $L(\mathbf{m}) = \exp(-\sum_i |g^i(\mathbf{m}) - d^i_{obs}| / \sigma^i)$. If $\mathbf{C}_D$ represents the covariance operator describing estimated data uncertainties and uncertainty correlations, assuming a Gaussian distribution gives (equation (1.101)) $L(\mathbf{m}) = \exp(-\frac{1}{2}(\mathbf{g}(\mathbf{m}) - \mathbf{d}_{obs})^t \mathbf{C}_D^{-1} (\mathbf{g}(\mathbf{m}) - \mathbf{d}_{obs}))$. Some other examples are given in chapter 1.

Tarantola (2005)

Aside

Quantifying the "quality" of the fit ...

Smith and Booker (1988)

Spearman statistic to assess "whiteness" of data fit

$$D = \sum_{i=1}^{N} (S_i - R_i)^2$$

low values, positive correlation, high values, negative correlation;
based on rankings

Aside

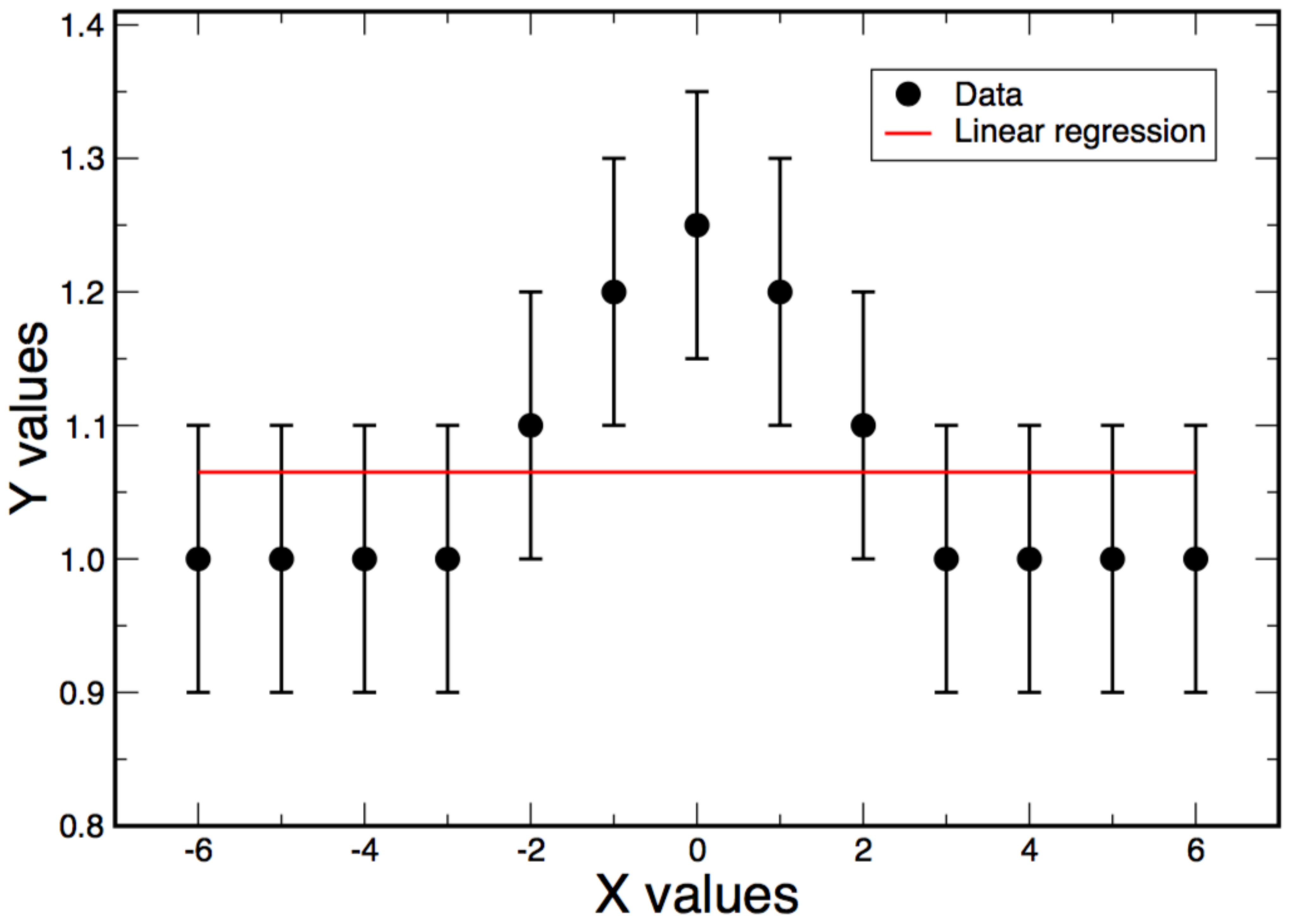
Quantifying the "quality" of the fit ...

Jones (2019)

Durbin-Watson statistic

$$dw_1 = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=i}^{n}(e_i - \bar{e})^2}$$

auto-correlated?



Include Durbin-Watson statistic in inversion:

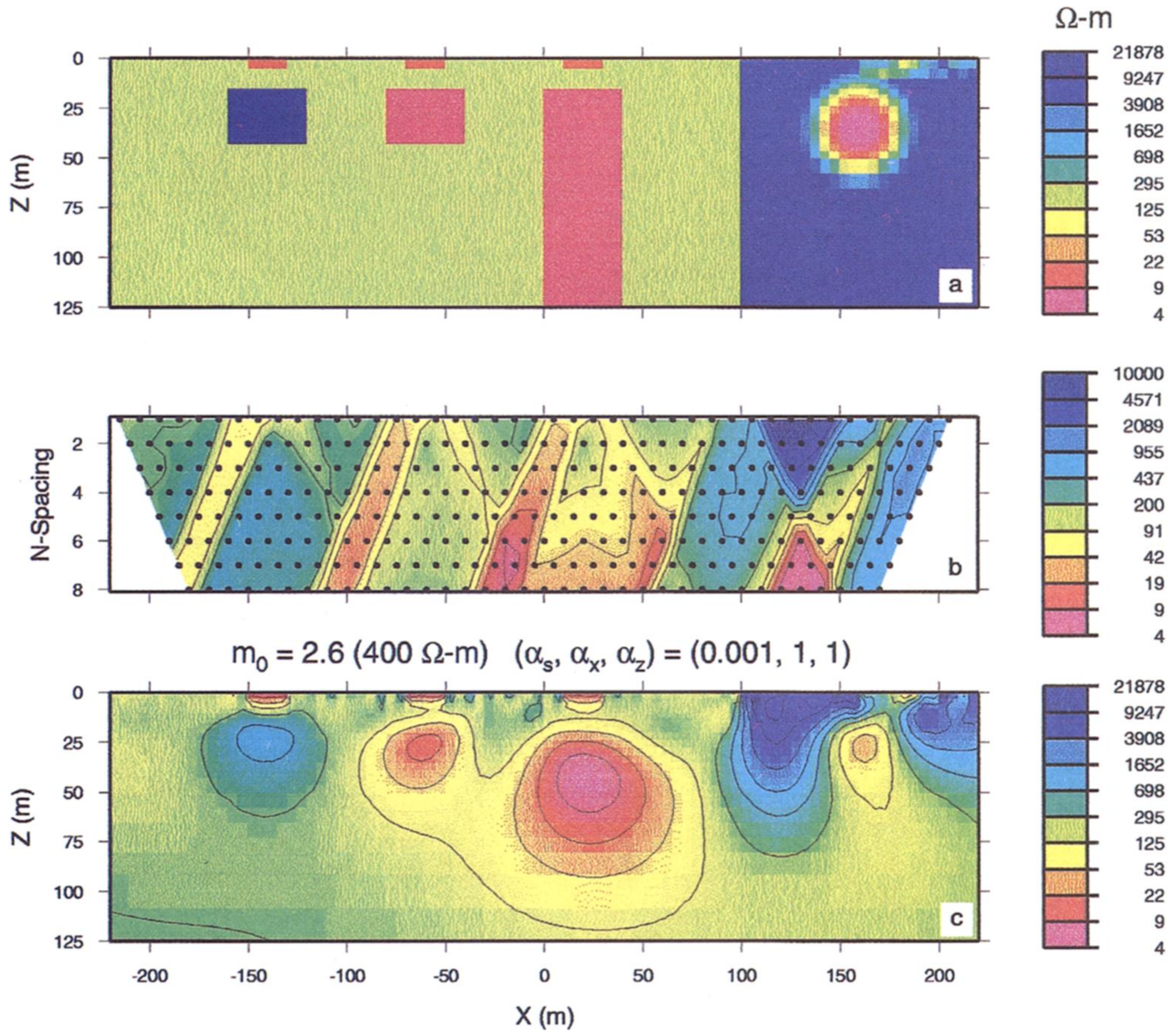
$$\Phi = \lambda_1 nRMS + \lambda_2(dw_1 - 2)^2 + \lambda_3 S(m)$$

Quantifying how well the data from our candidate model reproduce the observations.

Minimize the measure of data misfit ... optimization ... ?

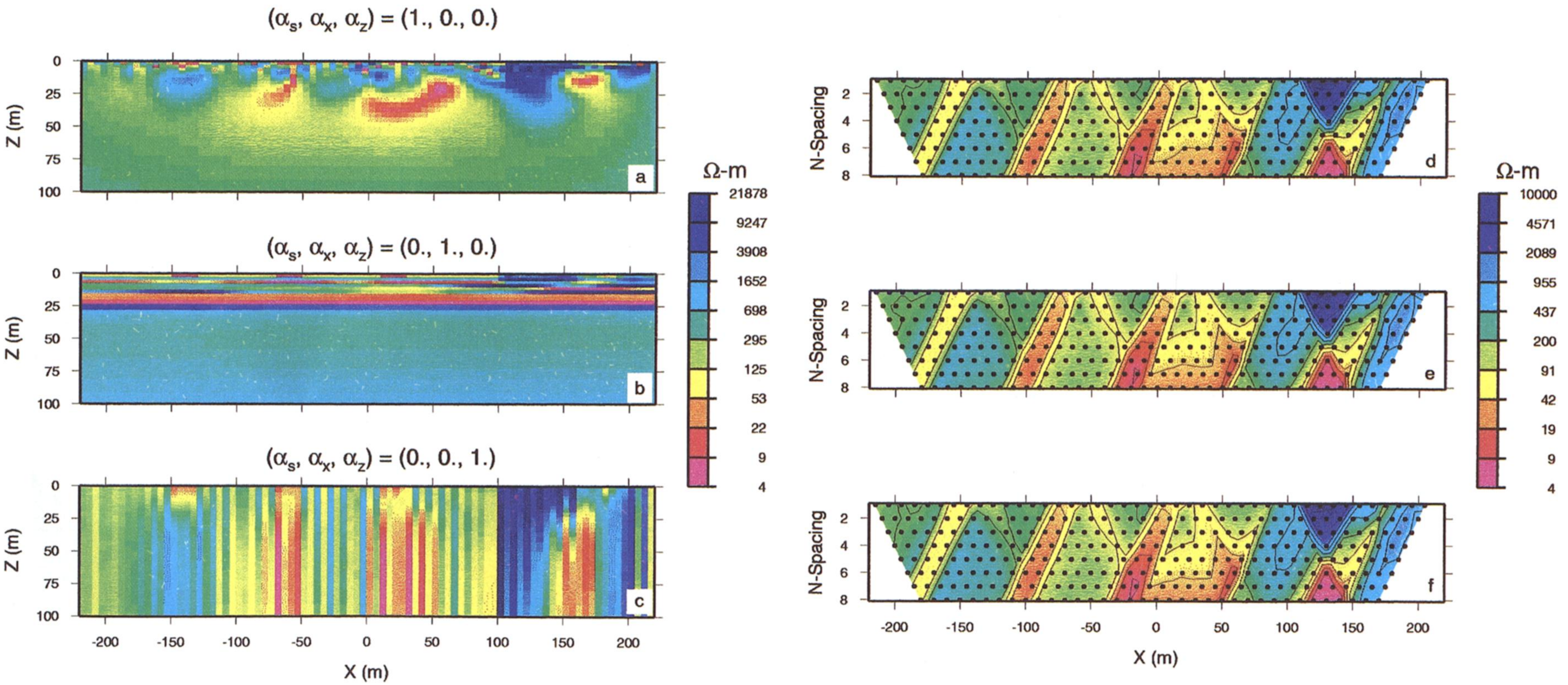
(Just curve fitting!)

Oh ...



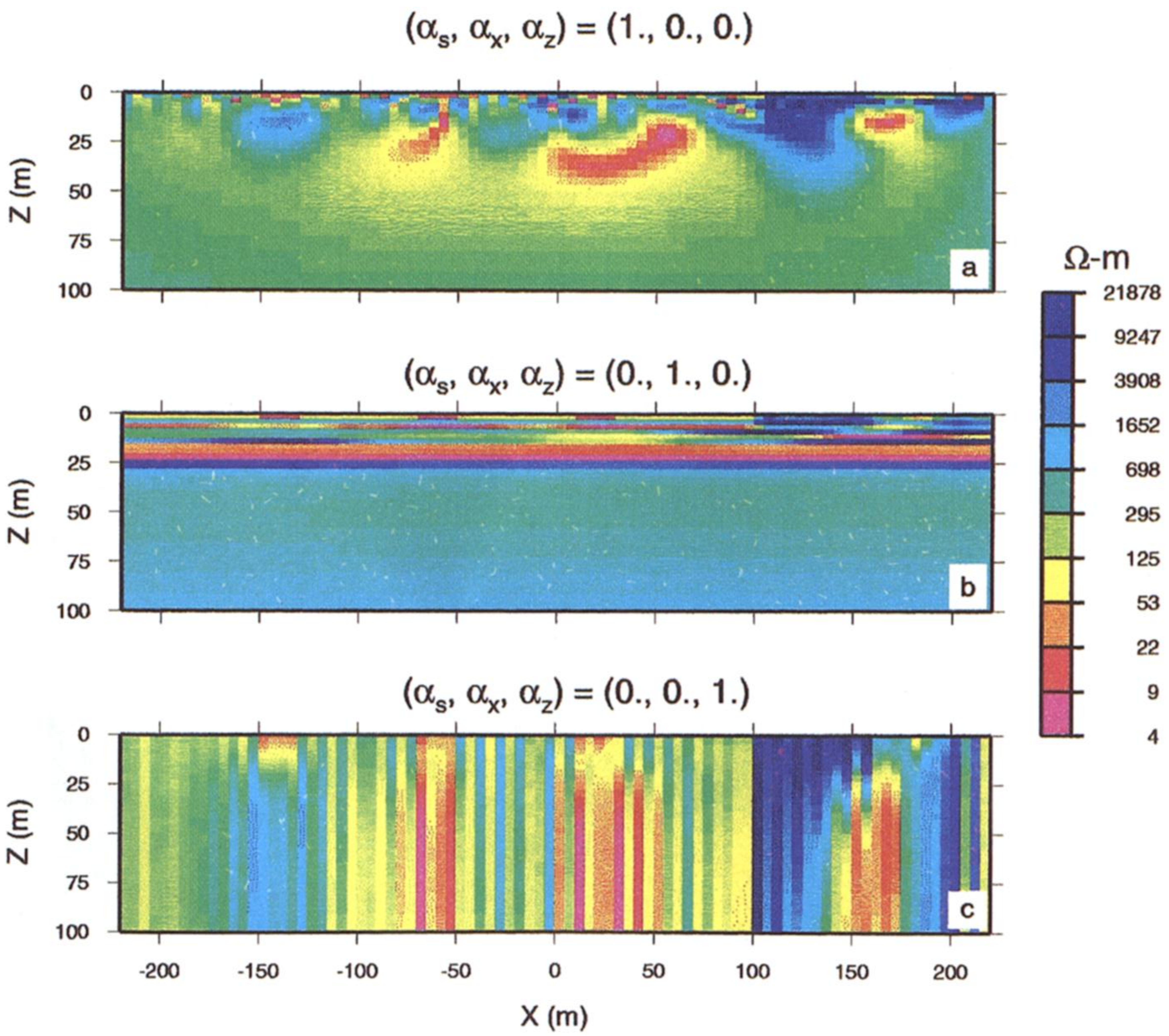
Oldenburg and Li (1999)

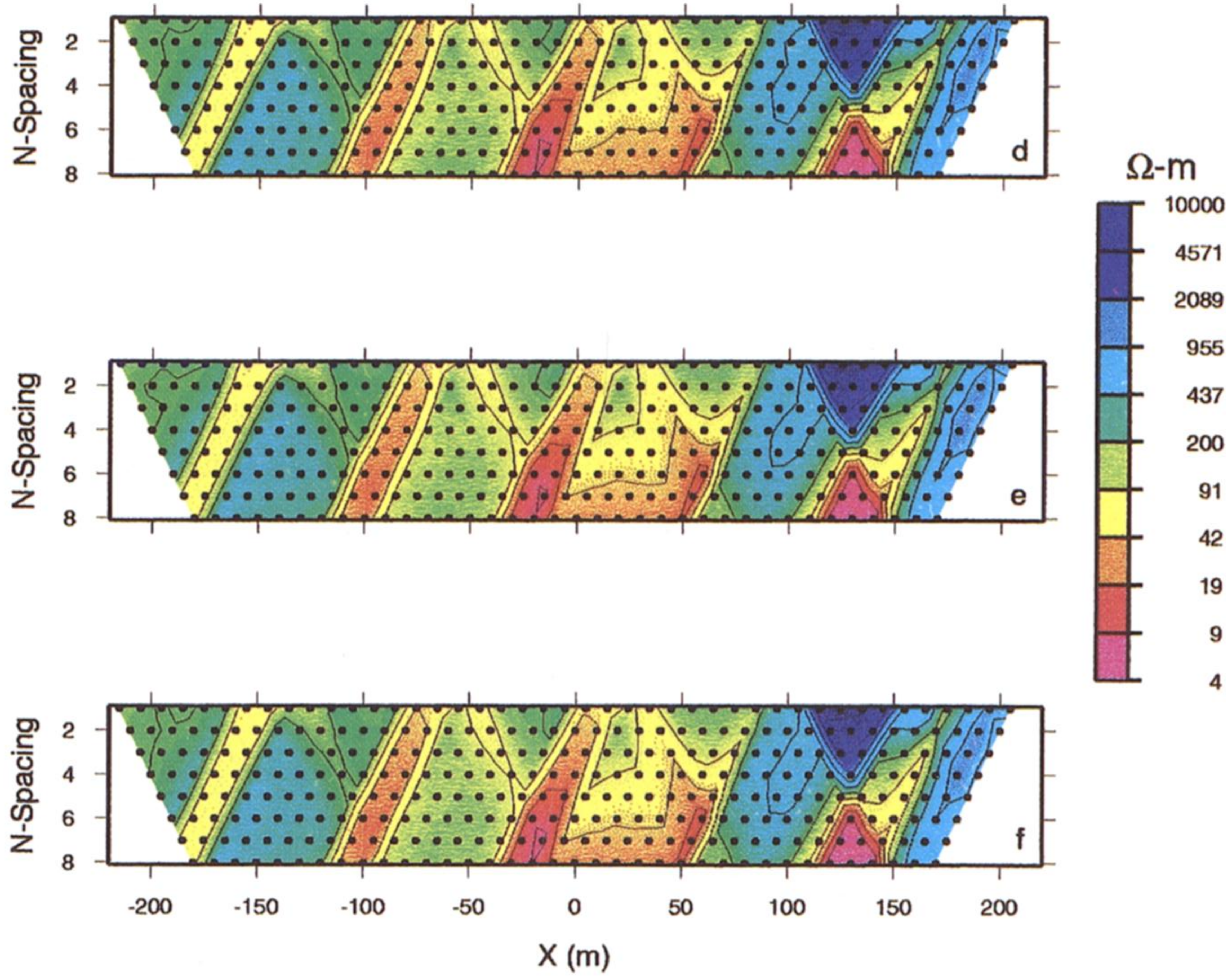
Oh ... Non-uniqueness.



Oldenburg and Li (1999)

Aside Appraisal ?





Oldenburg and Li (1999)

Non-uniqueness! Well then, let's deal with that.

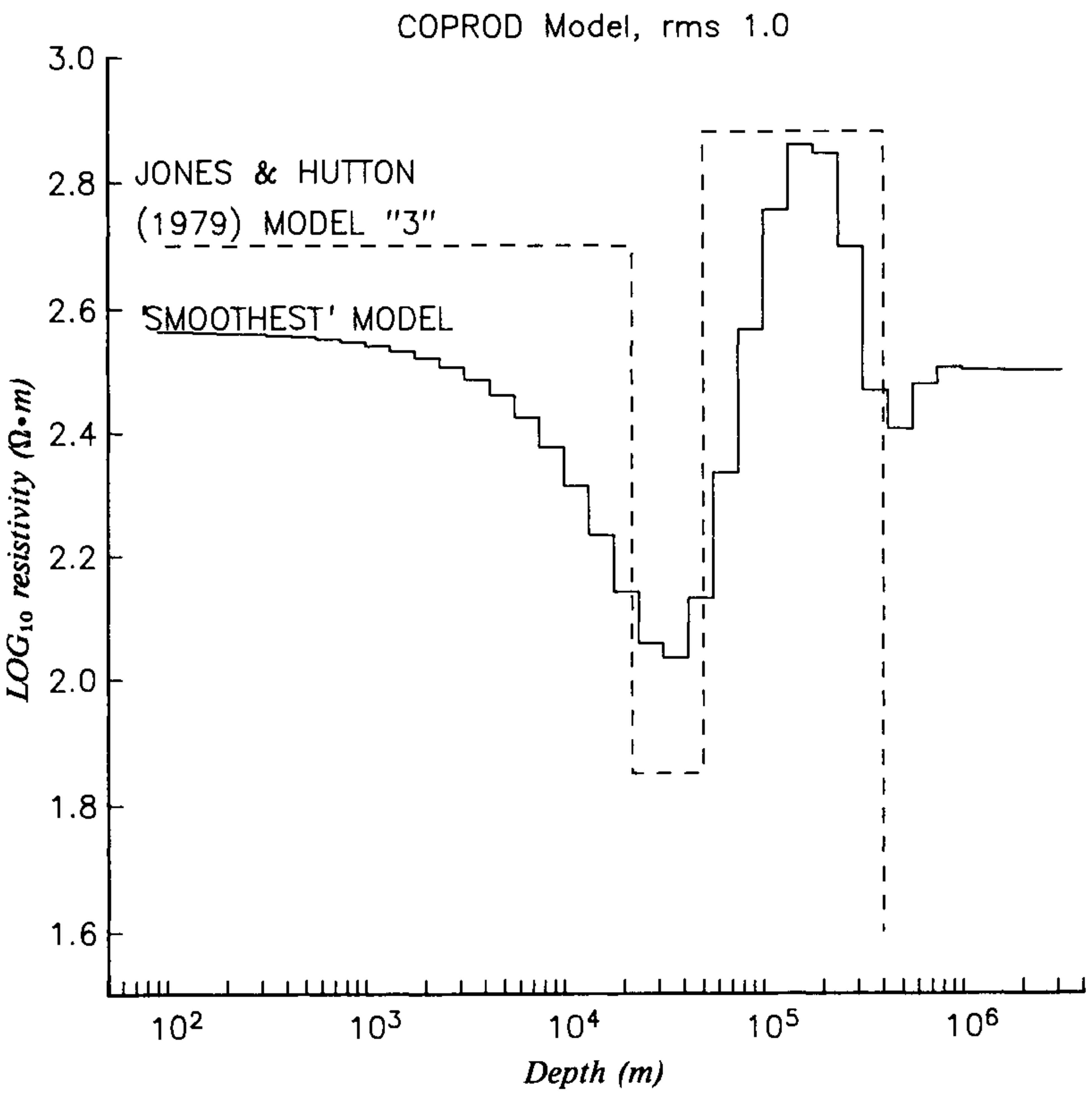
Constable et al. (1987)

$$U = \| \partial \mathbf{m} \|^2 + \mu^{-1}\{\| \mathbf{Wd} - \mathbf{WF}[\mathbf{m}] \|^2 - X_*^2\}.$$

"Occam's inversion"

model roughness

smoothest model, simplest model



Non-uniqueness! Well then, let's deal with that.

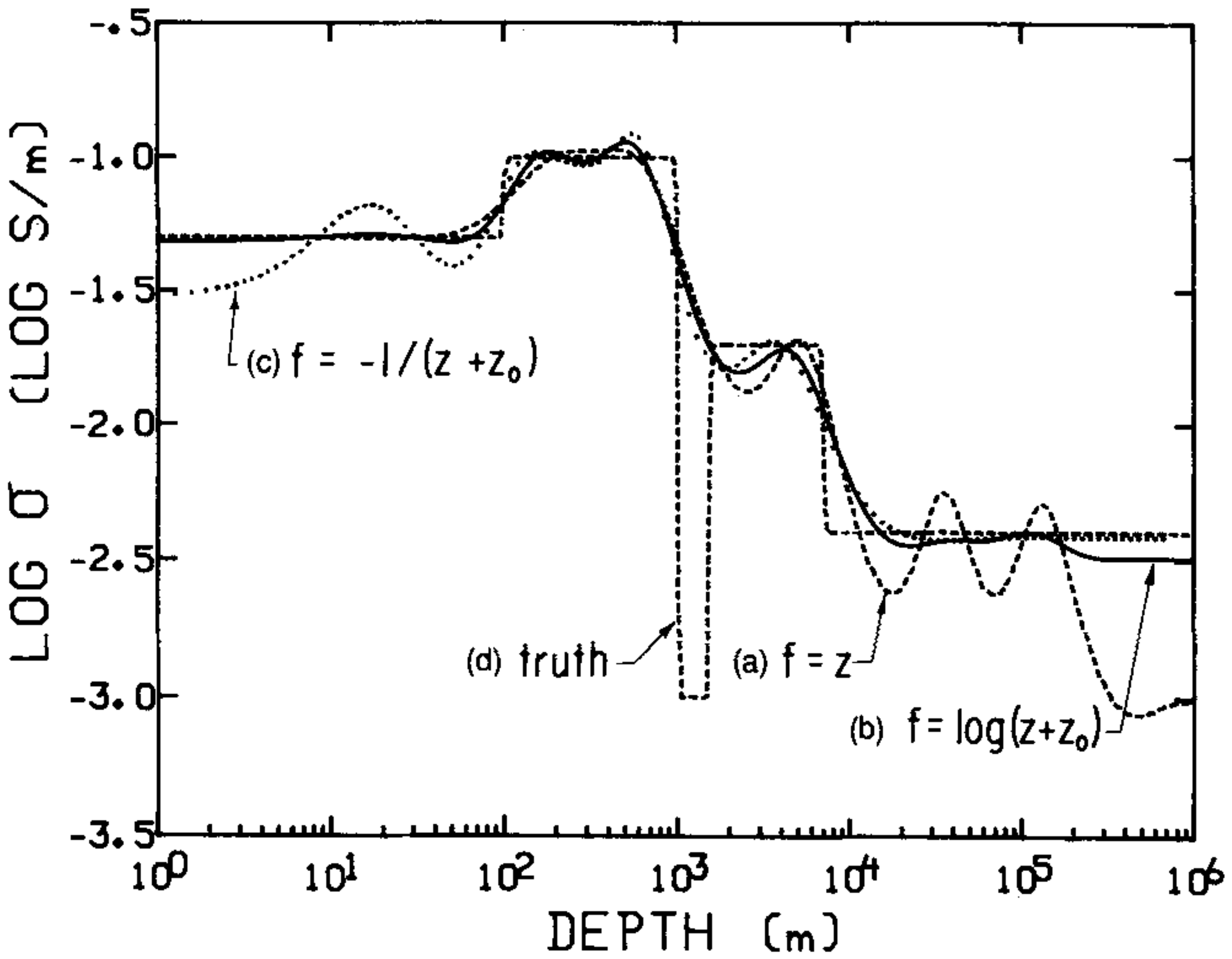
Smith and Booker (1988)

$$\chi^2 = \sum_{i=1}^{2N} \left[\frac{\Delta\gamma_i}{\varepsilon_i}\right]^2, \qquad F(m, f) = \int_0^\infty \left[\frac{dm}{df(z)}\right]^2 df(z)$$

$$W(m_1, \chi_t^2, \beta_t) = F(m_1, f) + \beta_t \chi_t^2$$

"minimum-structure" inversion

flattest model



"Occam's inversion", "minimum-structure" inversion:

A combination of ...

$$\Phi \;=\; \phi_d \,+\, \gamma\,\phi_m$$

a measure of how well the observations are reproduced (small value is good), and

a measure of whatever-we-think-will-give-us-a-good-model (small value good).

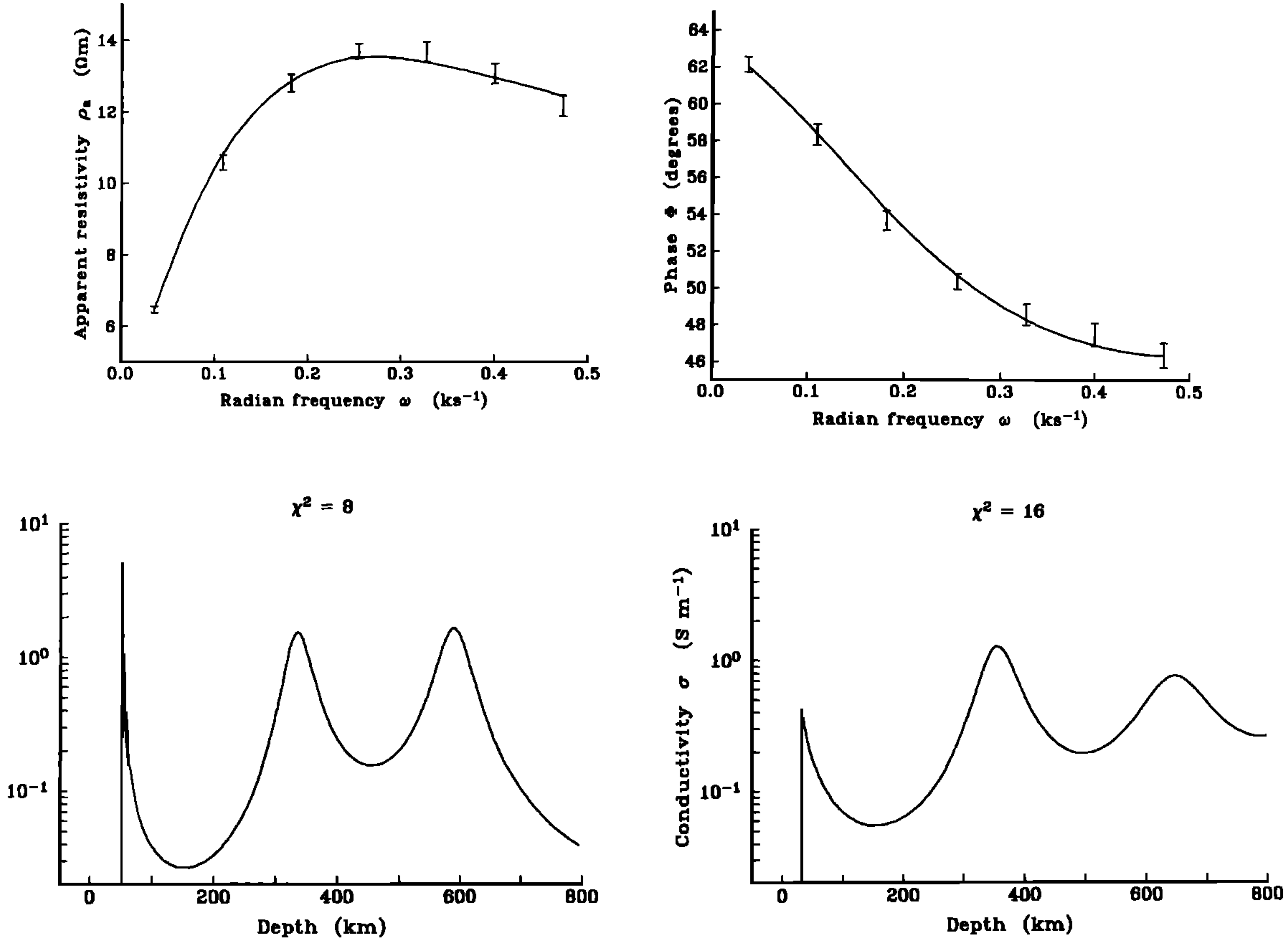
"Occam's inversion", "minimum-structure" inversion:

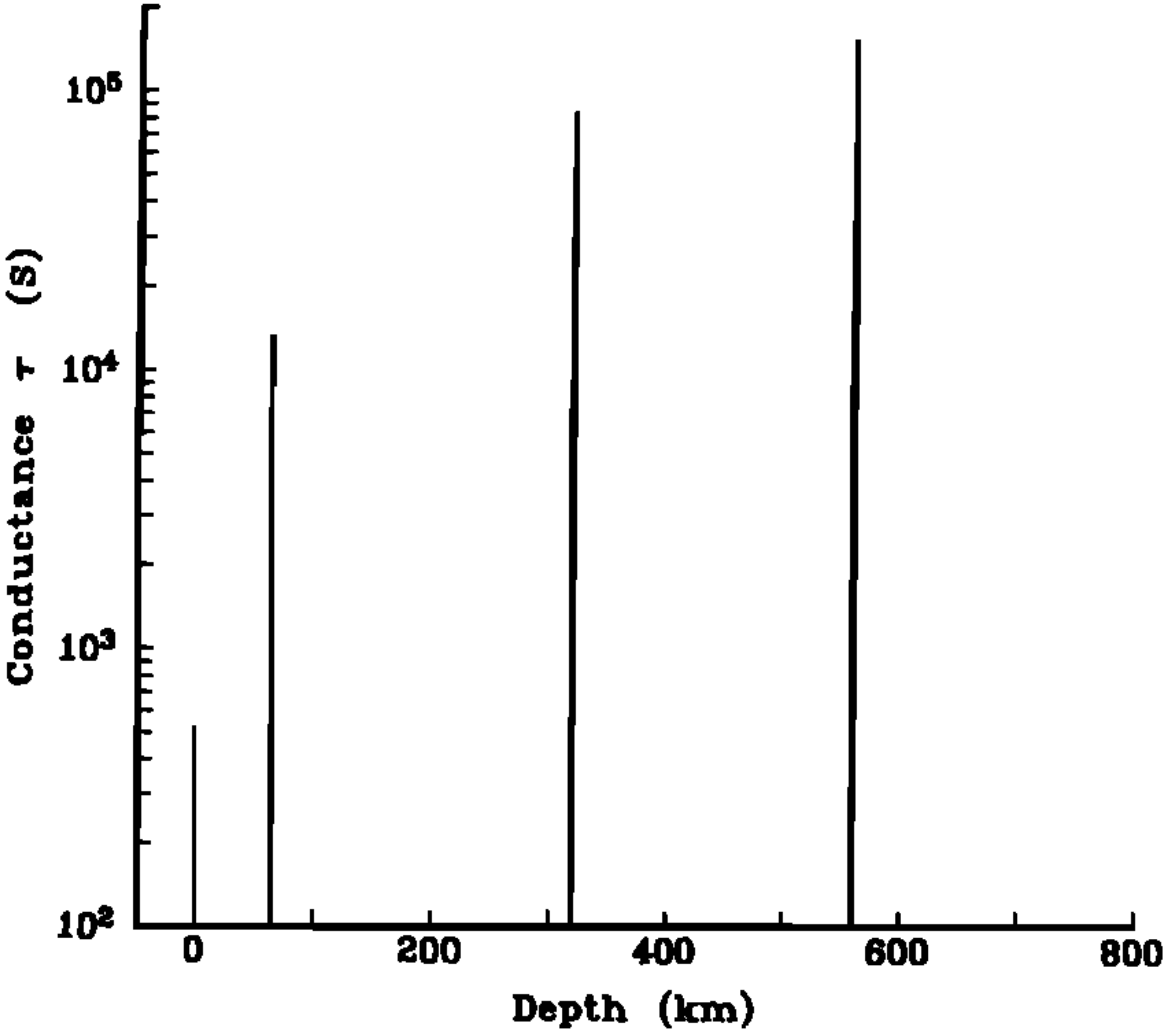
Has proved to be very successful: everyone uses this approach (gravity, magnetic, DC/IP, seismic travel-time tomography; and FWI is getting there).

Arguably most important aspect is that chances of a useful model from any one run are very high (compare with needing to re-start parameter estimation algorithms from lots of different initial models). Reliable, robust.

Aside

"D+" models of Parker (1980)

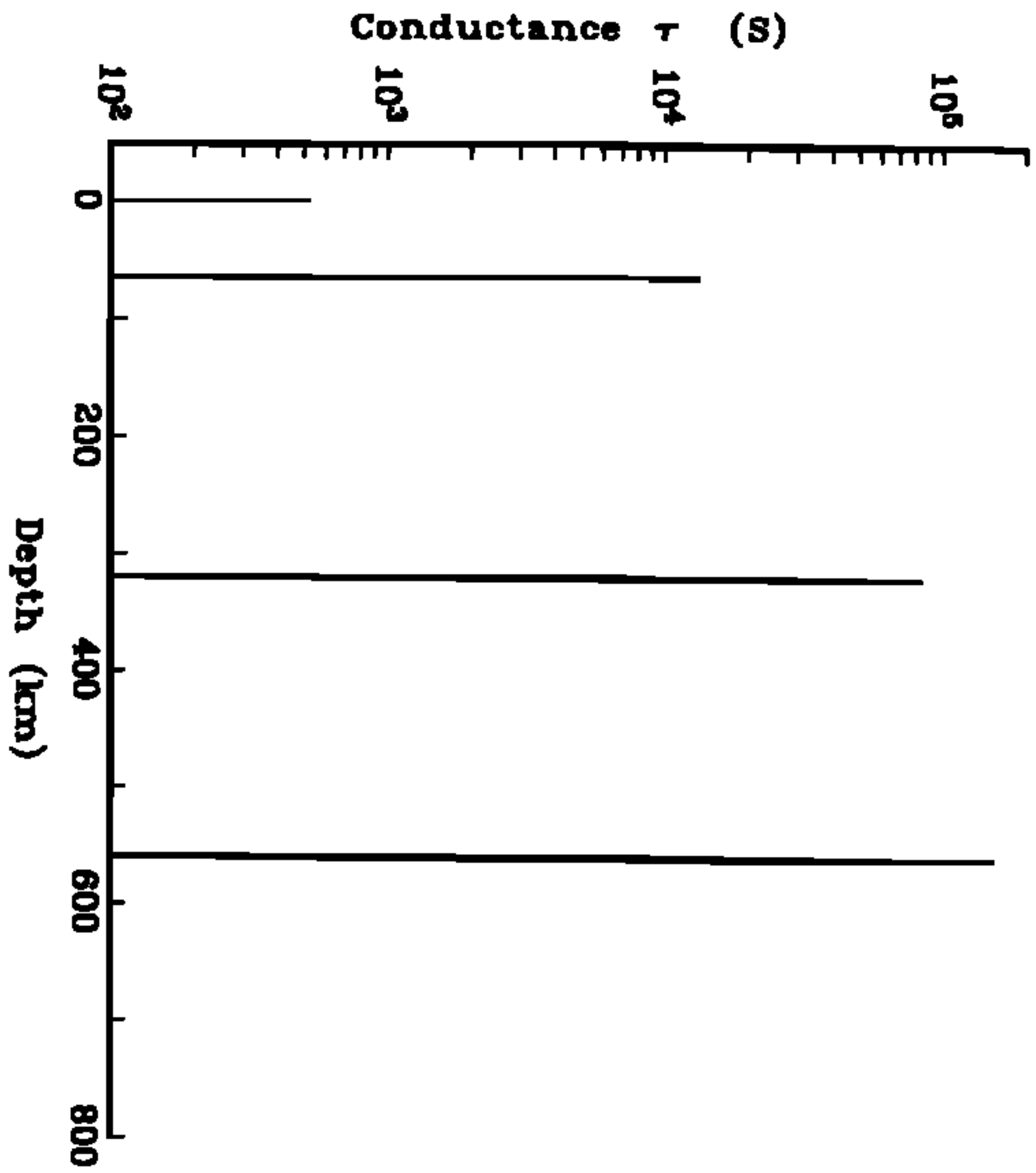




Parker and Whaler (1981)

Aside

"D+" models of Parker (1980)



Parker and Whaler (1981)

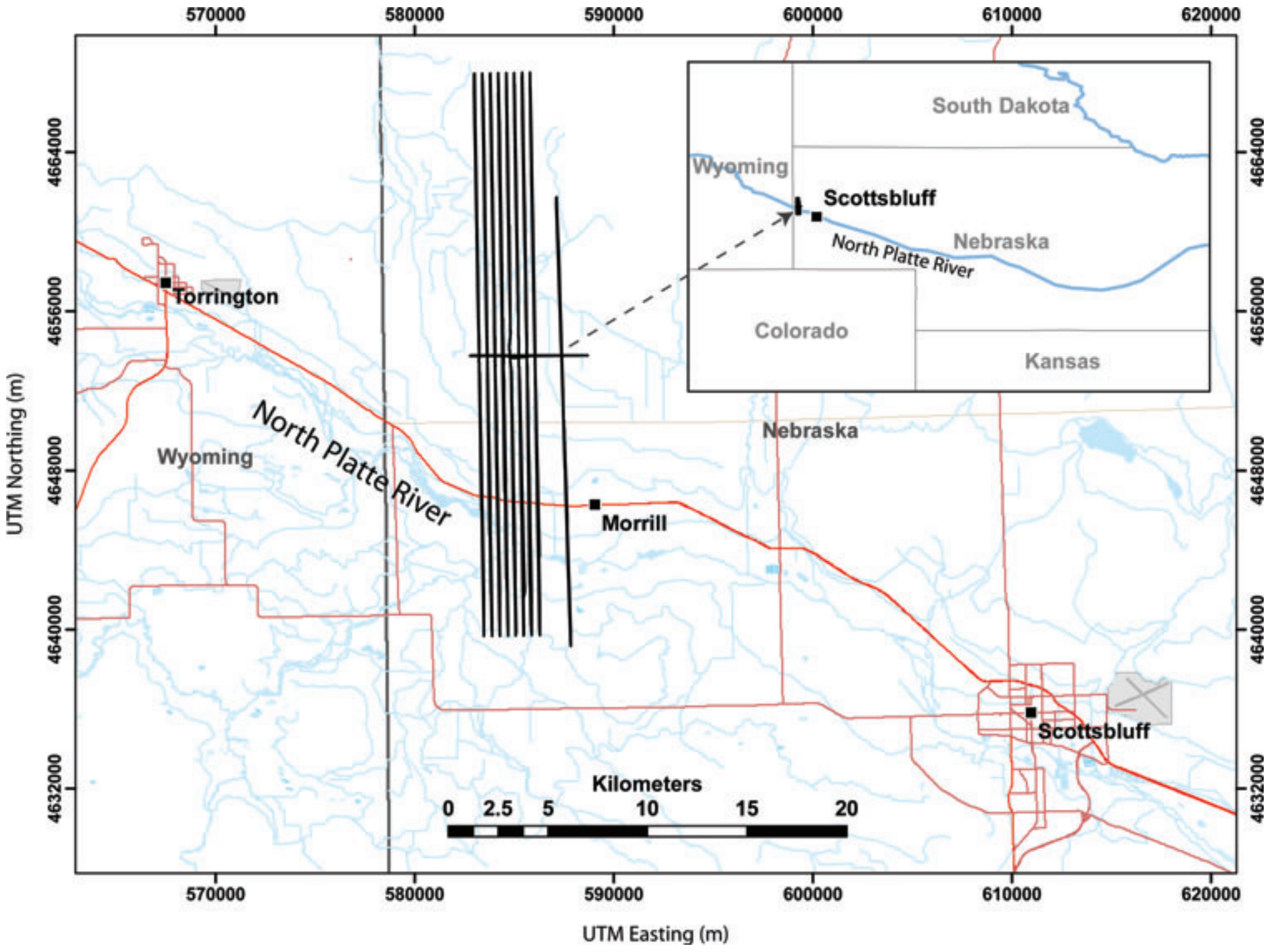
Rather than finding a unique model that reproduces the data ...

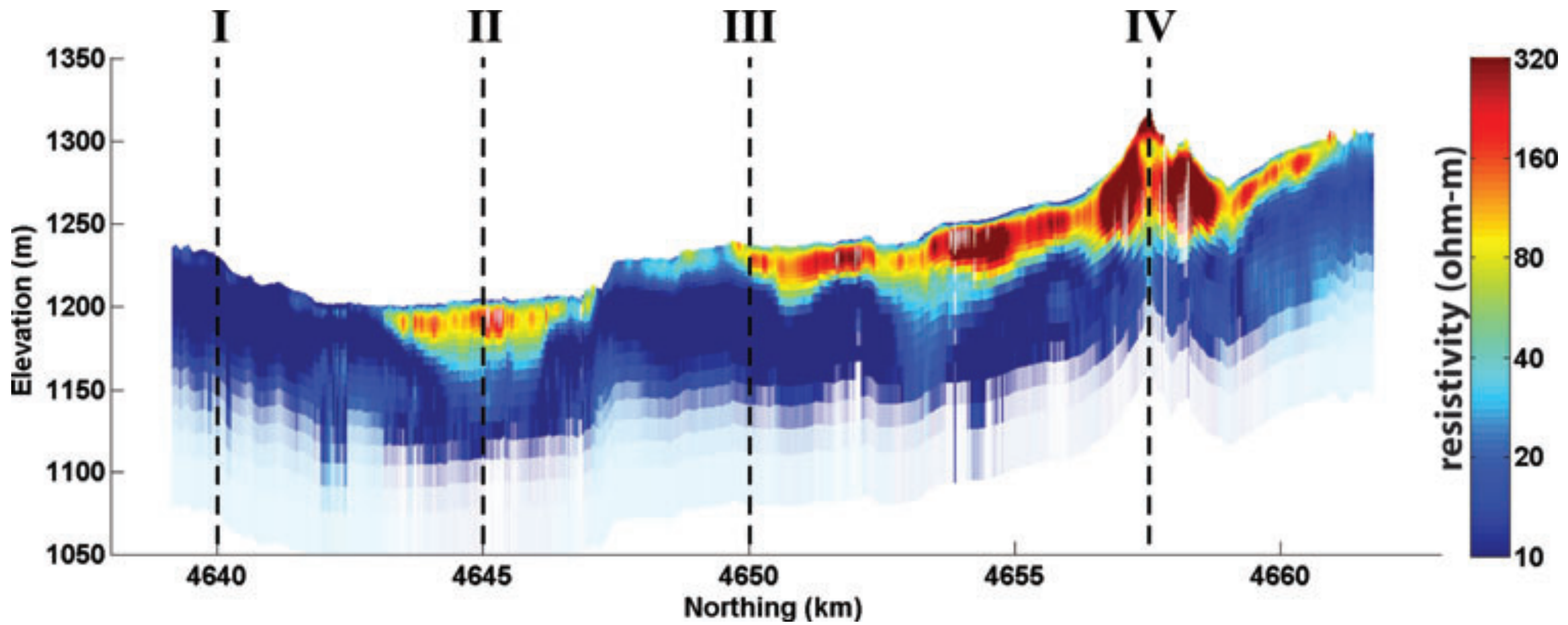
$$\phi_d \;=\; \sum_{i=1}^{M} \left(\frac{d_i^{\text{obs}} - F[\mathbf{m}^*]_i}{\sigma_i} \right)^2$$

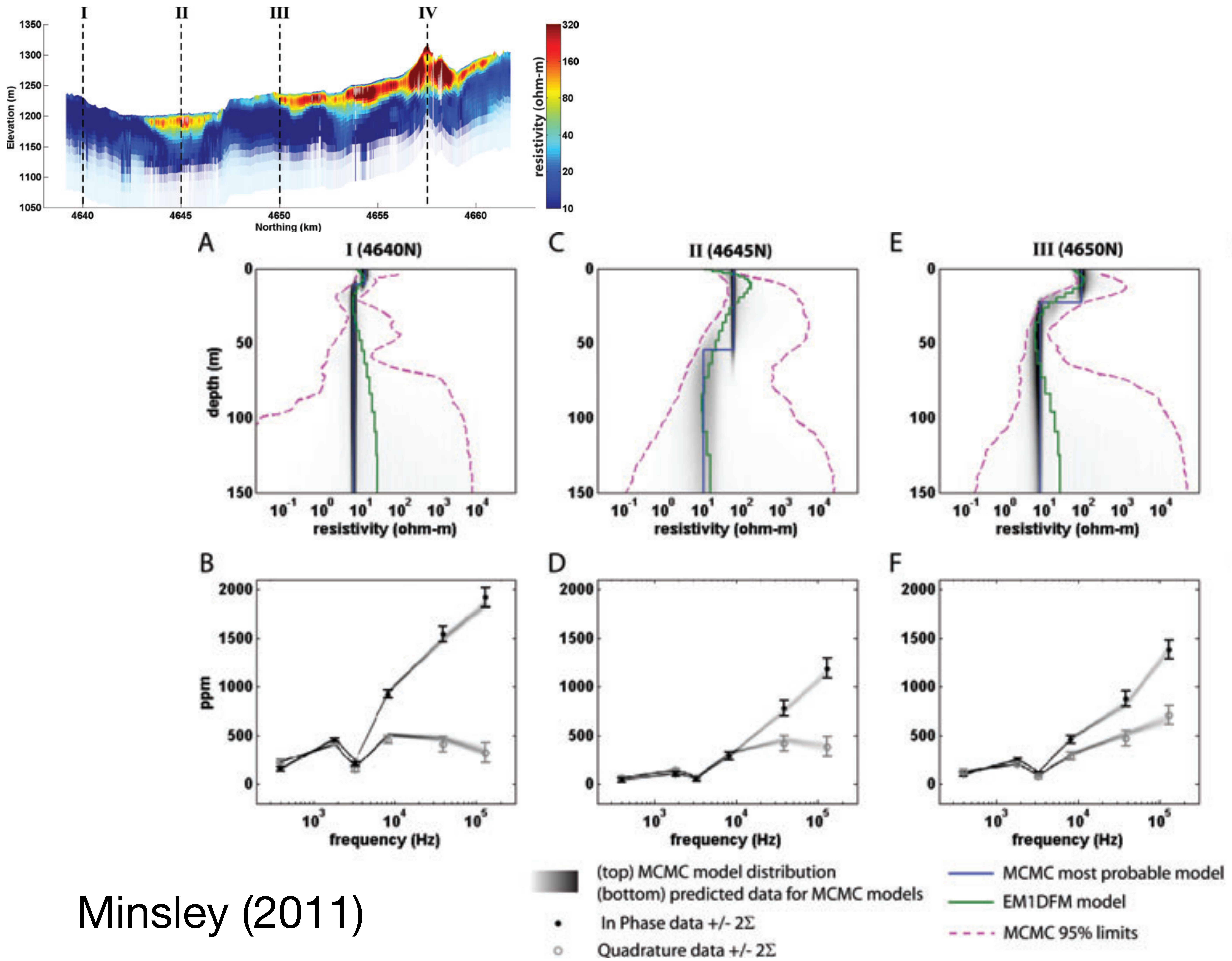
Finding all models that give a suitable misfit ... sampling.

Bayesian Markov chain Monte Carlo algorithm for model assessment

Minsley (2011)

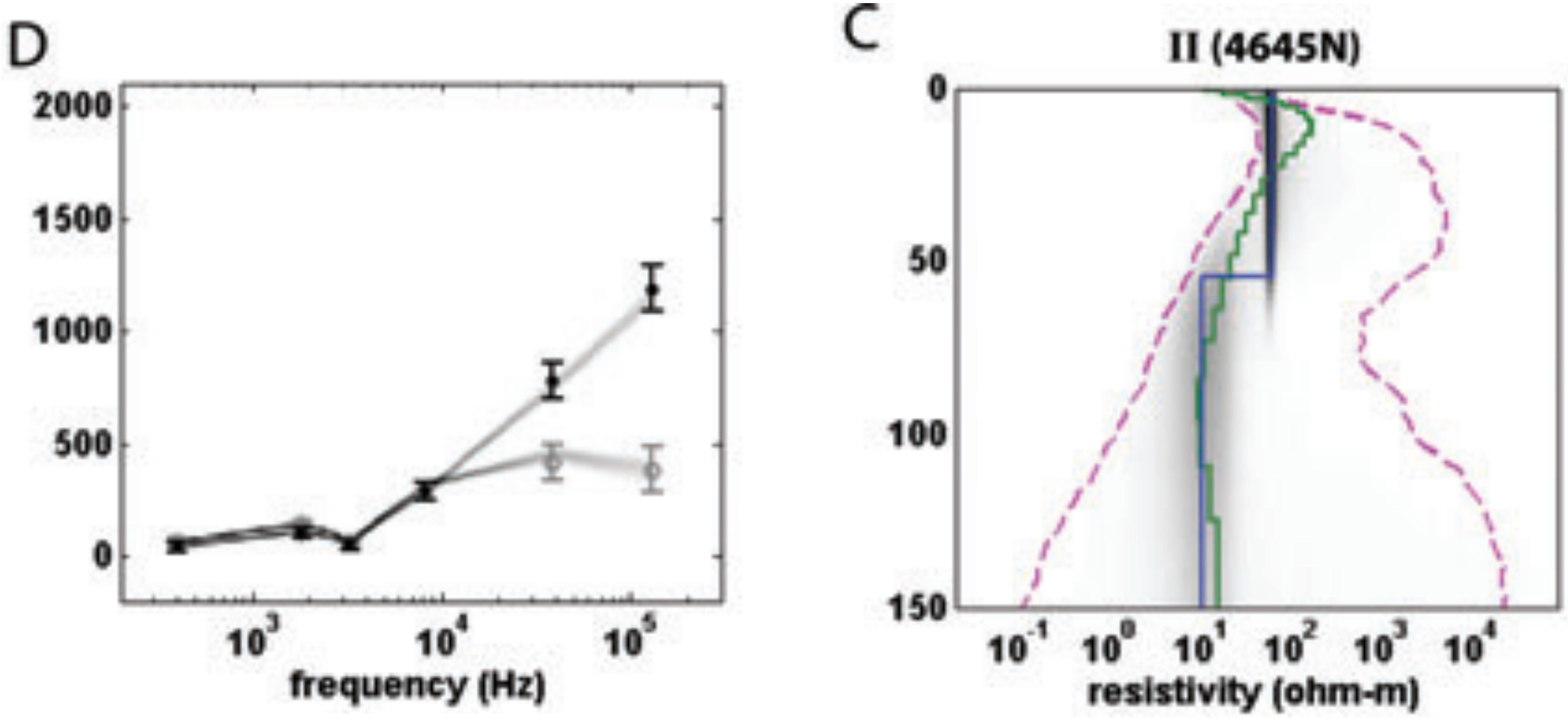






Minsley (2011)

Bayesian Markov chain Monte Carlo algorithm for model assessment



Minsley (2011)

Outline

→ ~~Inversion~~ optimization background:

- why optimization
- data misfit
- minimizing data misfit, non-uniqueness
- measure of model somethingorother
- descent-based, gradient-based optimization (linearization)
- or sampling and selection of collections of models we like

Forward modelling

Sensitivities

Descent/gradient/derivative/linearization-based algorithms

The Conclusion

Future work/thoughts

- different models, different approaches?
- D+ ?

Outline

~~Inversion~~ optimization background:

- why optimization
- data misfit
- minimizing data misfit, non-uniqueness
- measure of model somethingorother
- descent-based, gradient-based optimization (linearization)
- or sampling and selection of collections of models we like

→ Forward modelling

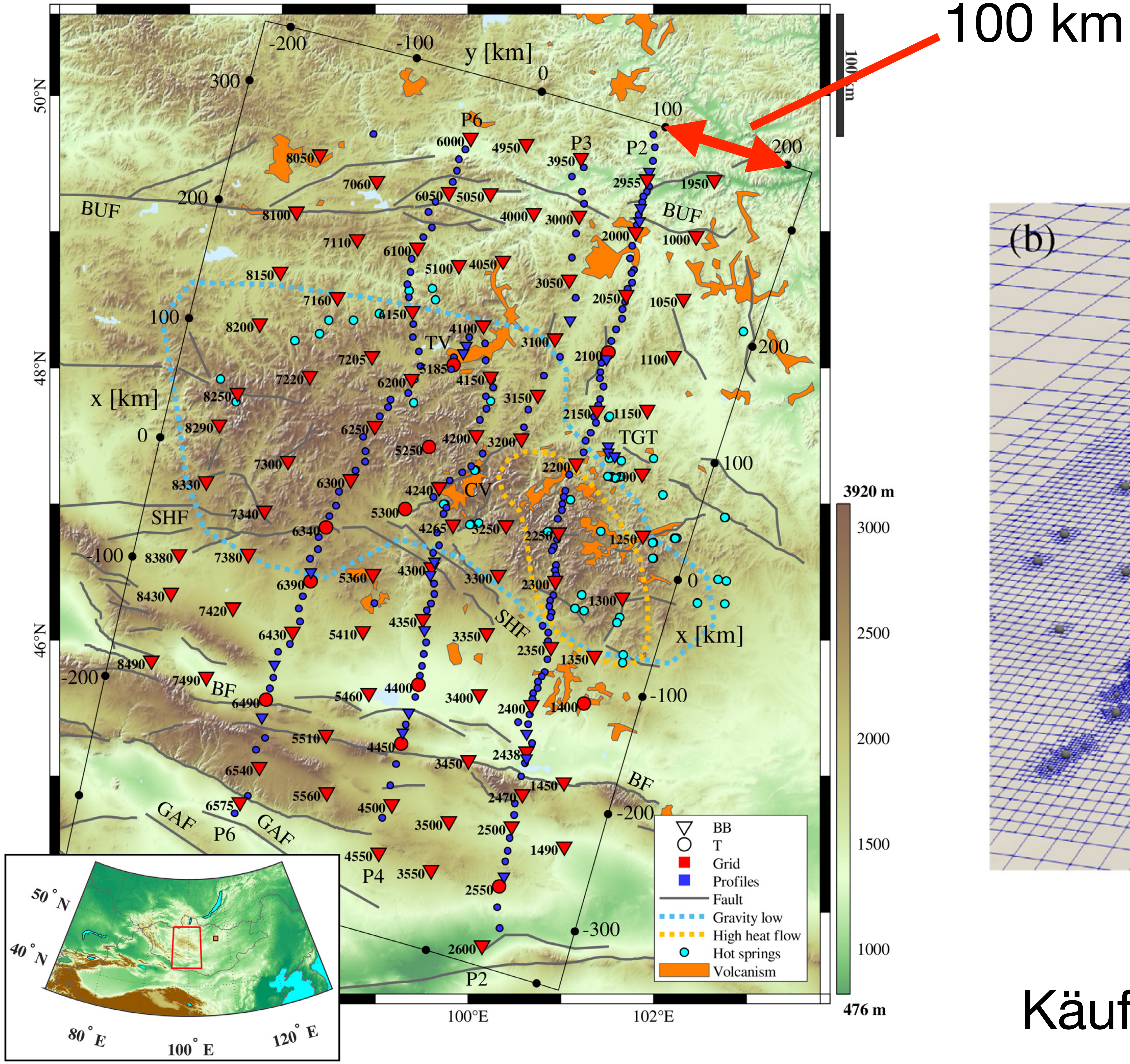
Sensitivities

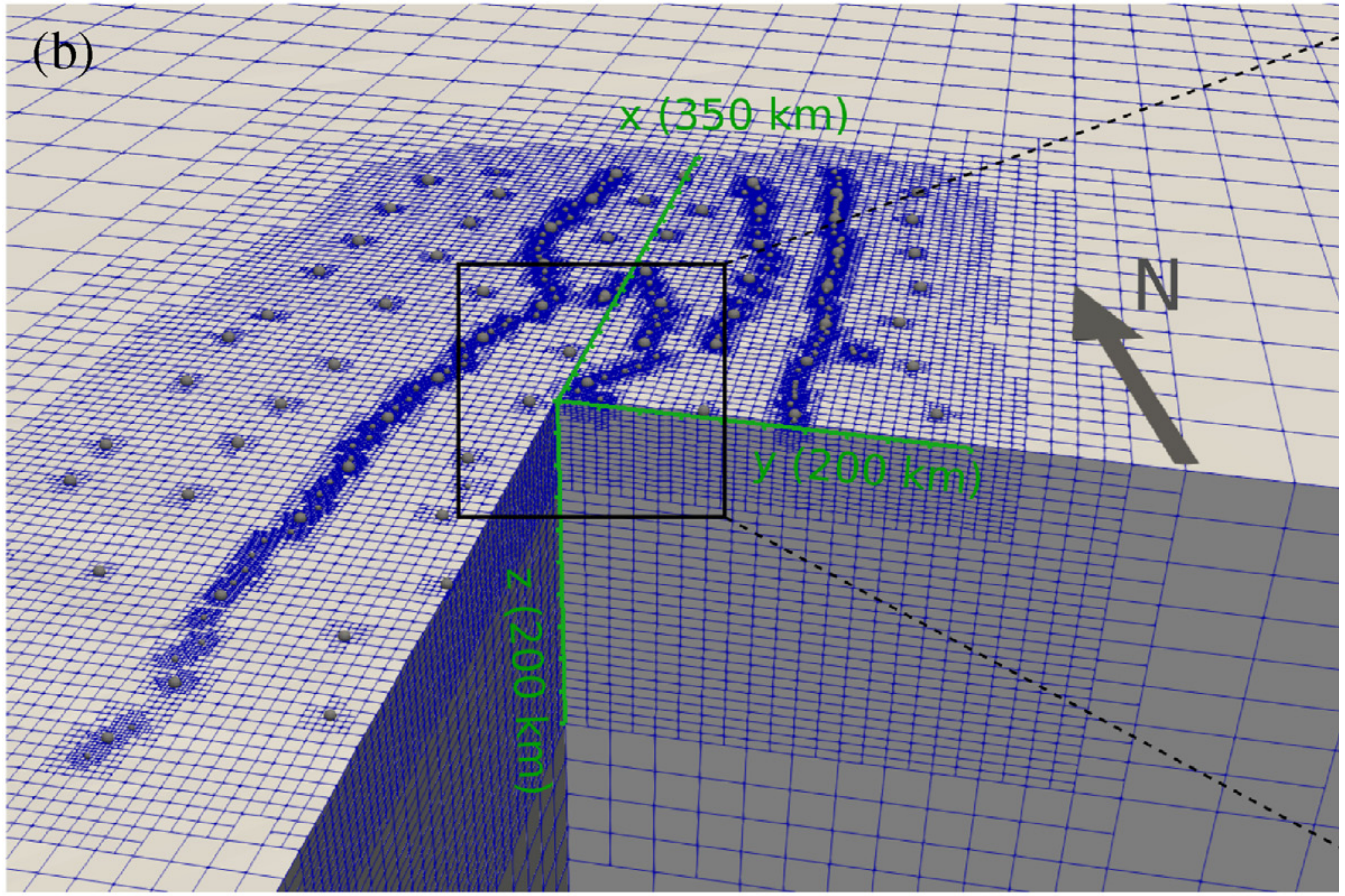
Descent/gradient/derivative/linearization-based algorithms

The Conclusion

Future work/thoughts

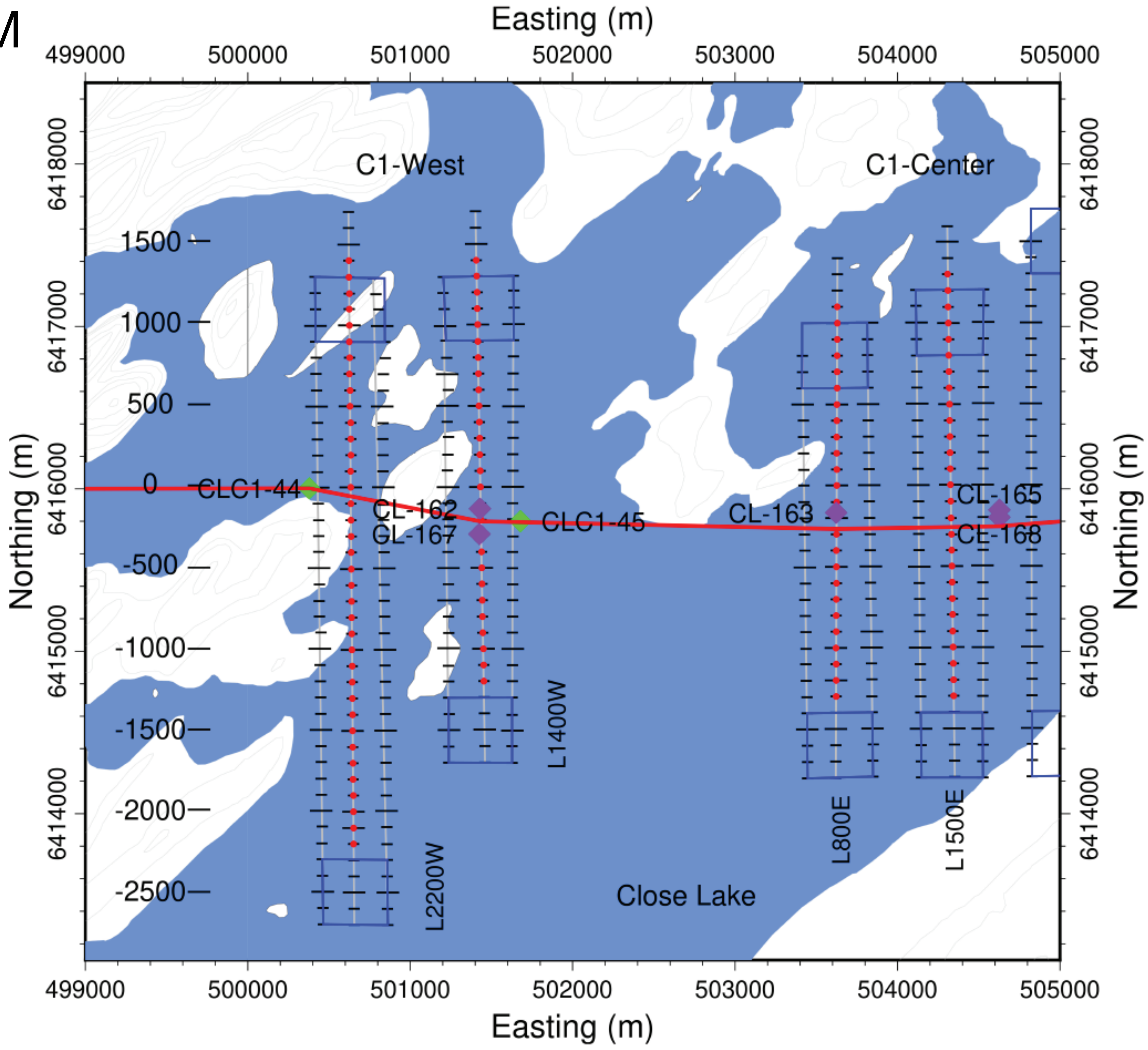
- different models, different approaches?
- D+ ?





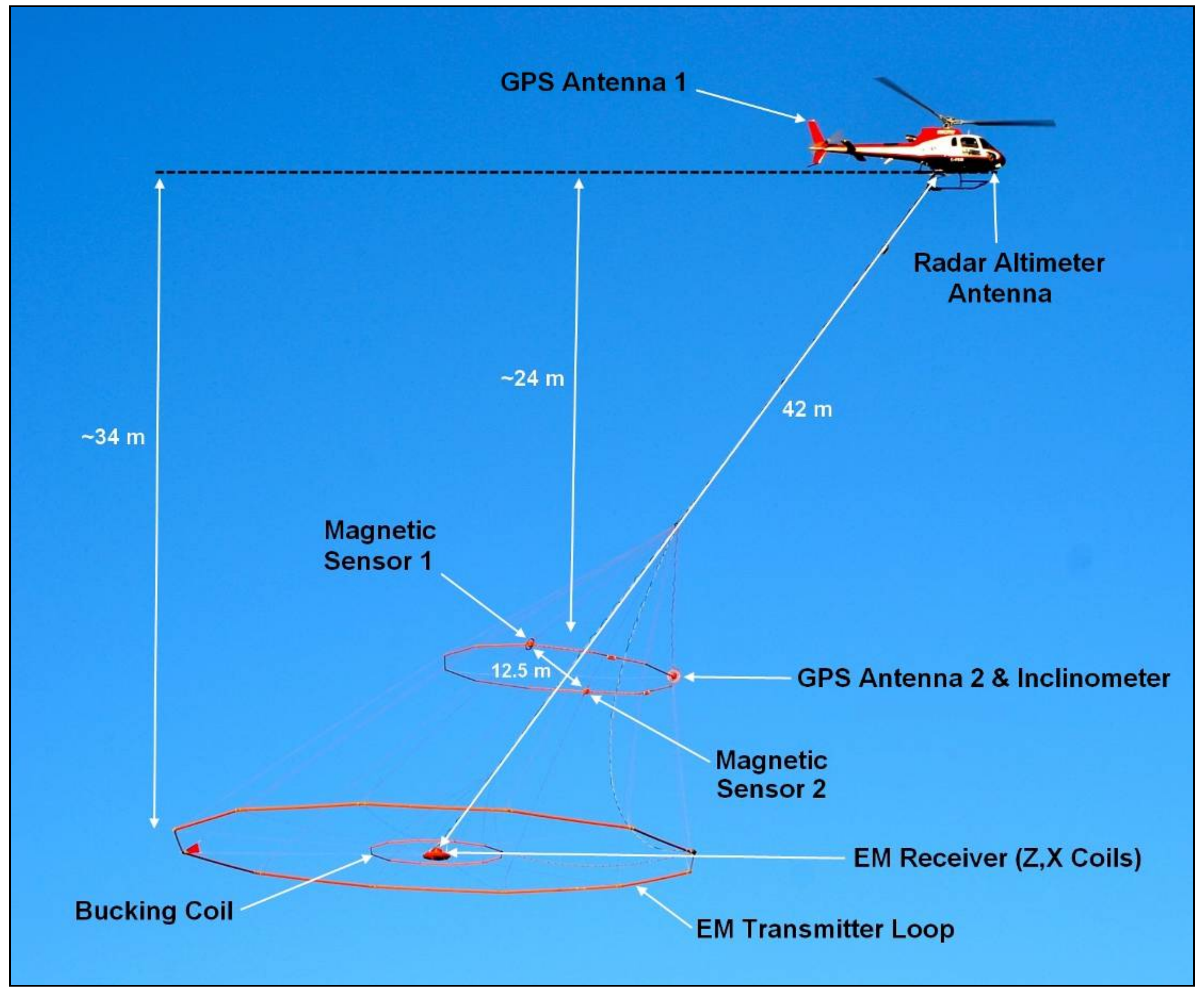
Käufl et al. (2020)

Moving-loop TEM



64 Tx-Rx pairs in C1-West grid

Lu et al. (2021)



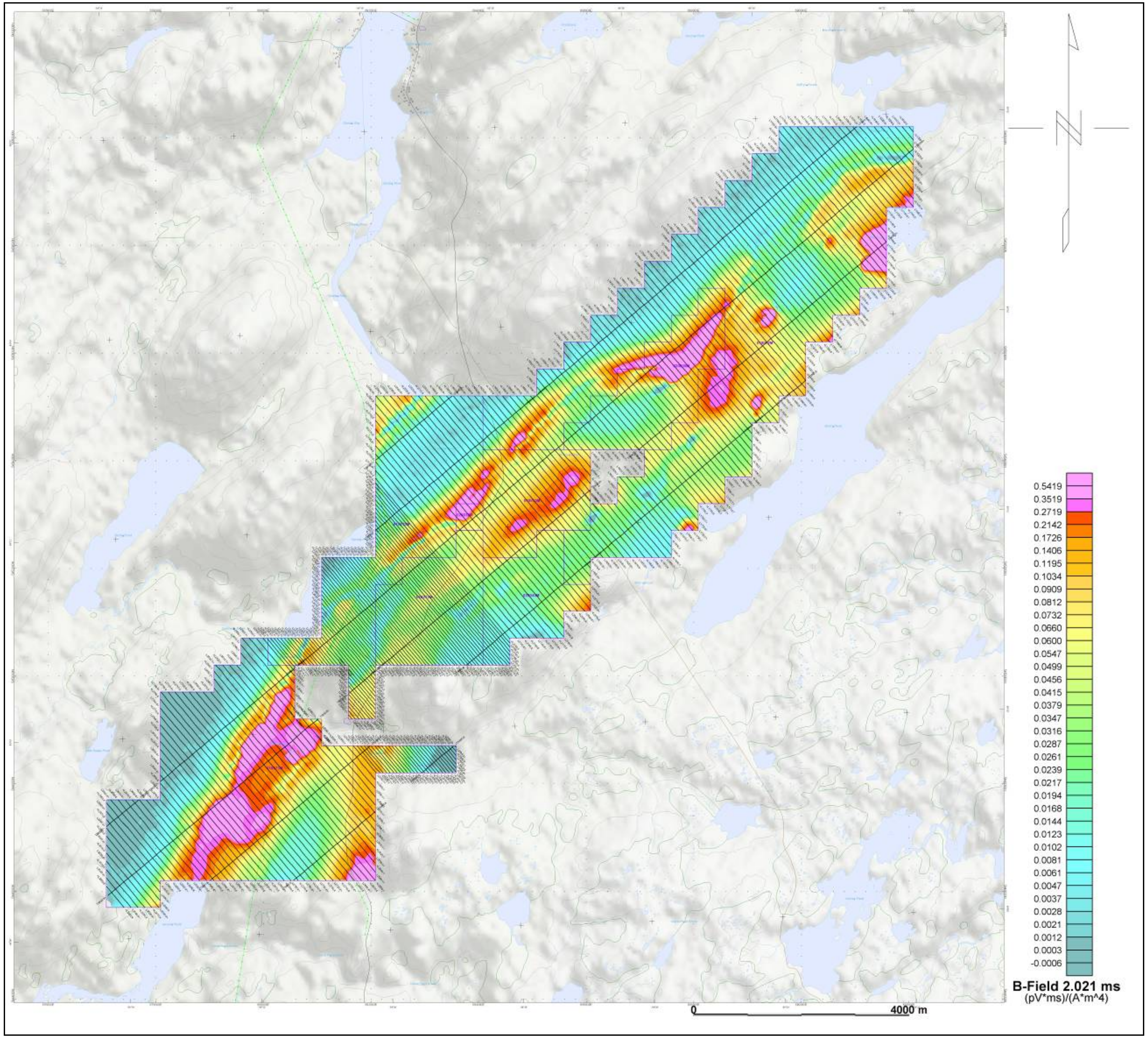
50m & 100m line spacing

900 line-km

Tx-Rx location every ~2m

vertical and in-line components

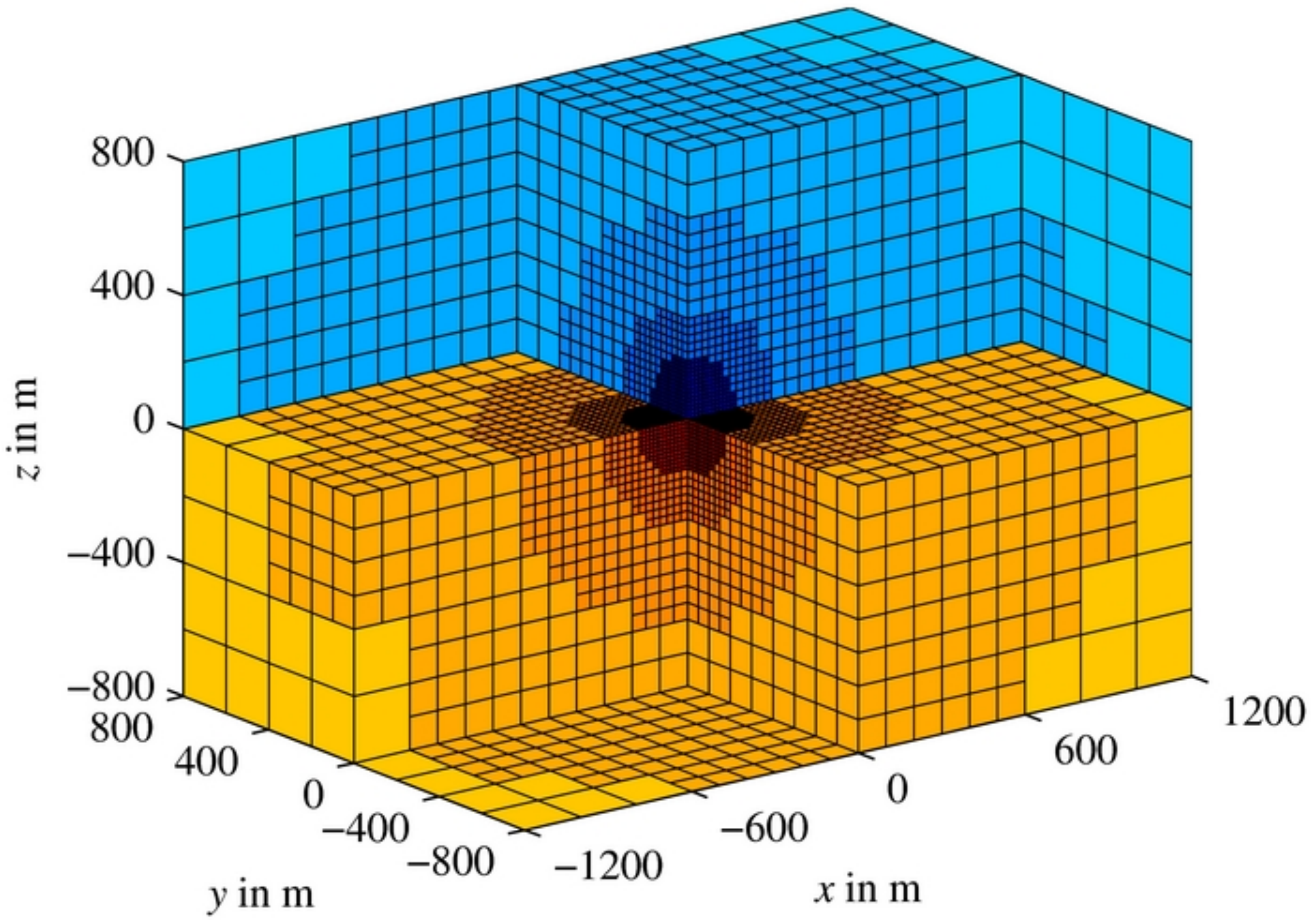
45 time channels



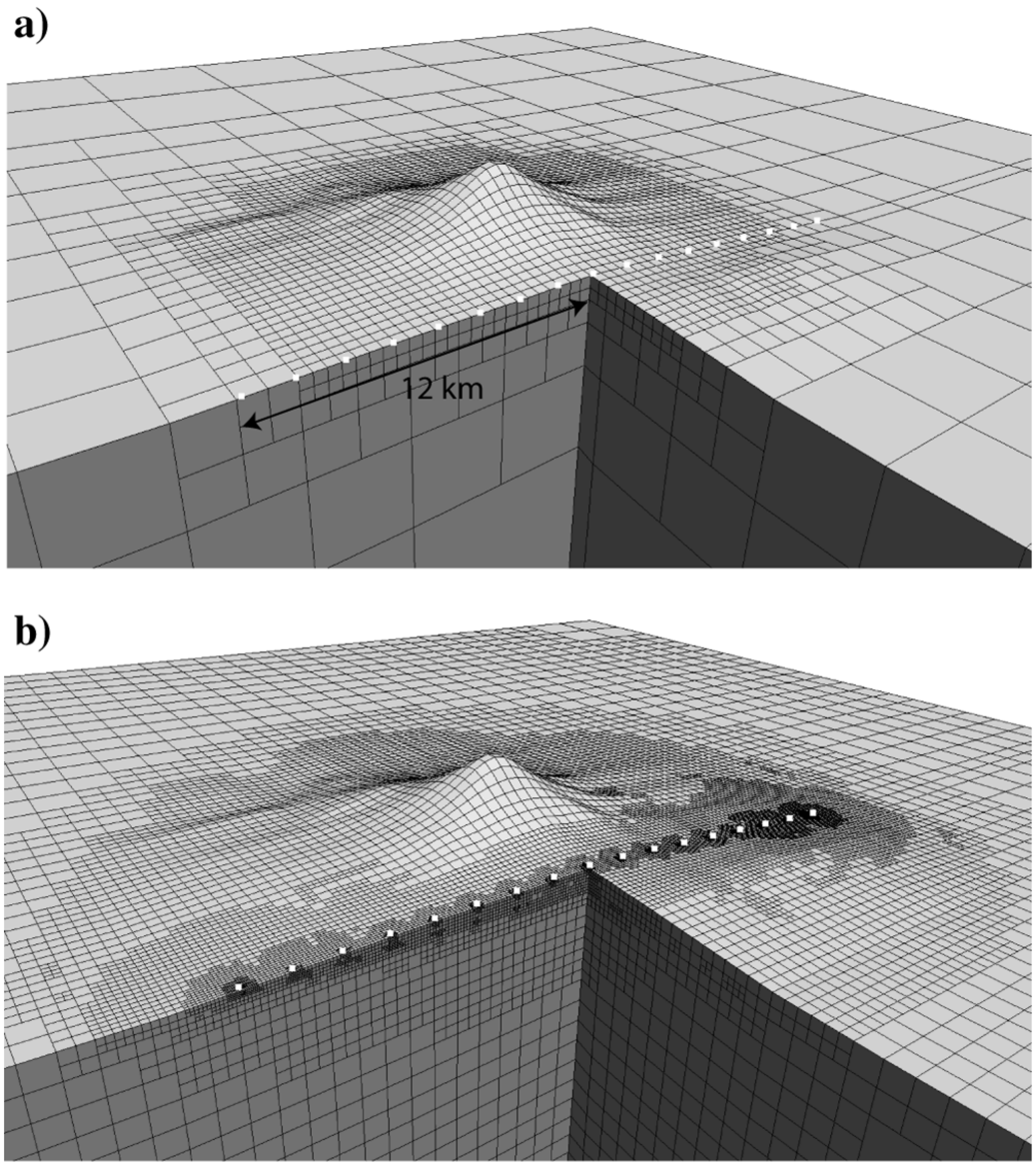
Efficient meshing: unstructured tetrahedral

Ansari et al. (2017)

Efficient meshing: OcTree, non-conforming

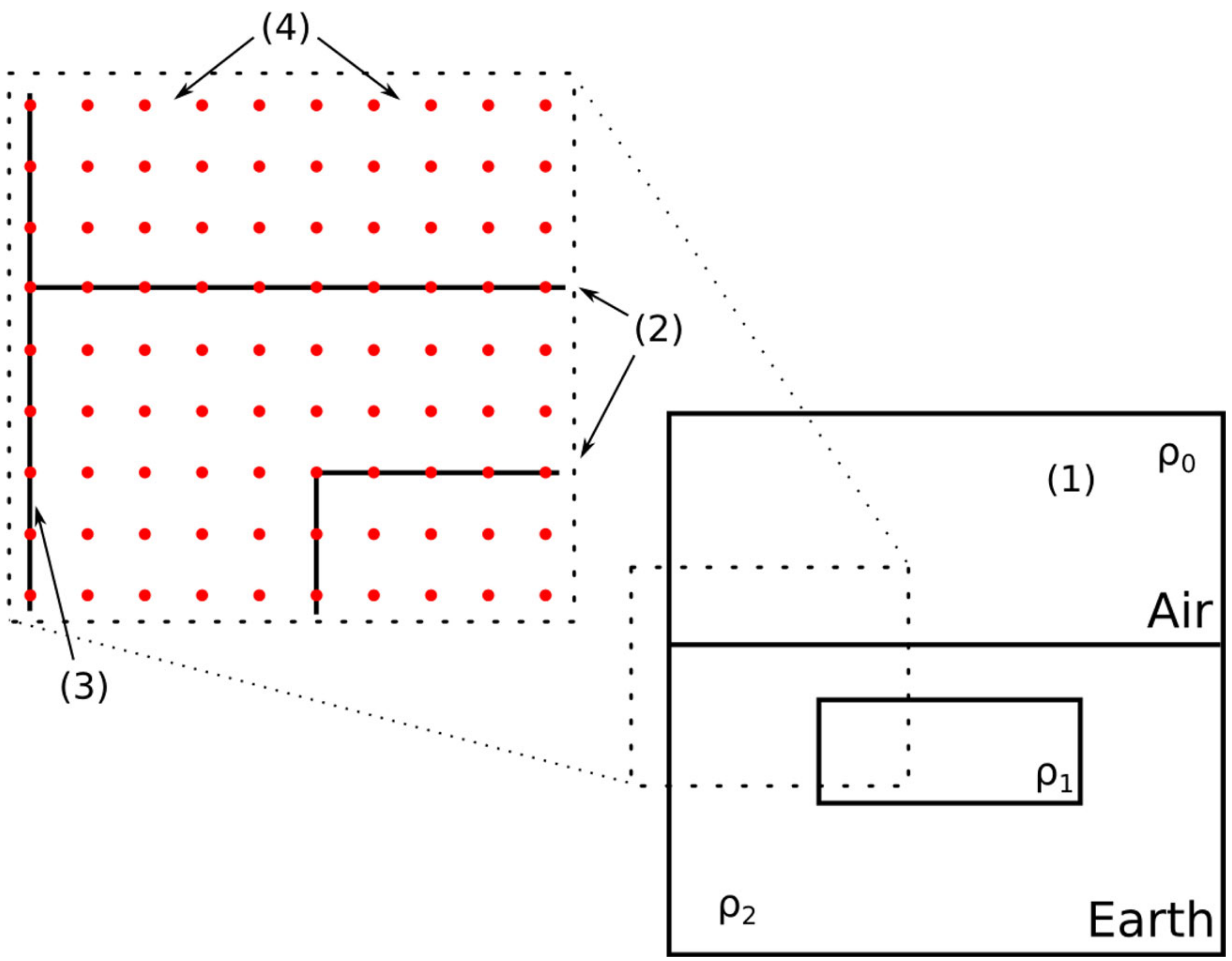


Haber and Schwarzbach (2014)

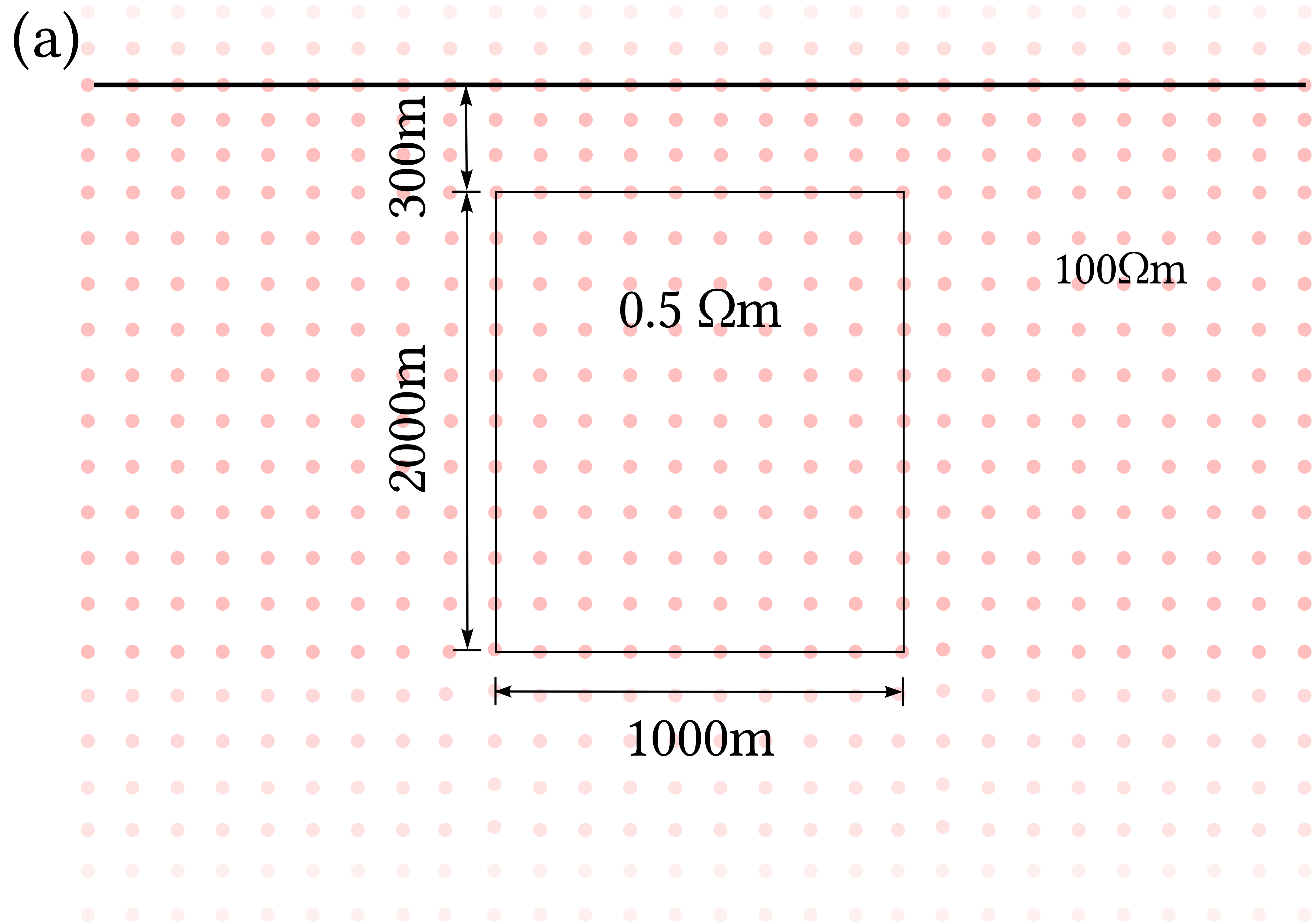


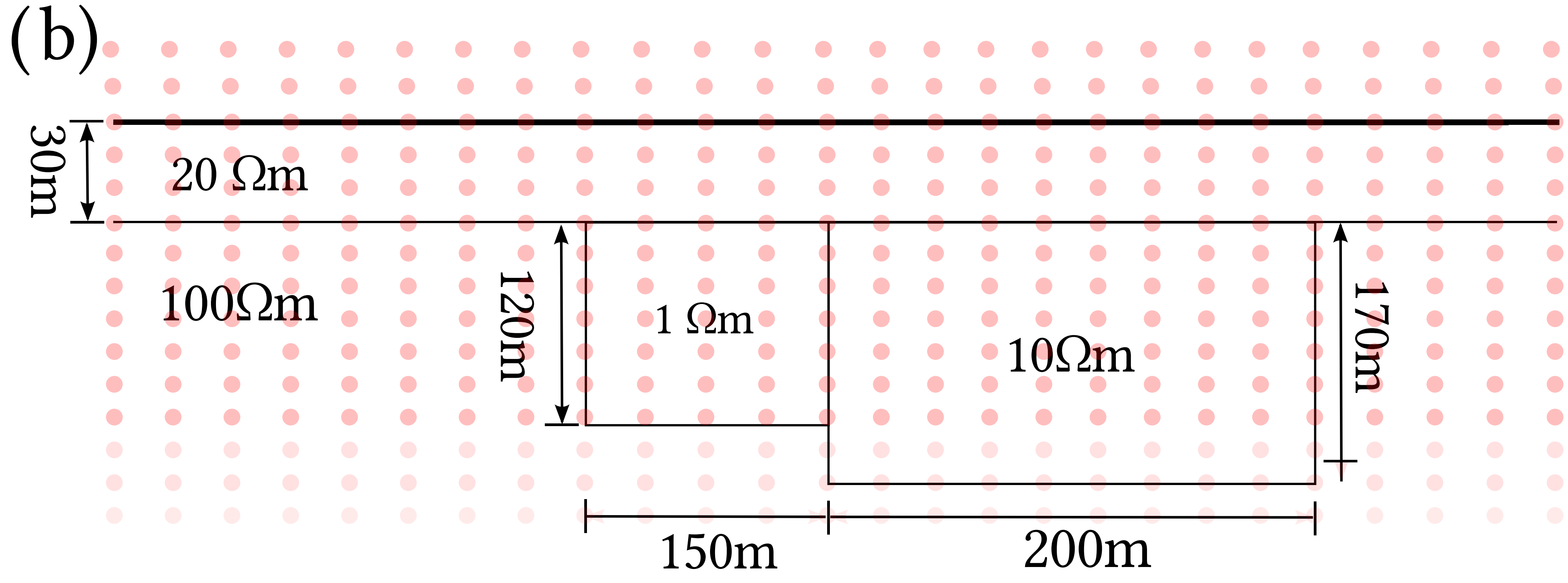
Grayver and Kolev (2015)

Decouple model and computational meshes: "meshfree"

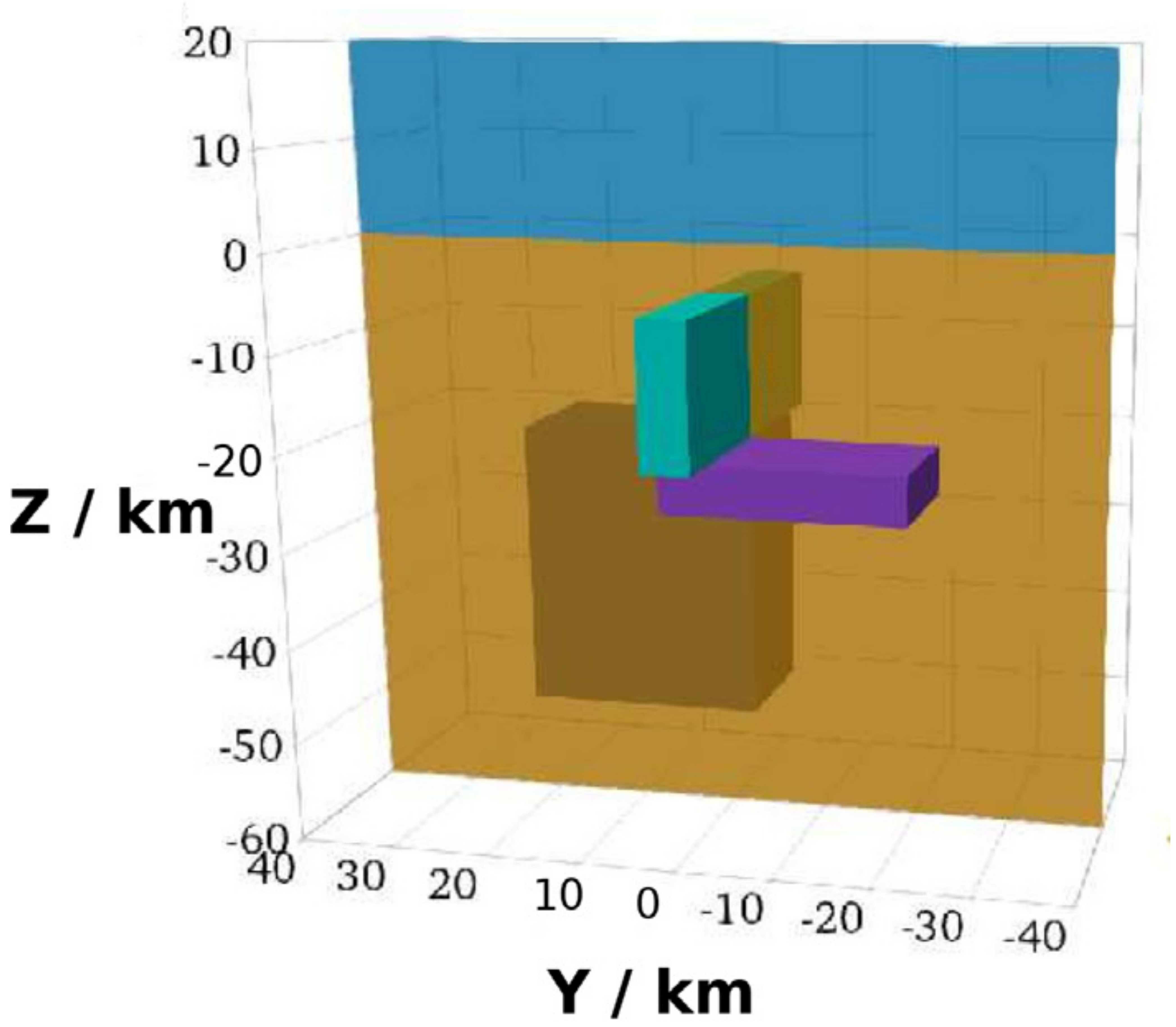


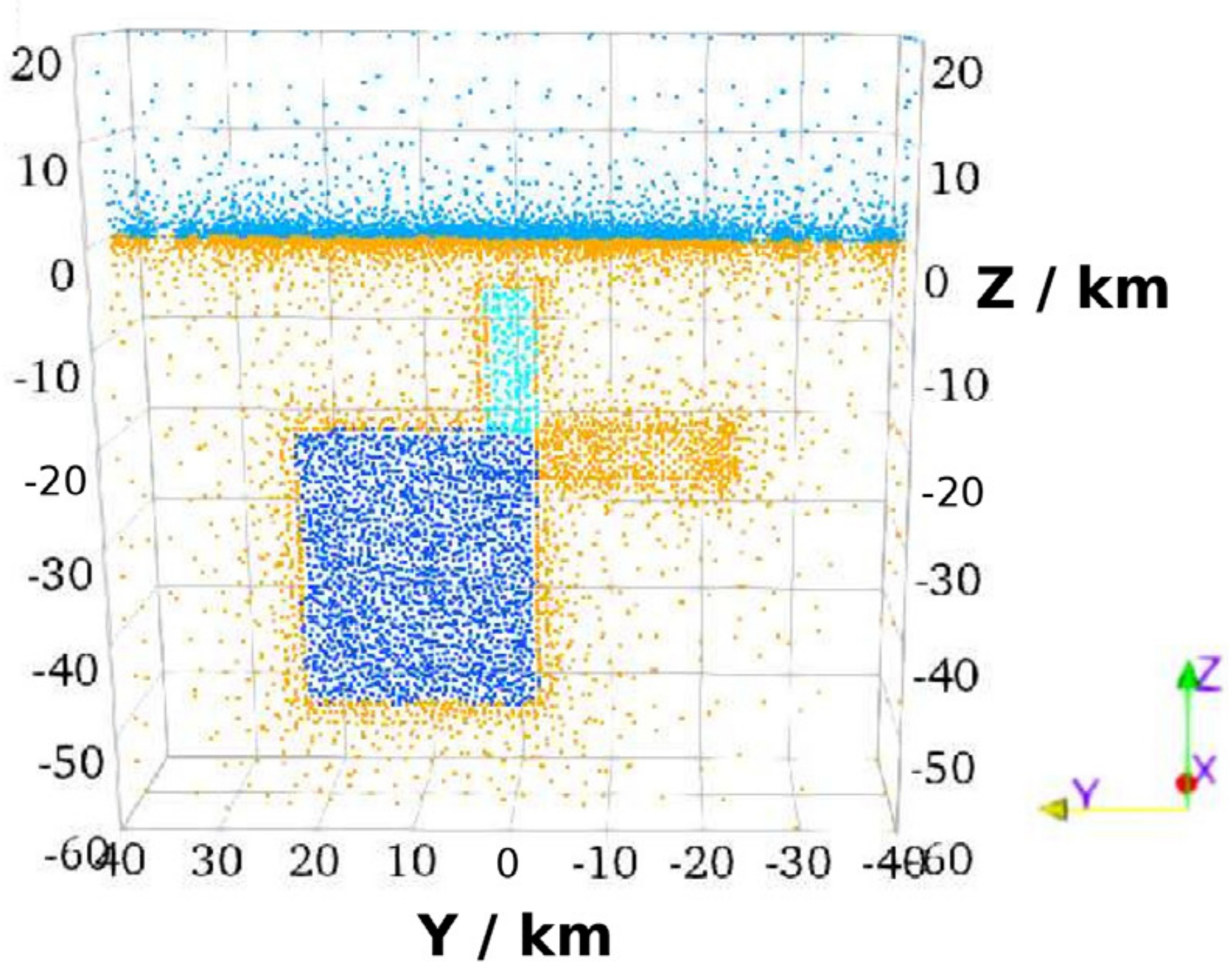
Wittke and Tezkan (2014)





Decouple model and computational meshes: "meshfree"

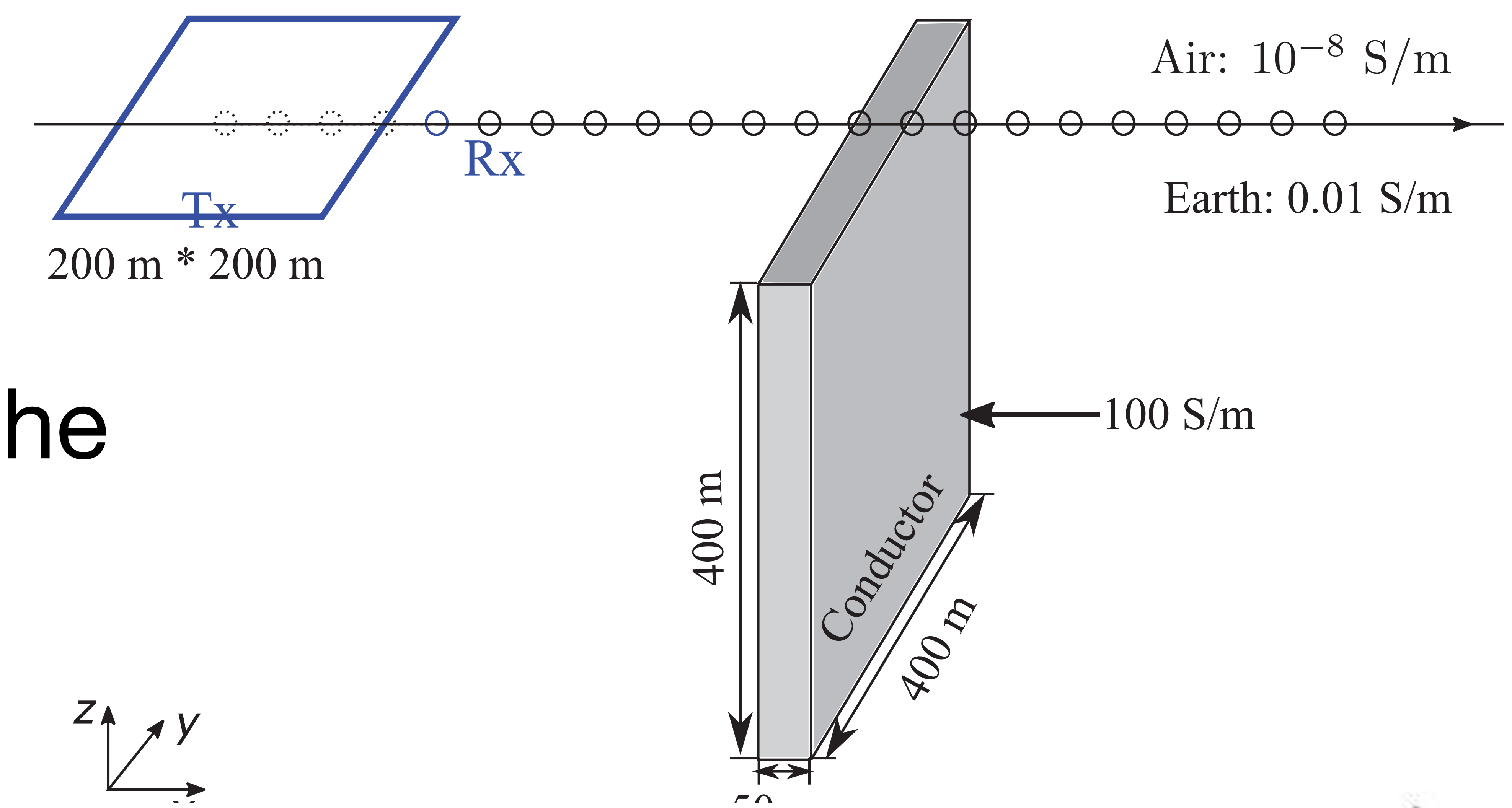


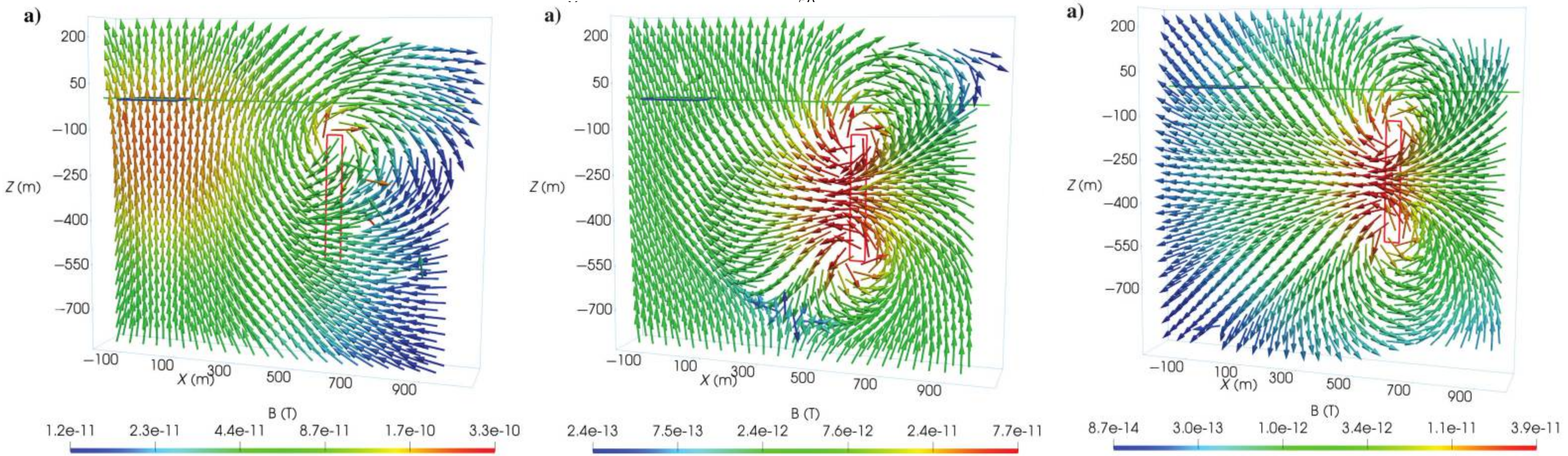


Long and Farquharson (2020)

Aside

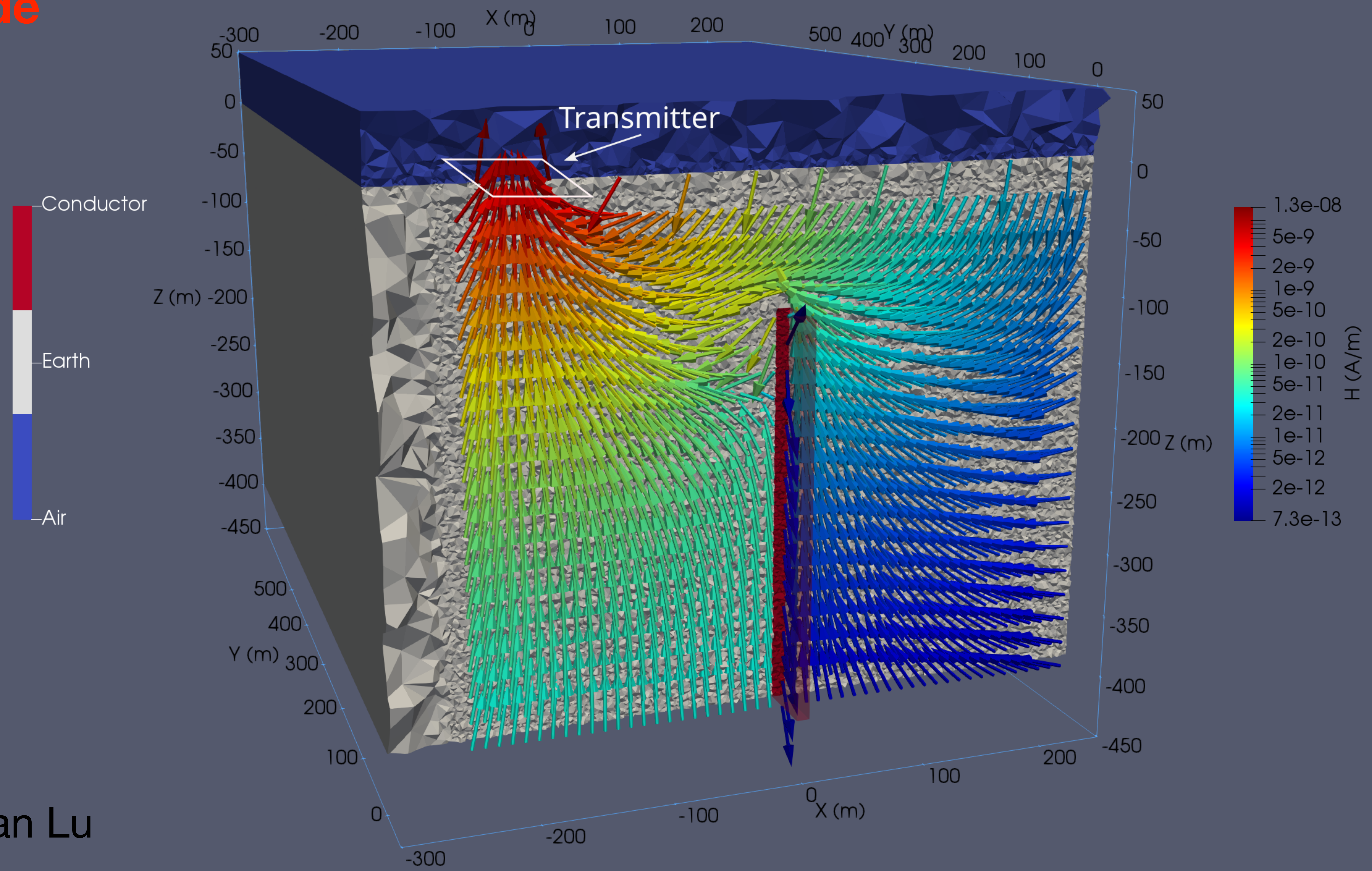
Visualizing the fields in the subsurface.





Lu and Farquharson (2020)

Aside



Xushan Lu

Outline

~~Inversion~~ optimization background:

- why optimization
- data misfit
- minimizing data misfit, non-uniqueness
- measure of model somethingorother
- descent-based, gradient-based optimization (linearization)
- or sampling and selection of collections of models we like

→ Forward modelling

Sensitivities

Descent/gradient/derivative/linearization-based algorithms

The Conclusion

Future work/thoughts

- different models, different approaches?
- D+ ?

Outline

~~Inversion~~ optimization background:

- why optimization
- data misfit
- minimizing data misfit, non-uniqueness
- measure of model somethingorother
- descent-based, gradient-based optimization (linearization)
- or sampling and selection of collections of models we like

Forward modelling

→ Sensitivities

Descent/gradient/derivative/linearization-based algorithms

The Conclusion

Future work/thoughts

- different models, different approaches?
- D+ ?

"Perturbation" approach:

$$\frac{\partial F_j(\mathbf{m})}{\partial m_k} \approx \frac{F_j(\mathbf{m} + \Delta\mathbf{m}_k) - F_j(\mathbf{m})}{\Delta m_k}.$$

Compute Jacobian **one column at a time** using forward-modelling routine.

Requires N forward modellings (where N is the number of model parameters).

McGillivray and Oldenburg (1990)

"Sensitivity equation" approach: McGillivray at al. (1994)

Forward modelling, matrix equation:

$$\mathbf{A}\,\mathbf{u} = \mathbf{s} \qquad \mathbf{A} = \mathbf{A}(\mathbf{m})$$

Differentiate w.r.t. model parameter:

$$\frac{\partial}{\partial m_j}\{\mathbf{A}\,\mathbf{u}\} = \frac{\partial \mathbf{s}}{\partial m_j},$$

$$\frac{\partial \mathbf{A}}{\partial m_j}\,\mathbf{u} + \mathbf{A}\,\frac{\partial \mathbf{u}}{\partial m_j} = 0,$$

$$\mathbf{A}\,\frac{\partial \mathbf{u}}{\partial m_j} = -\frac{\partial \mathbf{A}}{\partial m_j}\,\mathbf{u}. \qquad \frac{\partial \mathbf{u}}{\partial m_j} = -\mathbf{A}^{-1}\,\frac{\partial \mathbf{A}}{\partial m_j}\,\mathbf{u}.$$

Compute Jacobian **one column at a time** using forward-modelling routine.

Requires N forward modellings (where N is the number of model parameters).

(But what if a direct solver has been used for the forward problem, and the factorization available?)
(What if only one column of the Jacobian matrix is required?)

"Adjoint equation" approach

McGillivray at al. (1994)

Forward modelling, PDE:

$$L\,u \;=\; s \quad\rightarrow\quad L\,\frac{\partial u}{\partial m_j} \;=\; -\frac{\partial L}{\partial m_j}\,u \qquad L\,g(\mathbf{r};\mathbf{r}_o) \;=\; \delta(\mathbf{r}-\mathbf{r}_o)$$

Green's function solution:

$$\int_V \left\{ g\,L\,\frac{\partial u}{\partial m_j} \;-\; \frac{\partial u}{\partial m_j}\,L\,g \right\} dv \;=\; \int_V \left\{ -g\,\frac{\partial L}{\partial m_j}\,u \;-\; \frac{\partial u}{\partial m_j}\,\delta(\mathbf{r}-\mathbf{r}_o) \right\} dv,$$

$$0 \;=\; \int_V \left\{ -g\,\frac{\partial L}{\partial m_j}\,u \right\} dv \;-\; \left.\frac{\partial u}{\partial m_j}\right|_{\mathbf{r}_o},$$

$$\left.\frac{\partial u}{\partial m_j}\right|_{\mathbf{r}_o} \;=\; -\int_{V_j} \left\{ g(\mathbf{r};\mathbf{r}_o)\,u \right\} dv,$$

$$\left.\frac{\partial u}{\partial \mathbf{m}}\right|_{\mathbf{r}_o} \;=\; -\int_{V} \left\{ g(\mathbf{r};\mathbf{r}_o)\,u \right\} dv.$$

$$\frac{\partial H_z(\mathbf{x}_0)}{\partial \sigma_k} = \int_D \tilde{\mathbf{E}} \cdot \mathbf{E}\,\psi_k(\mathbf{x})\,dv.$$

Compute Jacobian **one row at a time** using forward-modelling routine.

Requires M forward modellings (where M is the number of data).

"Implicit", "pseudo-forward modelling" approach: Mackie and Madden (1993)

Matrix equation to be solved for model update (for example):

$$\{\mathbf{J}^{n-1\,T}\mathbf{W}_d^T\mathbf{W}_d\;\mathbf{J}^{n-1} + \alpha_r\beta^n\;\mathbf{W}_r^T\mathbf{W}_r + \alpha_s\beta^n\;\mathbf{W}_s^T\mathbf{W}_s\}\;\delta\,\mathbf{m}^n$$
$$= \mathbf{J}^{n-1\,T}\mathbf{W}_d^T\mathbf{W}_d(\mathbf{d}^t - \mathbf{d}^{n-1}) - \alpha_r\beta^n\;\mathbf{W}_r^T\mathbf{W}_r\;\mathbf{m}^{n-1} + \alpha_s\beta^n\;\mathbf{W}_s^T\mathbf{W}_s(\mathbf{m}^f - \mathbf{m}^{n-1}),$$

Iterative solution requires results of: $\mathbf{y} \;=\; \mathbf{J}\,\mathbf{x}$

From "sensitivity equation" approach: $\mathbf{J}_j \;=\; \mathbf{A}^{-1}\,\tilde{\mathbf{u}}$

But do this using forward-solving routine: $\mathbf{A}\,\mathbf{J}_j \;=\; \tilde{\mathbf{u}}$

Compute product of Jacobian with vector using forward-modelling routine.

Don't have to construct and store Jacobian matrix.

(Trade memory for computations.)

Outline

~~Inversion~~ optimization background:

- why optimization
- data misfit
- minimizing data misfit, non-uniqueness
- measure of model somethingorother
- descent-based, gradient-based optimization (linearization)
- or sampling and selection of collections of models we like

Forward modelling

→ Sensitivities

Descent/gradient/derivative/linearization-based algorithms

The Conclusion

Future work/thoughts

- different models, different approaches?
- D+ ?

Outline

~~Inversion~~ optimization background:

- why optimization
- data misfit
- minimizing data misfit, non-uniqueness
- measure of model somethingorother
- descent-based, gradient-based optimization (linearization)
- or sampling and selection of collections of models we like

Forward modelling

Sensitivities

→ Descent/gradient/derivative/linearization-based algorithms

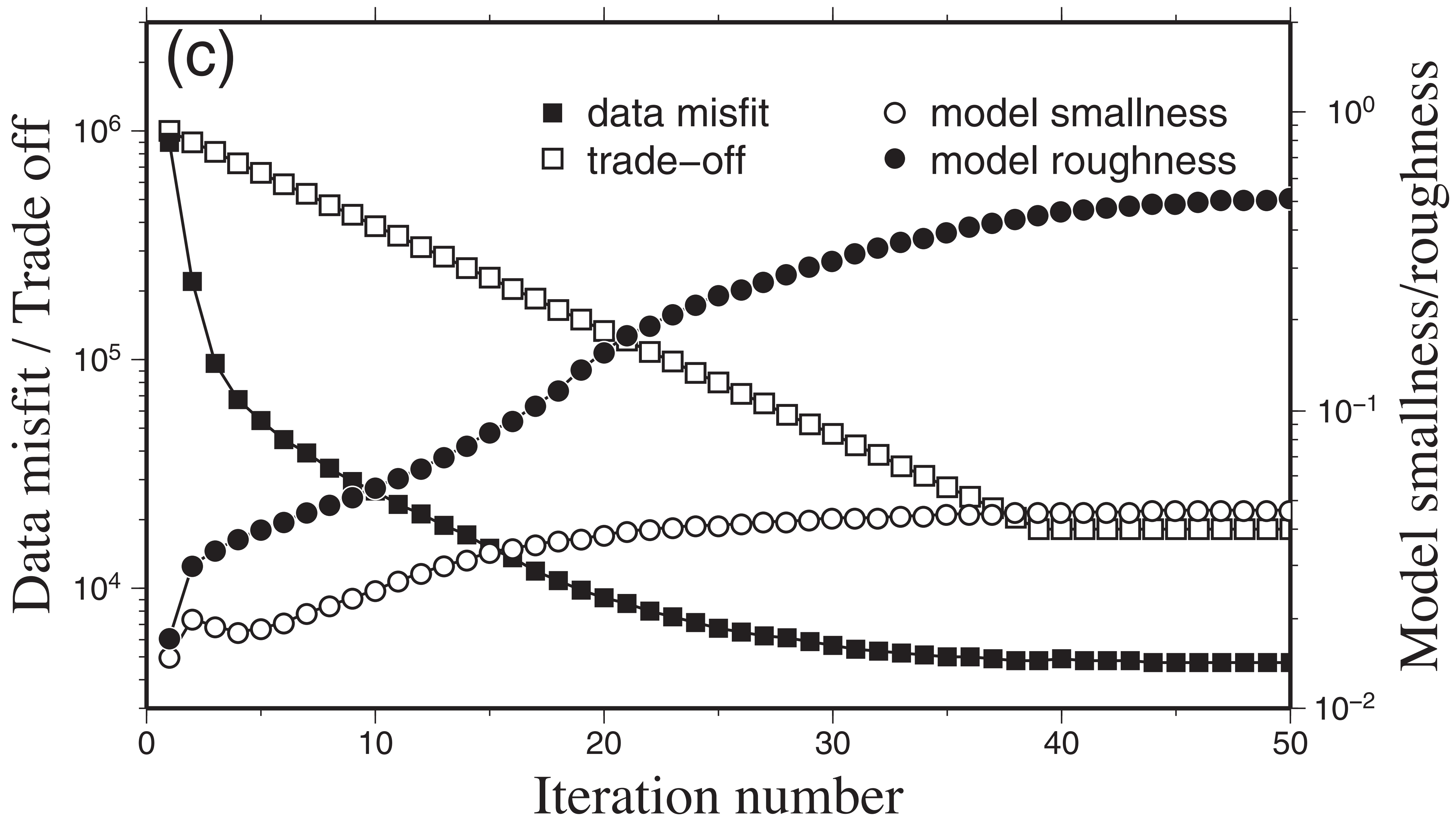
The Conclusion

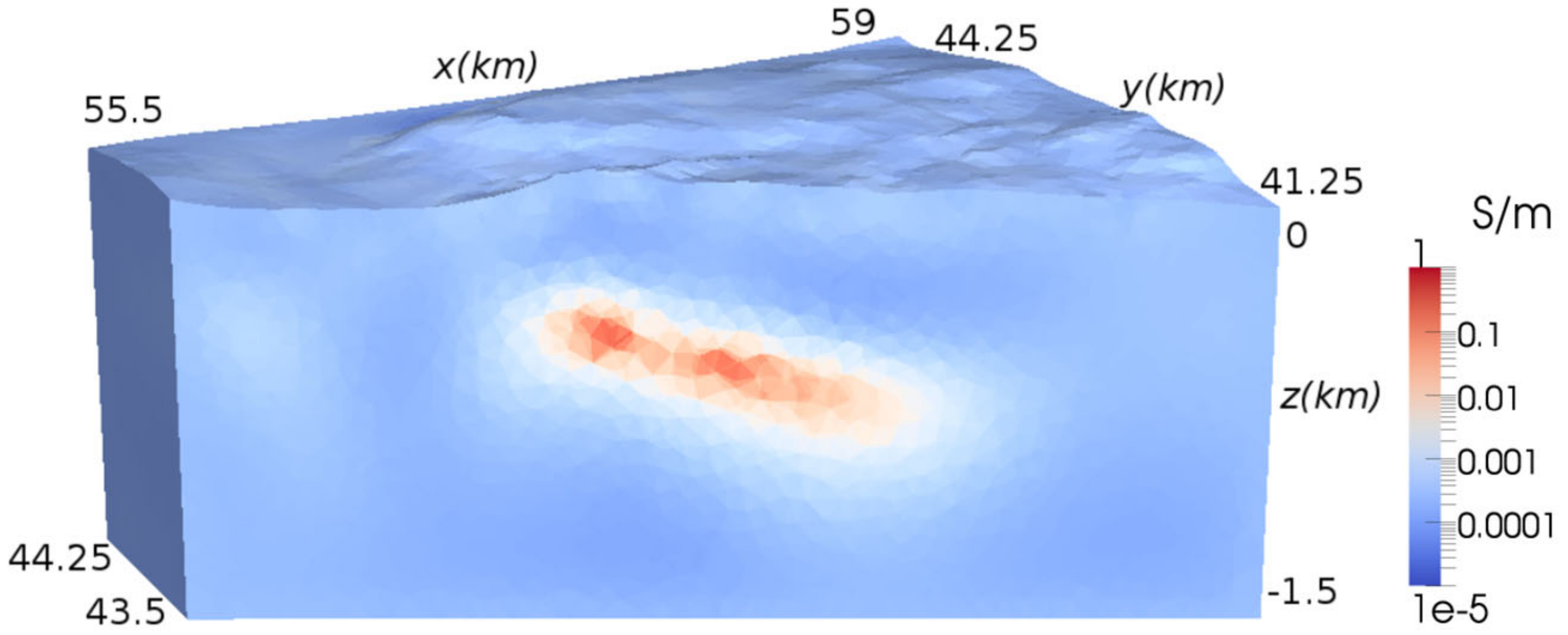
Future work/thoughts

- different models, different approaches?
- D+ ?

Descent-based, gradient-based optimization, e.g., Gauss-Newton:

$$\{\mathbf{J}^{n-1\,T}\mathbf{W}_d^T\mathbf{W}_d\ \mathbf{J}^{n-1} + \alpha_r\beta^n\ \mathbf{W}_r^T\mathbf{W}_r + \alpha_s\beta^n\ \mathbf{W}_s^T\mathbf{W}_s\}\ \delta\ \mathbf{m}^n$$
$$= \mathbf{J}^{n-1\,T}\mathbf{W}_d^T\mathbf{W}_d(\mathbf{d}^t - \mathbf{d}^{n-1}) - \alpha_r\beta^n\ \mathbf{W}_r^T\mathbf{W}_r\ \mathbf{m}^{n-1} + \alpha_s\beta^n\ \mathbf{W}_s^T\mathbf{W}_s(\mathbf{m}^f - \mathbf{m}^{n-1}),$$





Jahandari and Farquharson (2017)

Descent-based, gradient-based optimization:

NLCG, Rodi and Mackie (2001):

- A variant of conjugate gradients applied directly to the function being minimized.
- Avoids the Hessian matrix, thus needs fewer forward modelling, and so faster than GN.
- Any less "powerful" than GN?

Data space, Siripunvaraporn and Egbert (2000):

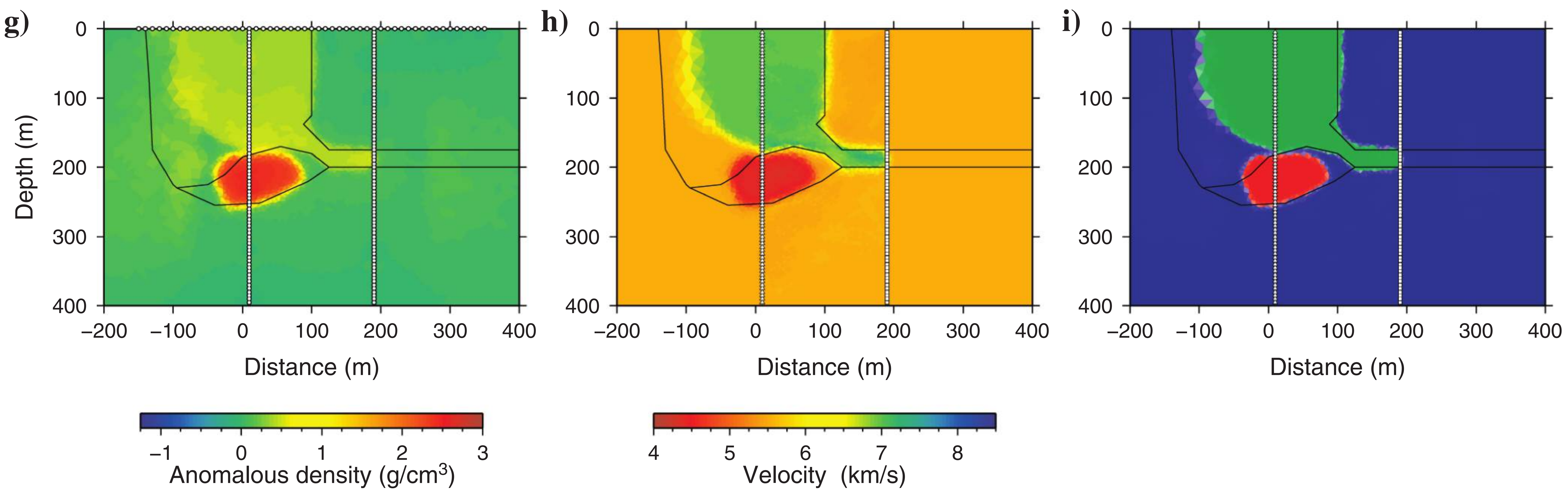
- Transform GN matrix equation from $N \times N$ to $M \times M$ (N is number of model parameters).
- Smaller matrix to invert/solve.
- How expensive are the matrix operations for the transformation?

Descent-based, gradient-based optimization, complex objective function:

$$\Phi(\mathbf{m}_1, \mathbf{m}_2) = \lambda_1 \Phi_{d1}(\mathbf{m}_1) + \lambda_2 \Phi_{d2}(\mathbf{m}_2) + \Phi_{m1}(\mathbf{m}_1) + \Phi_{m2}(\mathbf{m}_2) + \rho \Psi(\mathbf{m}_1, \mathbf{m}_2).$$

Design an objective/cost/penalty function that gives us what we want.

Then go ahead and minimize (!).



Carter-McAuslan et al. (2017)

Outline

~~Inversion~~ optimization background:

- why optimization
- data misfit
- minimizing data misfit, non-uniqueness
- measure of model somethingorother
- descent-based, gradient-based optimization (linearization)
- or sampling and selection of collections of models we like

Forward modelling

Sensitivities

→ Descent/gradient/derivative/linearization-based algorithms

The Conclusion

Future work/thoughts

- different models, different approaches?
- D+ ?

Outline

~~Inversion~~ optimization background:

- why optimization
- data misfit
- minimizing data misfit, non-uniqueness
- measure of model somethingorother
- descent-based, gradient-based optimization (linearization)
- or sampling and selection of collections of models we like

Forward modelling

Sensitivities

Descent/gradient/derivative/linearization-based algorithms

→ The Conclusion

Future work/thoughts

- different models, different approaches?
- D+ ?

The Conclusion

"Occam", minimum-structure inversion is very effective — it gets the job done — and hence useful and popular.

Outline

- ~~Inversion~~ optimization background:
  - why optimization
  - data misfit
  - minimizing data misfit, non-uniqueness
  - measure of model somethingorother
  - descent-based, gradient-based optimization (linearization)
  - or sampling and selection of collections of models we like
- Forward modelling
- Sensitivities
- Descent/gradient/derivative/linearization-based algorithms
- → The Conclusion
- Future work/thoughts
  - different models, different approaches?
  - D+ ?

Outline

~~Inversion~~ optimization background:

- why optimization
- data misfit
- minimizing data misfit, non-uniqueness
- measure of model somethingorother
- descent-based, gradient-based optimization (linearization)
- or sampling and selection of collections of models we like

Forward modelling

Sensitivities

Descent/gradient/derivative/linearization-based algorithms

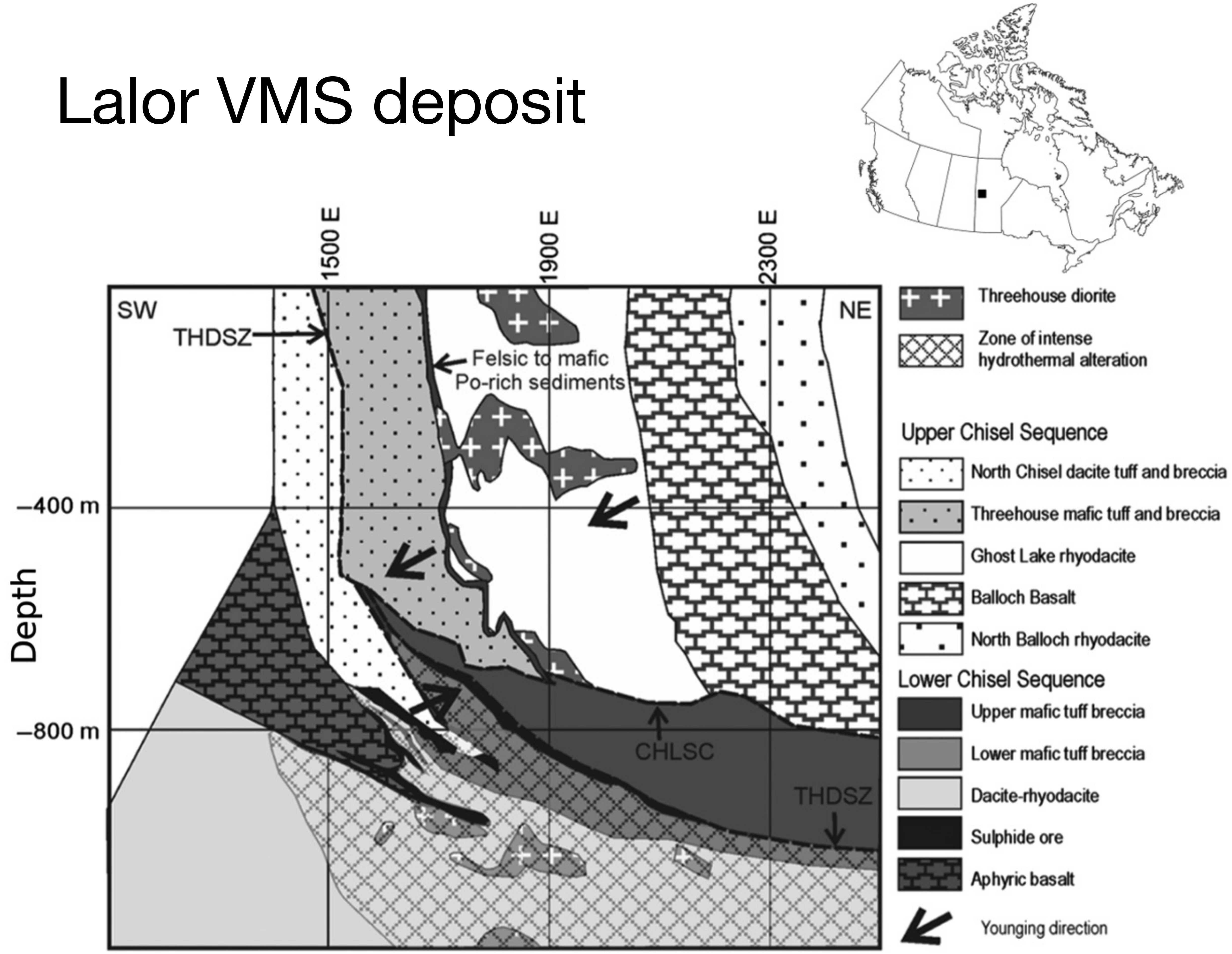
The Conclusion

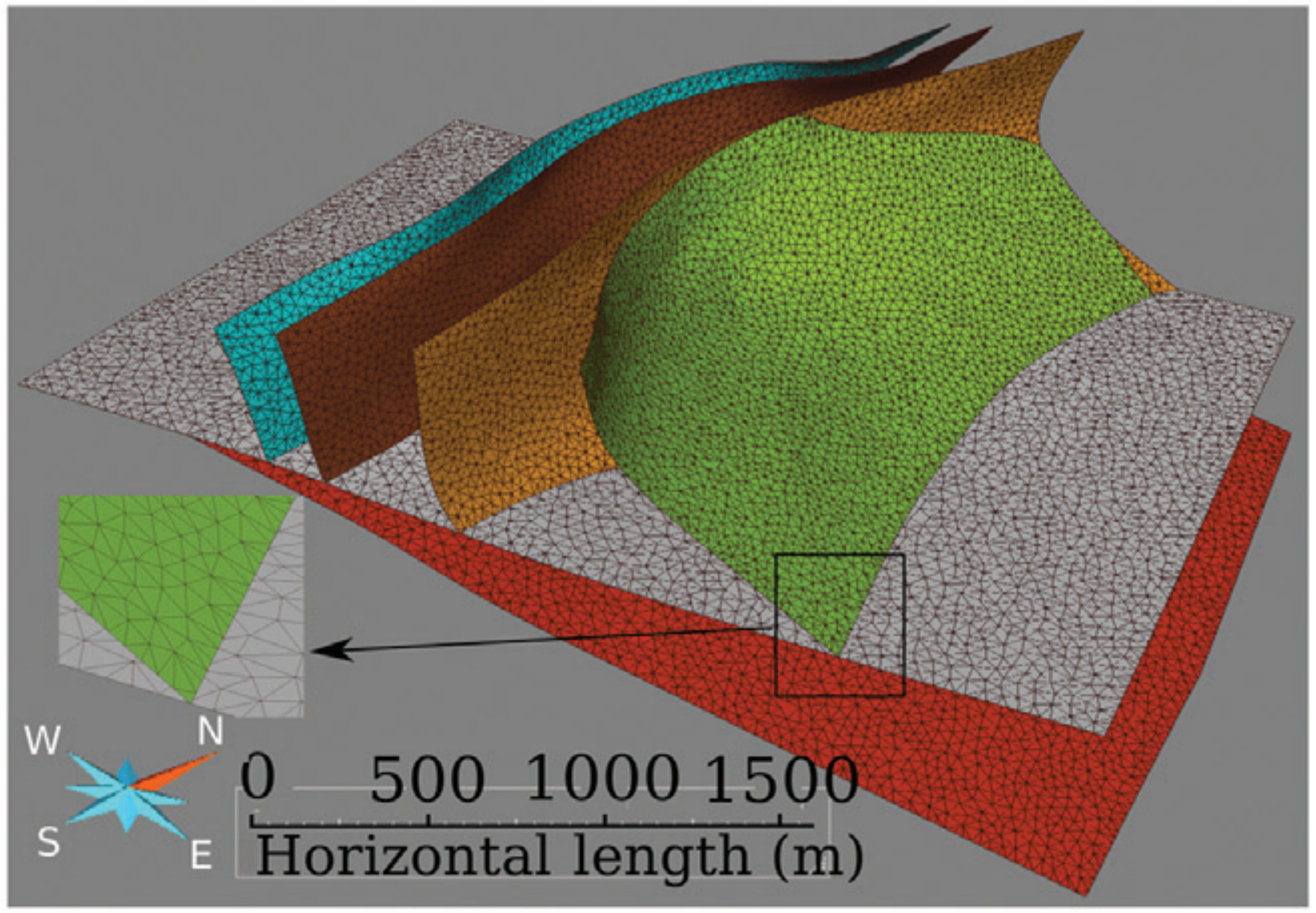
→ Future work/thoughts

- different models, different approaches?
- D+ ?

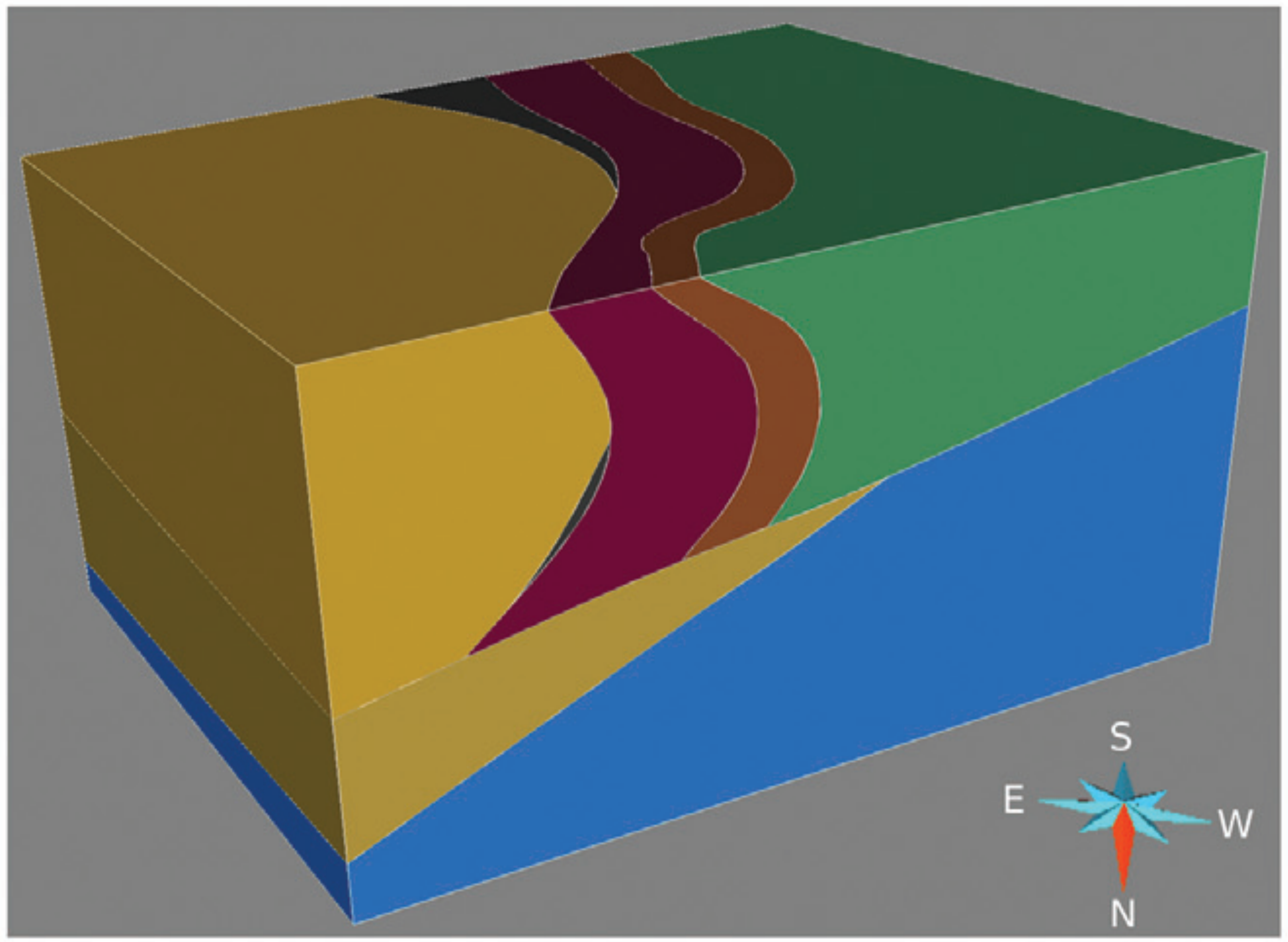
3-D computer *geology* models

Lalor VMS deposit





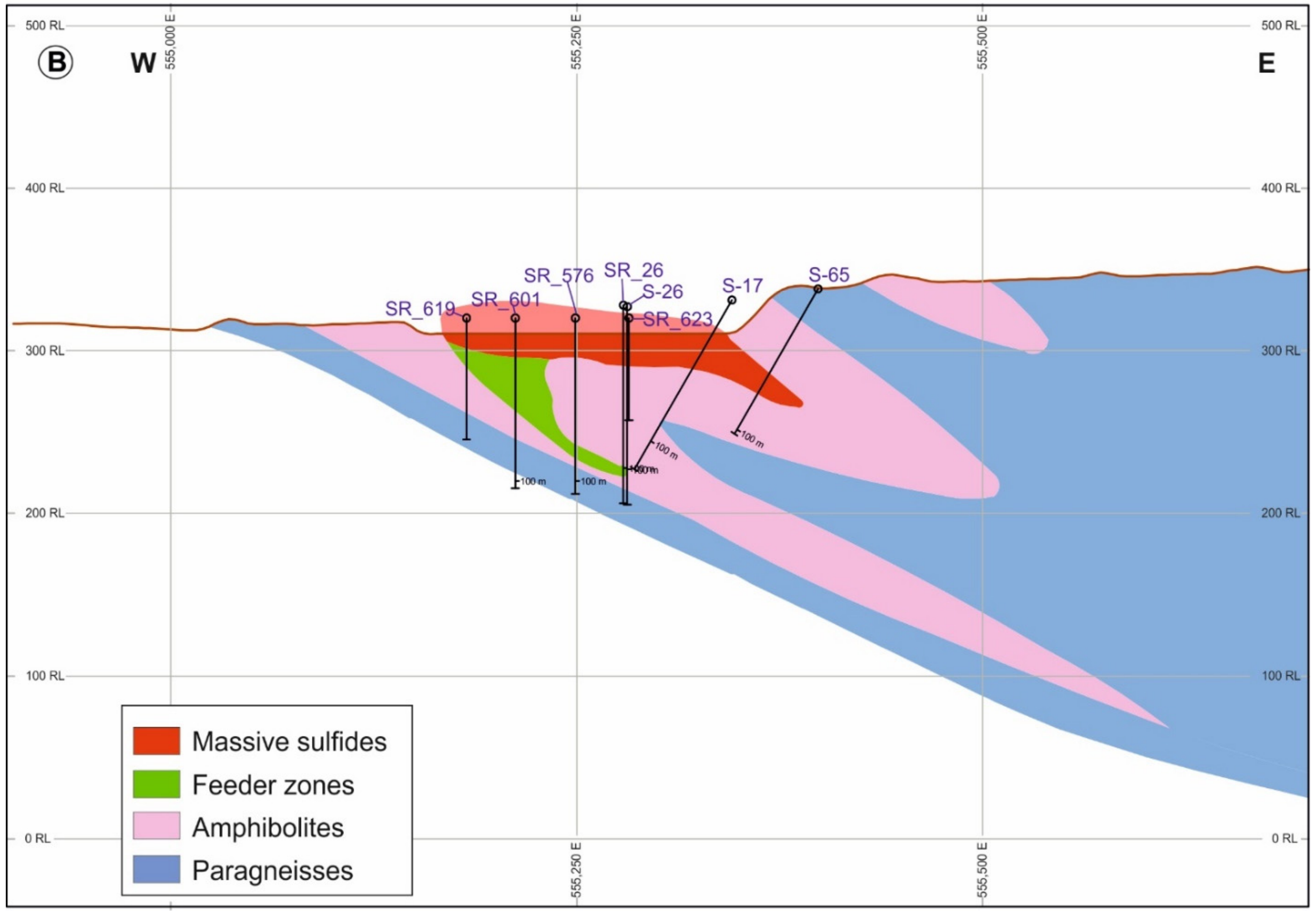
c)

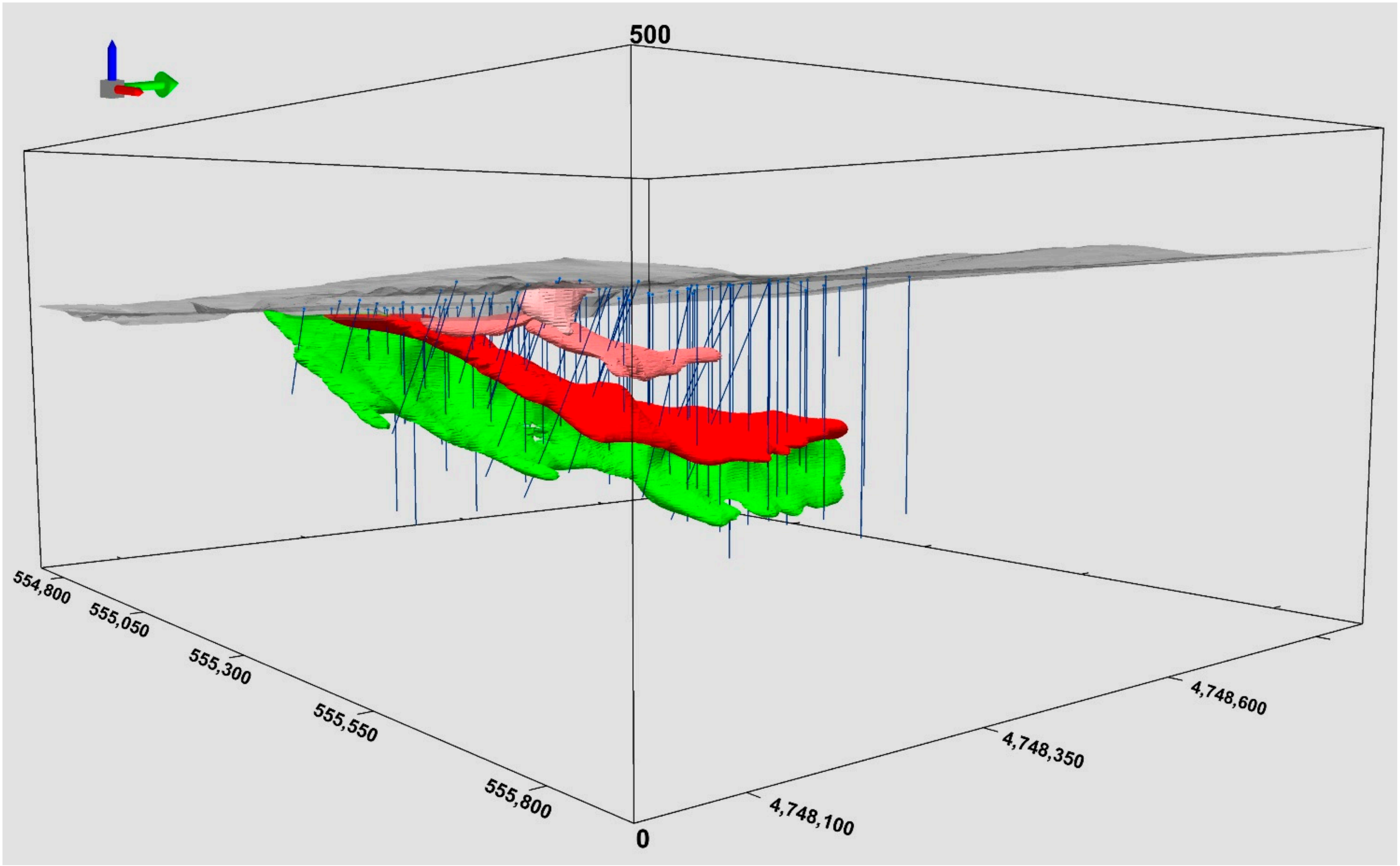


Ansari et al. (2020)

3-D computer *geology* models

Touro VMS deposit, NW Spain

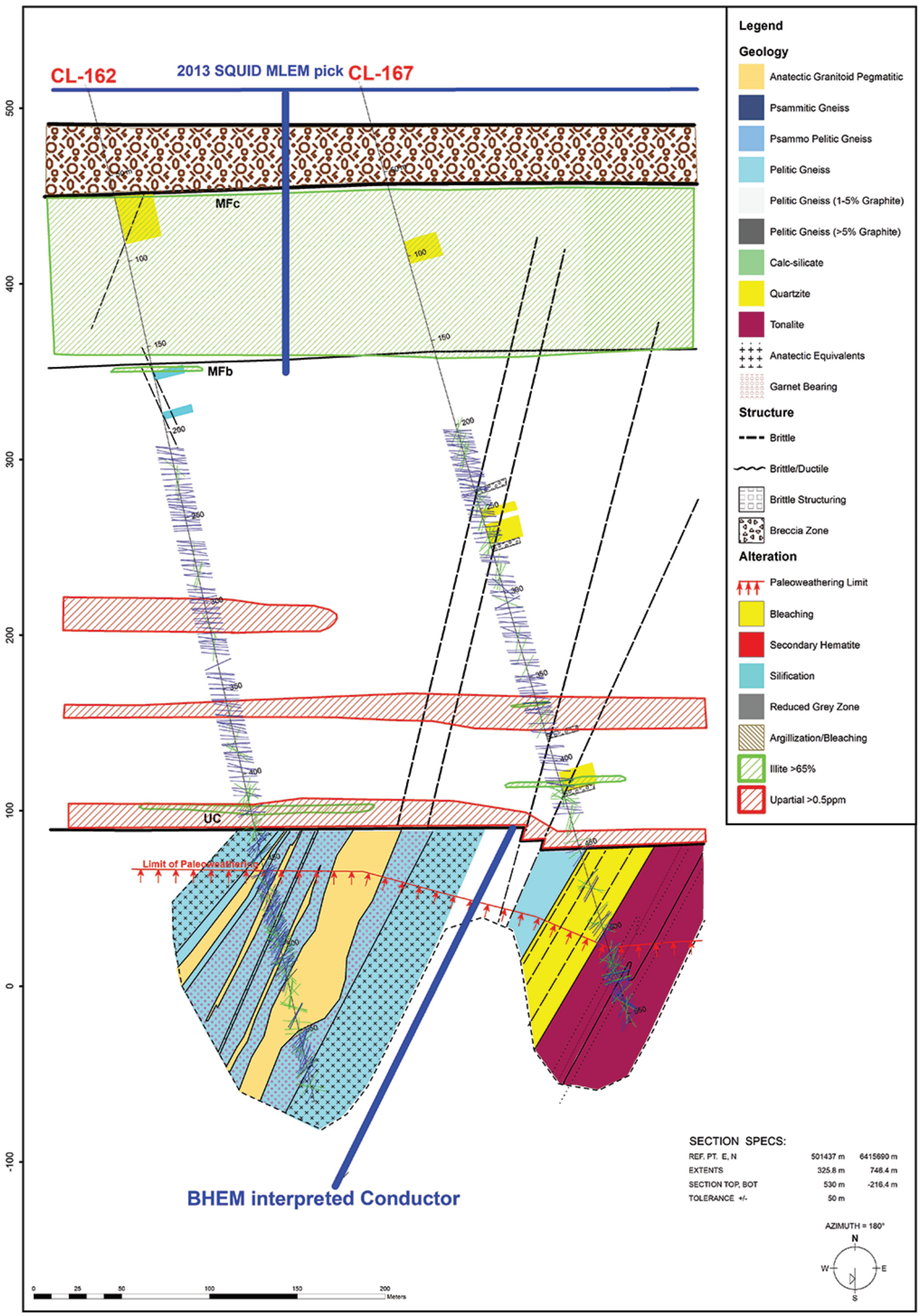




Arias et al. (2021)

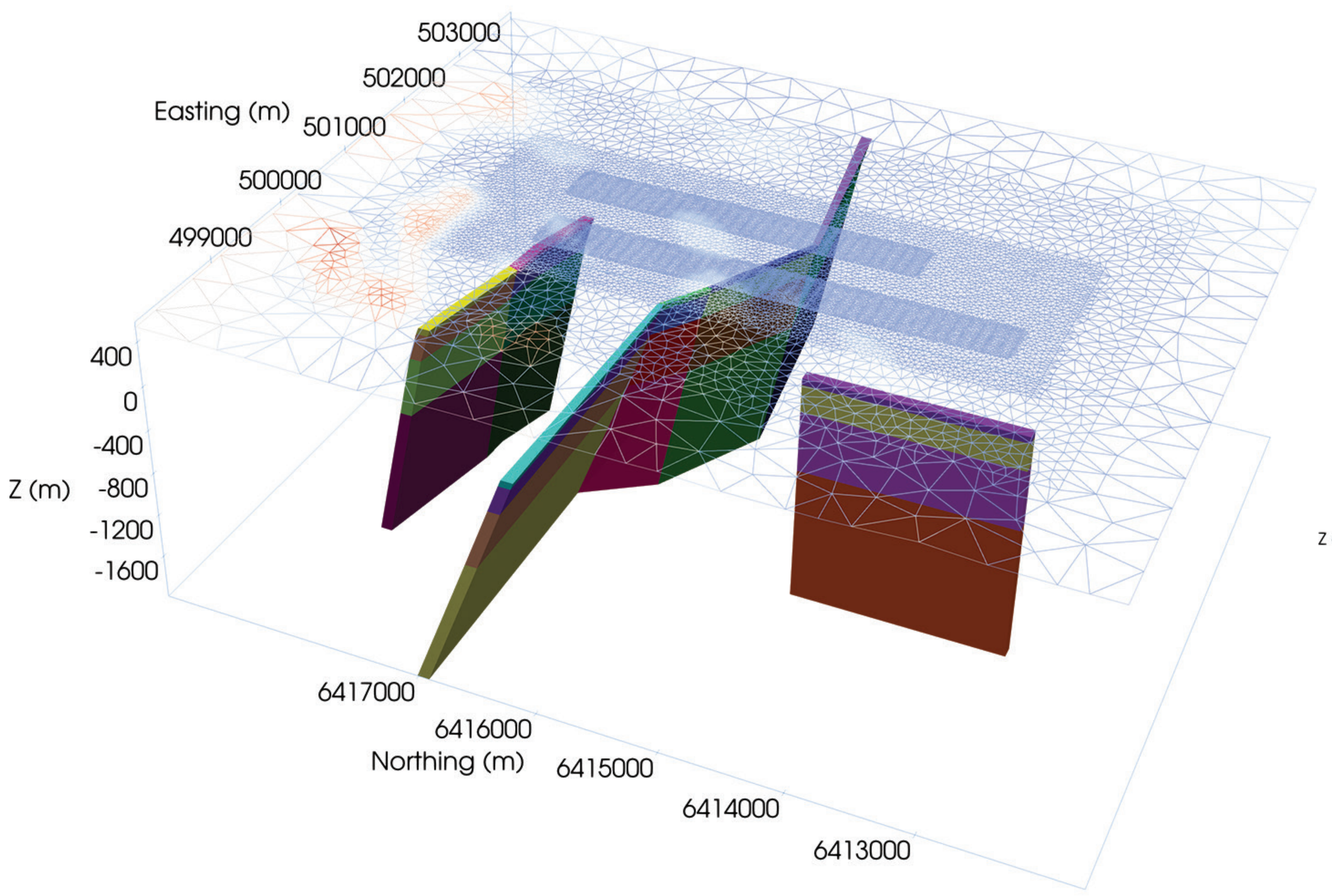
3-D computer *geology* models

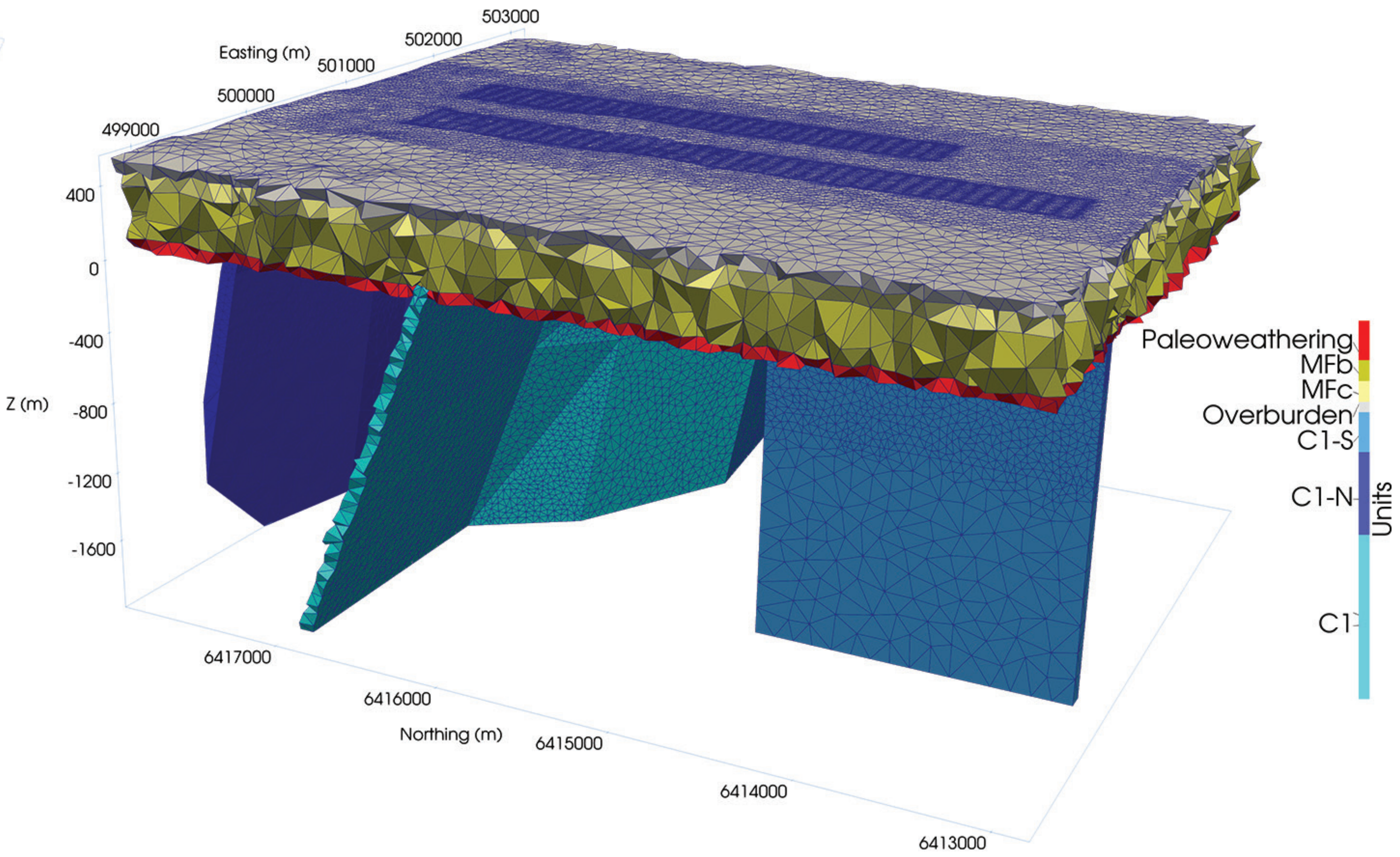
Uranium exploration,
Athabasca Basin
(geological section).



Lu et al. (2021)

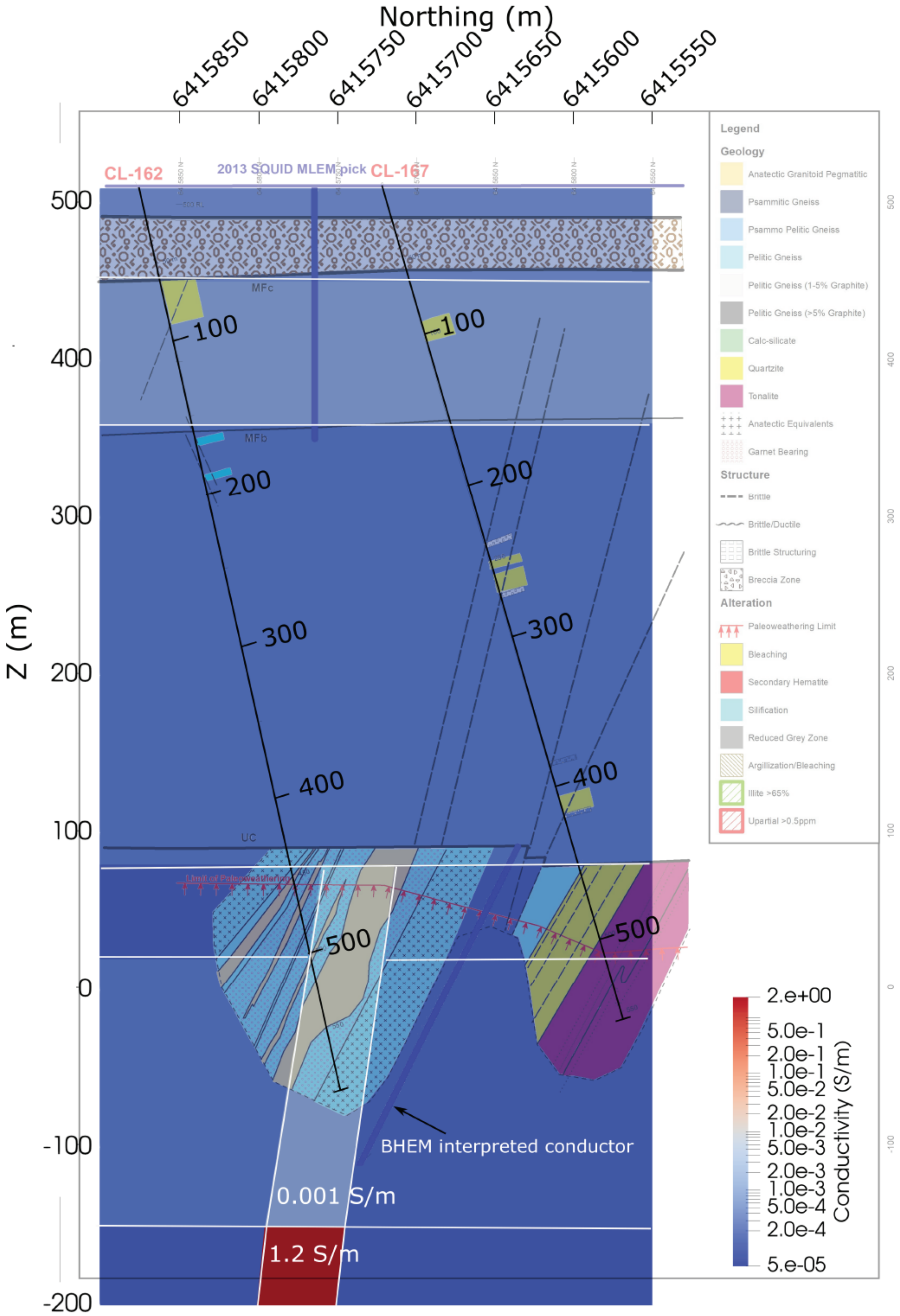
Inversion using same kind of computer Earth model as the geologists ...

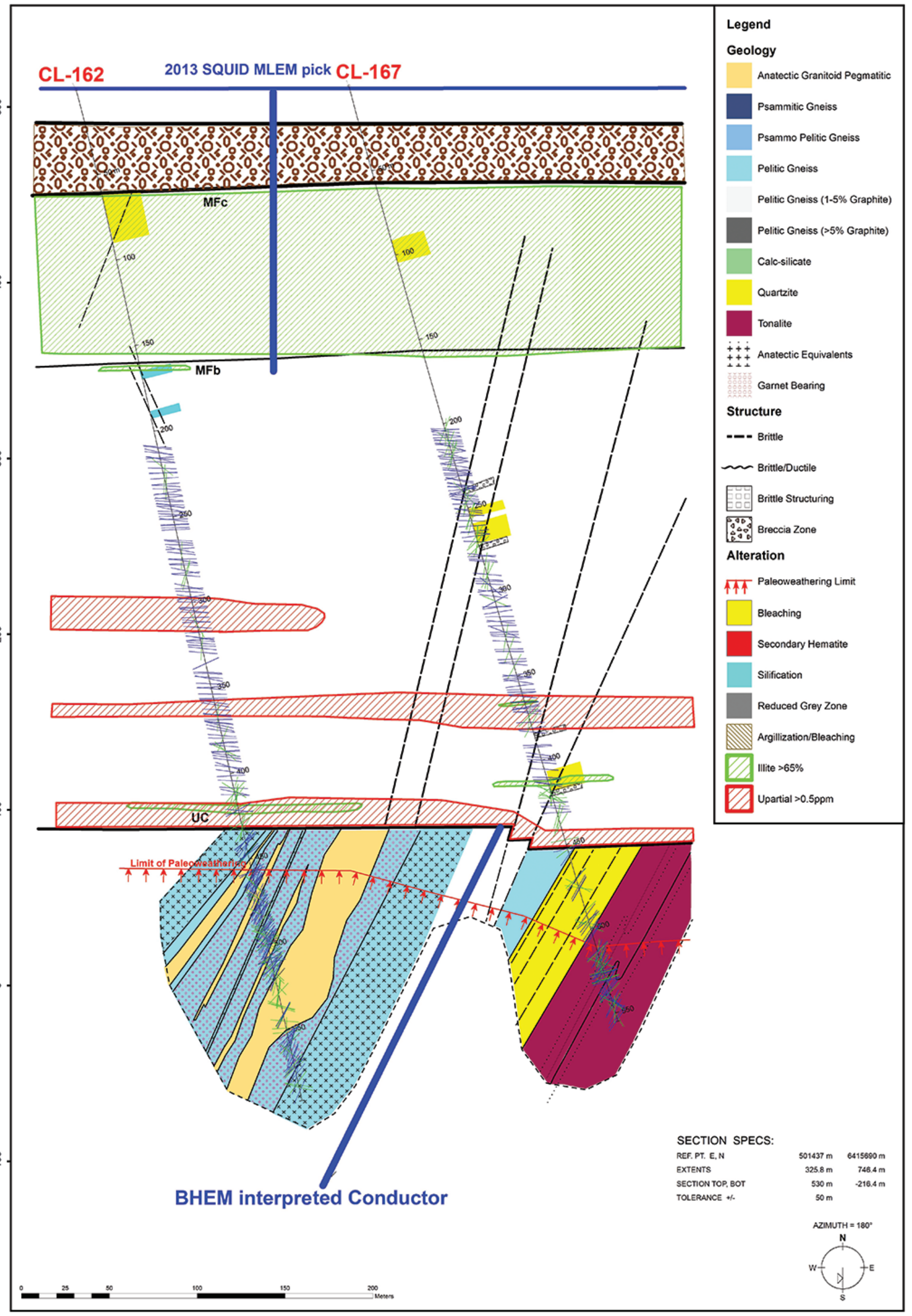




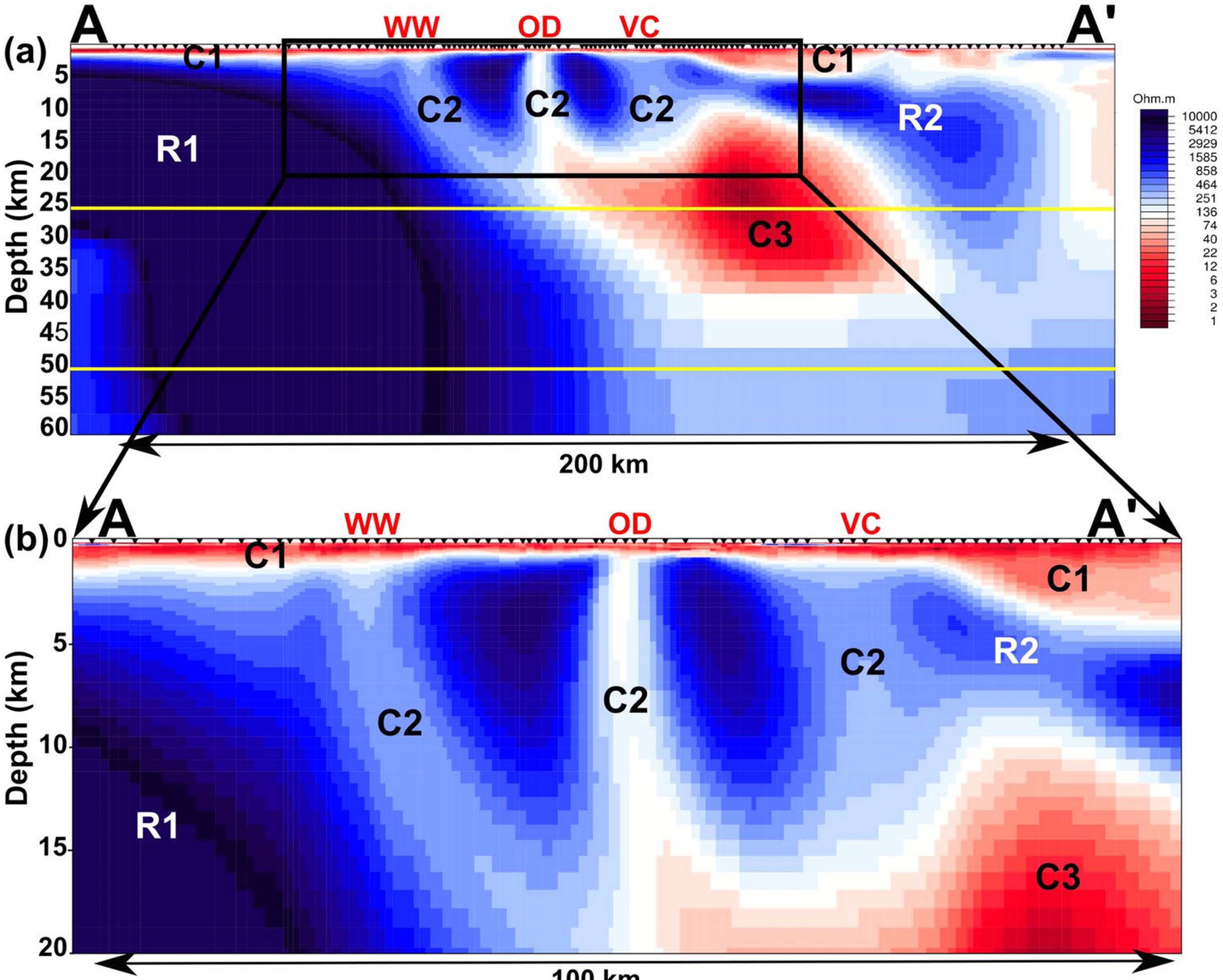
Lu et al. (2021)

Surface geometry inversion, global optimization, sampling ...

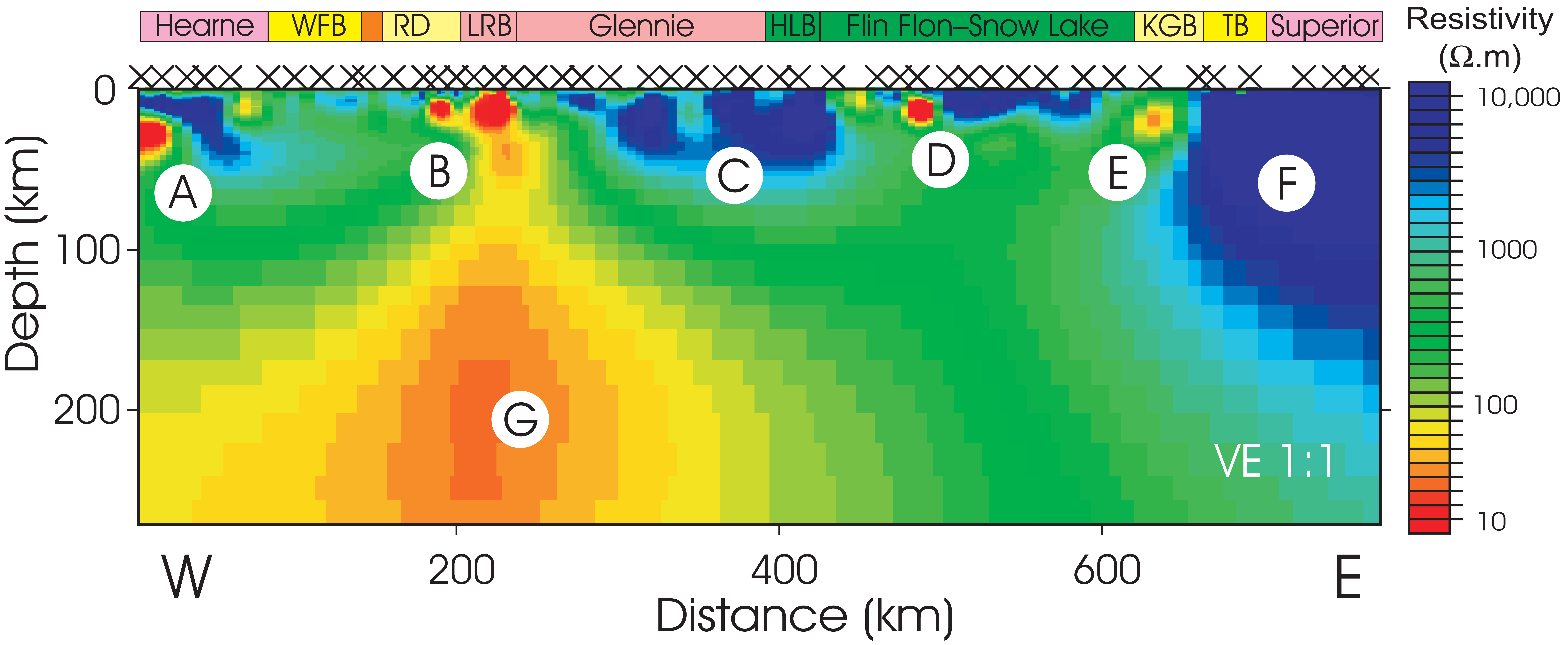




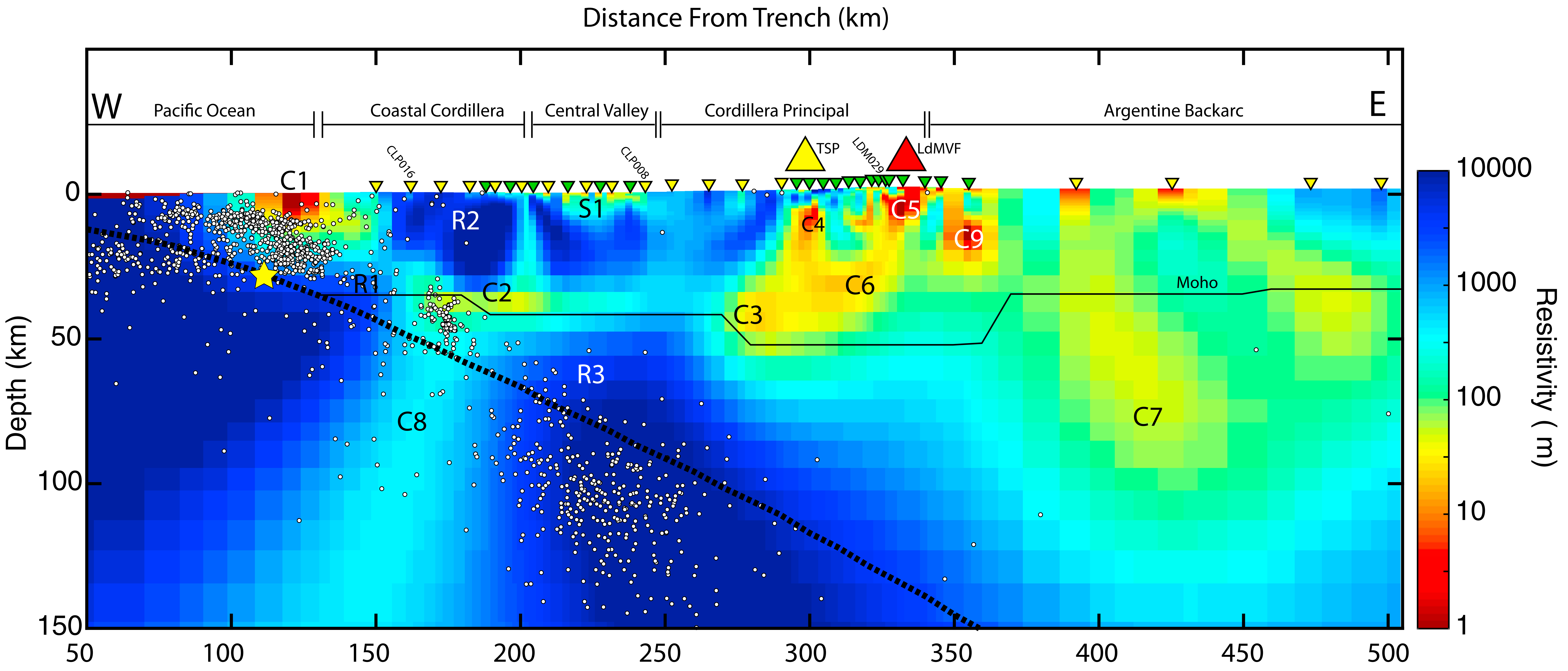
Lu et al. (2021)



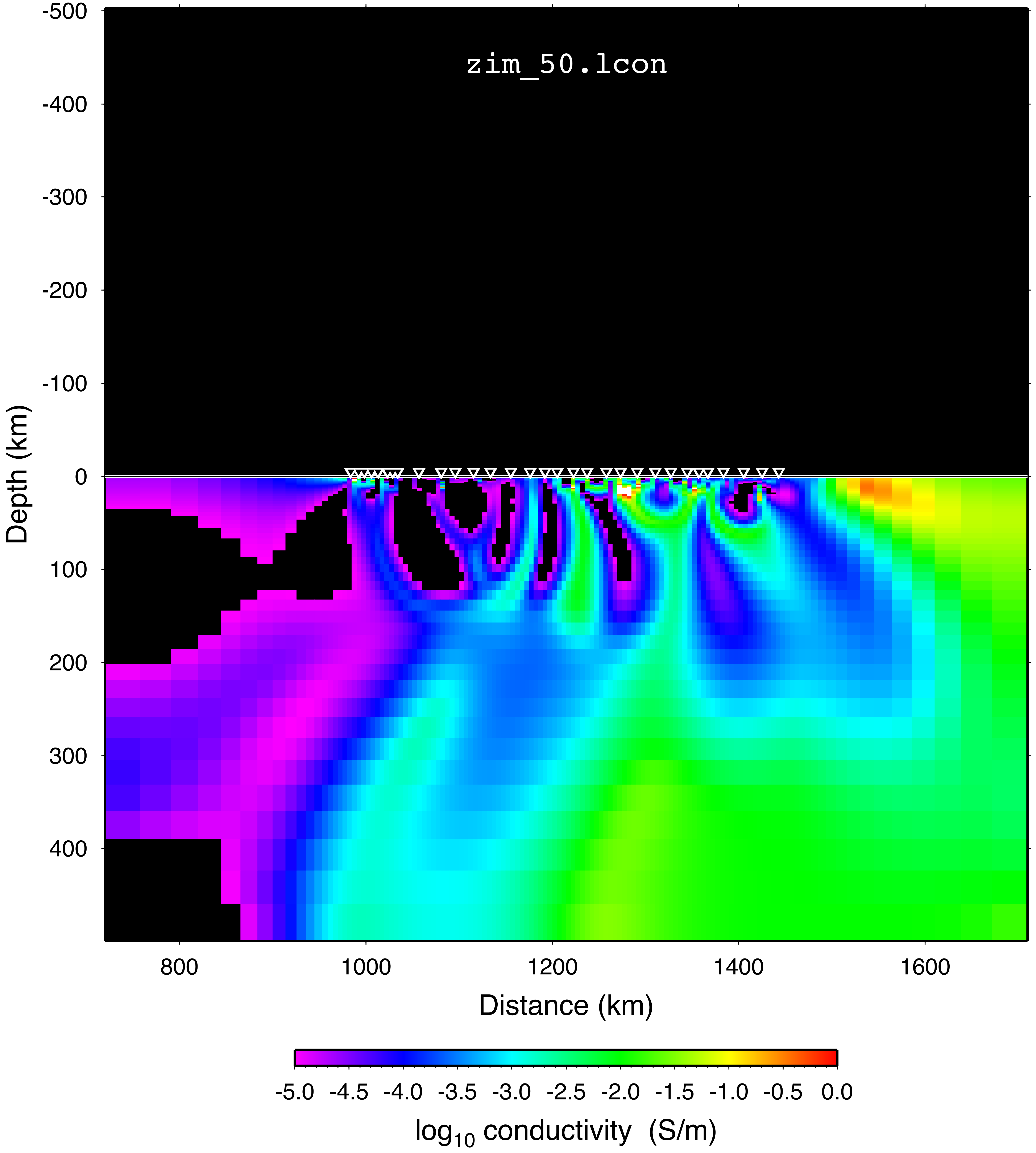
Heinson et al. (2018)



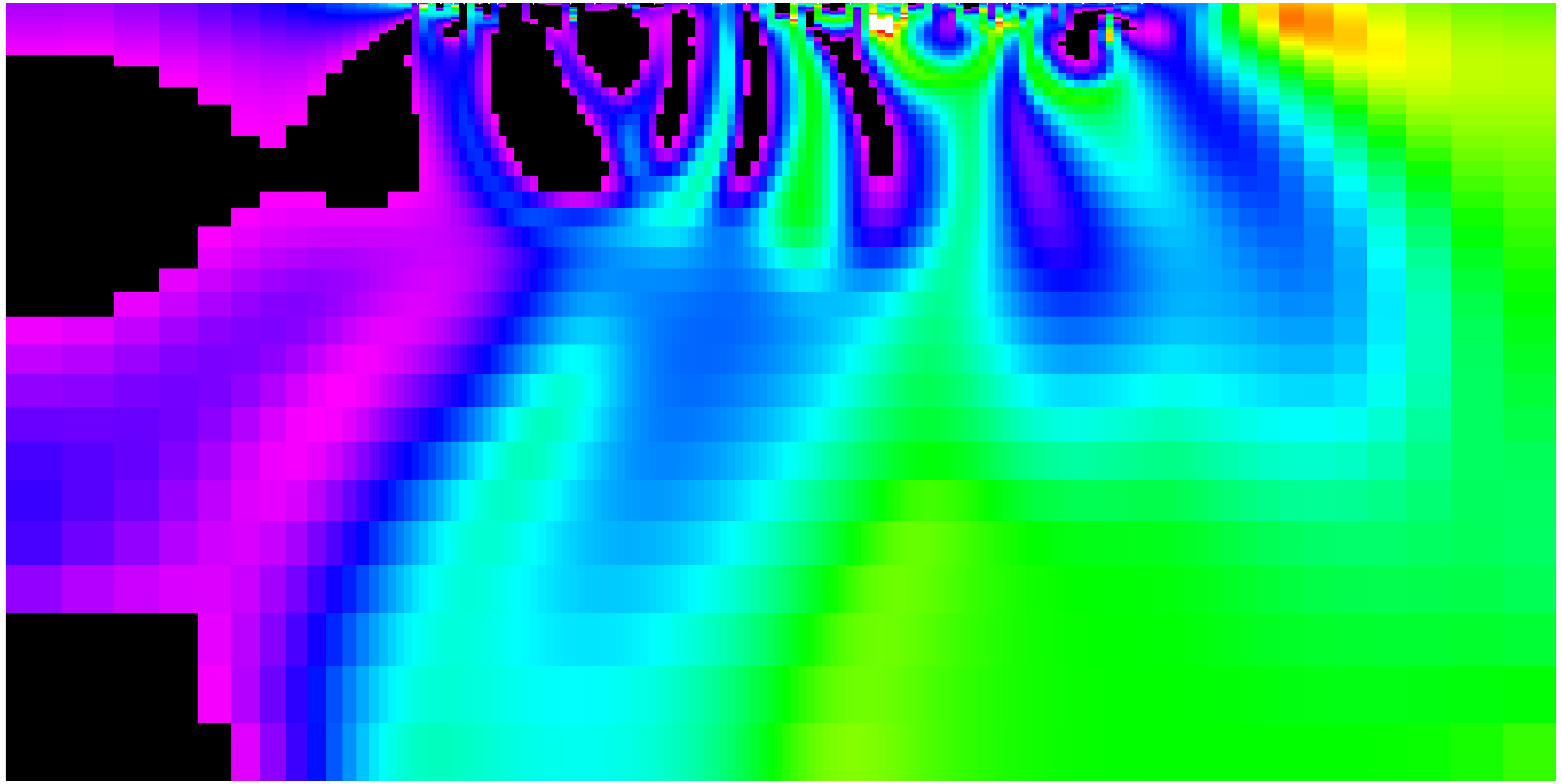
Jones et al. (2005)

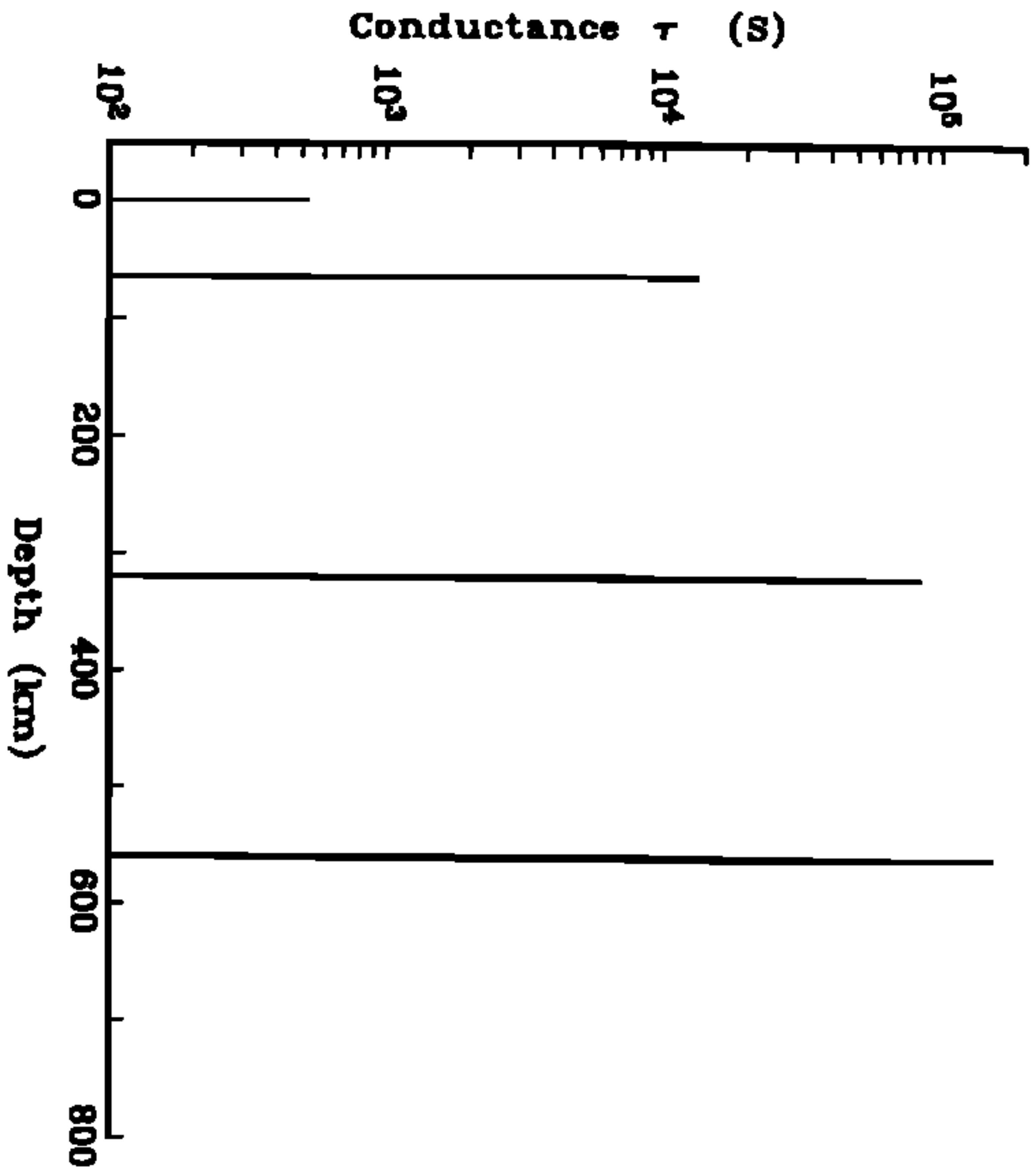


Cordell et al. (2019)


zim_50.lcon
Depth (km)
-500
-400
-300
-200
-100
0
100
200
300
400
800
1000
1200
1400
1600
Distance (km)
-5.0 -4.5 -4.0 -3.5 -3.0 -2.5 -2.0 -1.5 -1.0 -0.5 0.0
log_{10} conductivity (S/m)

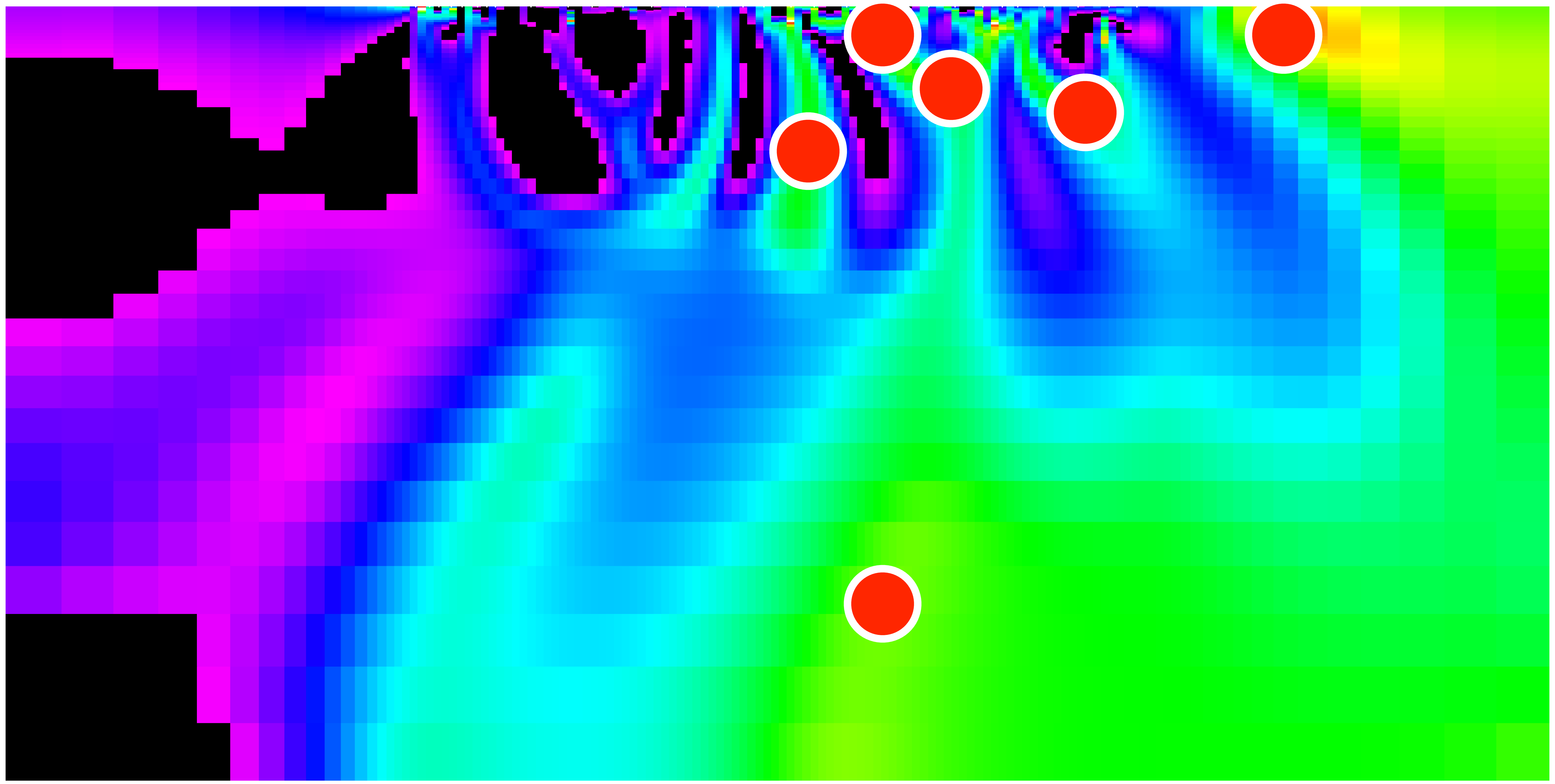
2-D D+ model ?

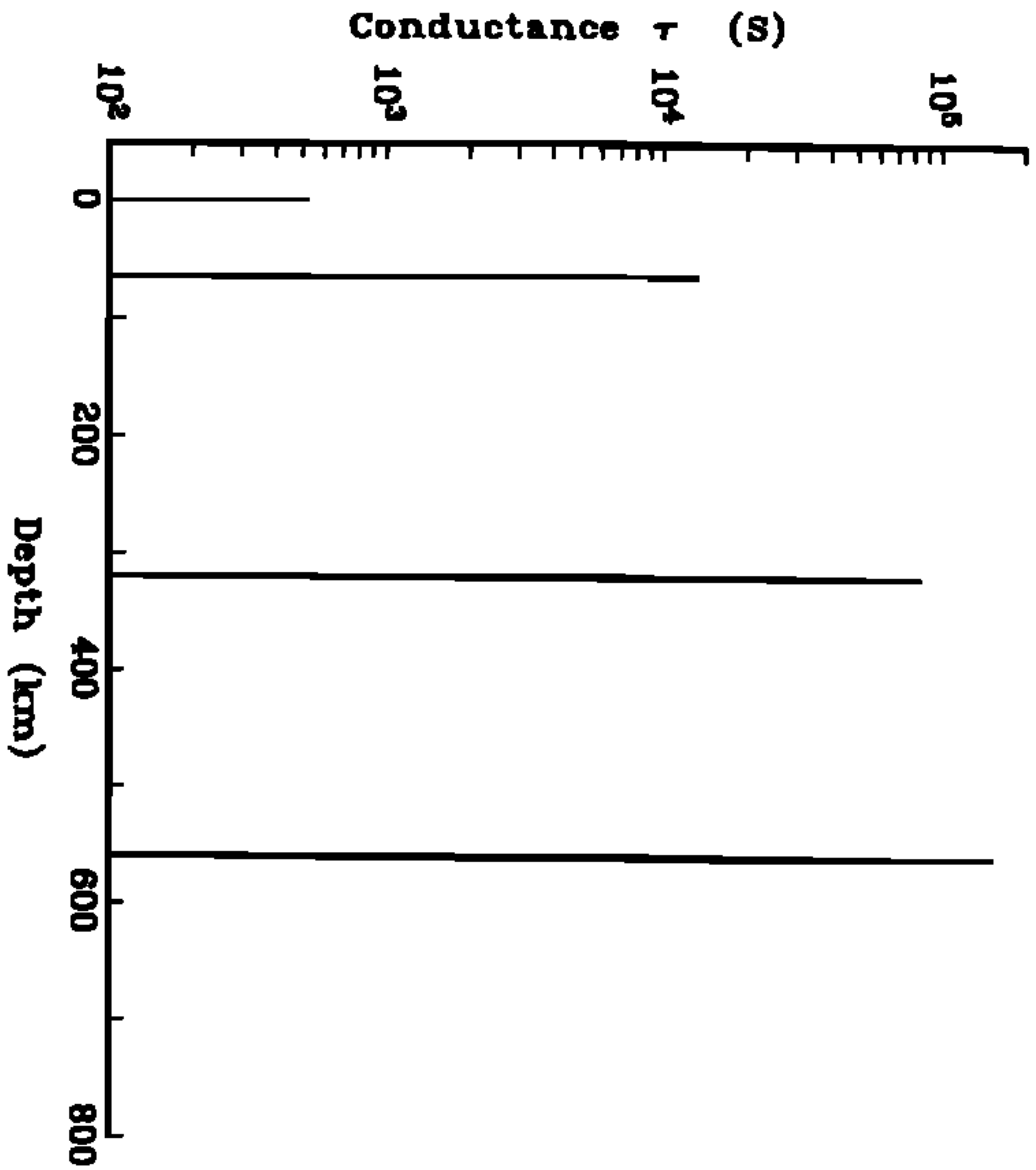




2-D D+ model ?

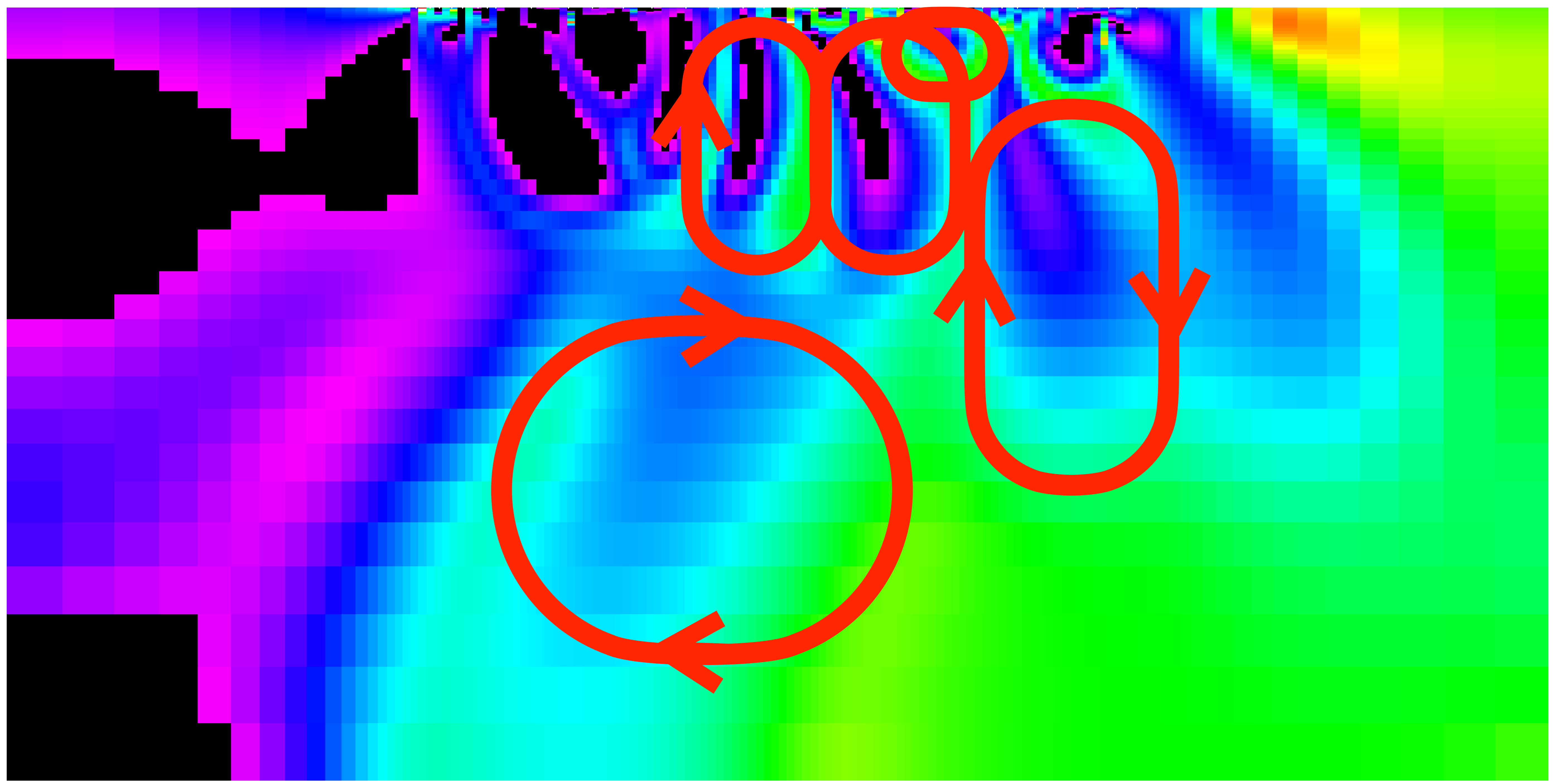
TE-model/E-polarization ?

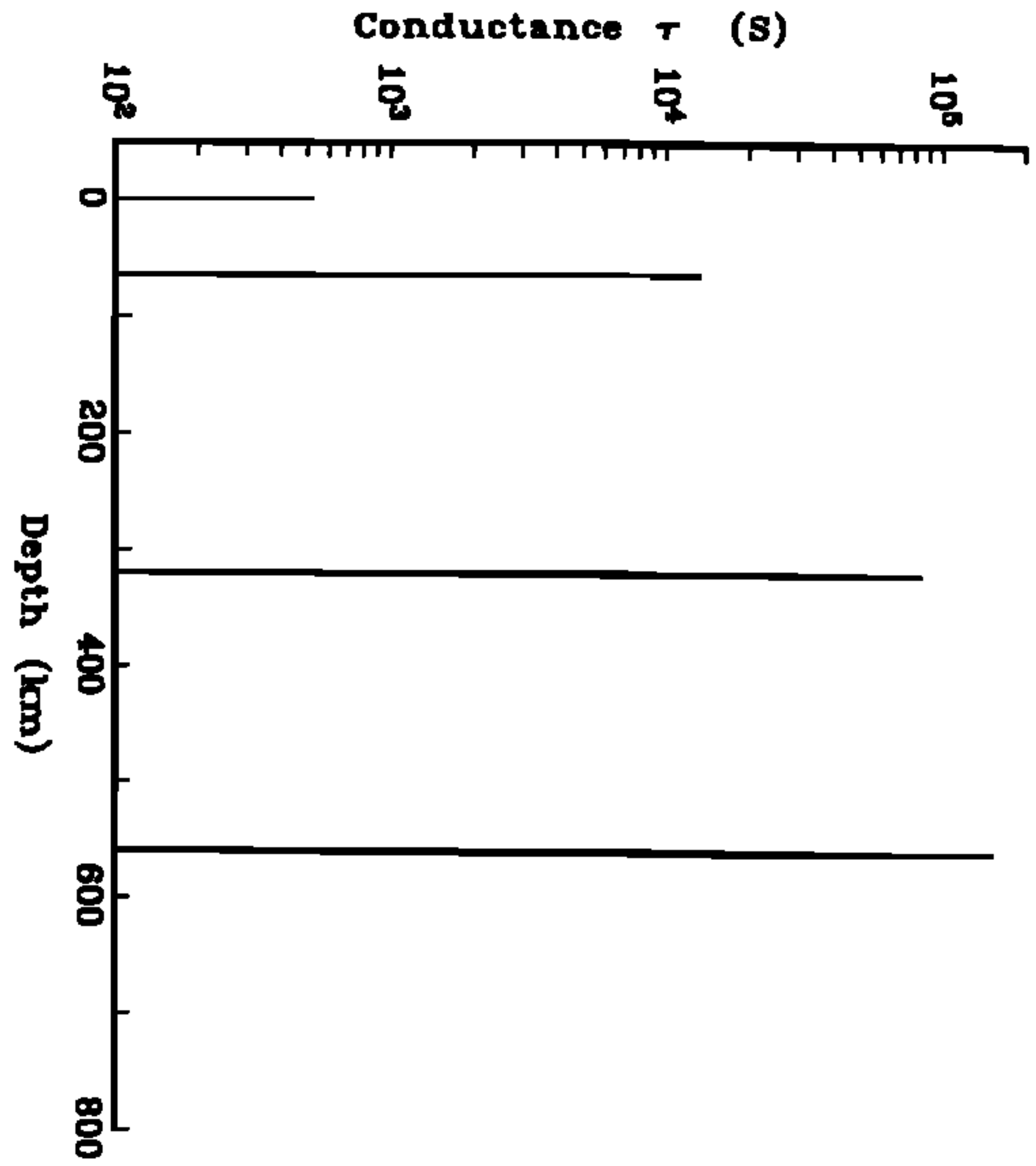




2-D D+ model ?

TM-model/B-polarization ?





Outline

~~Inversion~~ optimization background:

- why optimization
- data misfit
- minimizing data misfit, non-uniqueness
- measure of model somethingorother
- descent-based, gradient-based optimization (linearization)
- or sampling and selection of collections of models we like

Forward modelling

Sensitivities

Descent/gradient/derivative/linearization-based algorithms

The Conclusion

→ Future work/thoughts

- different models, different approaches?
- D+ ?

Outline

~~Inversion~~ optimization background:

- why optimization
- data misfit
- minimizing data misfit, non-uniqueness
- measure of model somethingorother
- descent-based, gradient-based optimization (linearization)
- or sampling and selection of collections of models we like

Forward modelling

Sensitivities

Descent/gradient/derivative/linearization-based algorithms

The Conclusion

Future work/thoughts

- different models, different approaches?
- D+ ?

The Conclusion

"Occam", minimum-structure inversion is very effective — it gets the job done — and hence useful and popular.

The End.

MEMORIAL
UNIVERSITY