SOCIETY OF EXPLORATION GEOPHYSICISTS

Geophysical Inversion: which model do you want?

Steven Constable

Scripps Institution of Oceanography

SOCIETY OF EXPLORATION GEOPHYSICISTS

Acknowledgements

- SEG and SEG Foundation
- Sponsored by Statoil
- Sponsored by Paradigm
- Scripps Institution of Oceanography,
  Seafloor Electromagnetic Methods Consortium

SOCIETY OF EXPLORATION GEOPHYSICISTS

SEG Membership Benefits

Join Today!

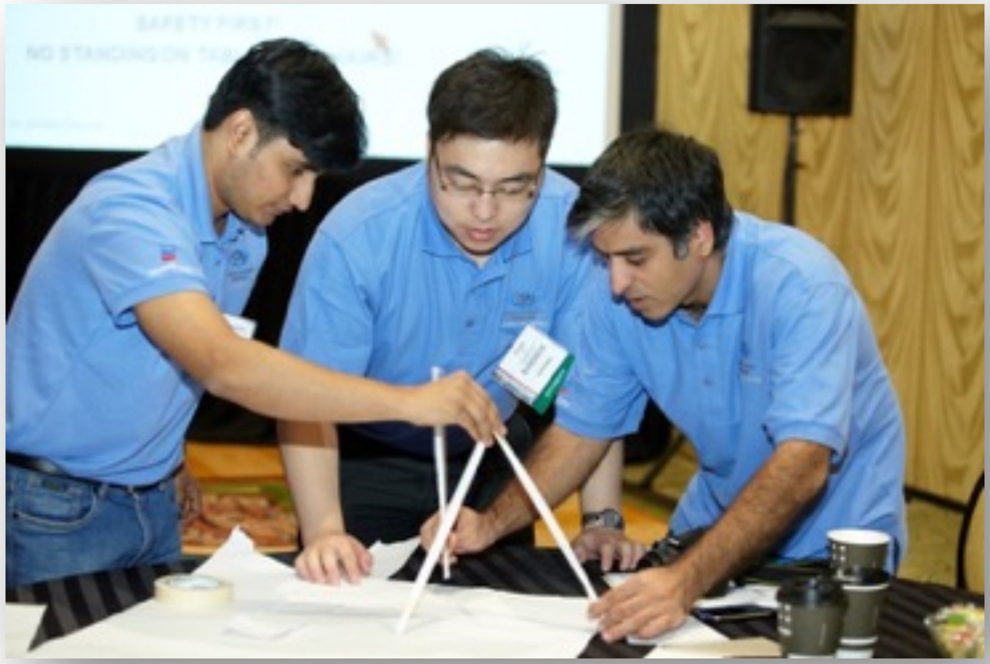
- SEG Digital Library – full text articles
- Technical Journals in Print and Online
- Networking Opportunities
- Membership Discounts on
  - Continuing Education Training Courses
  - Publications (35% off list price)
  - Workshops and Meetings

Membership materials are available today!

Join Online at seg.org/About-SEG/Membership

Student Opportunities

- Sponsored Membership
- Student Chapter Programs
- SEG/Chevron Student Leadership Symposium
- SEG/ExxonMobil Student Education Program
- Challenge Bowl
- Scholarships
- Field Camp Grants
- Geoscientists *Without* Borders®
- Student Expos & IGSCs
- Honorary and Distinguished Lecturers
- SEG Online



Learn more at seg.org/Education/Students-Early-Career

Section/Associated Society Opportunities

- Host DL, HL, and DISC Programs
- Council Representation
- Annual Meeting Booth Discount
- Best Papers presented at SEG Annual Meetings
- Joint Conferences, Workshops, and Forums in partnership with SEG

For more information, including a list of benefits, please visit:
www.seg.org/resources/sections-societies

Geophysical Inversion: which model do you want?

Steven Constable

Scripps Institution of Oceanography

There is an old joke…

The production manager asked a geologist, engineer, and geophysicist what 2 + 2 was.

The production manager asked a geologist, engineer, and geophysicist what 2 + 2 was.

The geologist thought for a bit and then said "*somewhere between 3 and 5*".

The production manager asked a geologist, engineer, and geophysicist what 2 + 2 was.

The geologist thought for a bit and then said "*somewhere between 3 and 5*".

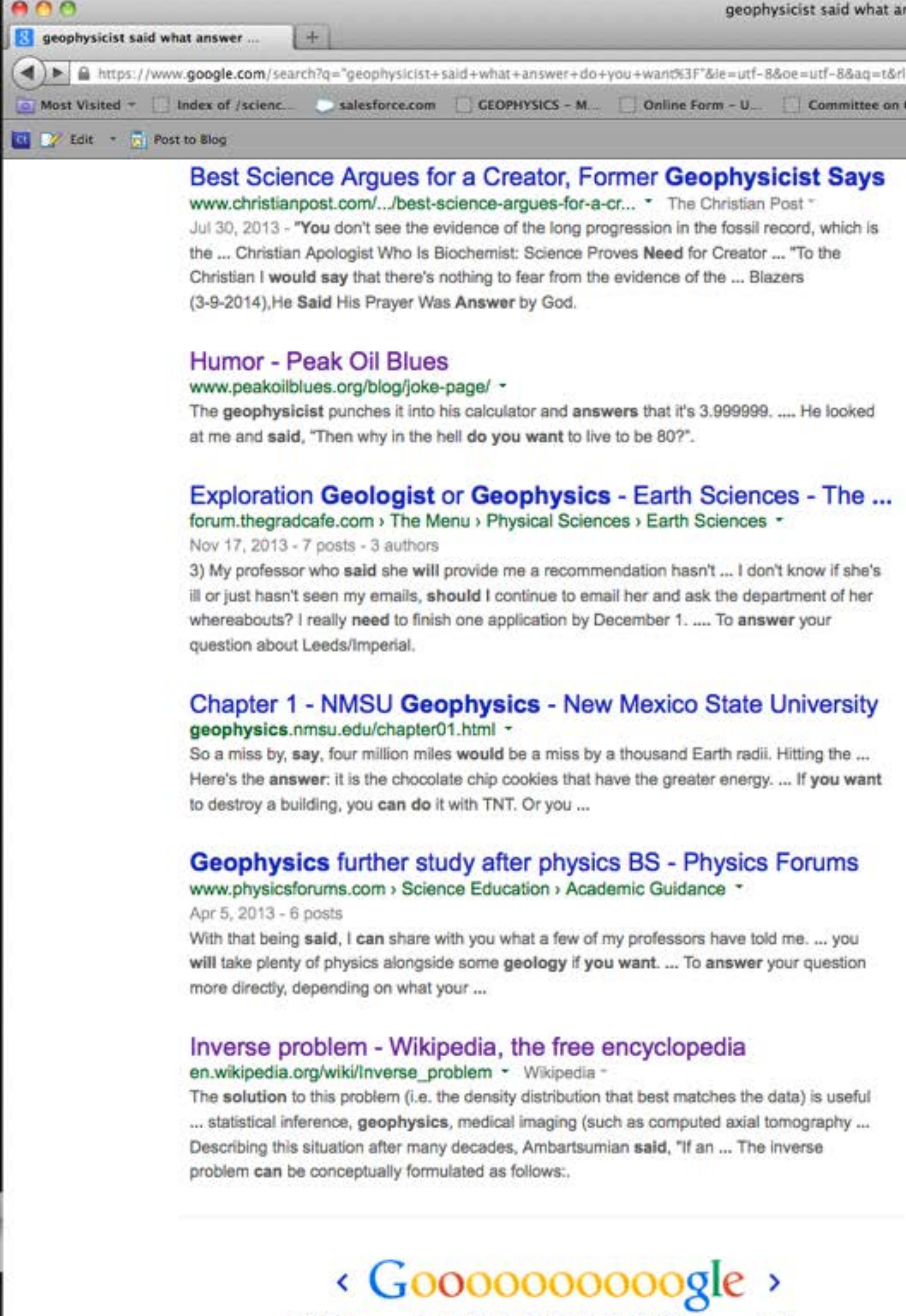
The engineer fiddled with a calculator and said "*3.9999999*".

The production manager asked a geologist, engineer, and geophysicist what 2 + 2 was.

The geologist thought for a bit and then said "*somewhere between 3 and 5*".

The engineer fiddled with a calculator and said "*3.9999999*".

The geophysicist looked her in the eye and asked "*what answer do you want*"


geophysicist said what answer do you want? - Google Search

geophysicist said what answer ...

https://www.google.com/search?q="geophysicist+said+what+answer+do+you+want%3F"&ie=utf-8&oe=utf-8&aq=t&rls=org.mozilla:en-US:official&client=firefox-a&channel

"geophysicist said what answer do you want?"

Most Visited | Index of /scienc... | salesforce.com | GEOPHYSICS - M... | Online Form - U... | Committee on C... | :: SIO Ship Opera... | Web of Knowled...

Edit | Post to Blog

Best Science Argues for a Creator, Former Geophysicist Says
www.christianpost.com/.../best-science-argues-for-a-cr... - The Christian Post
Jul 30, 2013 - "You don't see the evidence of the long progression in the fossil record, which is the ... Christian Apologist Who Is Biochemist: Science Proves Need for Creator ... "To the Christian I would say that there's nothing to fear from the evidence of the ... Blazers (3-9-2014),He Said His Prayer Was Answer by God.

Humor - Peak Oil Blues
www.peakoilblues.org/blog/joke-page/
The geophysicist punches it into his calculator and answers that it's 3.999999. .... He looked at me and said, "Then why in the hell do you want to live to be 80?".

Exploration Geologist or Geophysics - Earth Sciences - The ...
forum.thegradcafe.com › The Menu › Physical Sciences › Earth Sciences
Nov 17, 2013 - 7 posts - 3 authors
3) My professor who said she will provide me a recommendation hasn't ... I don't know if she's ill or just hasn't seen my emails, should I continue to email her and ask the department of her whereabouts? I really need to finish one application by December 1. .... To answer your question about Leeds/Imperial.

Chapter 1 - NMSU Geophysics - New Mexico State University
geophysics.nmsu.edu/chapter01.html
So a miss by, say, four million miles would be a miss by a thousand Earth radii. Hitting the ... Here's the answer: it is the chocolate chip cookies that have the greater energy. ... If you want to destroy a building, you can do it with TNT. Or you ...

Geophysics further study after physics BS - Physics Forums
www.physicsforums.com › Science Education › Academic Guidance
Apr 5, 2013 - 6 posts
With that being said, I can share with you what a few of my professors have told me. ... you will take plenty of physics alongside some geology if you want. ... To answer your question more directly, depending on what your ...

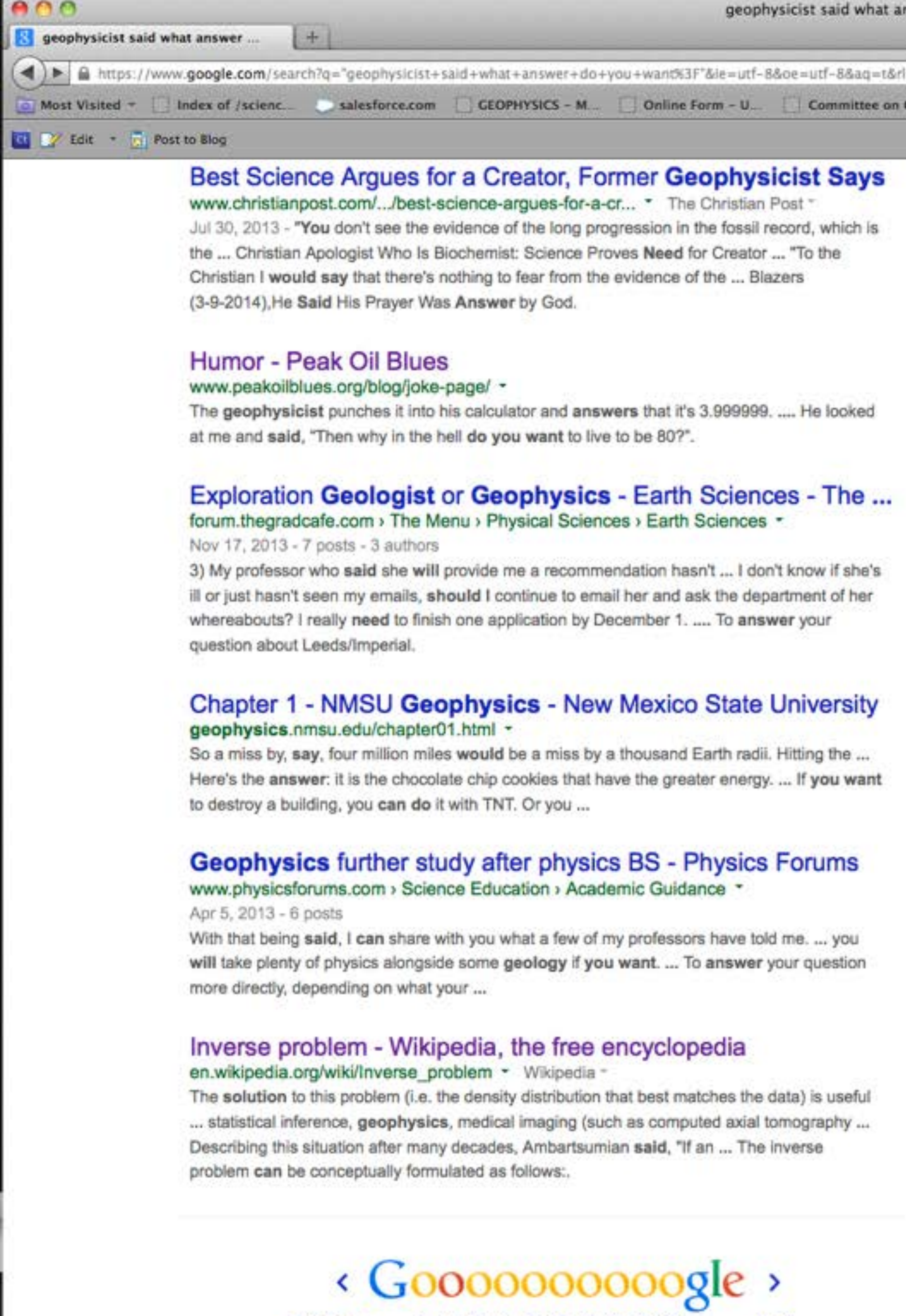
Inverse problem - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Inverse_problem - Wikipedia
The solution to this problem (i.e. the density distribution that best matches the data) is useful ... statistical inference, geophysics, medical imaging (such as computed axial tomography ... Describing this situation after many decades, Ambartsumian said, "If an ... The inverse problem can be conceptually formulated as follows:.

‹ Goooooooooogle ›

It is notable that when once searching the web for this joke, I not only got the joke page of an oil price blog, but also the Wikipedia entry for "Inverse problem".

This says it all…

Today we will explore how the inverse problem can give you whatever model you want.



geophysicist said what answer do you want? - Google Search

Best Science Argues for a Creator, Former Geophysicist Says
www.christianpost.com/.../best-science-argues-for-a-cr... The Christian Post
Jul 30, 2013 - "You don't see the evidence of the long progression in the fossil record, which is the ... Christian Apologist Who Is Biochemist: Science Proves Need for Creator ... "To the Christian I would say that there's nothing to fear from the evidence of the ... Blazers (3-9-2014),He Said His Prayer Was Answer by God.

Humor - Peak Oil Blues
www.peakoilblues.org/blog/joke-page/
The geophysicist punches it into his calculator and answers that it's 3.999999. .... He looked at me and said, "Then why in the hell do you want to live to be 80?".

Exploration Geologist or Geophysics - Earth Sciences - The ...
forum.thegradcafe.com › The Menu › Physical Sciences › Earth Sciences
Nov 17, 2013 - 7 posts - 3 authors
3) My professor who said she will provide me a recommendation hasn't ... I don't know if she's ill or just hasn't seen my emails, should I continue to email her and ask the department of her whereabouts? I really need to finish one application by December 1. .... To answer your question about Leeds/Imperial.

Chapter 1 - NMSU Geophysics - New Mexico State University
geophysics.nmsu.edu/chapter01.html
So a miss by, say, four million miles would be a miss by a thousand Earth radii. Hitting the ... Here's the answer: it is the chocolate chip cookies that have the greater energy. ... If you want to destroy a building, you can do it with TNT. Or you ...

Geophysics further study after physics BS - Physics Forums
www.physicsforums.com › Science Education › Academic Guidance
Apr 5, 2013 - 6 posts
With that being said, I can share with you what a few of my professors have told me. ... you will take plenty of physics alongside some geology if you want. ... To answer your question more directly, depending on what your ...

Inverse problem - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Inverse_problem Wikipedia
The solution to this problem (i.e. the density distribution that best matches the data) is useful ... statistical inference, geophysics, medical imaging (such as computed axial tomography ... Describing this situation after many decades, Ambartsumian said, "If an ... The inverse problem can be conceptually formulated as follows:.

‹ Goooooooooogle ›

It is notable that when once searching the web for this joke, I not only got the joke page of an oil price blog, but also the Wikipedia entry for “Inverse problem”.

This says it all…

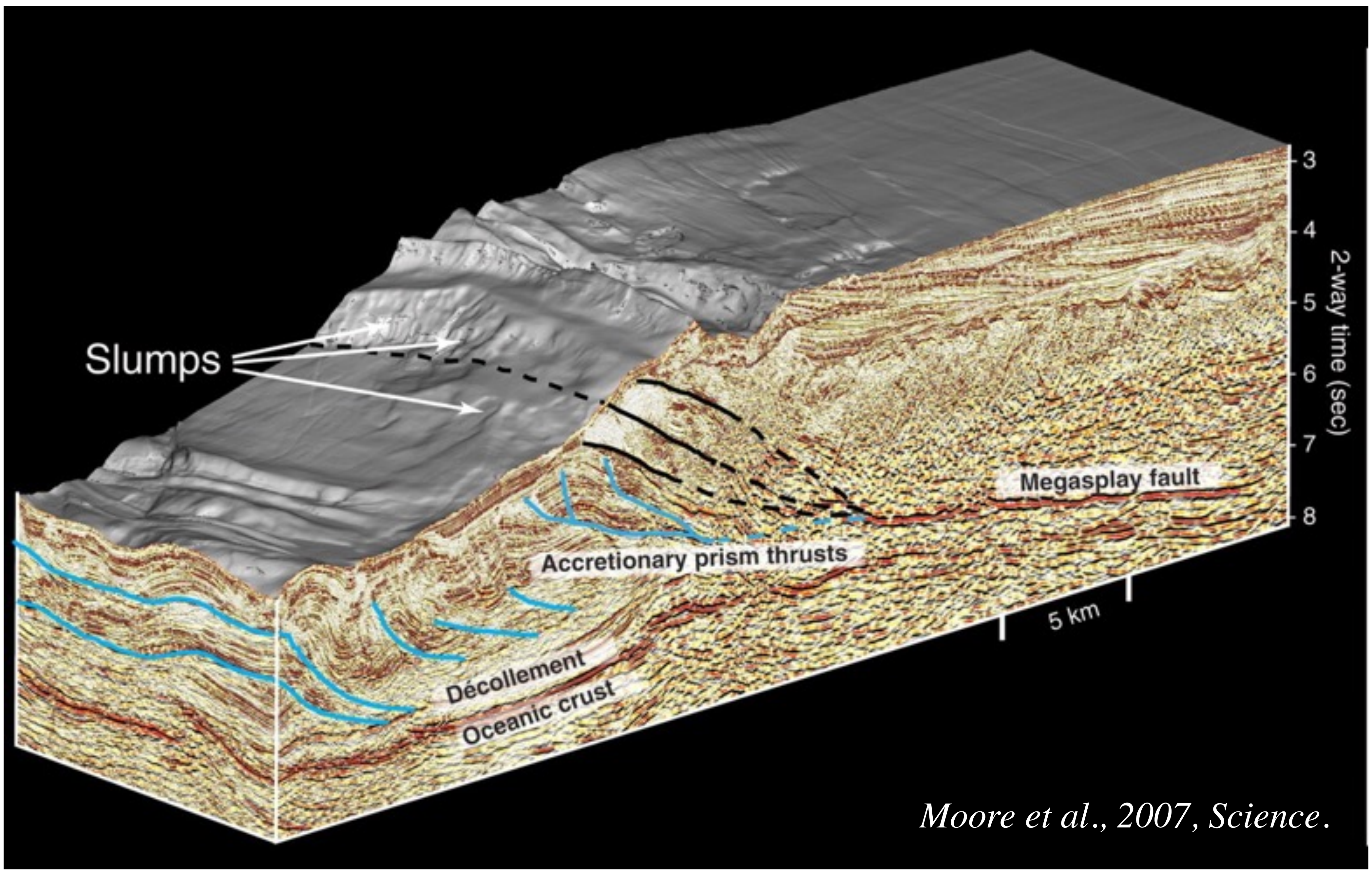
Today we will explore how the inverse problem can give you whatever model you want.

Well, almost.

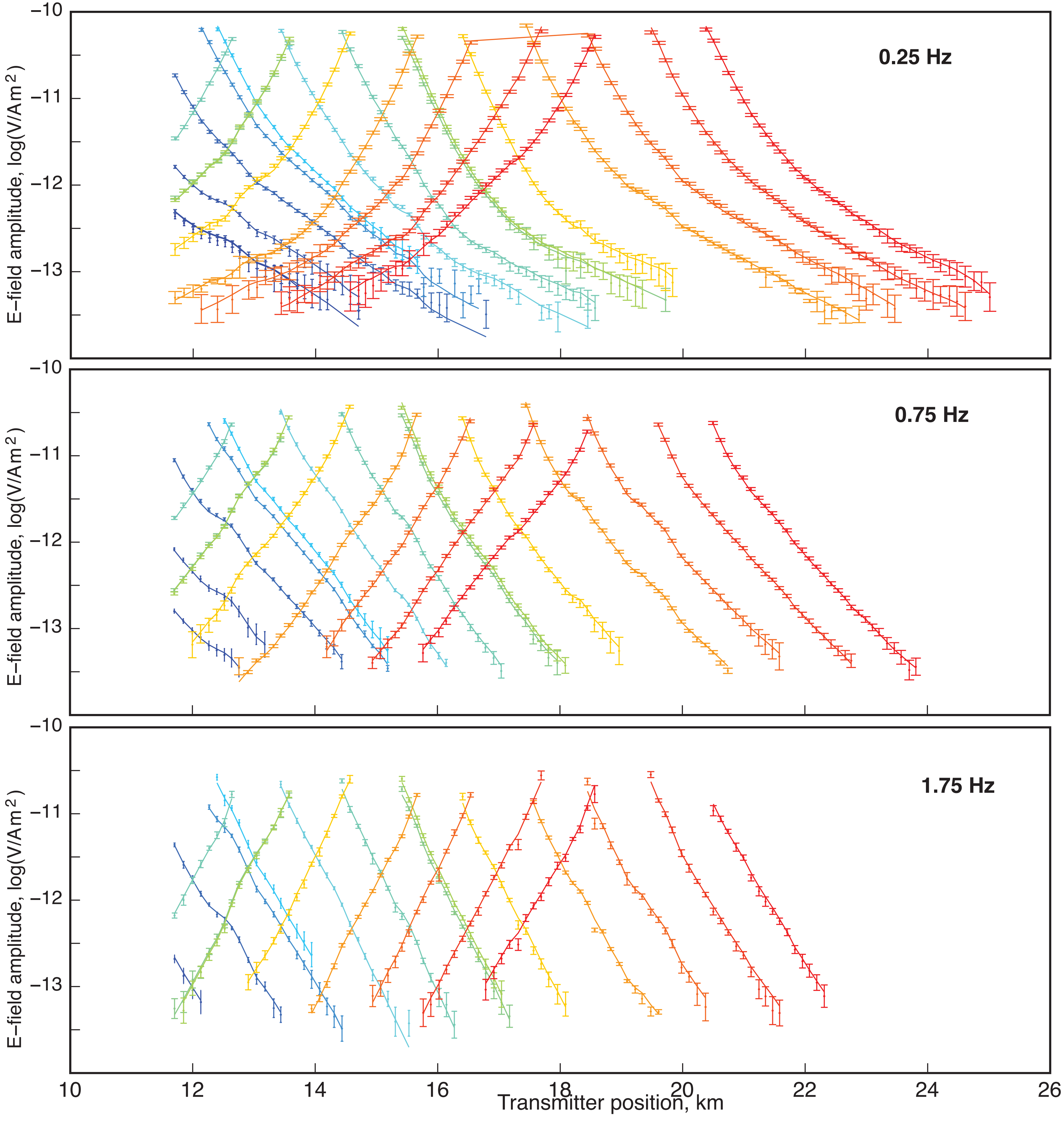
With seismic reflection images you can often see the geology in the data.

For EM and potential field methods you need inversion to recover something that can be interpreted as geology.

Same is true for seismic tomography and full waveform inversion.



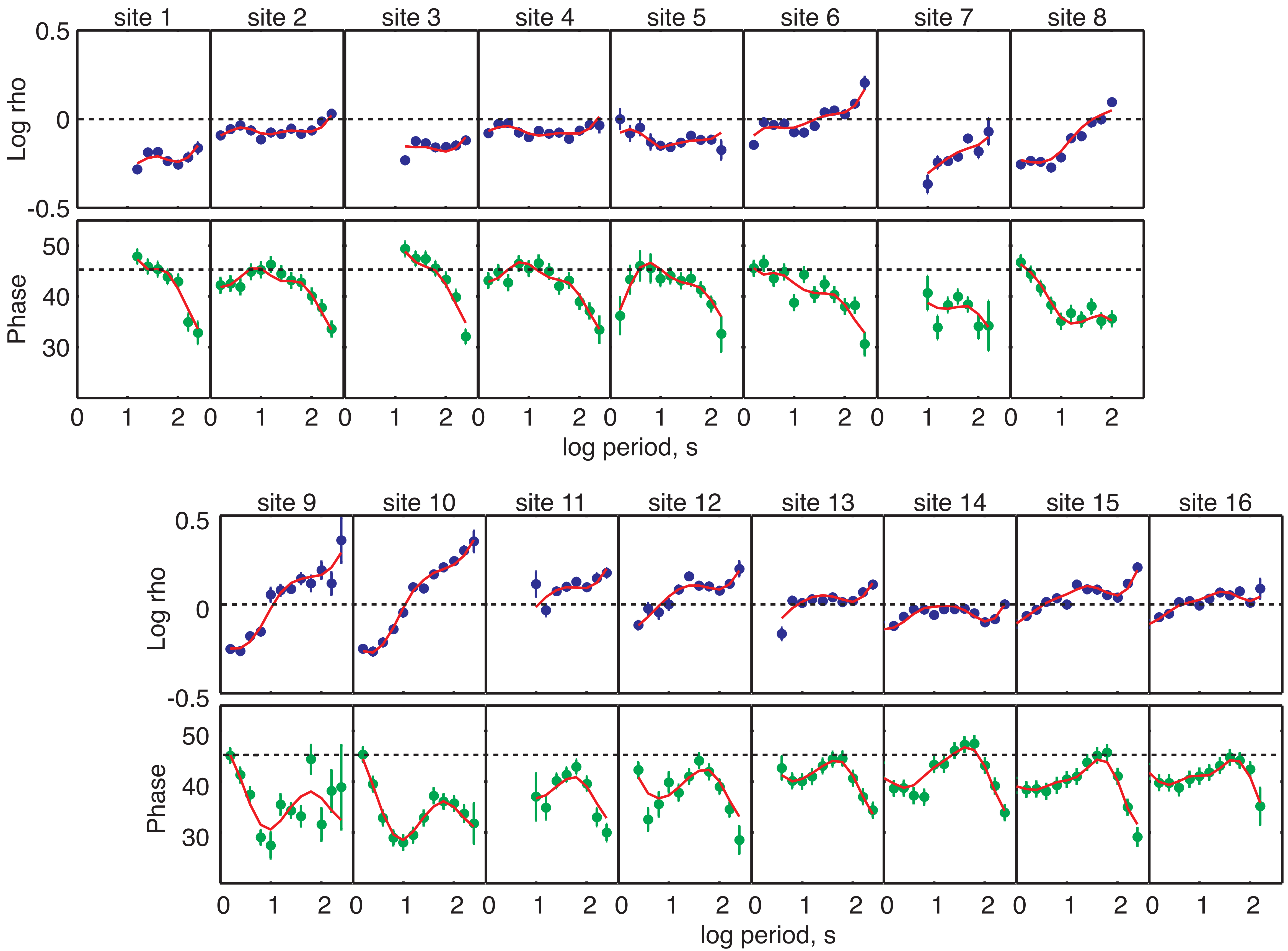
Moore et al., 2007, Science.



These are some marine controlled source electromagnetic CSEM data. You can't say much about geology just by looking at them.

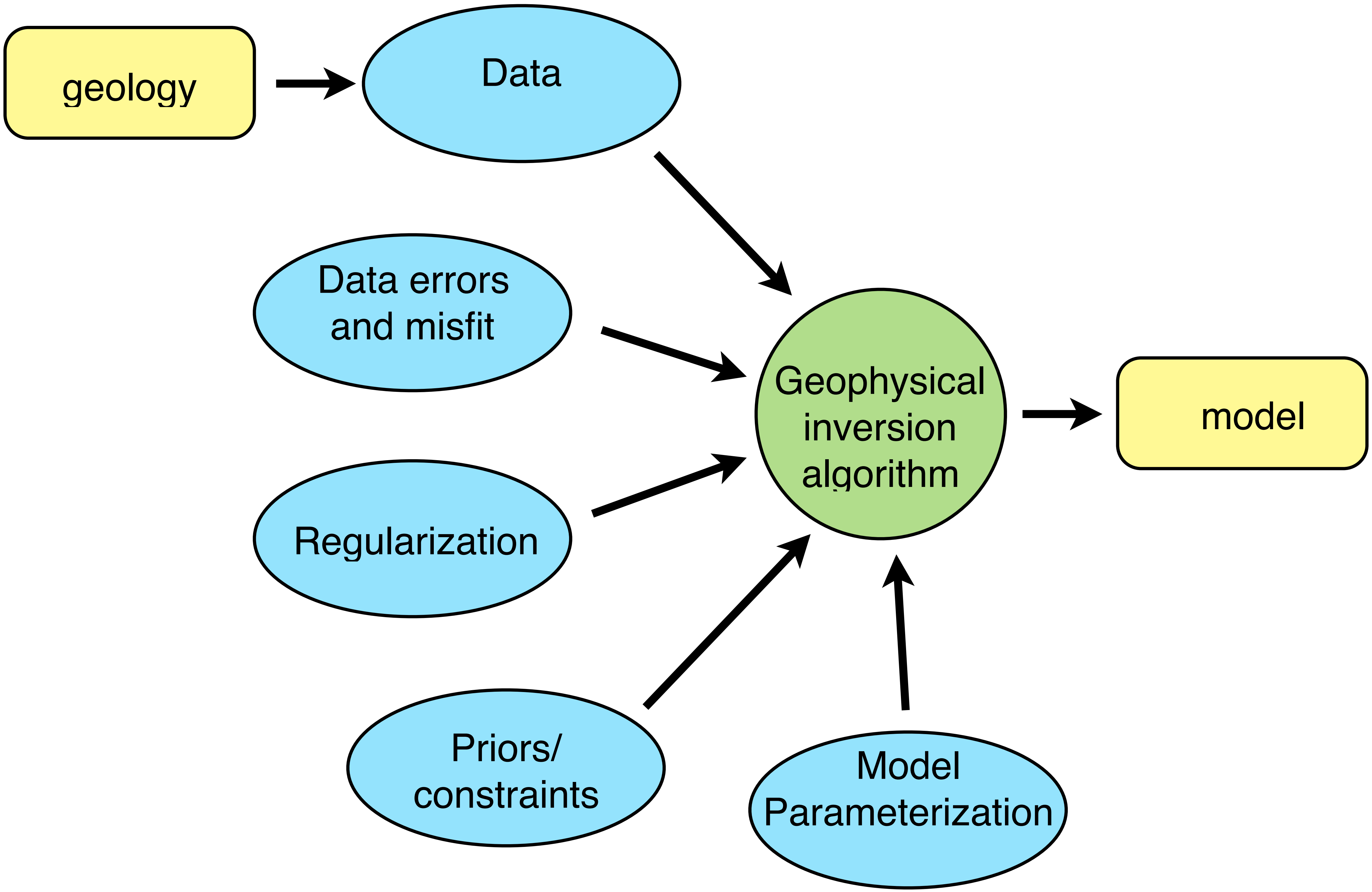
Constable, Orange, and Key, 2015

The same is true of magnetotelluric (MT) data…



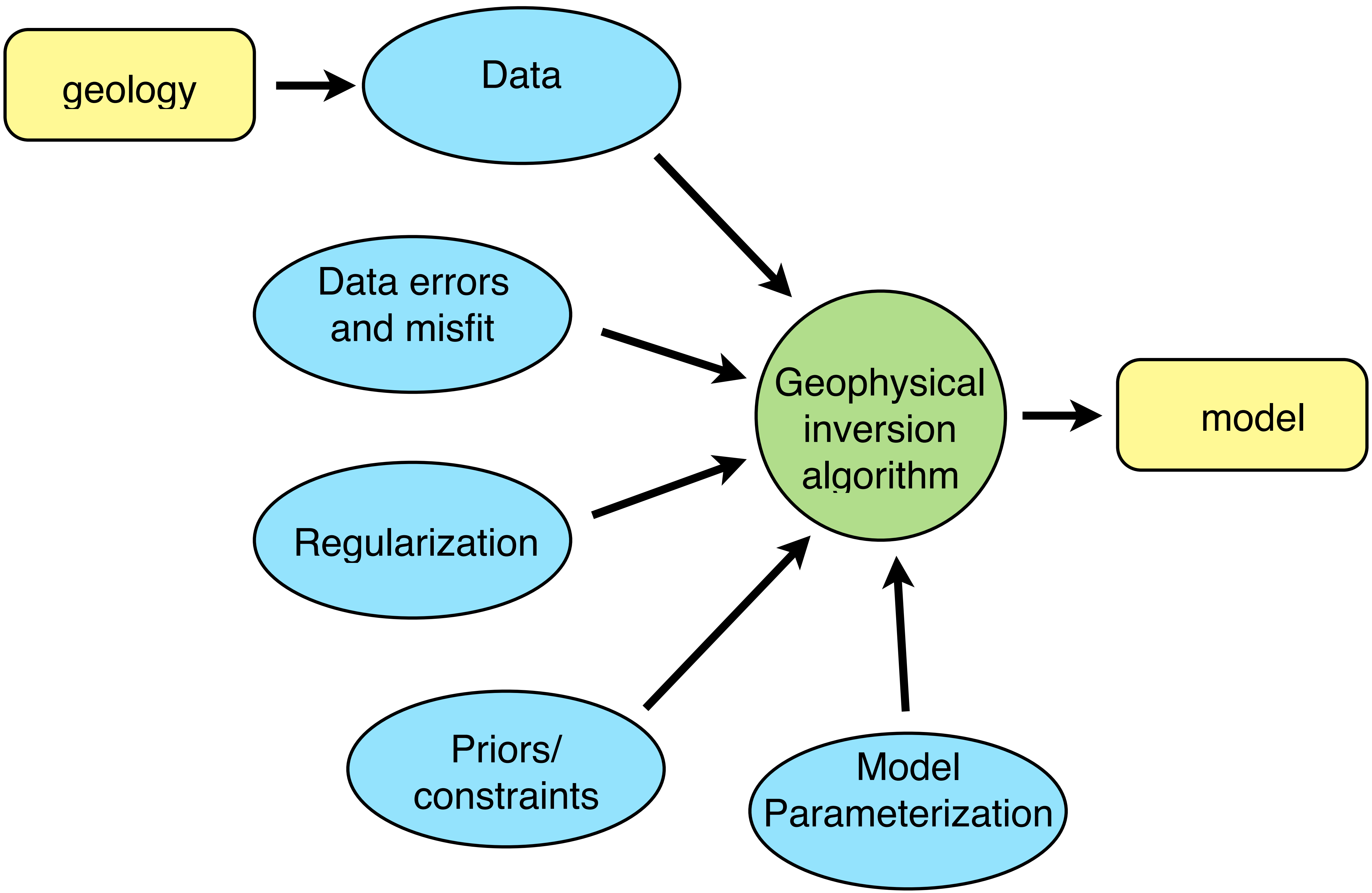
Constable, Orange, and Key, 2015

With many modern inversion algorithms available, it is all-so-easy to input data and turn the crank to get a model.



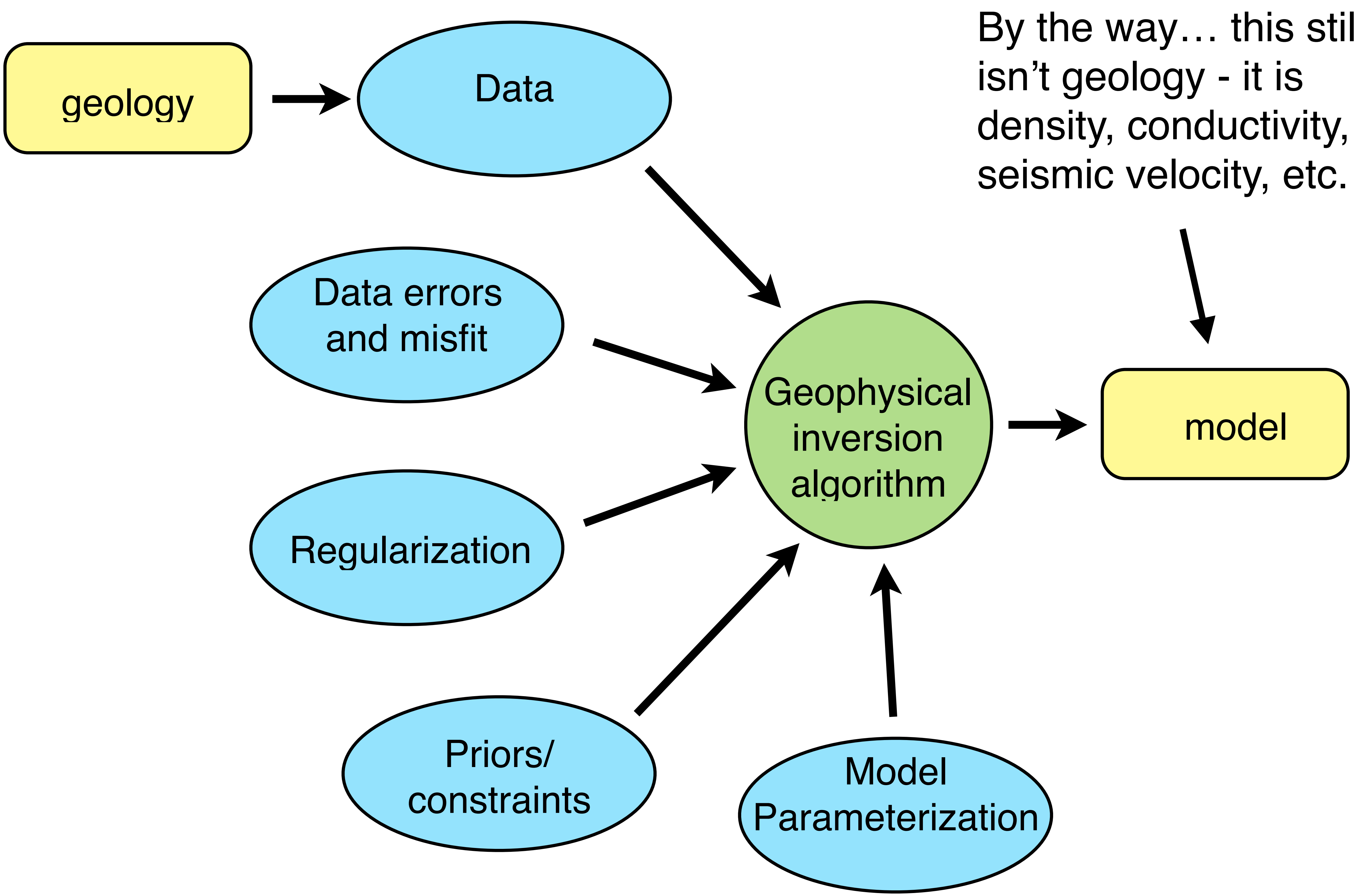
With many modern inversion algorithms available, it is all-so-easy to input data and turn the crank to get a model.

One of the main messages of my talk is that models from geophysical inversion depend on much more than the data inputs:

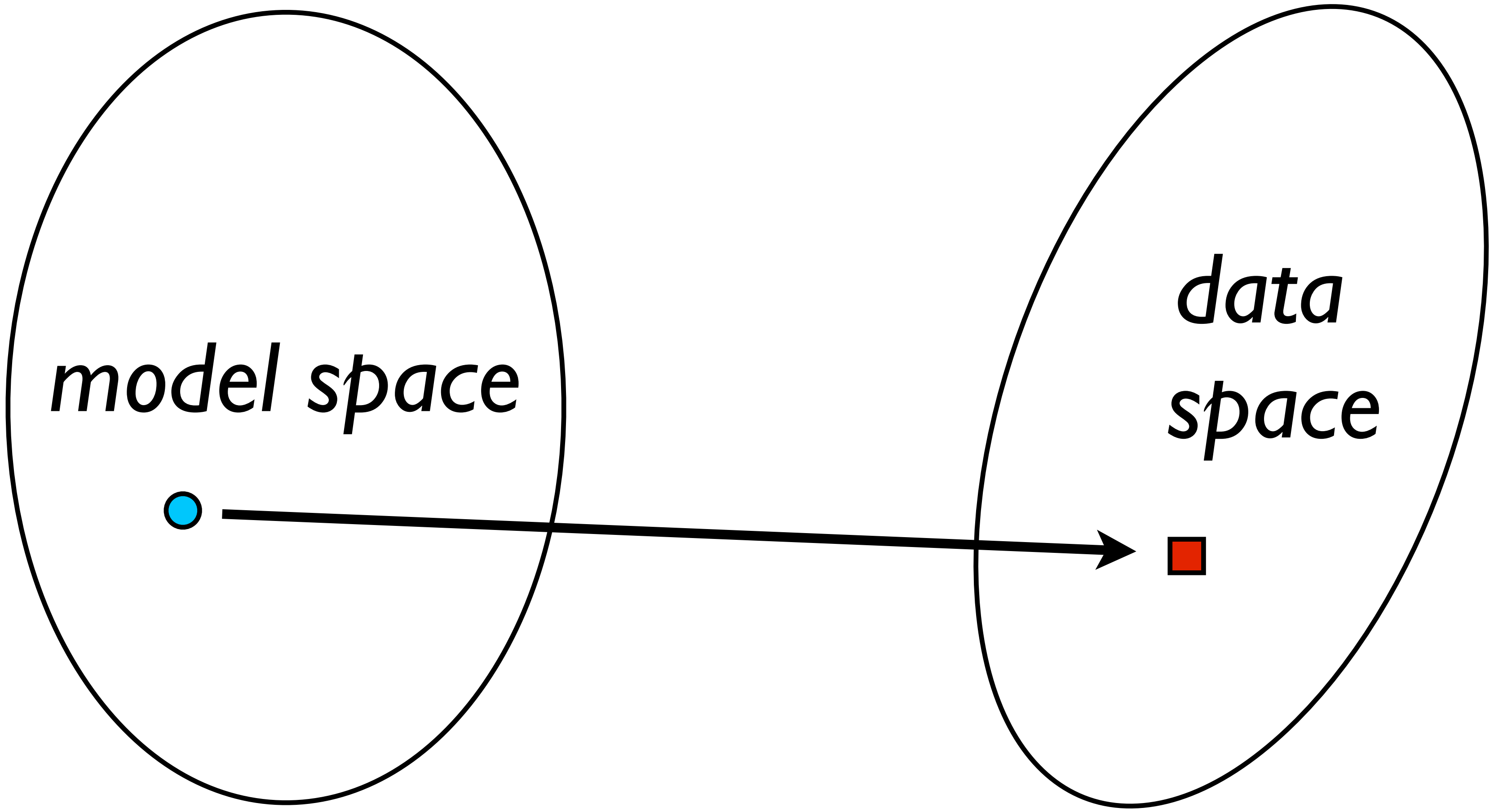


With many modern inversion algorithms available, it is all-so-easy to input data and turn the crank to get a model.

One of the main messages of my talk is that models from geophysical inversion depend on much more than the data inputs:



Forward modeling:



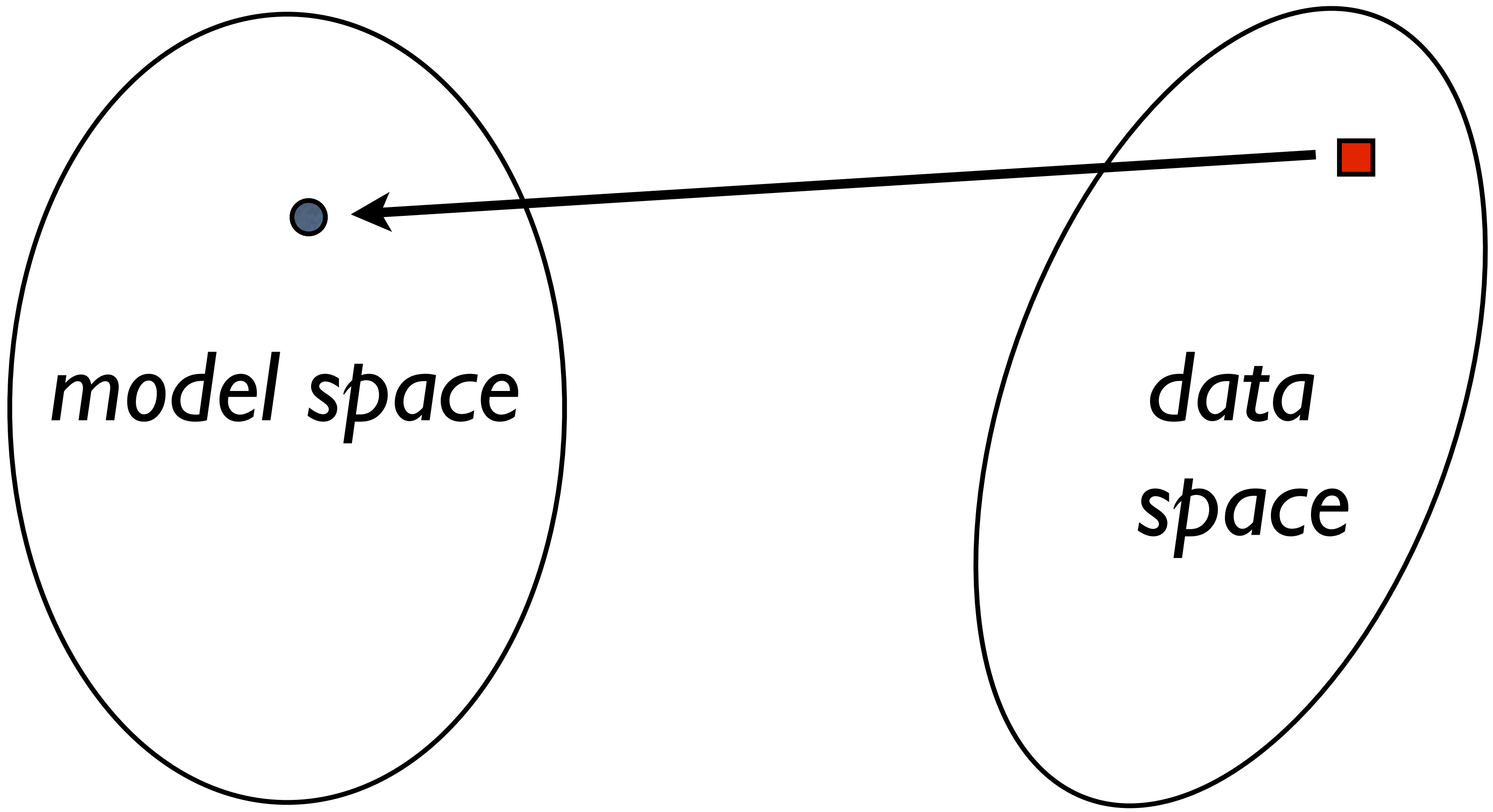
$$\hat{\mathbf{d}} = f(\mathbf{x}, \mathbf{m})$$ Some forward functional f

$$\mathbf{m} = (m_1, m_2, \ldots\ldots, m_N)$$ Model parameters (layers, blocks, ...)

$$\mathbf{x} = (x_1, x_2, x_3, \ldots\ldots, x_{kM})$$ Independent variables (freqs., locations, ...)

$$\hat{\mathbf{d}} = (\hat{d}_1, \hat{d}_2, \hat{d}_3, \ldots\ldots, \hat{d}_M)$$ Predicted data (gravity, magnetic, electric, ...)

Inverse modeling:



Given real (observed) data $\mathbf{d} = (d_1, d_2, d_3, ..., d_M)$

with errors $\sigma = (\sigma_1, \sigma_2, ..., \sigma_M)$

find an $\mathbf{m}$

There are several approaches to inversion:

Stochastic

- Monte Carlo, Markov Chains
- Genetic Algorithms
- Simulated annealing, etc.
  - (Bayesian Searches)

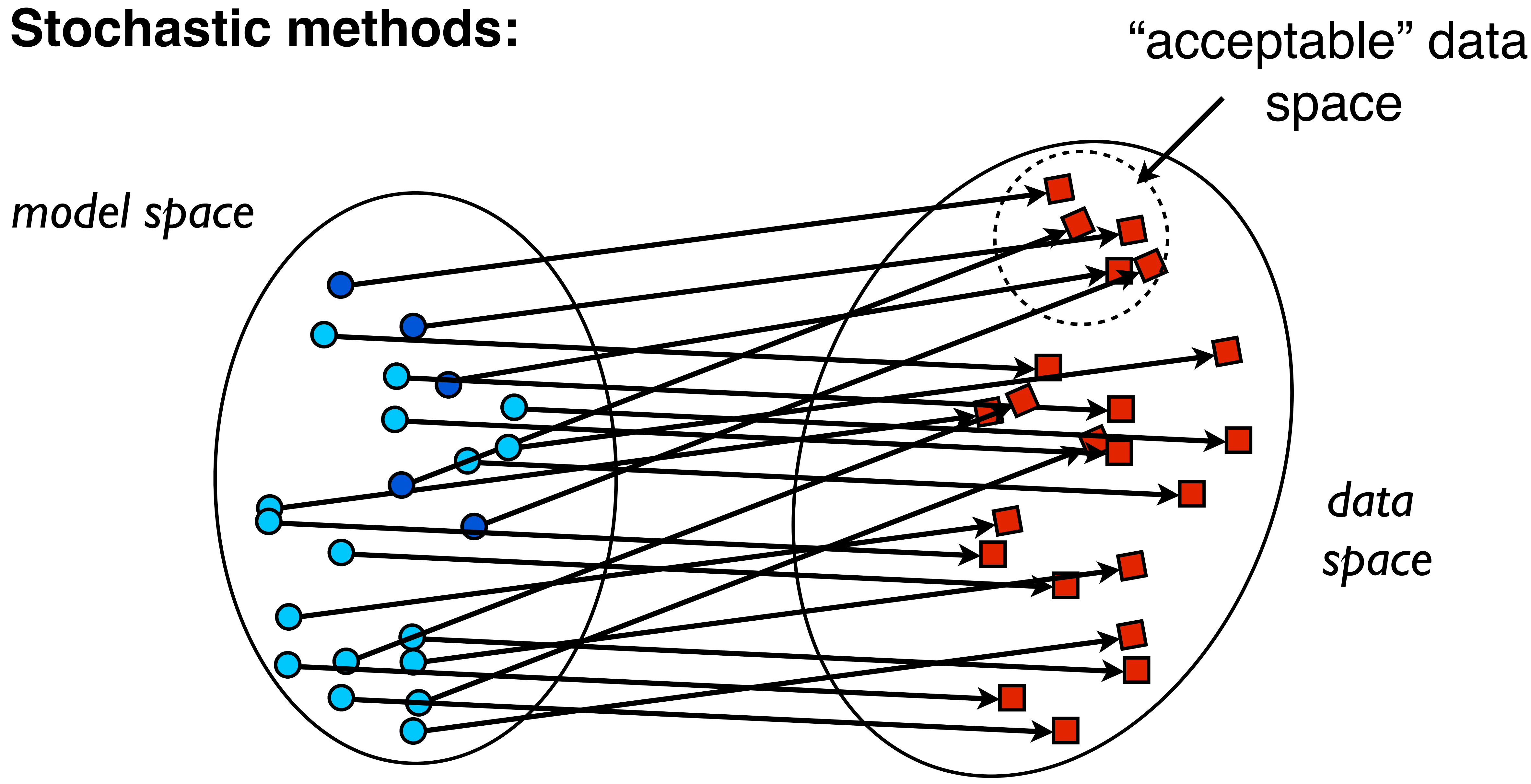
Deterministic

- Newton Algorithms
- Steepest descent
- Conjugate Gradients
- Quadratic (and Linear) Programming, etc.

Analytical

- D+ (1D MT)
- Bilayer (1D resistivity)
- Ideal body theory in gravity and magnetism

Stochastic methods:

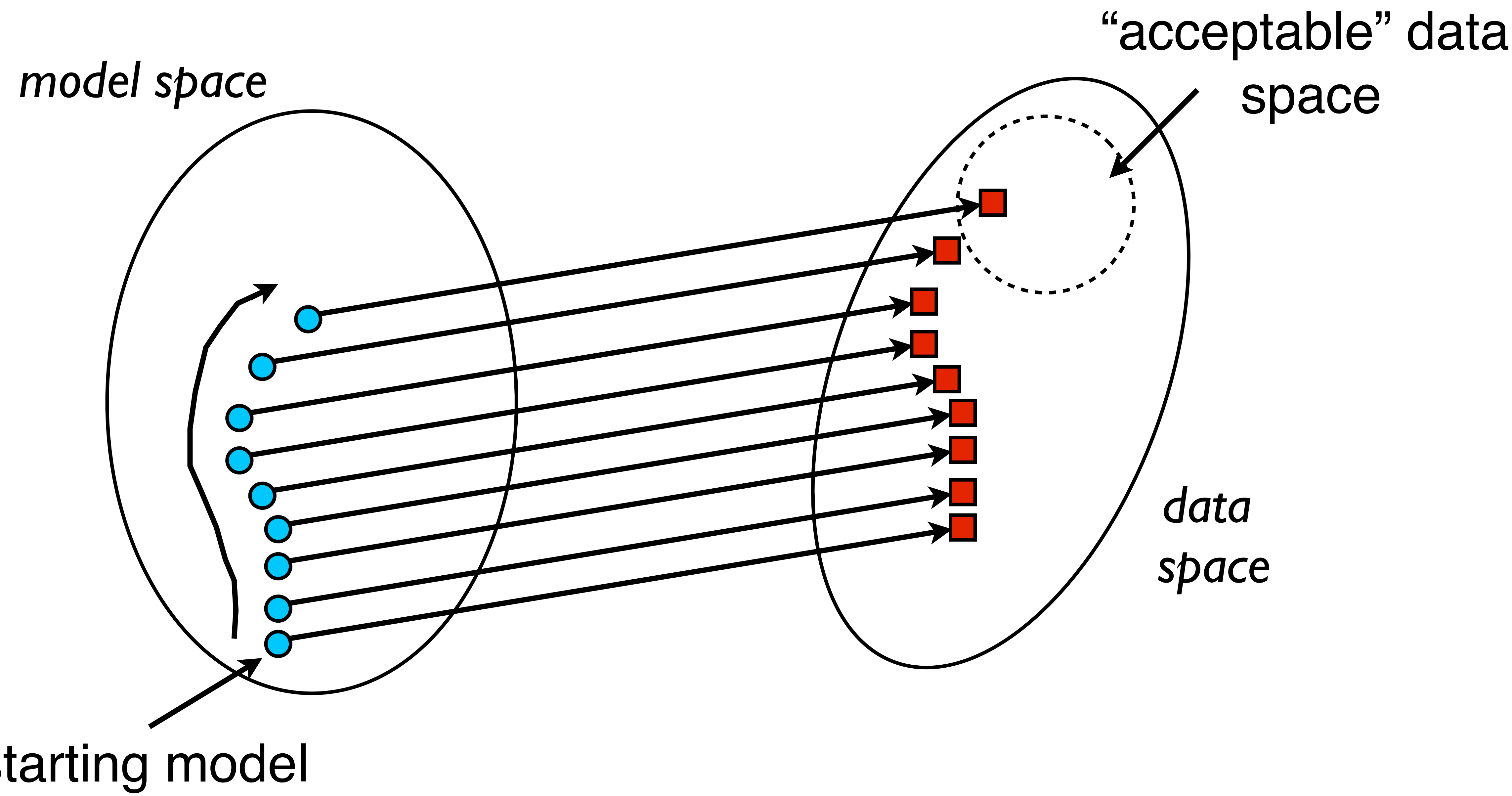


A useful approach, largely restricted to simple problems (because millions of models required), with most of the subtlety in model generation methods.

The advantages are that (i) only forward calculations are made and (ii) some statistics can be obtained on model parameters. Best for sparsely parameterized models. One needs to be careful that bounds on explored model space don't unduly influence the outcome.

Deterministic

Newton Algorithms
Steepest descent
Conjugate Gradients



The direction of the search is determined by how changing parts of the model affects the fit to the data.

Analytical

e.g. D+ (1D MT) and Bilayer (DC resistivity)

my data

and across the insulating interval $z_k < z < z_{k+1}$ we find

$$E_{k+1} = E_k + (z_{k+1} - z_k) D_k^+ \tag{52}$$

$$= E_k + (z_{k+1} - z_k) D_{k+1}^- \tag{53}$$

Define the admittance just above the k-th conductor in the usual way

$$C_k = -E_k / D_k^- . \tag{54}$$

Then by means of equations (50), (51) and (53) we can eliminate the E_k and $D_k^{\pm}$ as we did for uniform layers (although C_k is not continuous):

$$C_k = \frac{E_k}{-D_k^-} = \frac{E_k}{\mathrm{i}\omega\mu_0\tau_k E_k - D_k^+} = \frac{1}{\mathrm{i}\omega\mu_0\tau_k - D_k^+/E_k} \tag{55}$$

$$= \frac{1}{\mathrm{i}\omega\mu_0\tau_k - D_{k+1}^-/E_k} = \frac{1}{\mathrm{i}\omega\mu_0\tau_k - \dfrac{D_{k+1}^-}{E_{k+1} - (z_{k+1} - z_k) D_{k+1}^-}} . \tag{56}$$

Finally, dividing by D_{k+1}^- in the bottom tier we find the connection between the admittance at one level to the one above:

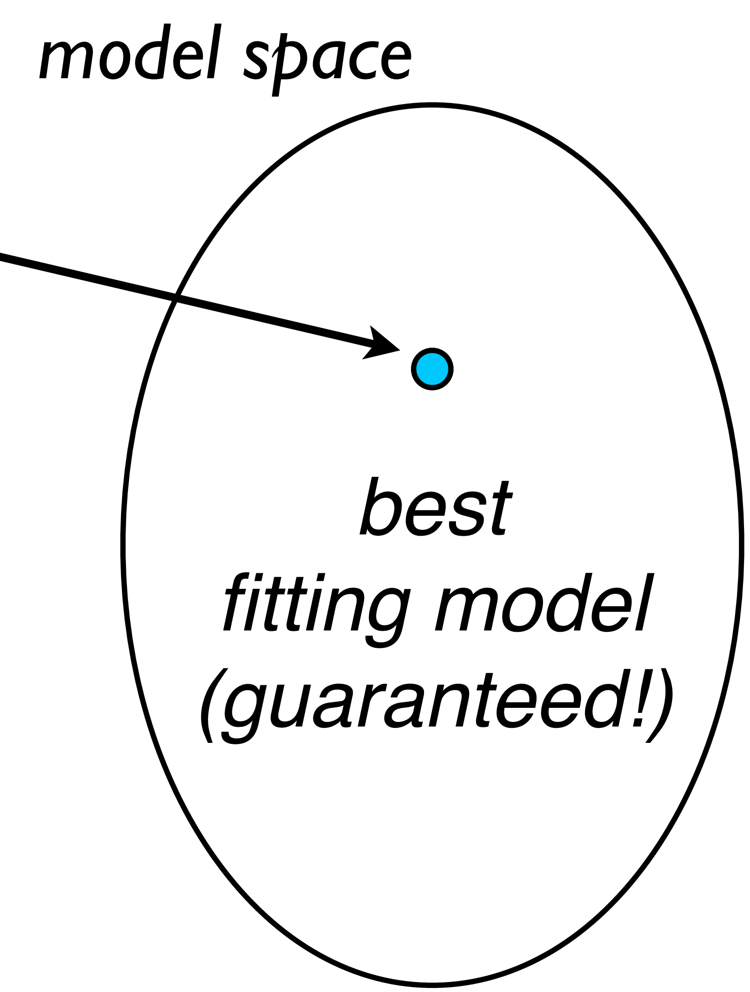
$$C_k = \frac{1}{\mathrm{i}\omega\mu_0\tau_k + \dfrac{1}{z_{k+1} - z_k + C_{k+1}}} . \tag{57}$$

We could solve (48) by recurring upwards in the familiar way, starting with $E(H) = 0 = C_{K+1}$, to get the value of $E(0)$ and hence of $C_1 = c(\omega)$. But now we do something different: we substitute repeatedly from the top, and we get a magnificent **continued fraction** for the admittance:

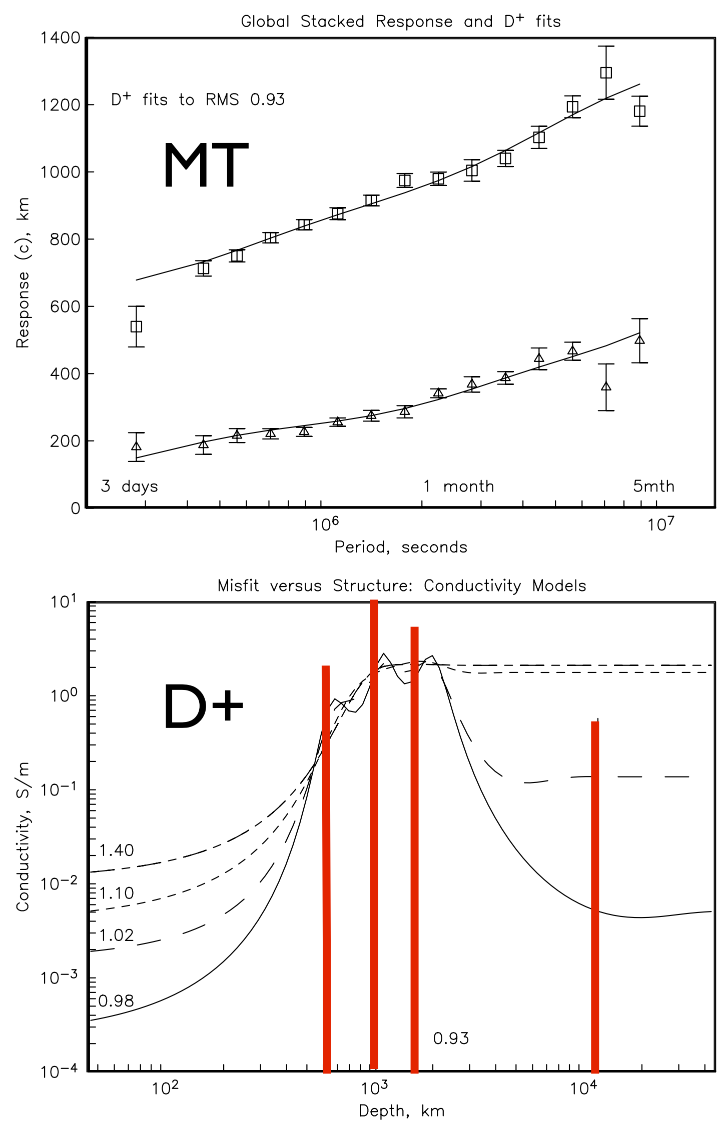
$$c(\omega) = z_1 + \cfrac{1}{\mathrm{i}\omega\mu_0\tau_1 + \cfrac{1}{z_2 - z_1 + \cfrac{1}{\mathrm{i}\omega\mu_0\tau_2 + \cfrac{1}{z_3 - z_2 + \cfrac{1}{\mathrm{i}\omega\mu_0\tau_3 + \cdots \cfrac{1}{H - z_K}}}}}} . \tag{58}$$

The initial z_1 allows us to put an insulator at $z = 0$, rather than a conducting sheet at the surface. While not exactly the same as the continued fractions described in the introduction, (58) can be rearranged by similar elementary algebra to be a *finite* partial fraction expansion:

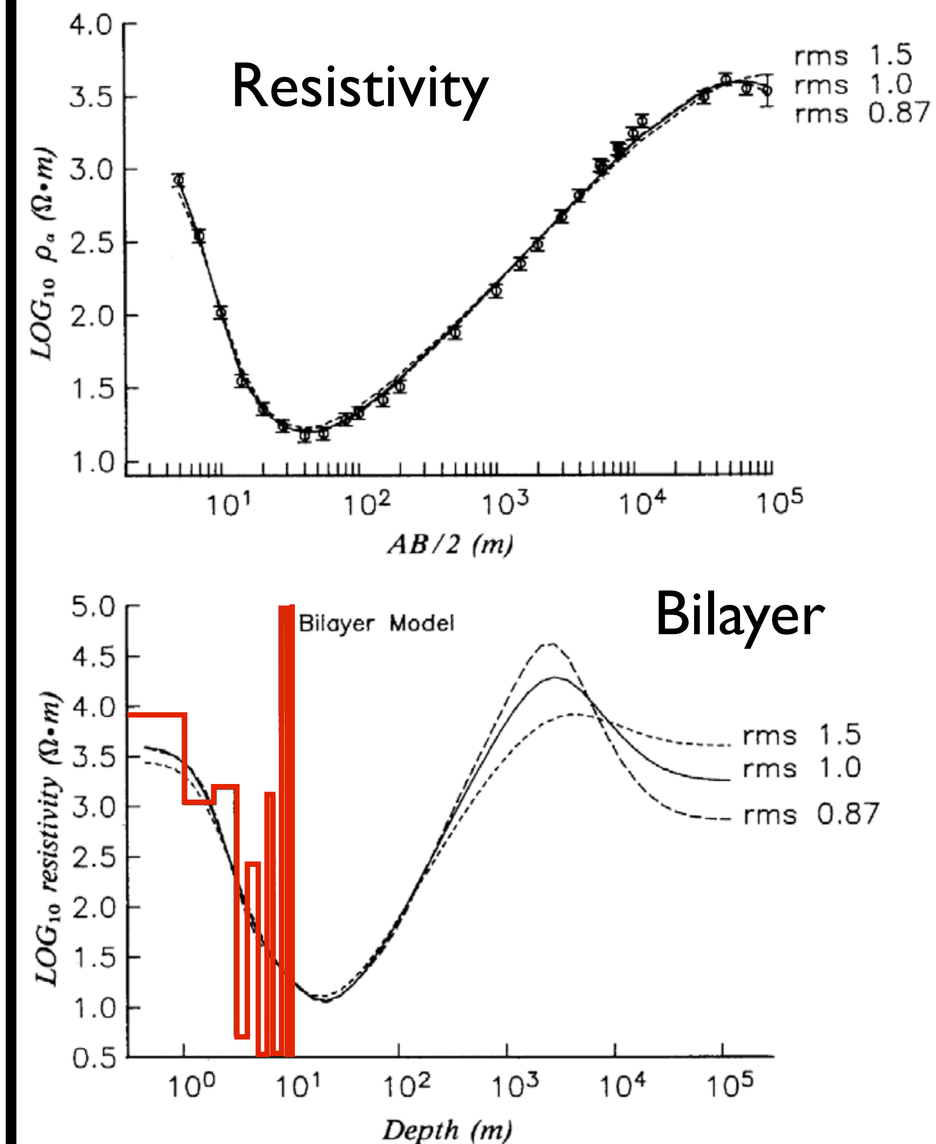
$$c(\omega) = z_1 + \sum_{k=1}^{K} \frac{\alpha_k}{\lambda_k + \mathrm{i}\omega} . \tag{59}$$



These solutions are guaranteed best fitting but pathological.



MT: Delta functions of σ

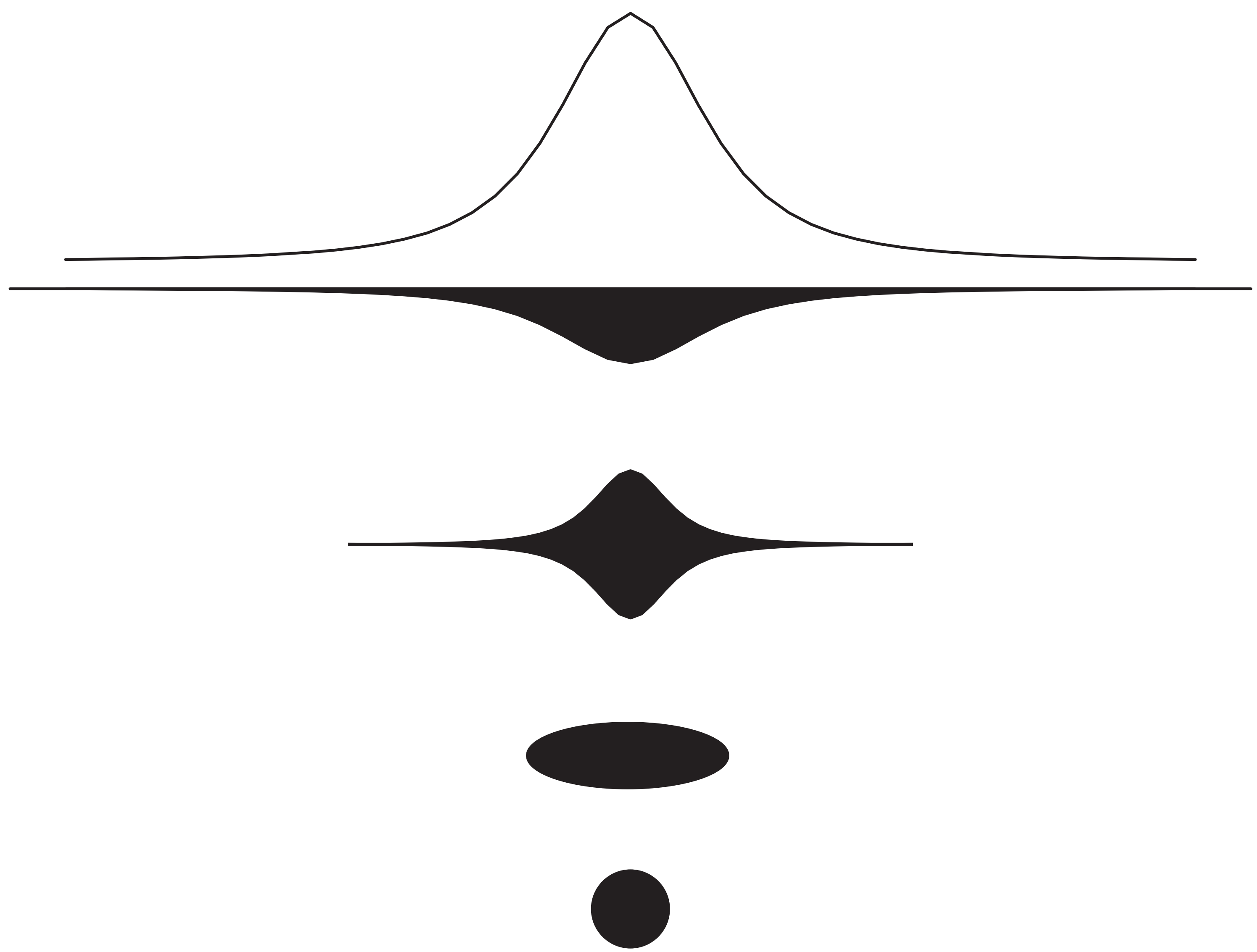


Resistivity: Arbitrarily thin surface layers

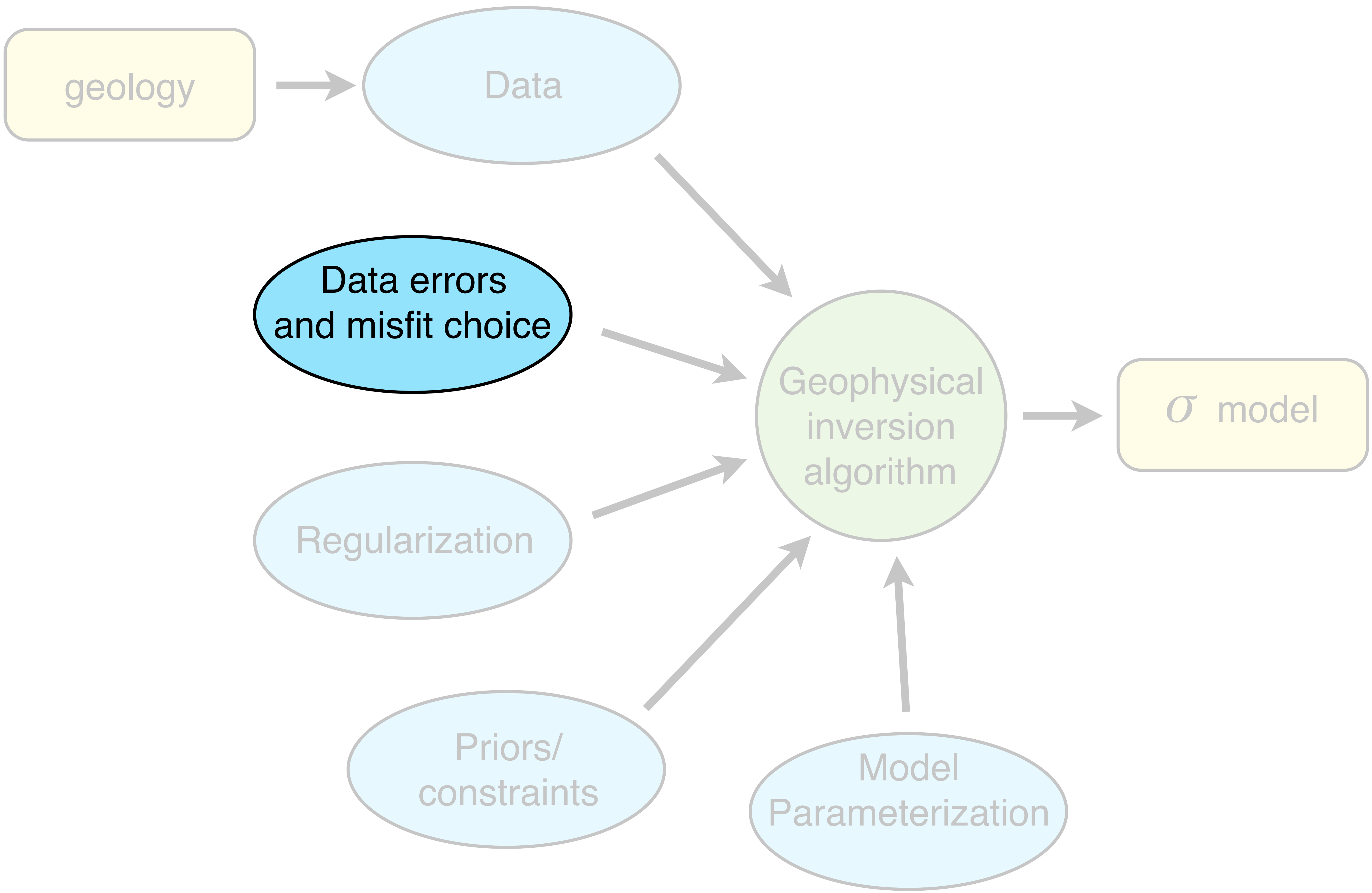
We don’t know for sure, but least squares (LS) fits to higher dimensional models are probably also pathological.

But we are pretty sure that true LS solutions are maximally “rough”.

What we have talked about so far is **model construction**. For a great many geophysicists this is what they think of when inversion is mentioned. More rigorous approaches try to obtain bounds on model properties - something that is true of all models. The classic example is total mass from gravity:



So what constitutes an “adequate” fit to the data?



For noisy data (read: **all** data), we need a measure of how well a given model fits. Sum of squares is the venerable way:

$$\chi^2 = \sum_{i=1}^{M} \frac{1}{\sigma_i^2} \left[d_i - f(x_i, \mathbf{m}) \right]^2$$

or

$$\chi^2 = ||\mathbf{W}\mathbf{d} - \mathbf{W}\hat{\mathbf{d}}||^2$$

where $\mathbf{W}$ is a diagonal of reciprocal data errors

$$\mathbf{W} = \mathrm{diag}(1/\sigma_1, 1/\sigma_2, ..., 1/\sigma_M) \quad .$$

For noisy data (read: **all** data), we need a measure of how well a given model fits. Sum of squares is the venerable way:

$$\chi^2 = \sum_{i=1}^{M} \frac{1}{\sigma_i^2} \left[d_i - f(x_i, \mathbf{m}) \right]^2$$

or

$$\chi^2 = ||\mathbf{Wd} - \mathbf{W}\hat{\mathbf{d}}||^2$$

where $\mathbf{W}$ is a diagonal of reciprocal data errors

$$\mathbf{W} = \text{diag}(1/\sigma_1, 1/\sigma_2, ..., 1/\sigma_M) \quad .$$

I like to remove the dependence on data number and use RMS:

$$\text{RMS} = \sqrt{\chi^2/M} \quad .$$

The instinctive approach as this point is to minimize χ^2.

This is least squares.

For noisy data (read: **all** data), we need a measure of how well a given model fits. Sum of squares is the venerable way:

$$\chi^2 = \sum_{i=1}^{M} \frac{1}{\sigma_i^2} \left[d_i - f(x_i, \mathbf{m}) \right]^2$$

or

$$\chi^2 = ||\mathbf{W}\mathbf{d} - \mathbf{W}\hat{\mathbf{d}}||^2$$

where $\mathbf{W}$ is a diagonal of reciprocal data errors

$$\mathbf{W} = \mathrm{diag}(1/\sigma_1, 1/\sigma_2, ..., 1/\sigma_M) \quad .$$

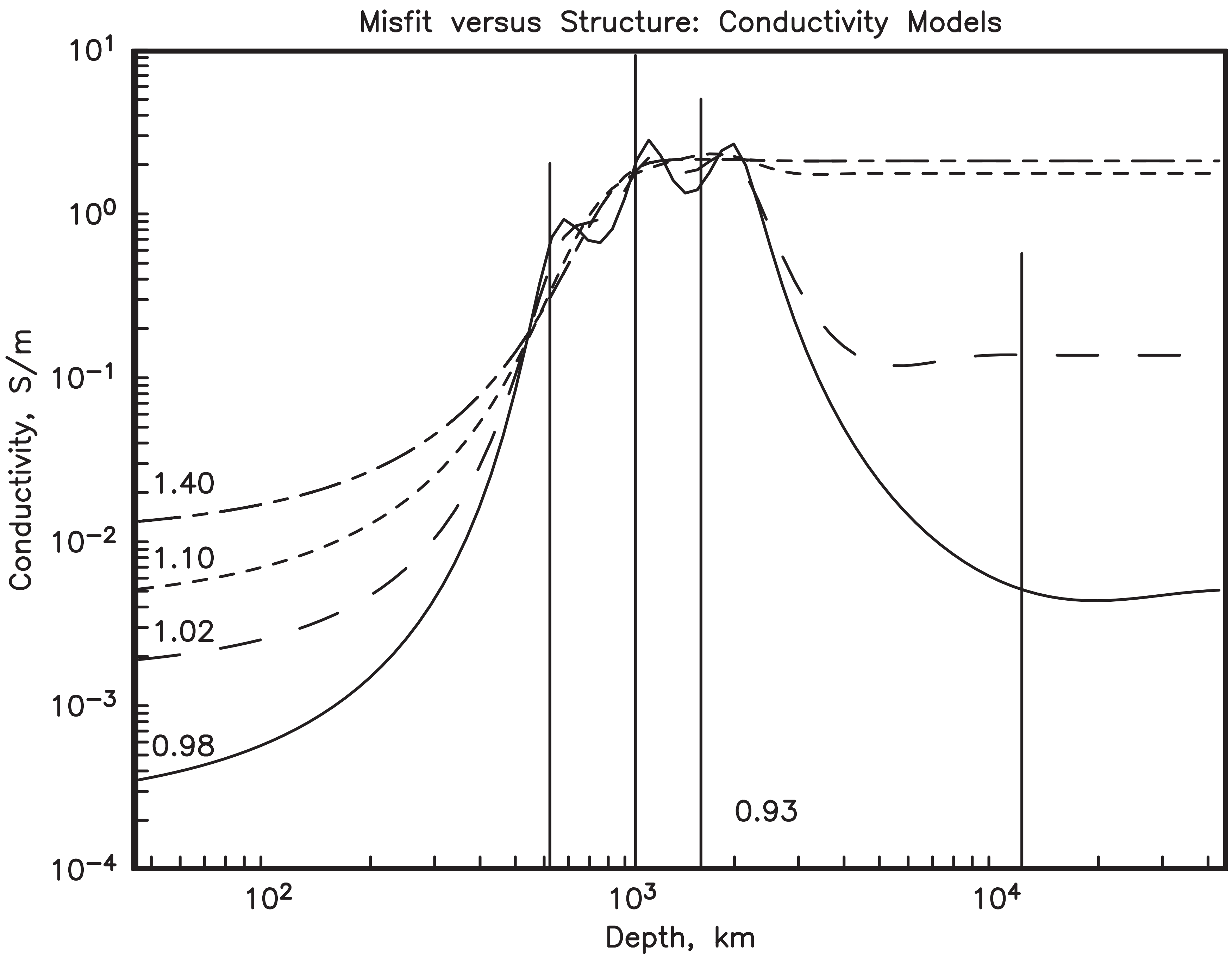
I like to remove the dependence on data number and use RMS:

$$\mathrm{RMS} = \sqrt{\chi^2/M} \quad .$$

The instinctive approach as this point is to minimize χ^2.

This is least squares. **In geophysics, this is dangerous!**

Why is this dangerous? Because as you try to approach the LS solution, your model tries to approach the maximally rough, pathological LS solutions, even if your model space does not contain delta functions, etc.



Existence and Uniqueness: Is there a solution to the inverse problem? Is there only one solution?

Finite noisy data for a linear problem (say, gravity)

An infinite number of solutions fit the data

Finite noisy data for a nonlinear problem

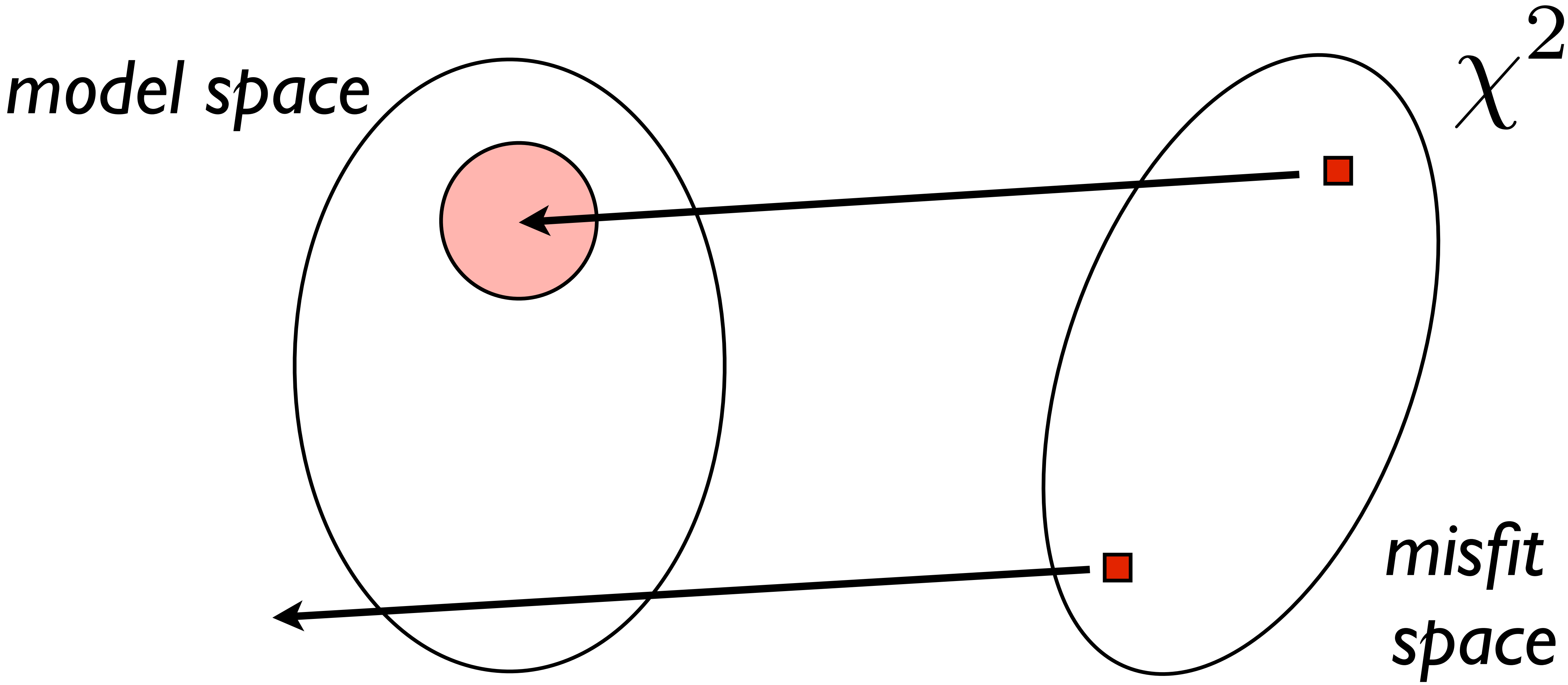
Either zero or an infinite number of solutions fit the data

Infinite exact data

A unique solution has been shown to exist for a few cases. Probably true in general but … who cares?

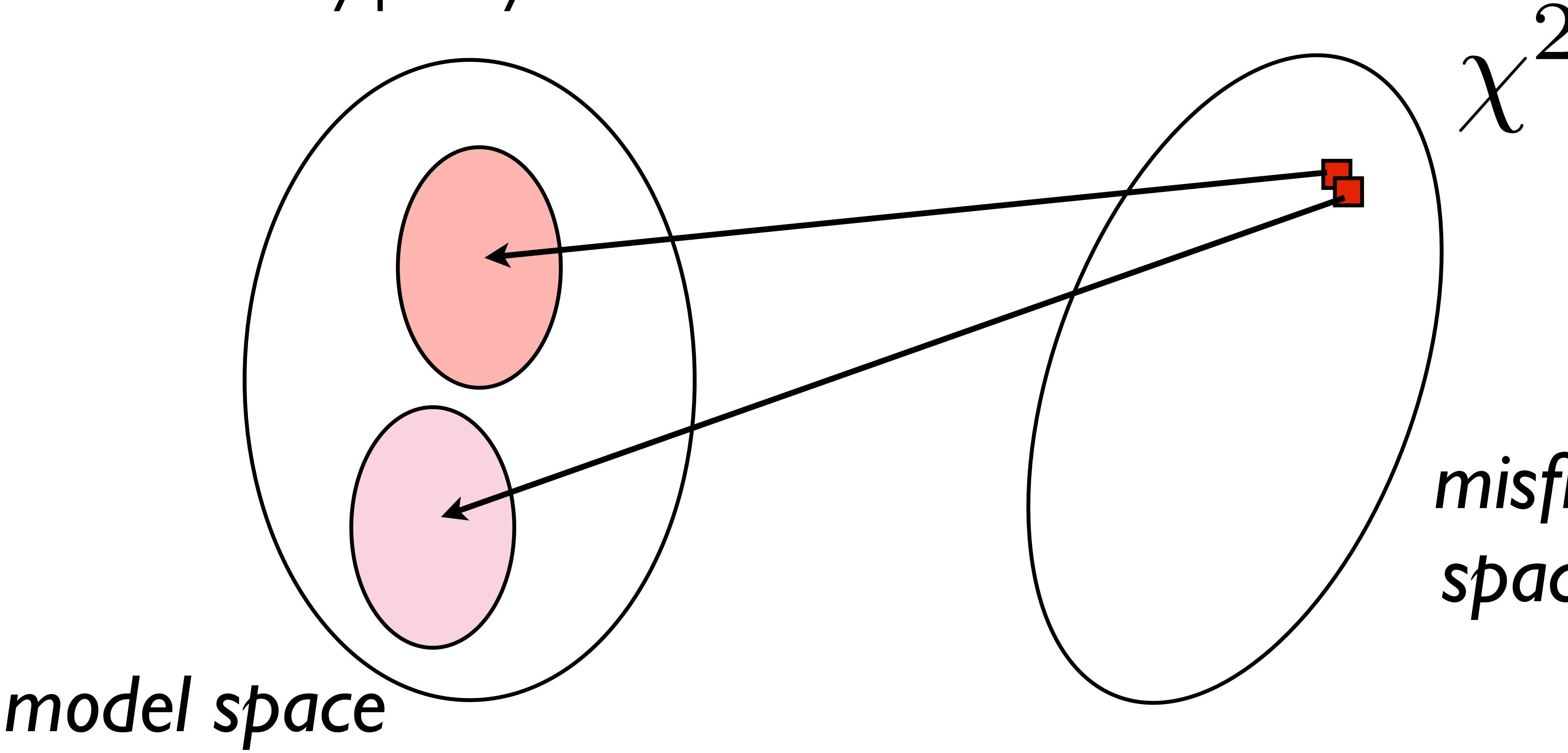
Some people think that we can approach infinite exact data with LOTS of VERY GOOD data. This is wrong. As Sven Treitel puts it, there is no such thing as being a little bit non-unique.

Geophysical inversion is non-unique:



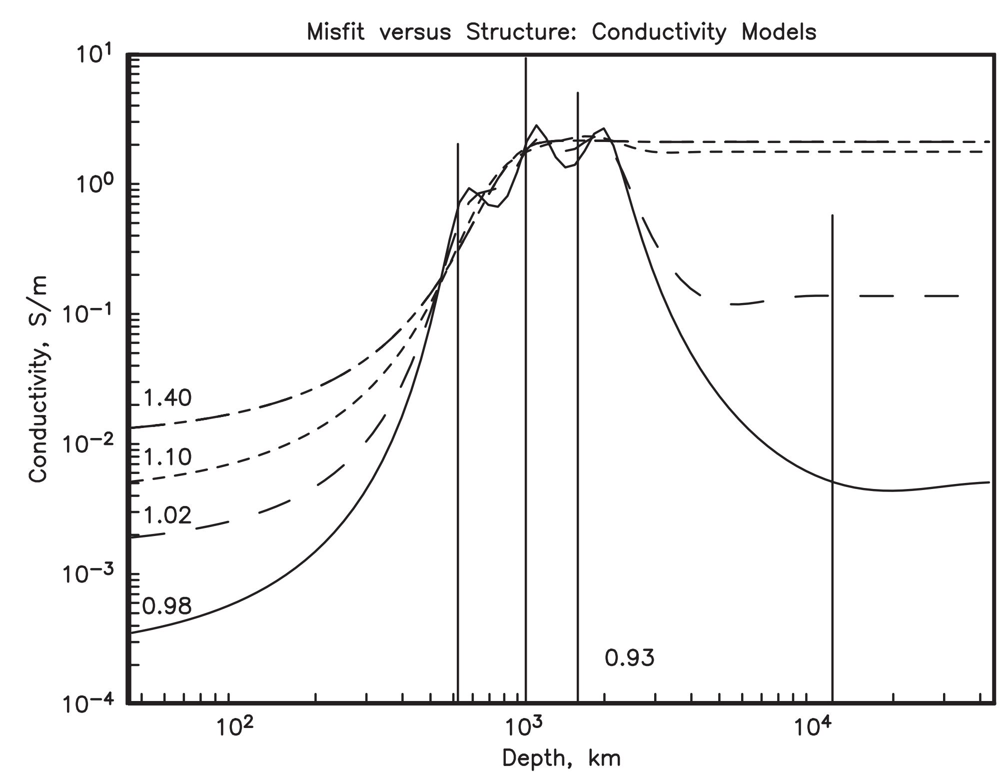
A single misfit will map into an infinite number of models (or none at all!).

It is also usually poorly constrained:



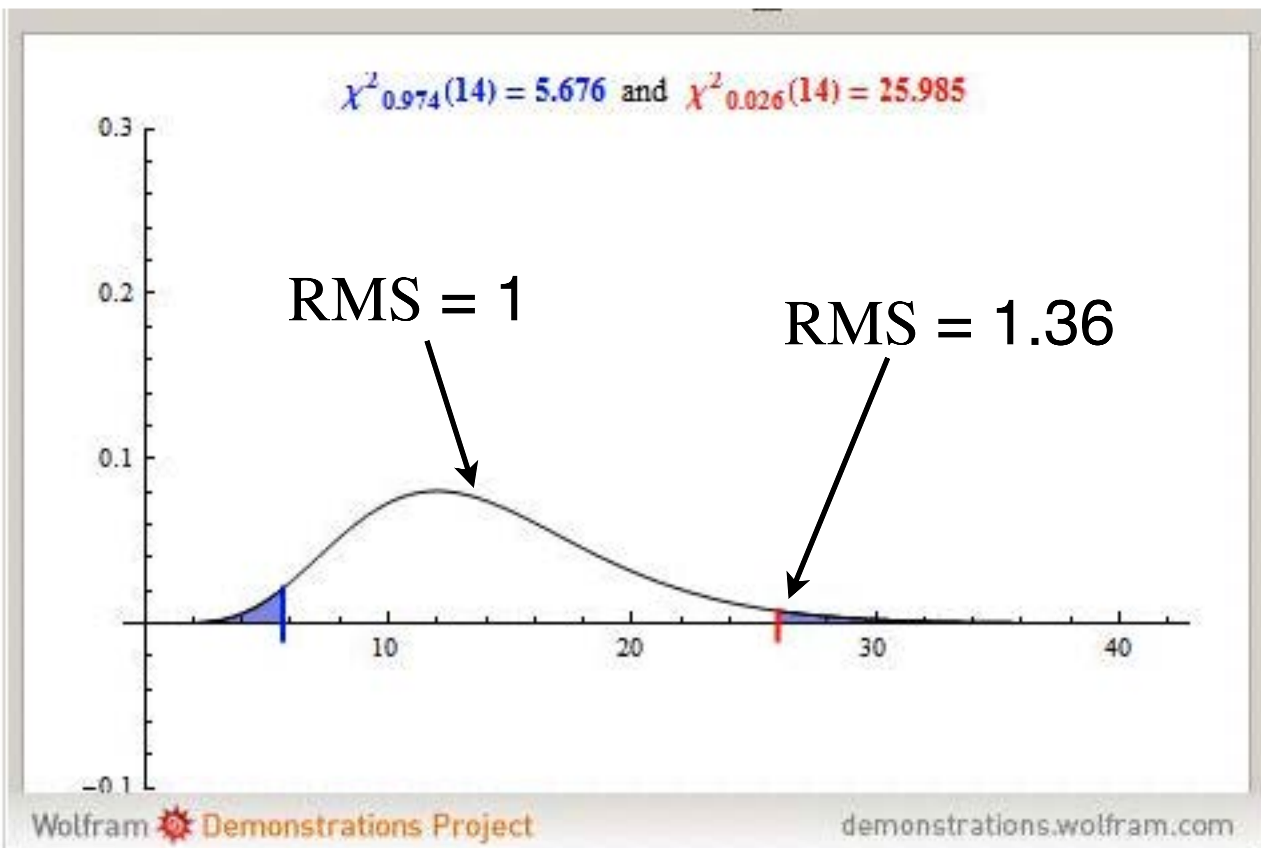
A small distance in χ^2 corresponds to a large distance in $\mathbf{m}$

(And don't forget: the minimum χ^2 is likely outside your model parameterization).



So what constitutes an adequate misfit?

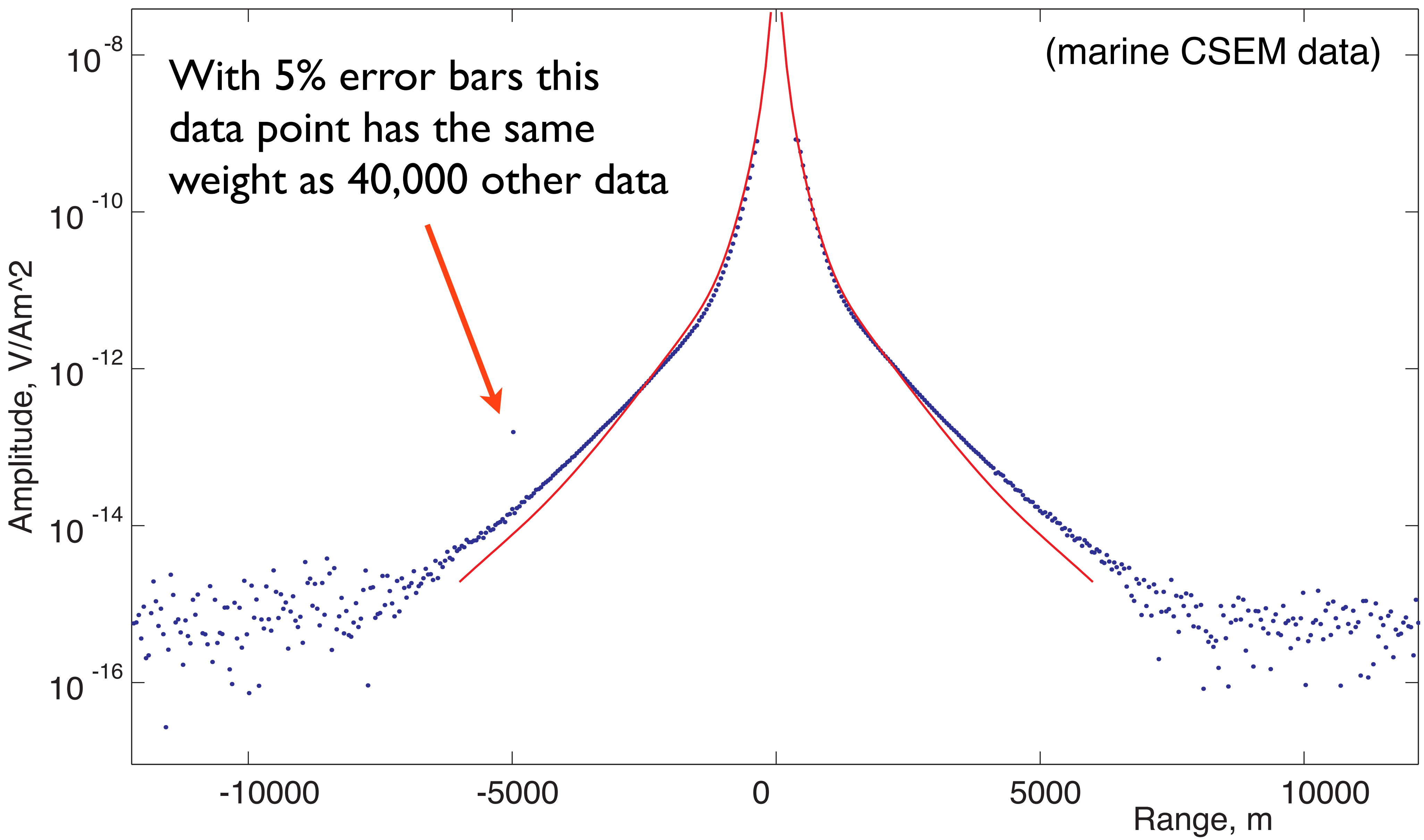
For zero-mean, Gaussian, independent errors, χ^2 is chi-squared distributed with M degrees of freedom. The expectation value is just M, which corresponds to RMS=1, and so this could be a reasonable target misfit. Or, one could look up the 95% (or other) confidence interval for chi-squared M.



χ^2 for 14 data. For large data sets, RMS=1 and $RMS_{95\%}$ are very much the same.

We could use other measures of fit, but the quadratic measure works with the mathematics of minimization, and for Gaussian errors has nice statistical properties (unbiased, maximum likelihood, minimum variance). But...

... sum-squared misfit measures are unforgiving of outliers:



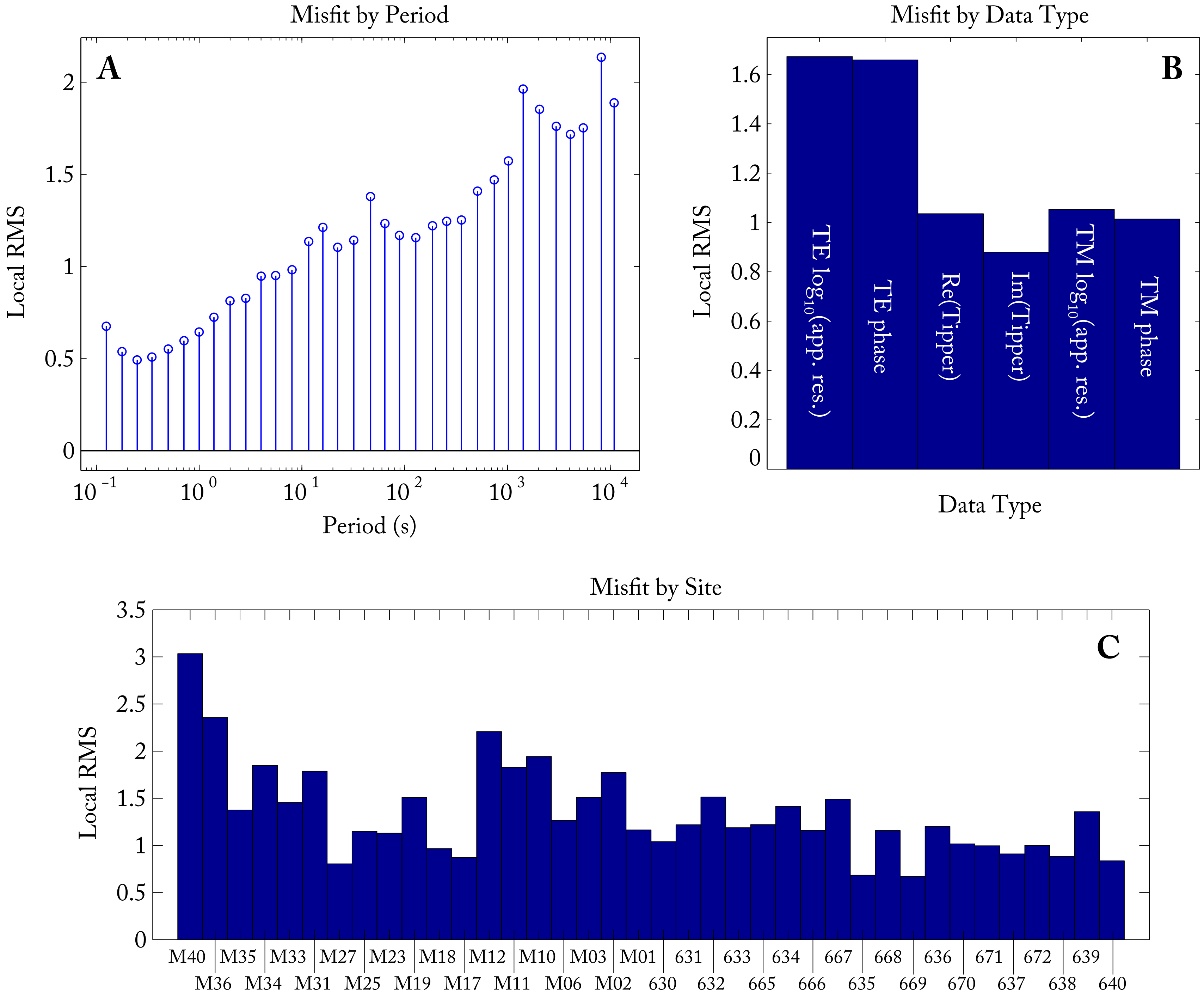
With Gaussian noise, the probability of a data point being misfit by 6 error bars is about one in a billion.

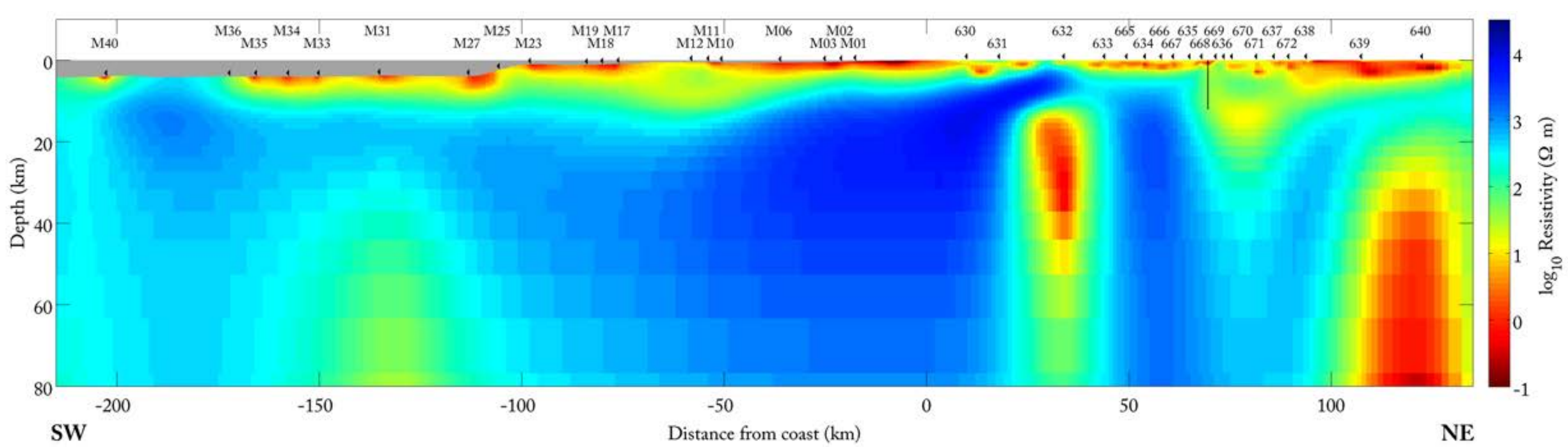
All through any inversion process you should monitor weighted residuals to ensure that there are no bad guys out there.

It is also a good idea to look at how the misfit is partitioned across the data:

Ideally it should be random, but in practice very rarely is.

Example is from MT data.





PhD thesis, Brent Wheelock, 2012.

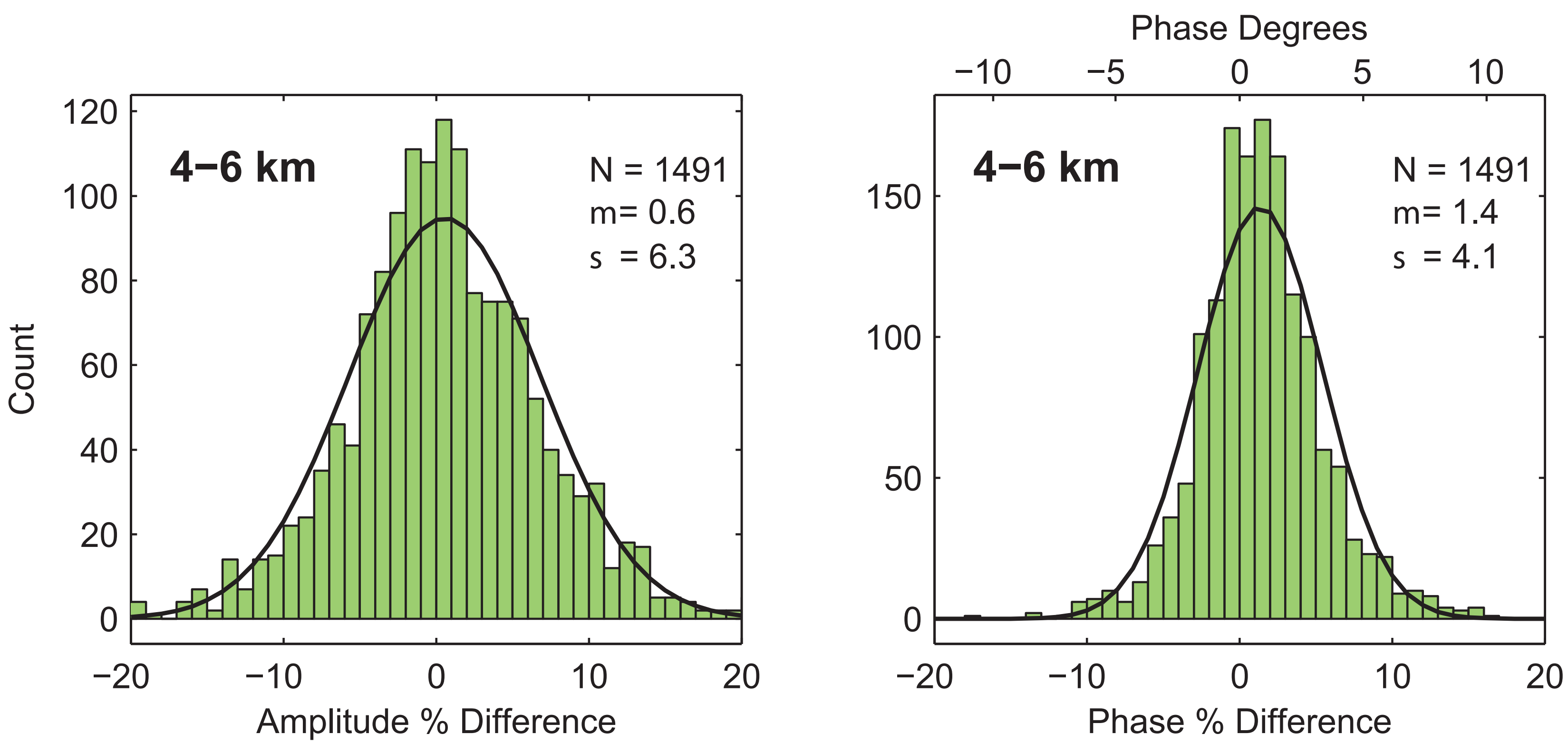
Errors come from

- statistical processing errors (spectral estimation for MT; stacking for CSEM and lots of other methods)

- systematic errors such as navigation errors and instrument calibrations, and

- “geological noise” (our inability to parameterize fine details of geology).

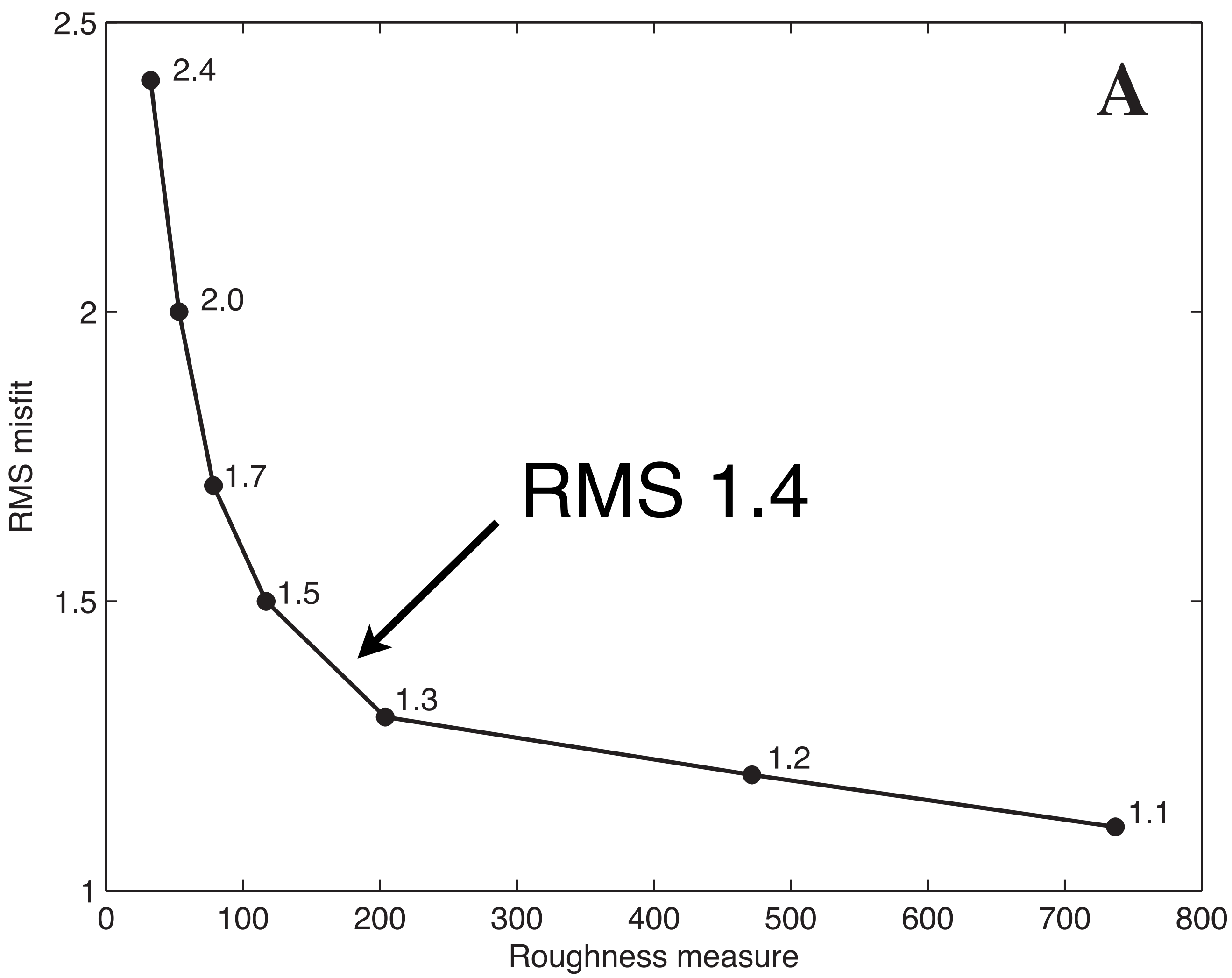
In practice, we only have a good handle on processing errors - everything else is lumped into a noise floor, which can be pretty arbitrary at times.



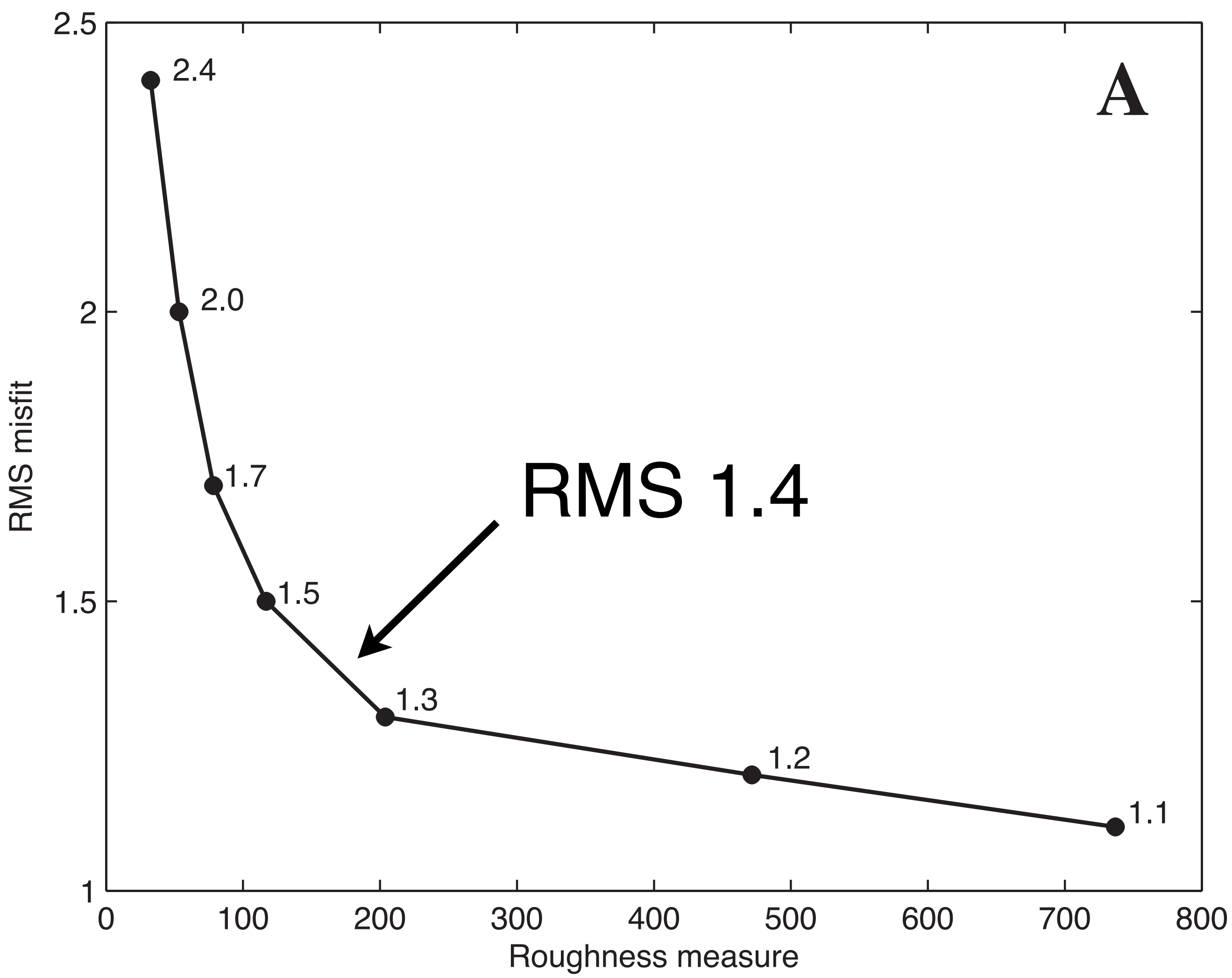
“errors” computed from actual repeat tows for marine CSEM

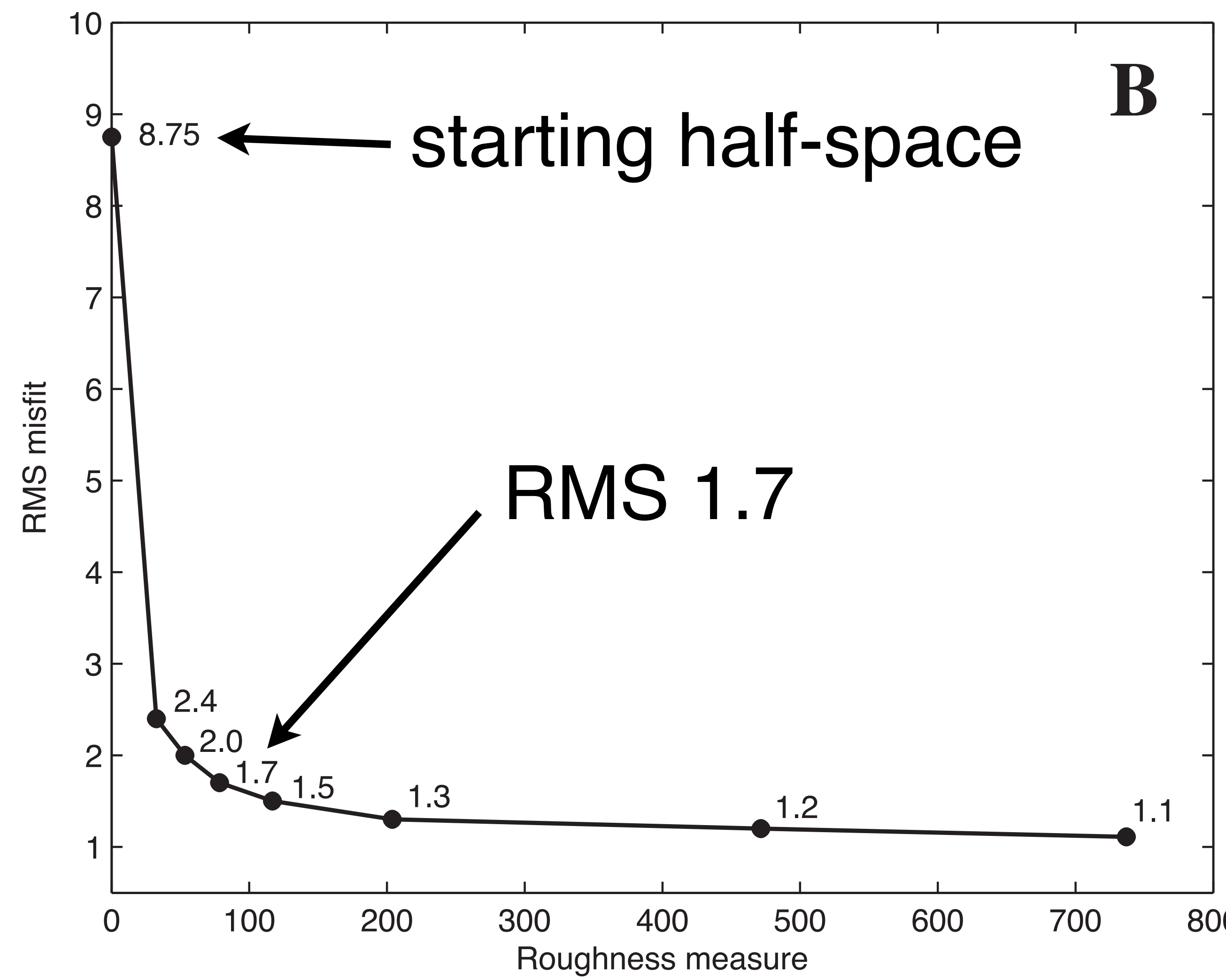
(modified from Myer et al., 2012)

So we are often left without statistical guidance and have to use judgement in determining an adequate fit. Some people like trade-off, or “L”-curves...

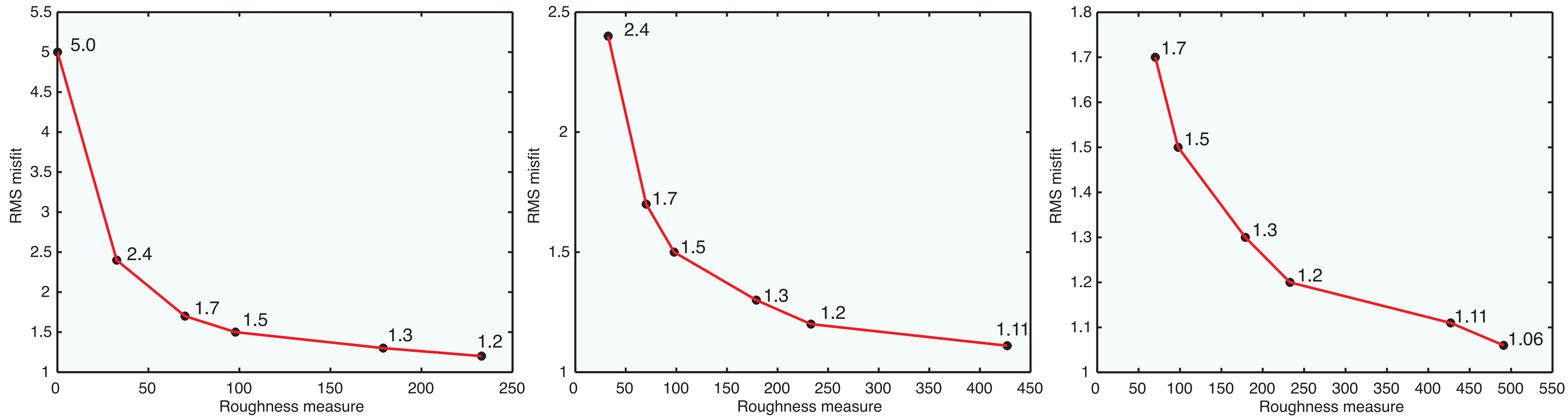


… but I am not one of them.





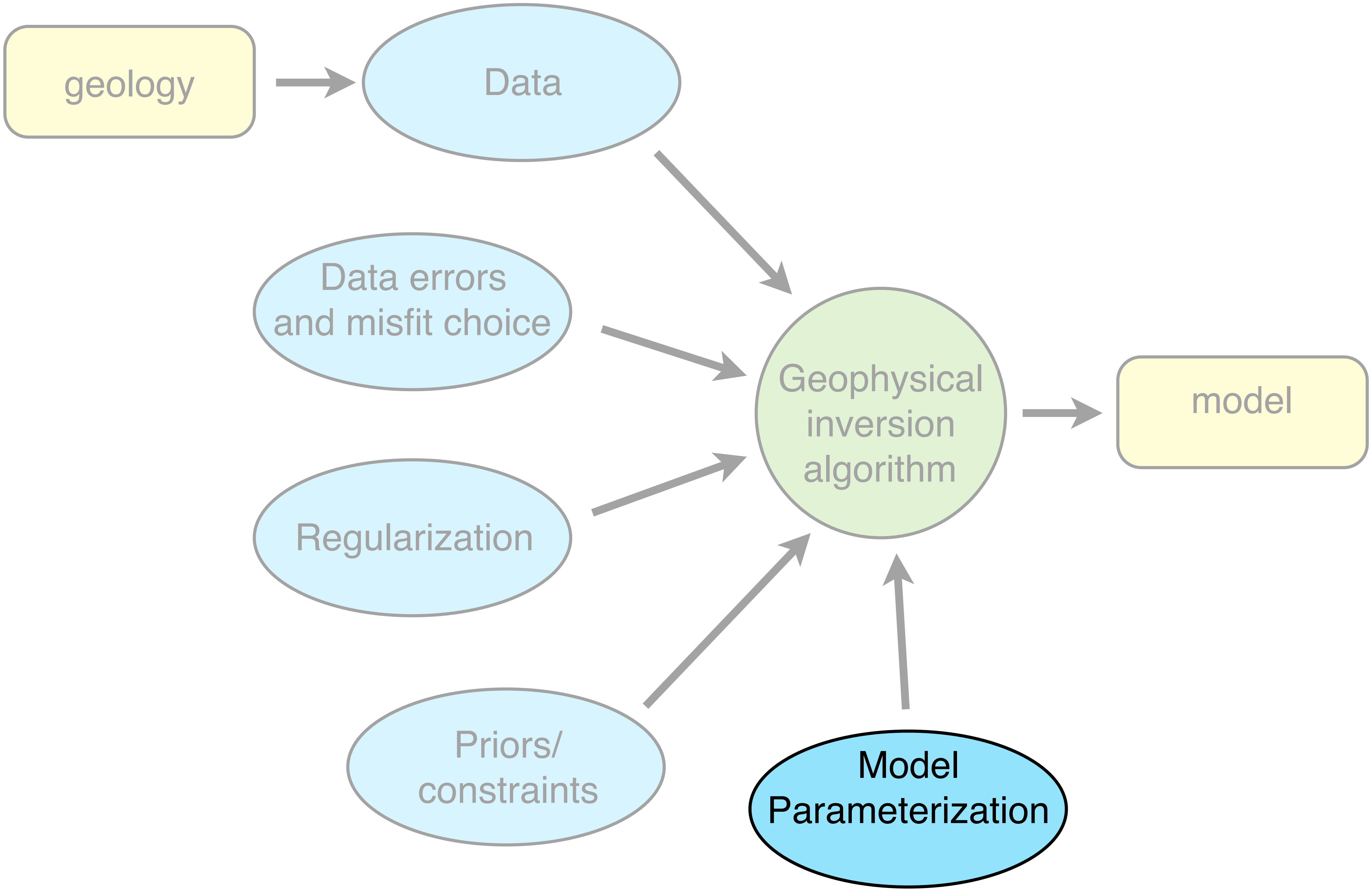
In fact you can get pretty well what you want simply by changing the range of the plot and the scaling of the axes.

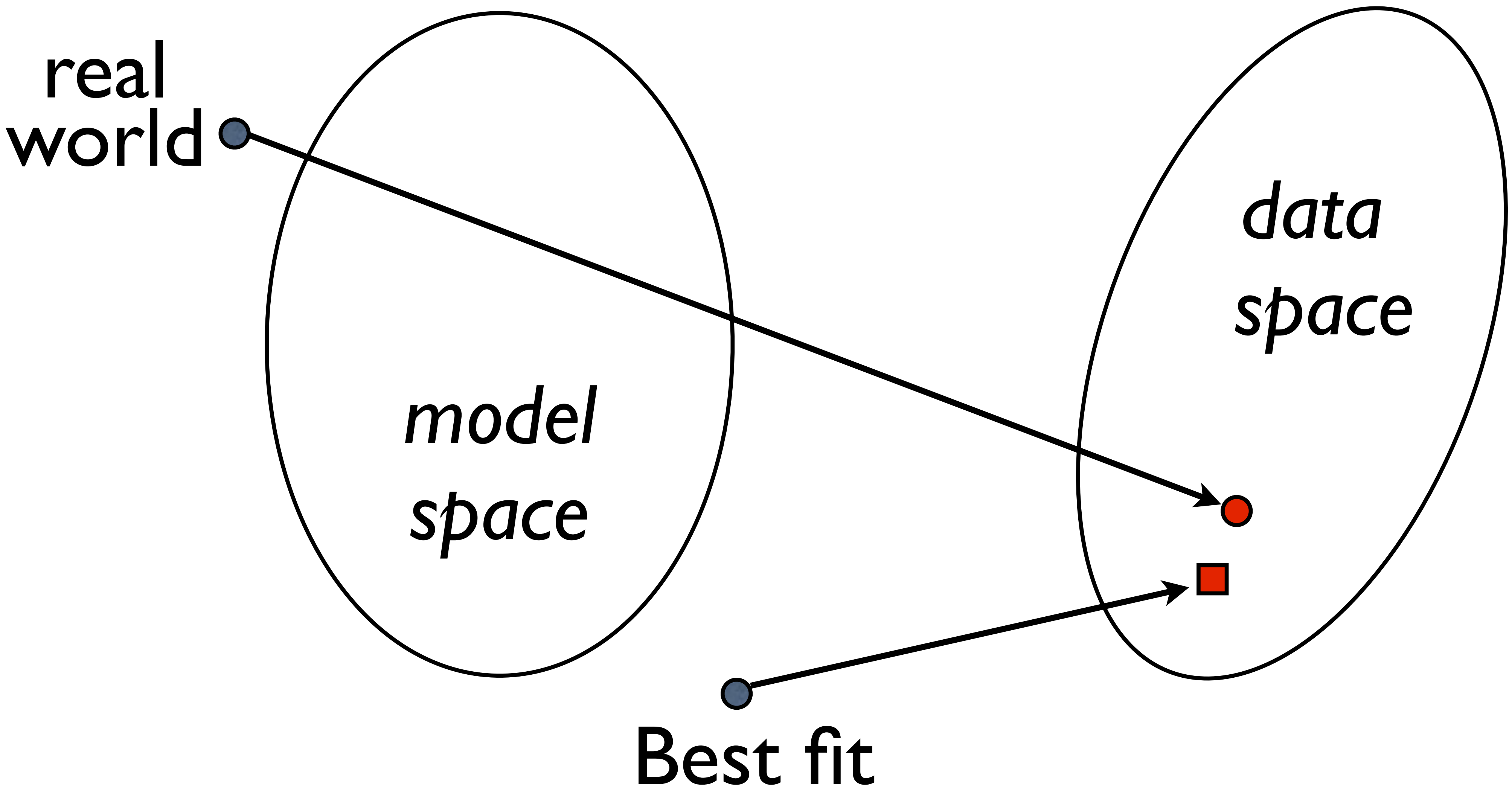


Constable, Orange, and Key, 2015

There really isn't an objective way to choose misfit level except through a good understanding of the data errors.

Model parameterization:



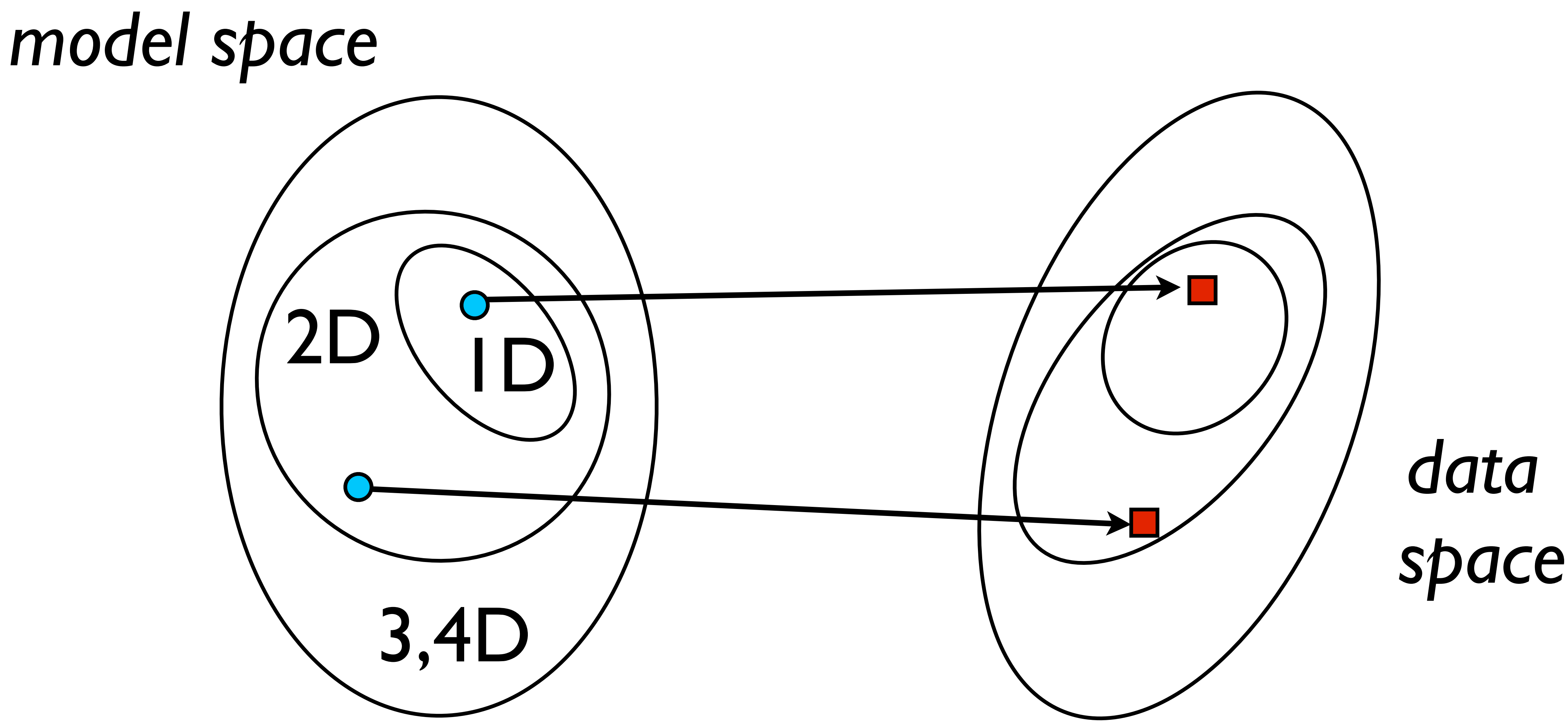


Even with your best efforts, the real world is unlikely to be captured by your model parameterization, and the best fitting model almost certainly won't be either. Understanding this can be important.

(And the best fitting model won't have the same misfit as the real world does, because of noise.)

Where in model space you are is determined by your parameterization - this also determines where in data space you can be.

In non-linear geophysical problems, even forward modeling can involve a challenging computational effort.



Sven Treitel once asked the question: "*Can our mathematics ever completely describe nature?*".

The trite answer, of course, is "*No*". However, it is more useful to understand the nature of the limitations:

Are the physics sufficient (e.g. scalar properties versus anisotropy)?

Is the forward computational machinery accurate? (e.g. finite difference calculations don't handle bathymetry well)

Is the dimensionality of model space large enough? (1D, 2D, 3D, 4D)

Is the discretization fine enough and the model size big enough?

One can rarely afford to blindly ensure these are all achieved, so intelligence and understanding must be applied, perhaps by trial and error.

Model space parameterization:

$$\hat{\mathbf{d}} = f(\mathbf{x}, \mathbf{m})$$ Some forward functional f

$$\mathbf{m} = (m_1, m_2, \ldots.., m_N)$$ Model parameters

Most geophysical properties cannot go negative, but your inversion scheme might well generate negative values in $\mathbf{m}$. The easiest way to handle this is by parameterizing as $\log(\mathbf{m})$, but there are other ways, such as NNLS.

Model space parameterization:

$$\hat{\mathbf{d}} = f(\mathbf{x}, \mathbf{m})$$

Some forward functional f

$$\mathbf{m} = (m_1, m_2, ..., m_N)$$

Model parameters

In the real world, N (model size) is infinite (even in 1D). How we proceed from here depends on whether N is small, moderately large, or infinite.

Small (sparse) parameterizations can be handled with parameterized inversions (e.g. Marquardt) or stochastic inversions. The concept of least squares fitting works because sparse models don't have the freedom to mimic the pathological true least squares solutions.

Model space parameterization:

$$\hat{\mathbf{d}} = f(\mathbf{x}, \mathbf{m})$$ Some forward functional f

$$\mathbf{m} = (m_1, m_2, ..., m_N)$$ Model parameters

In the real world, N (model size) is infinite (even in 1D). How we proceed from here depends on whether N is small, moderately large, or infinite.

Small (sparse) parameterizations can be handled with parameterized inversions (e.g. Marquardt) or stochastic inversions. The concept of least squares fitting works because sparse models don't have the freedom to mimic the pathological true least squares solutions.

Infinite N requires a real inverse theory mathematician. I am not one of them.

Model space parameterization:

$$\hat{\mathbf{d}} = f(\mathbf{x}, \mathbf{m})$$

Some forward functional f

$$\mathbf{m} = (m_1, m_2, \ldots\ldots, m_N)$$

Model parameters

In the real world, N (model size) is infinite (even in 1D). How we proceed from here depends on whether N is small, moderately large, or infinite.

Small (sparse) parameterizations can be handled with parameterized inversions (e.g. Marquardt) or stochastic inversions. The concept of least squares fitting works because sparse models don't have the freedom to mimic the pathological true least squares solutions.

Infinite N requires a real inverse theory mathematician. I am not one of them.

Most of the time geophysicists are working with moderately large N. Also, many geophysical problems are non-linear, so we will concentrate on that approach.

To invert non-linear forward problems we often linearize around a starting model:

$$\hat{\mathbf{d}} = f(\mathbf{m}_1) = f(\mathbf{m}_0 + \Delta\mathbf{m}) \approx f(\mathbf{m}_0) + \mathbf{J}\Delta\mathbf{m}$$

using a matrix of derivatives

$$J_{ij} = \frac{\partial f(x_i, \mathbf{m}_0)}{\partial m_j}$$

and a model perturbation

$$\Delta\mathbf{m} = \mathbf{m}_1 - \mathbf{m}_0 = (\delta m_1, \delta m_2, ..., \delta m_N)$$

Now our expression for χ^2 is

$$\chi^2 \approx ||\mathbf{W}\mathbf{d} - \mathbf{W}f(\mathbf{m}_0) + \mathbf{W}\mathbf{J}\Delta\mathbf{m}||^2$$

For a least squares solution we solve in the usual way by differentiating and setting to zero to get a linear system:

$$\beta = \alpha \Delta \mathbf{m}$$

where

$$\beta = (\mathbf{WJ})^T \mathbf{W}(\mathbf{d} - f(\mathbf{m_0}))$$

$$\alpha = (\mathbf{WJ})^T \mathbf{WJ} \quad .$$

So, given a starting model $\mathbf{m}_0$ we can find an update $\Delta \mathbf{m}$:

$$\Delta \mathbf{m} = \alpha^{-1} \beta$$

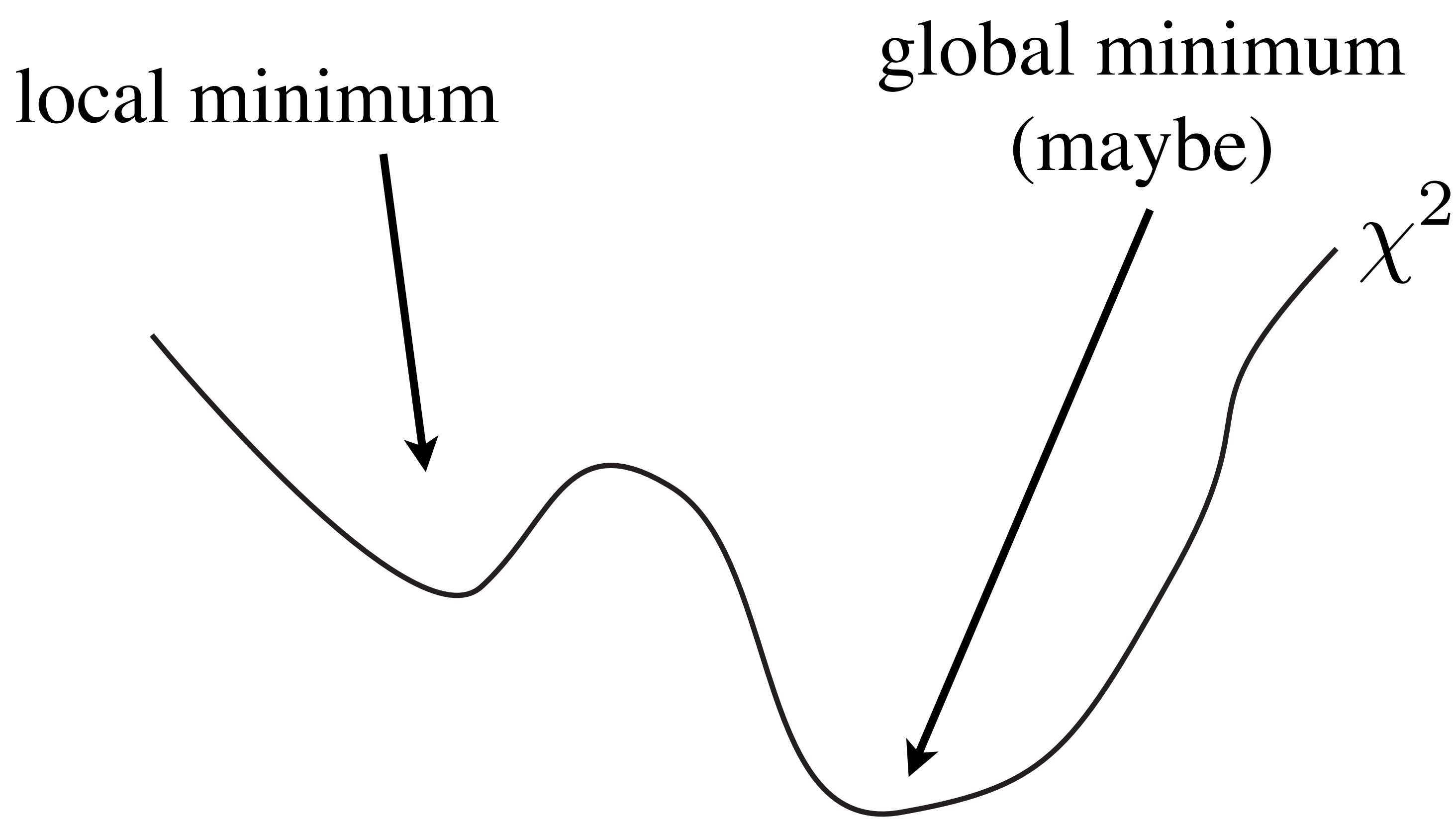
and iterate until we converge. (This is Gauss-Newton.)

Global versus local minima:

For nonlinear problems, there are no guarantees that Gauss-Newton will converge.

There are no guarantees that if it does converge the solution is a global one.

The solution might well depend on the starting model.

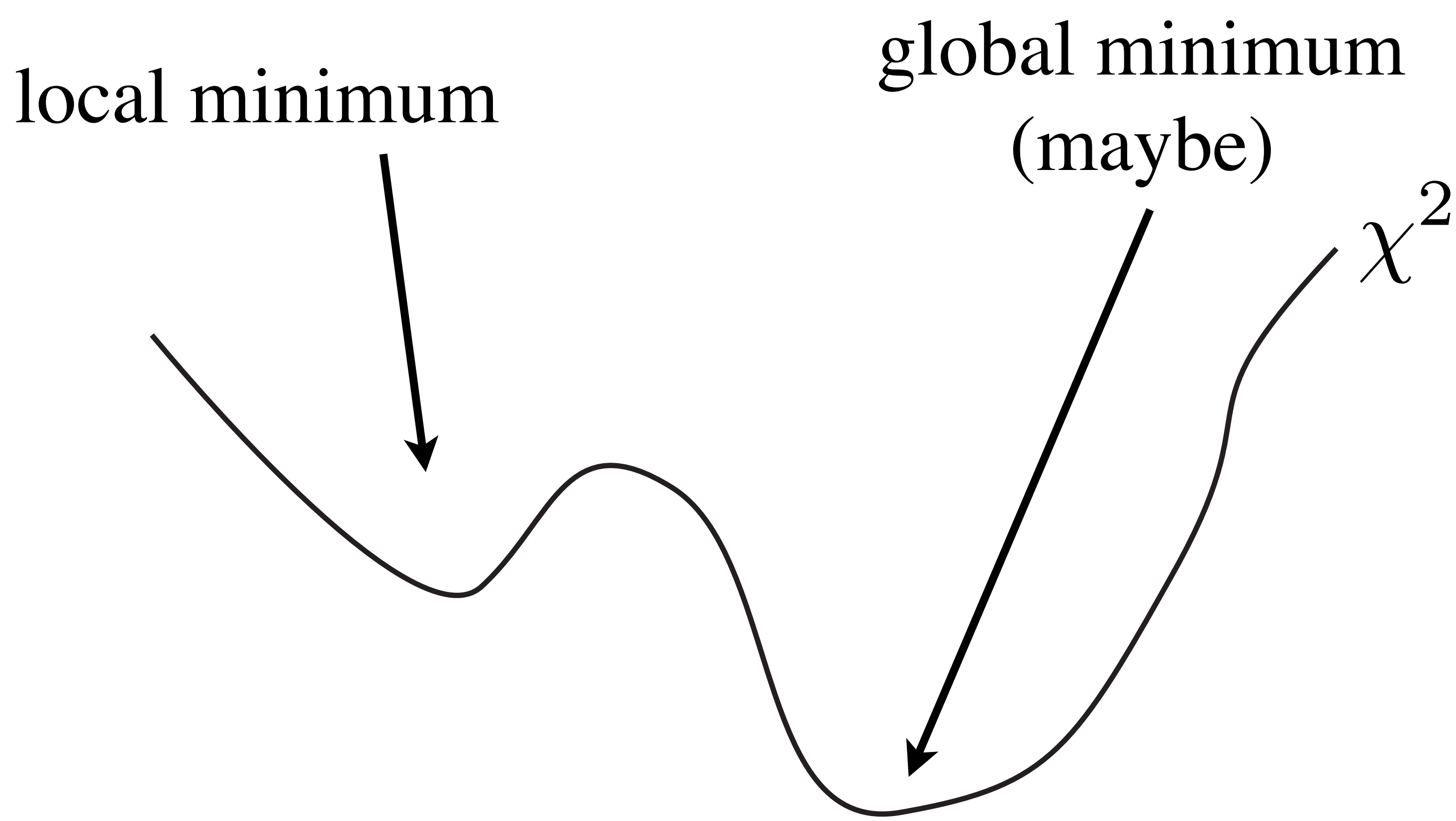


Global versus local minima:

For nonlinear problems, there are no guarantees that Gauss-Newton will converge.

There are no guarantees that if it does converge the solution is a global one.

The solution might well depend on the starting model.



Gauss-Newton only works for small N (it isn't even defined for $N > M$). If N gets too large then the solutions become unstable, oscillatory, and generally useless (they are probably trying to converge to D+ type solutions).

Almost all inversion today incorporates some type of regularization, which minimizes some aspect of the model as well as fit to data:

$$U = \left(||\mathbf{Wd} - \mathbf{W}f(\mathbf{m})||^2\right) + \mu||\mathbf{Rm}||^2$$

where $\mathbf{Rm}$ is some measure of the model and μ is a trade-off parameter or Lagrange multiplier. In 1D a typical $\mathbf{R}$ might be:

$$\mathbf{R}_1 = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & -1 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & -1 & 1 & 0 & \dots & 0 \\ & & \ddots & & & \ddots & \\ & & & & & -1 & 1 \end{pmatrix}$$

m_1	-1							
m_2	+1	-1						
m_3		+1	-1					
m_4			+1	-1				
m_5				+1	-1			
m_6					+1	-1		
m_7						+1	-1	
m_8							+1	

which extracts a measure of slope. This stabilizes the inversion, creates a single solution, allows $N > M$, and manufactures models with useful properties.

This is easily extended to 2D and 3D modeling.

The trade-off between roughness and misfit:

$$U = \left(||\mathbf{Wd} - \mathbf{W}f(\mathbf{m})||^2\right) + \mu||\mathbf{Rm}||^2$$

When μ is small, model roughness is ignored and we try to fit the data. When μ is large, we smooth the model at the expense of data fit.

The trade-off between roughness and misfit:

$$U = \left(||\mathbf{Wd} - \mathbf{W}f(\mathbf{m})||^2\right) + \mu||\mathbf{Rm}||^2$$

When μ is small, model roughness is ignored and we try to fit the data. When μ is large, we smooth the model at the expense of data fit.

One approach is to choose μ and minimize U by least squares, but picking μ *a priori* is simply choosing how rough your model is.

The trade-off between roughness and misfit:

$$U = \left(||\mathbf{Wd} - \mathbf{W}f(\mathbf{m})||^2\right) + \mu||\mathbf{Rm}||^2$$

When μ is small, model roughness is ignored and we try to fit the data. When μ is large, we smooth the model at the expense of data fit.

One approach is to choose μ and minimize U by least squares, but picking μ *a priori* is simply choosing how rough your model is.

We ought to have a decent idea of how well our data can be fit. This forms the basis of the “Occam” approach, where a target data misfit χ_*^2 is chosen:

$$U = \left(||\mathbf{Wd} - \mathbf{W}f(\mathbf{m})||^2 - \chi_*^2\right) + \mu||\mathbf{Rm}||^2$$

The trade-off between roughness and misfit:

$$U = \left(||\mathbf{Wd} - \mathbf{W}f(\mathbf{m})||^2\right) + \mu||\mathbf{Rm}||^2$$

When μ is small, model roughness is ignored and we try to fit the data. When μ is large, we smooth the model at the expense of data fit.

One approach is to choose μ and minimize U by least squares, but picking μ *a priori* is simply choosing how rough your model is.

We ought to have a decent idea of how well our data can be fit. This forms the basis of the "Occam" approach, where a target data misfit χ_*^2 is chosen:

$$U = \left(||\mathbf{Wd} - \mathbf{W}f(\mathbf{m})||^2 - \chi_*^2\right) + \mu||\mathbf{Rm}||^2$$

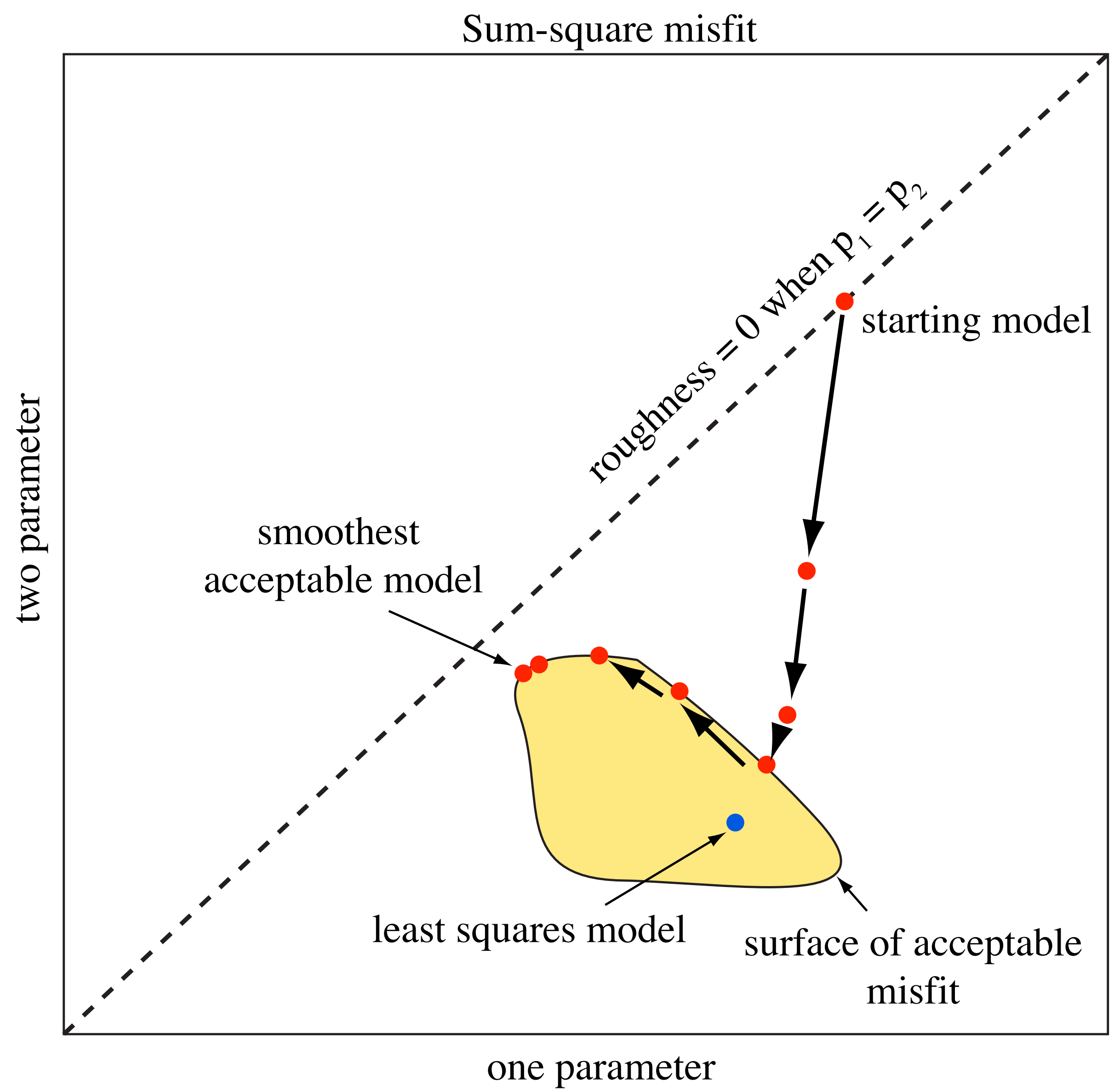
For linearized, iterative inversion we use

$$U = ||\mathbf{Rm}_1||^2 + \mu^{-1}\left(||\mathbf{Wd} - \mathbf{W}\big(f(\mathbf{m}_0) + \mathbf{J}(\mathbf{m}_1 - \mathbf{m}_0)\big)||^2 - \chi_*^2\right)$$

After differentiation and setting to zero we get an expression for a new model:

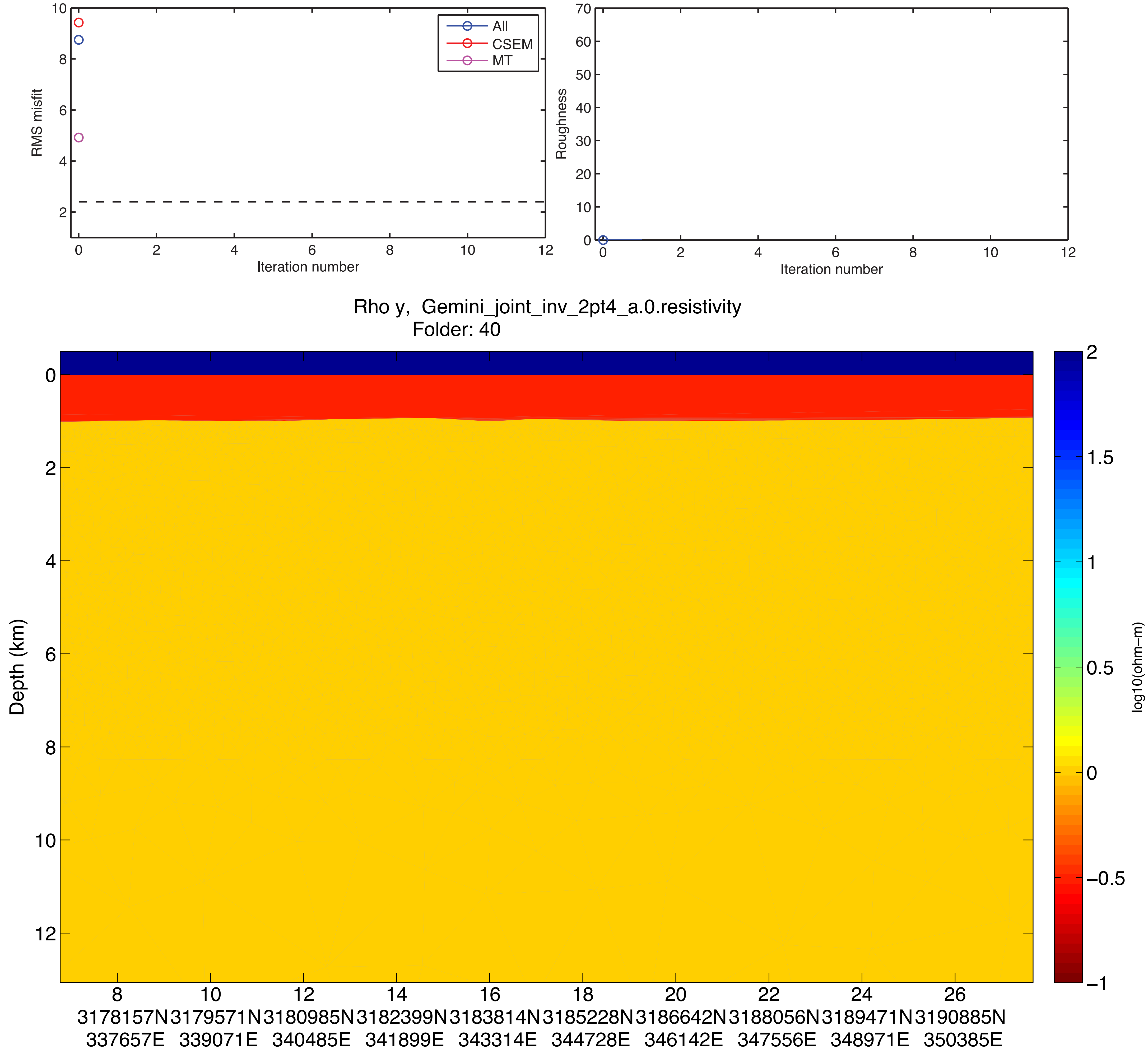
$$\mathbf{m}_1 = \left[\mu\mathbf{R}^T\mathbf{R} + (\mathbf{WJ})^T\mathbf{WJ}\right]^{-1}(\mathbf{WJ})^T\mathbf{W}(\mathbf{d} - f(\mathbf{m_0}) + \mathbf{Jm}_0) \quad .$$

If the Occam algorithm does not get hung up in a local minimum, it will converge to the smoothest model for a given misfit.

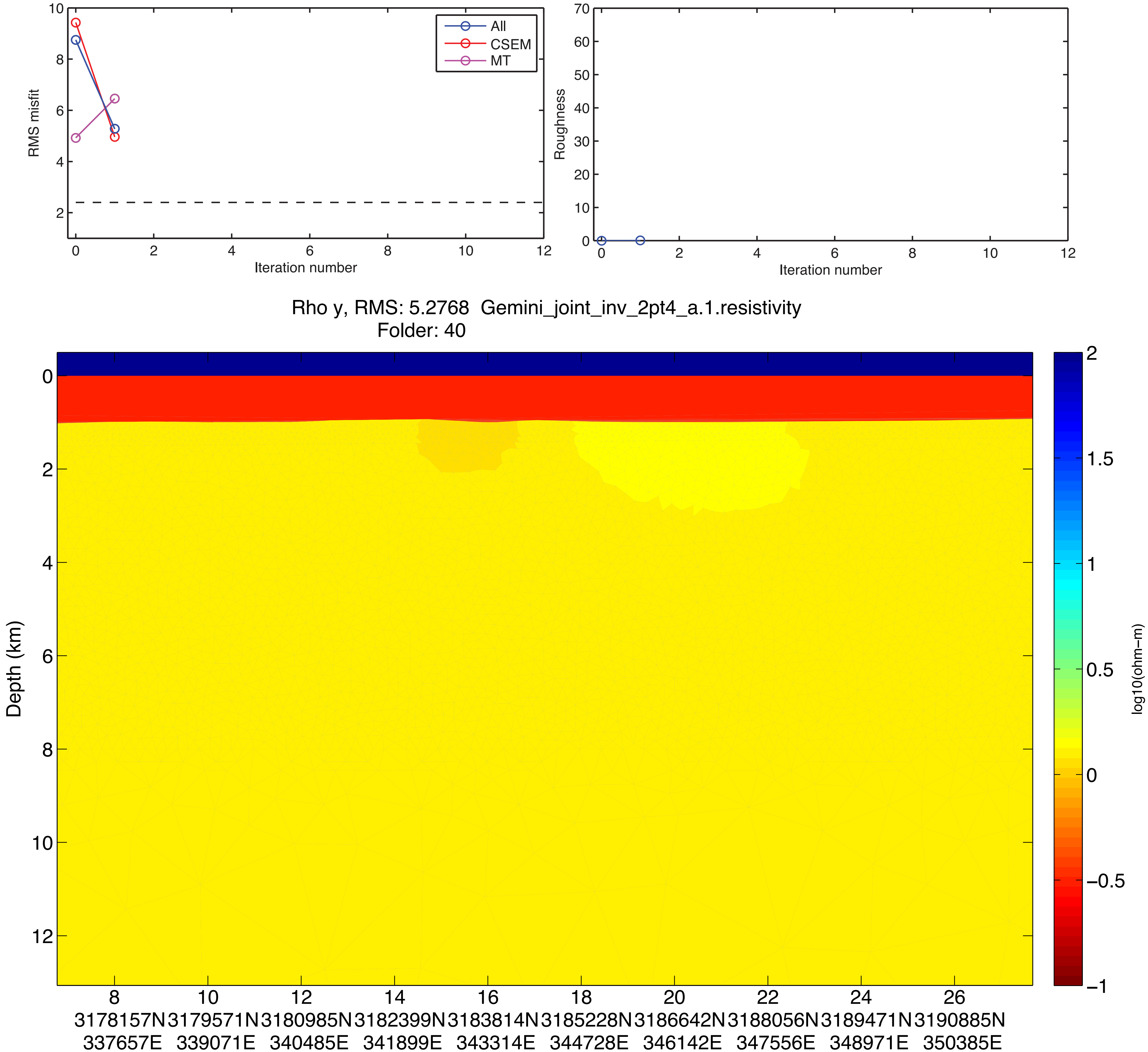


It is important to run the inversion to convergence, and not stop as soon as the target misfit is achieved.

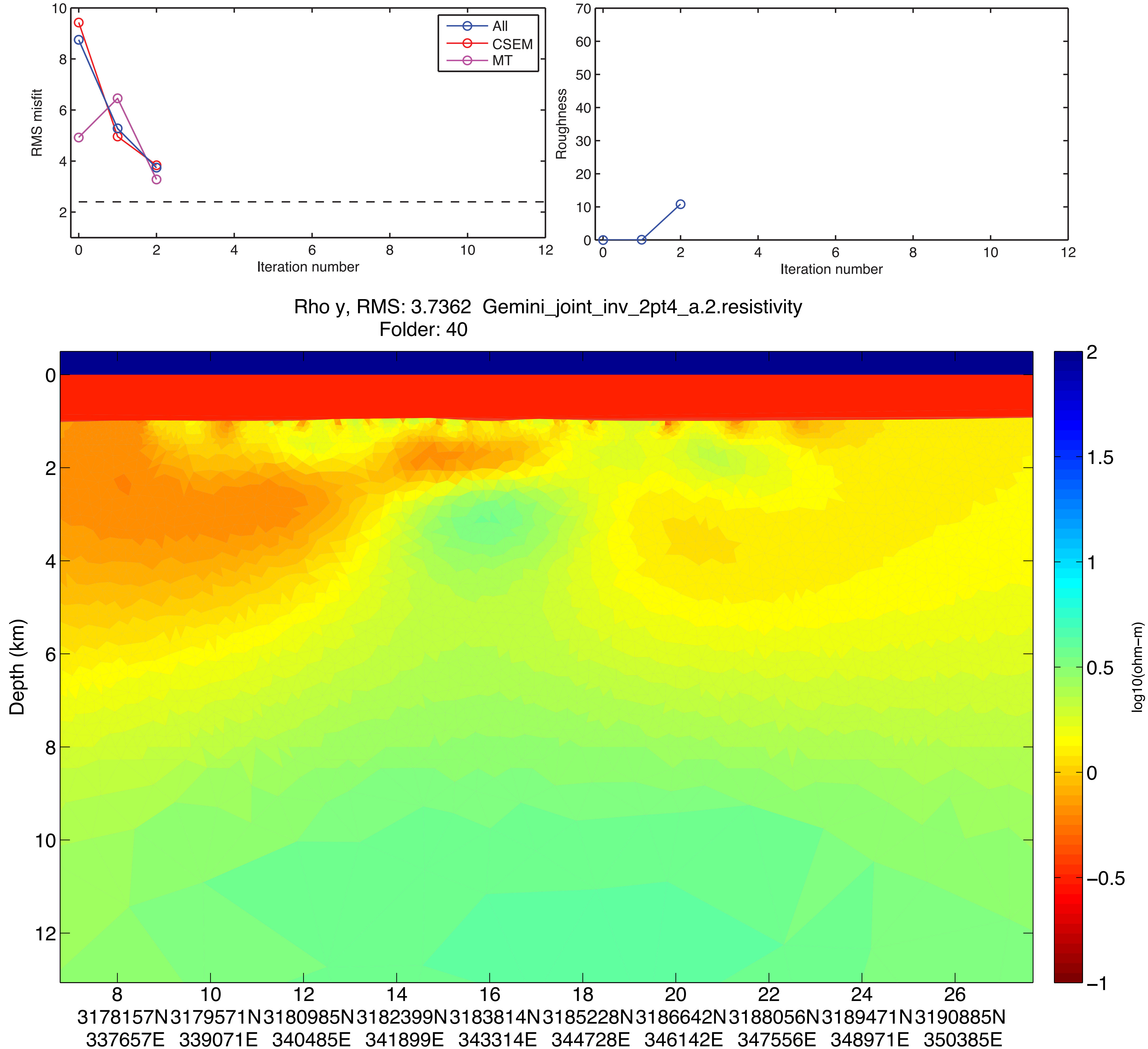
I will step through a joint 2D Occam inversion of marine CSEM (3 frequencies, no phase) and marine MT (Gemini salt prospect, Gulf of Mexico):



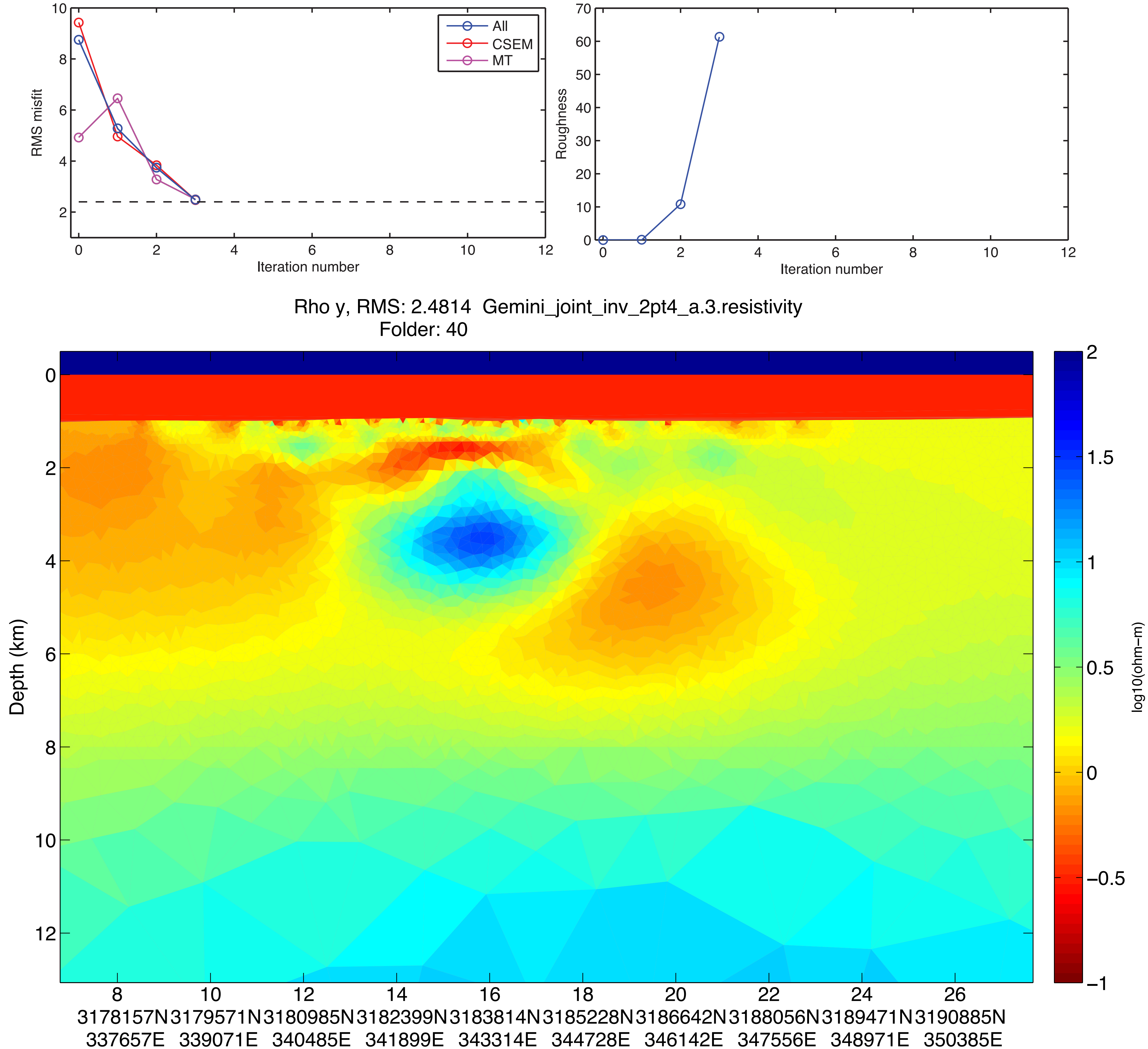
RMS misfit
Iteration number
All
CSEM
MT
Roughness
Iteration number
Rho y, Gemini_joint_inv_2pt4_a.0.resistivity
Folder: 40
Depth (km)
log10(ohm–m)
3178157N
337657E
3179571N
339071E
3180985N
340485E
3182399N
341899E
3183814N
343314E
3185228N
344728E
3186642N
346142E
3188056N
347556E
3189471N
348971E
3190885N
350385E



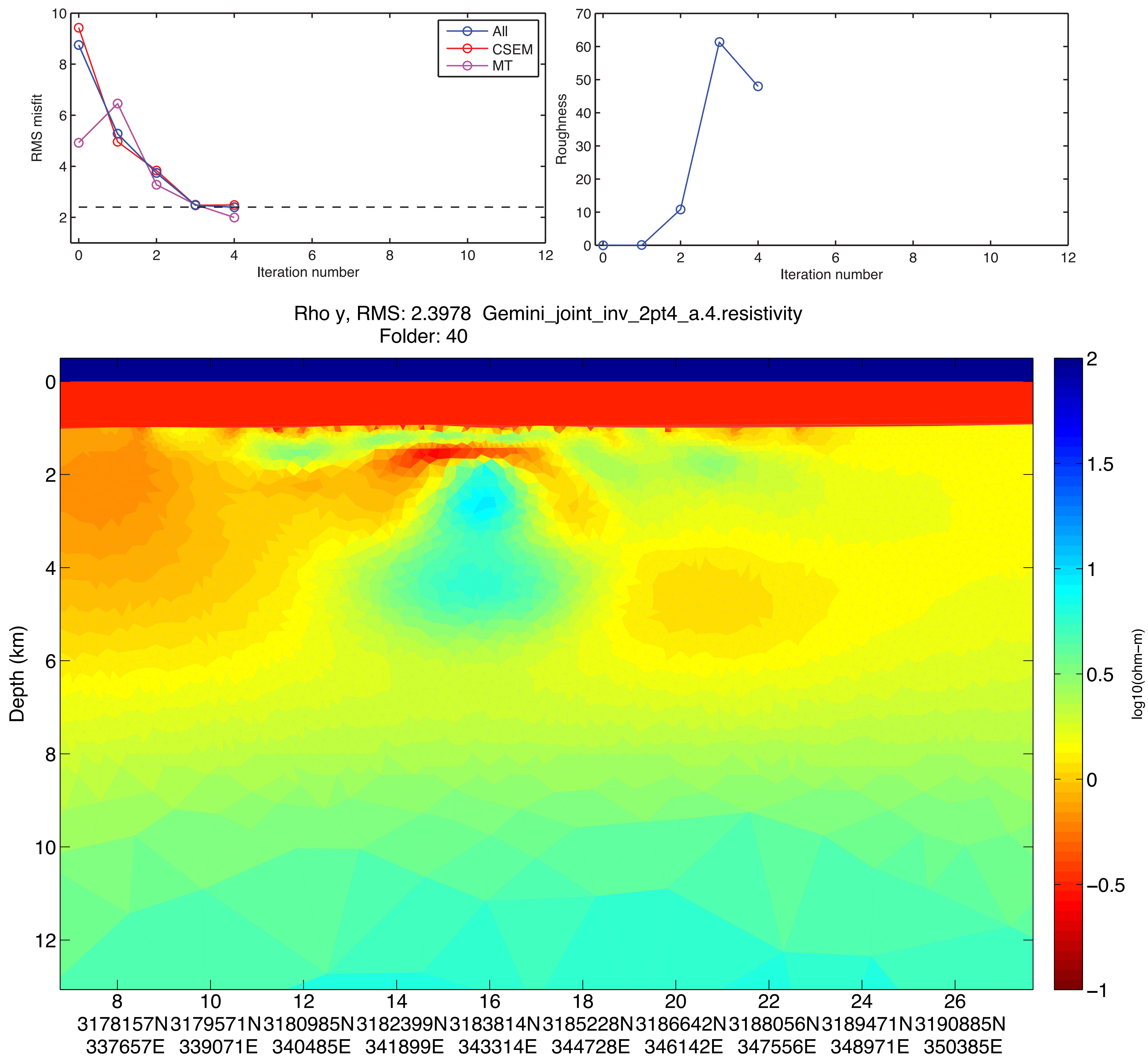
RMS misfit
Iteration number
All
CSEM
MT
Roughness
Iteration number
Rho y, RMS: 5.2768 Gemini_joint_inv_2pt4_a.1.resistivity
Folder: 40
Depth (km)
log10(ohm−m)
3178157N 337657E
3179571N 339071E
3180985N 340485E
3182399N 341899E
3183814N 343314E
3185228N 344728E
3186642N 346142E
3188056N 347556E
3189471N 348971E
3190885N 350385E

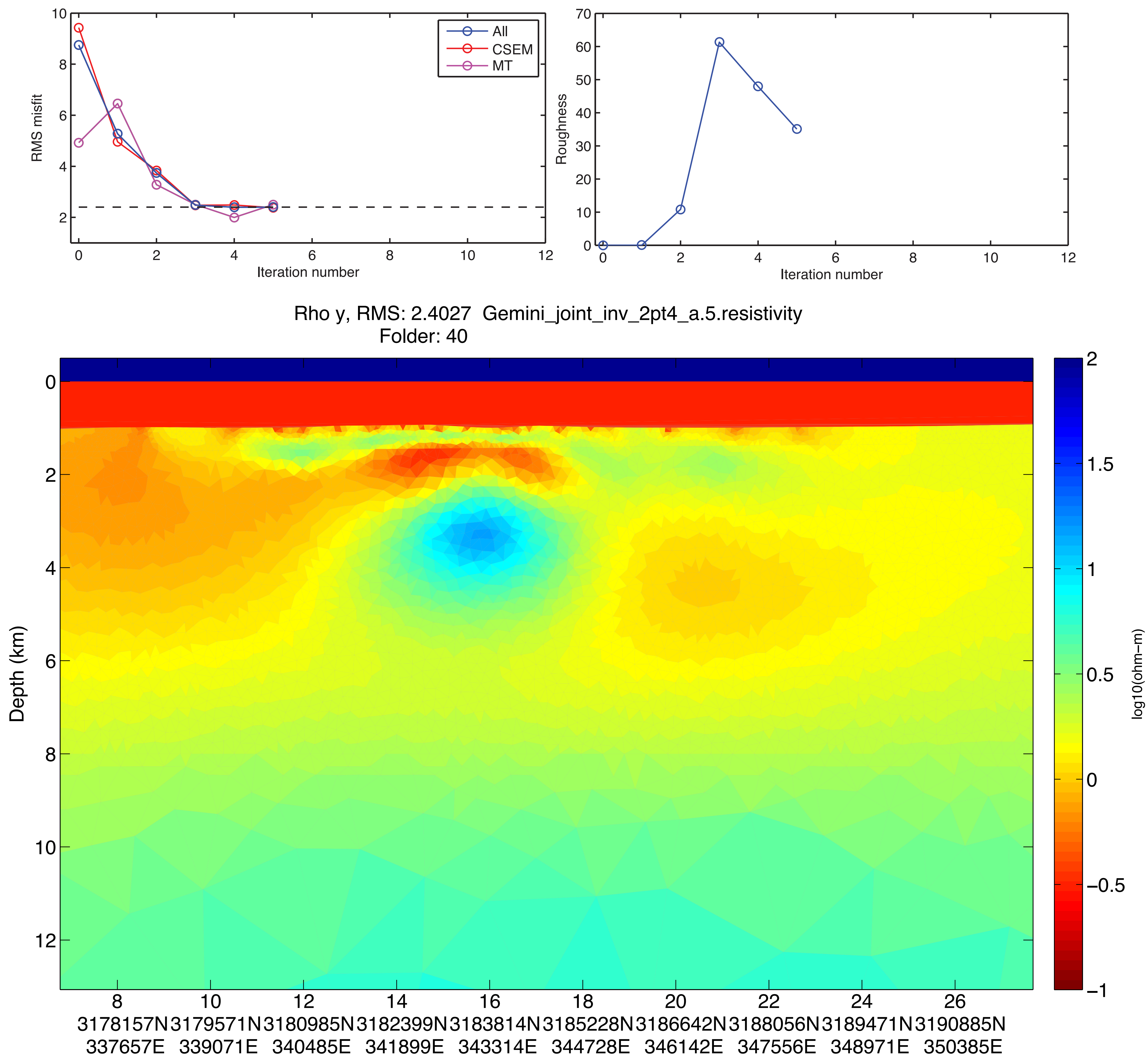


RMS misfit
Iteration number
All
CSEM
MT
Roughness
Iteration number
Rho y, RMS: 3.7362 Gemini_joint_inv_2pt4_a.2.resistivity
Folder: 40
Depth (km)
log10(ohm−m)
3178157N 337657E
3179571N 339071E
3180985N 340485E
3182399N 341899E
3183814N 343314E
3185228N 344728E
3186642N 346142E
3188056N 347556E
3189471N 348971E
3190885N 350385E

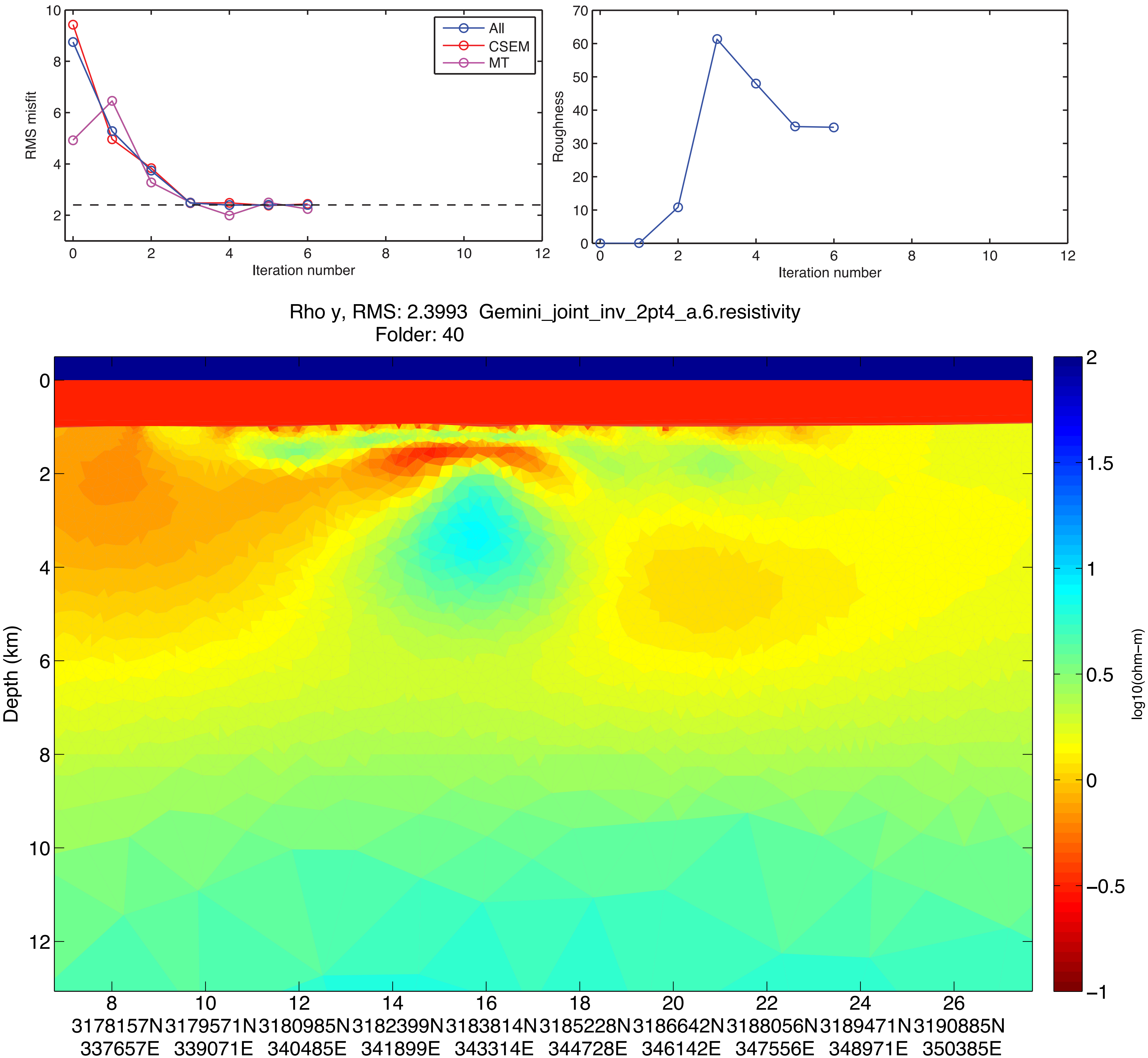


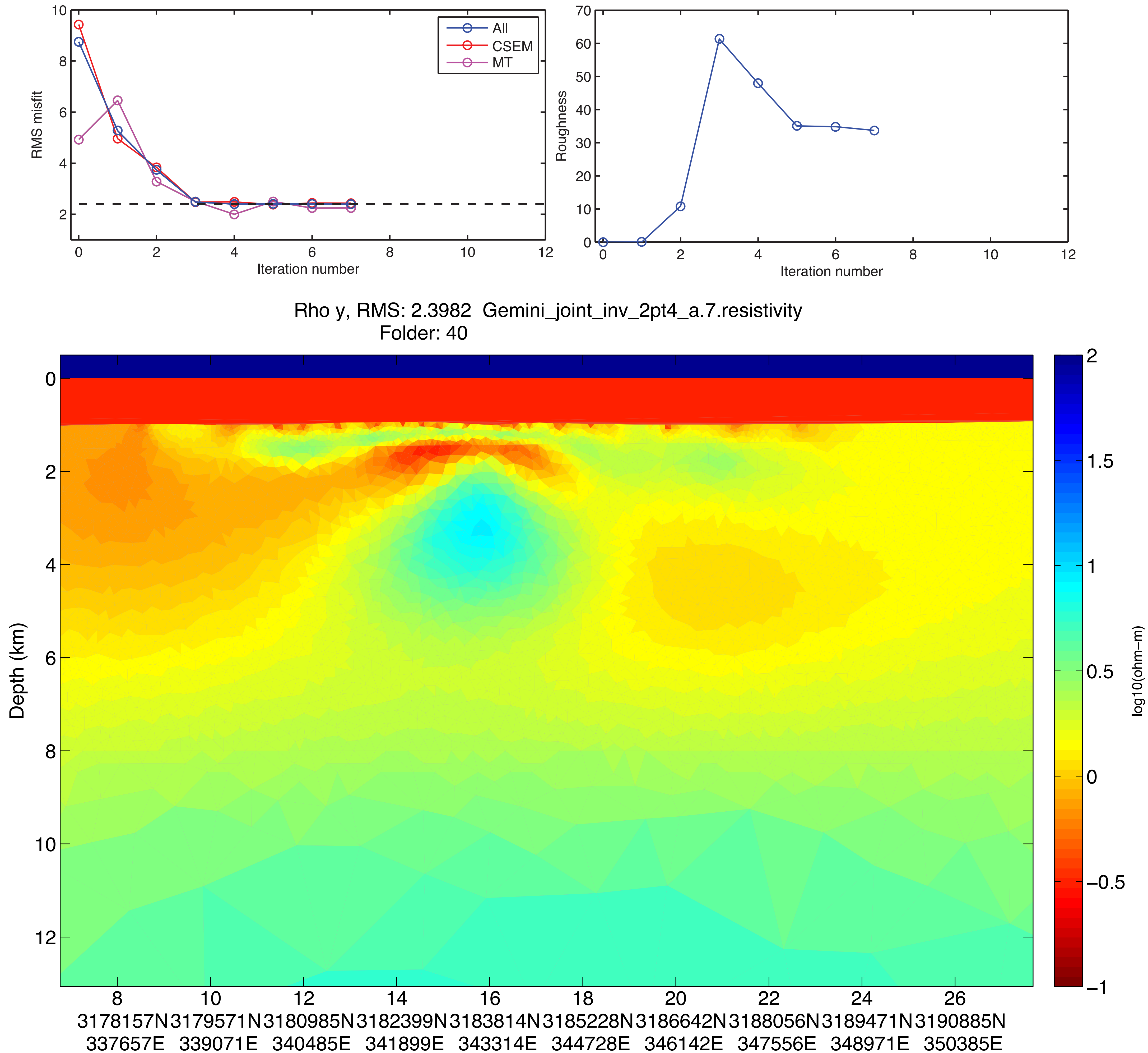
RMS misfit
Iteration number
All
CSEM
MT
Roughness
Iteration number
Rho y, RMS: 2.4814 Gemini_joint_inv_2pt4_a.3.resistivity
Folder: 40
Depth (km)
log10(ohm−m)
3178157N 337657E
3179571N 339071E
3180985N 340485E
3182399N 341899E
3183814N 343314E
3185228N 344728E
3186642N 346142E
3188056N 347556E
3189471N 348971E
3190885N 350385E


RMS misfit
Iteration number
All
CSEM
MT
Roughness
Iteration number
Rho y, RMS: 2.3978 Gemini_joint_inv_2pt4_a.4.resistivity
Folder: 40
Depth (km)
log10(ohm−m)
3178157N 3179571N 3180985N 3182399N 3183814N 3185228N 3186642N 3188056N 3189471N 3190885N
337657E 339071E 340485E 341899E 343314E 344728E 346142E 347556E 348971E 350385E

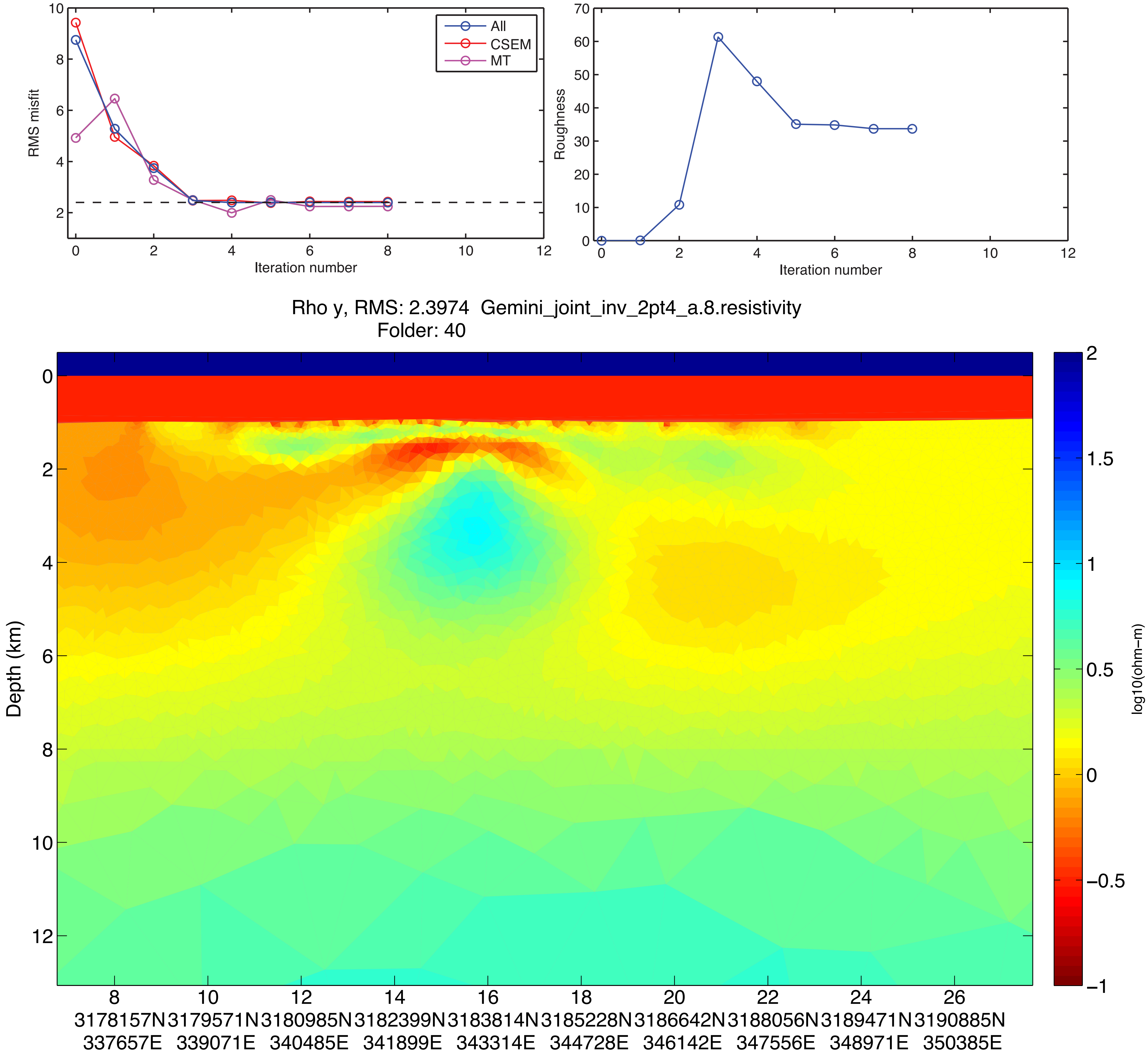


All
CSEM
MT
RMS misfit
Iteration number
Roughness
Iteration number
Rho y, RMS: 2.4027 Gemini_joint_inv_2pt4_a.5.resistivity
Folder: 40
Depth (km)
log10(ohm−m)
3178157N 3179571N 3180985N 3182399N 3183814N 3185228N 3186642N 3188056N 3189471N 3190885N
337657E 339071E 340485E 341899E 343314E 344728E 346142E 347556E 348971E 350385E

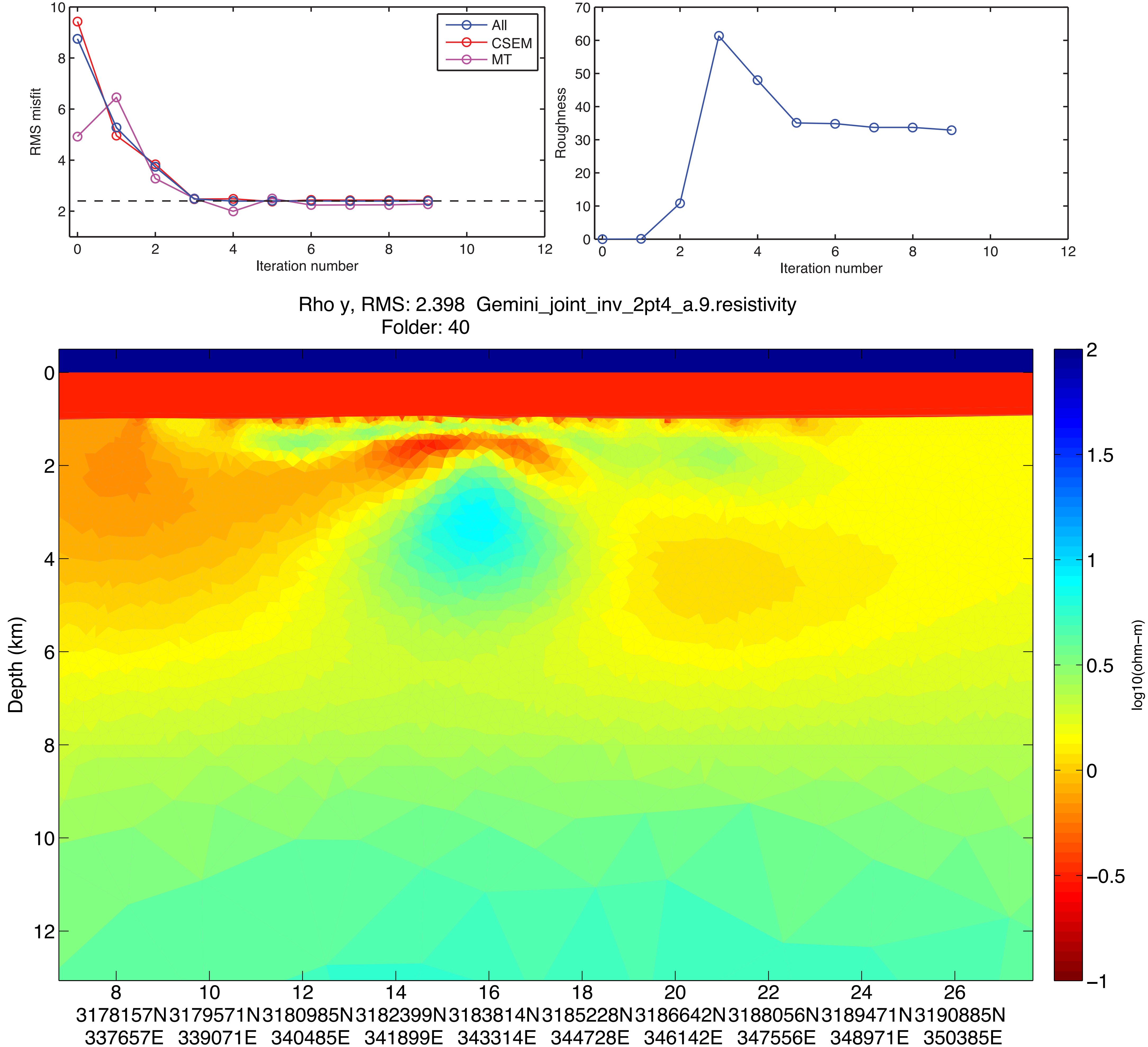

All
CSEM
MT
RMS misfit
Iteration number
Roughness
Iteration number
Rho y, RMS: 2.3993 Gemini_joint_inv_2pt4_a.6.resistivity
Folder: 40
Depth (km)
log10(ohm−m)
3178157N 337657E
3179571N 339071E
3180985N 340485E
3182399N 341899E
3183814N 343314E
3185228N 344728E
3186642N 346142E
3188056N 347556E
3189471N 348971E
3190885N 350385E



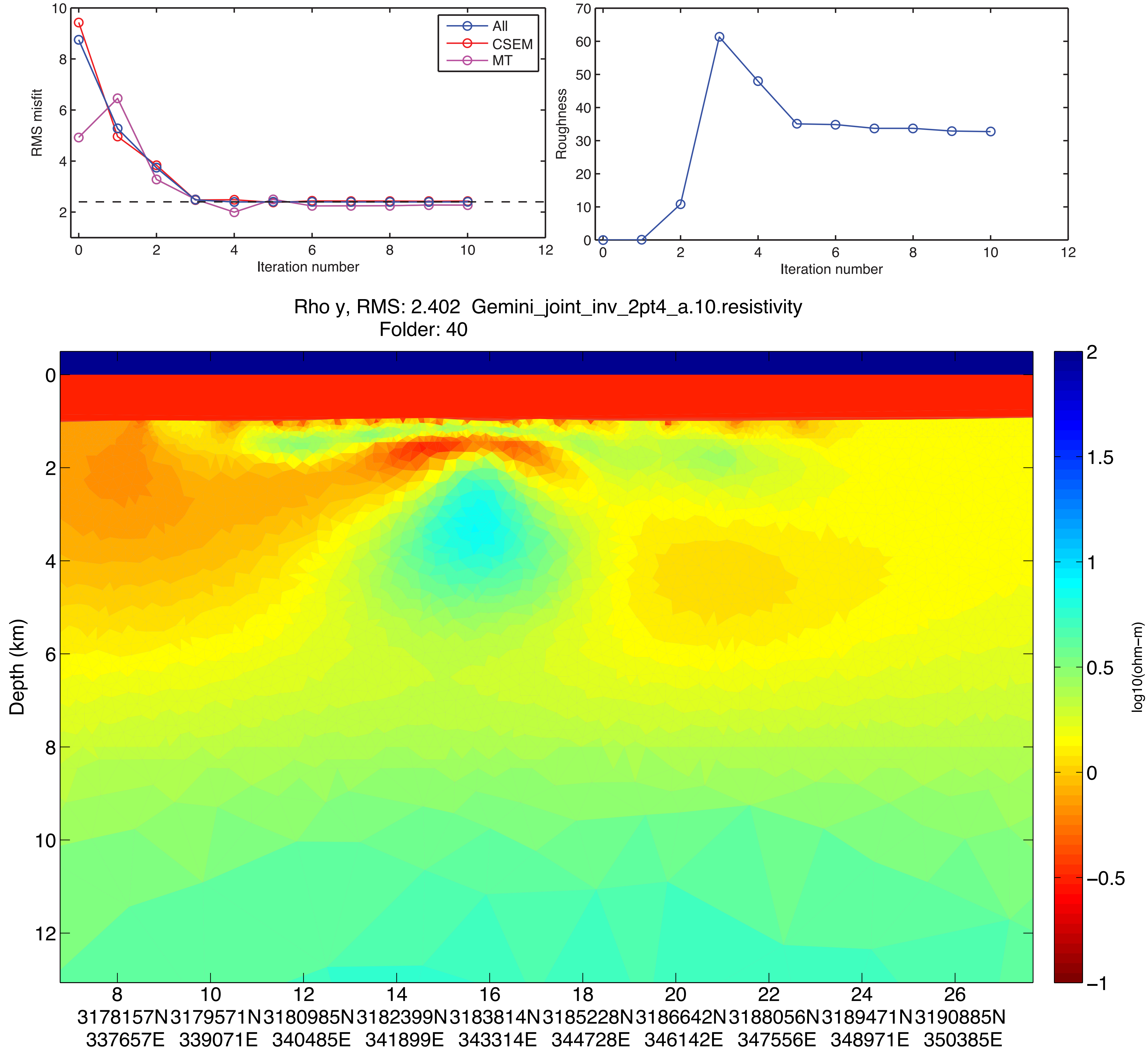
RMS misfit
Iteration number
All
CSEM
MT
Roughness
Iteration number
Rho y, RMS: 2.3982 Gemini_joint_inv_2pt4_a.7.resistivity
Folder: 40
Depth (km)
log10(ohm−m)
3178157N 337657E
3179571N 339071E
3180985N 340485E
3182399N 341899E
3183814N 343314E
3185228N 344728E
3186642N 346142E
3188056N 347556E
3189471N 348971E
3190885N 350385E



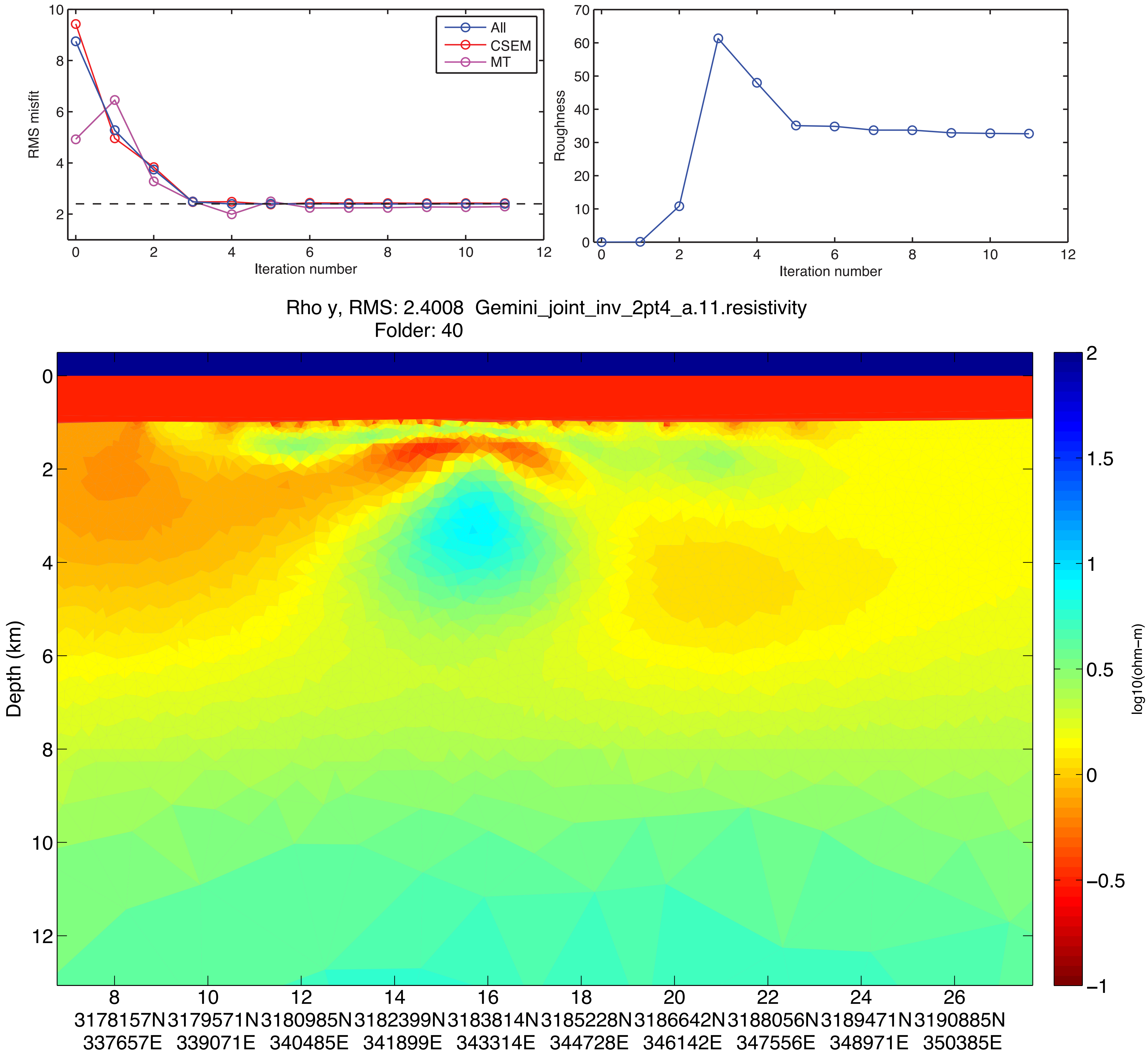
All
CSEM
MT
RMS misfit
Iteration number
Roughness
Iteration number
Rho y, RMS: 2.3974 Gemini_joint_inv_2pt4_a.8.resistivity
Folder: 40
Depth (km)
log10(ohm−m)
3178157N 3179571N 3180985N 3182399N 3183814N 3185228N 3186642N 3188056N 3189471N 3190885N
337657E 339071E 340485E 341899E 343314E 344728E 346142E 347556E 348971E 350385E



RMS misfit
Iteration number
All
CSEM
MT
Roughness
Iteration number
Rho y, RMS: 2.398 Gemini_joint_inv_2pt4_a.9.resistivity
Folder: 40
Depth (km)
log10(ohm−m)
3178157N 3179571N 3180985N 3182399N 3183814N 3185228N 3186642N 3188056N 3189471N 3190885N
337657E 339071E 340485E 341899E 343314E 344728E 346142E 347556E 348971E 350385E

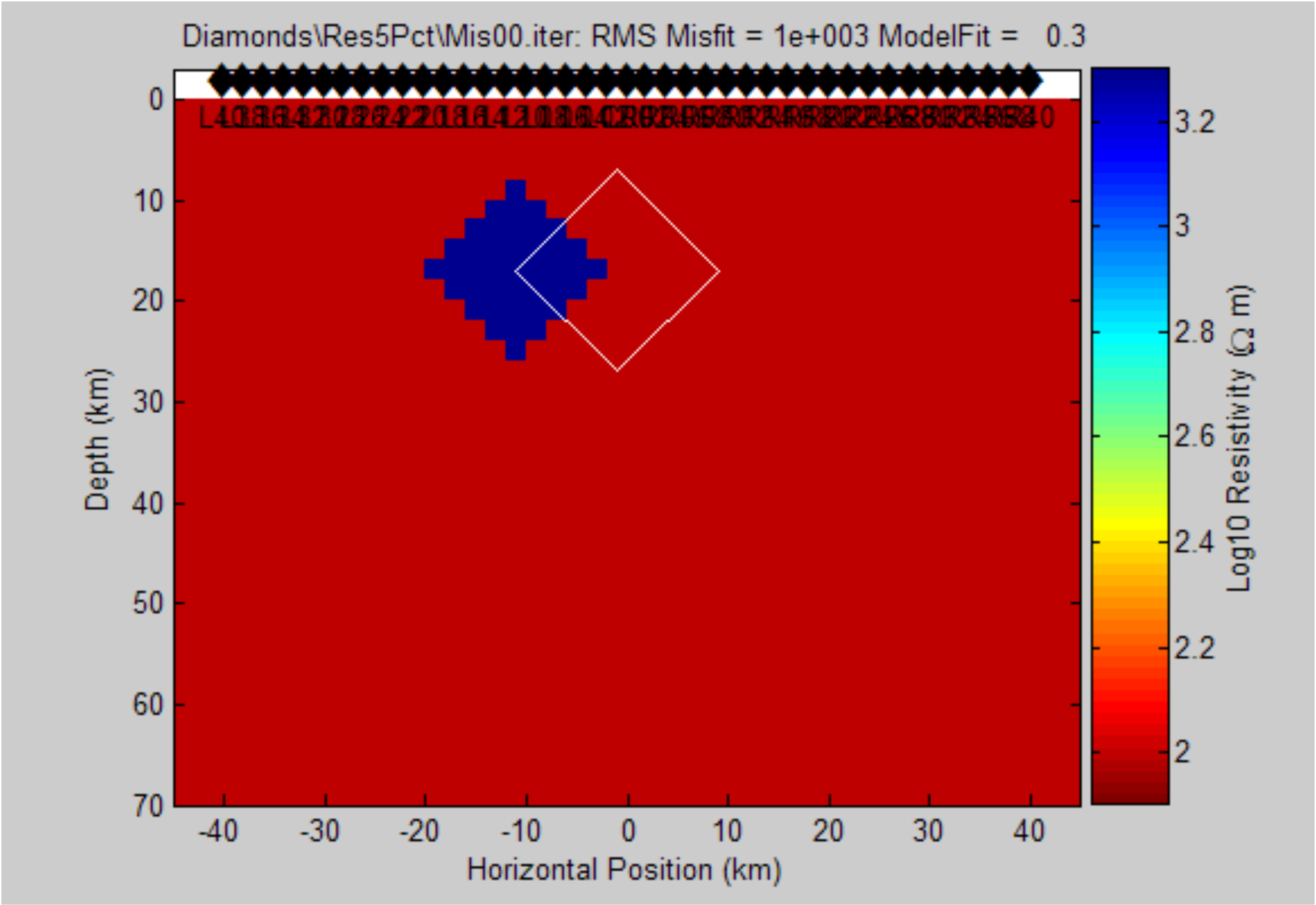


RMS misfit
Iteration number
All
CSEM
MT
Roughness
Iteration number
Rho y, RMS: 2.402 Gemini_joint_inv_2pt4_a.10.resistivity
Folder: 40
Depth (km)
log10(ohm−m)
3178157N 337657E
3179571N 339071E
3180985N 340485E
3182399N 341899E
3183814N 343314E
3185228N 344728E
3186642N 346142E
3188056N 347556E
3189471N 348971E
3190885N 350385E



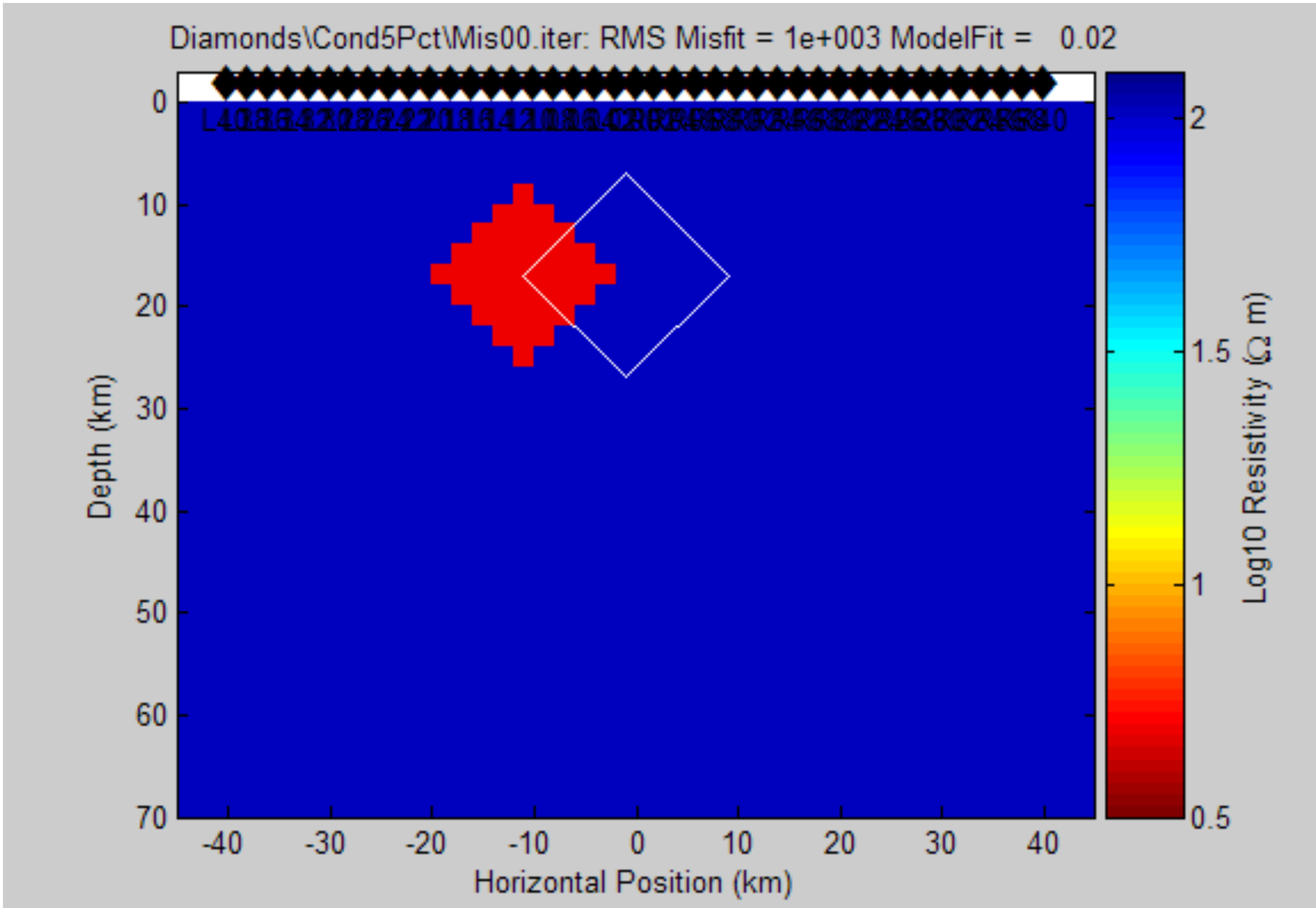
RMS misfit
Iteration number
All
CSEM
MT
Roughness
Iteration number
Rho y, RMS: 2.4008 Gemini_joint_inv_2pt4_a.11.resistivity
Folder: 40
Depth (km)
log10(ohm−m)
3178157N 337657E
3179571N 339071E
3180985N 340485E
3182399N 341899E
3183814N 343314E
3185228N 344728E
3186642N 346142E
3188056N 347556E
3189471N 348971E
3190885N 350385E

Why do we start from a half-space? Because $\mathbf{J}$ depends on $\mathbf{m}$.



MT: misaligned starting resistor - no harm done

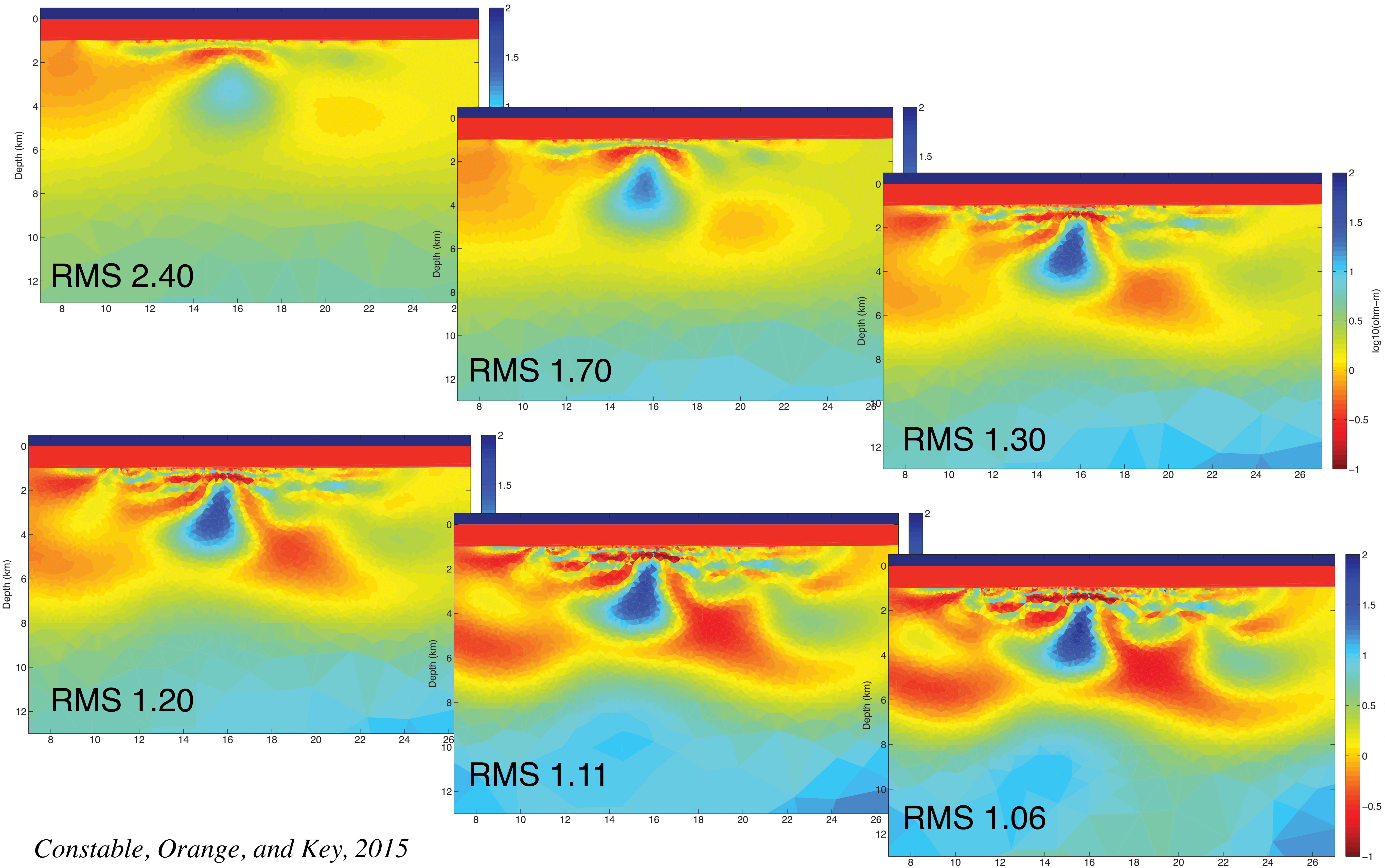
Courtesy David Myer.



misaligned starting conductor - forever trapped by **J**

Courtesy David Myer.

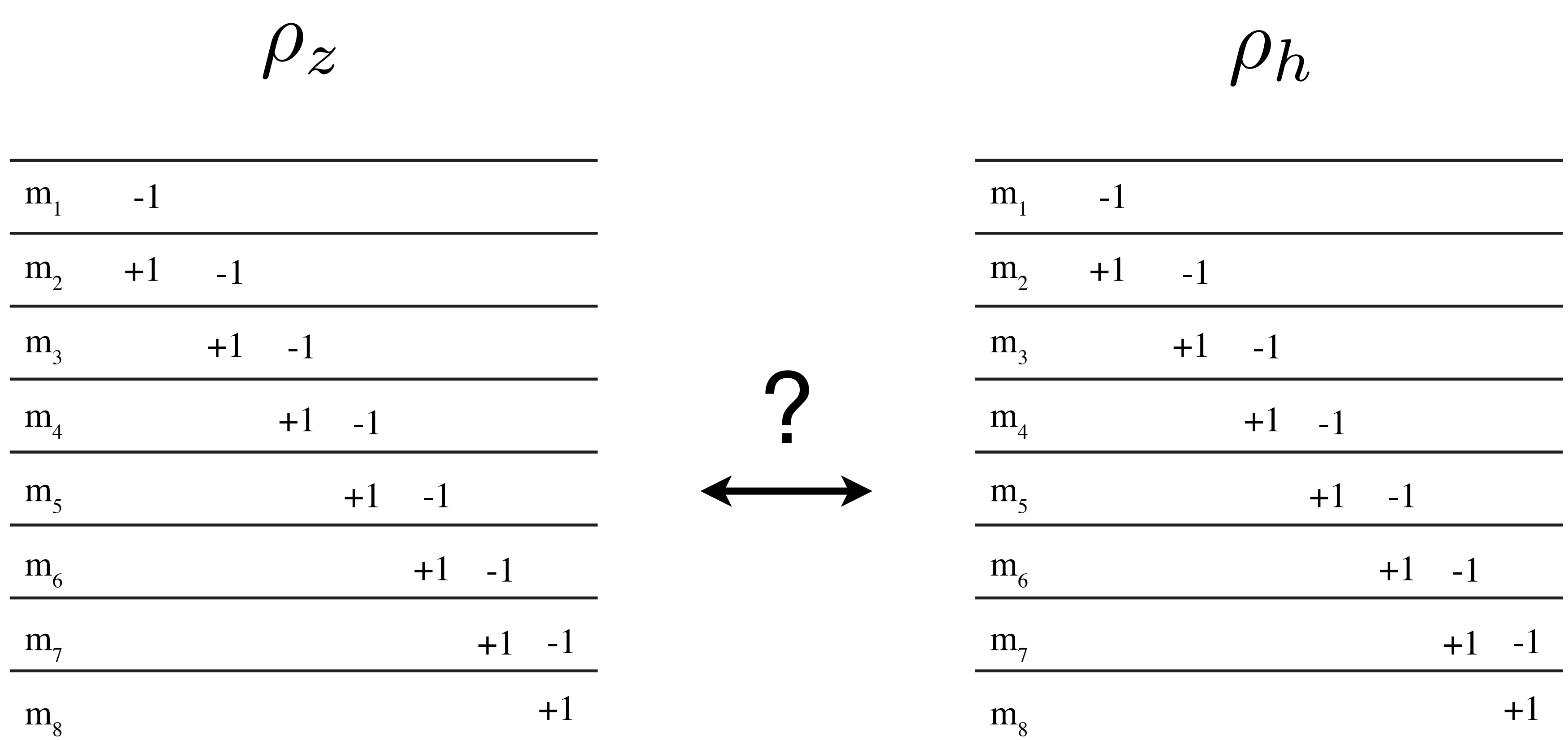
Even with well-estimated errors, choice of misfit can still be somewhat subjective.



Constable, Orange, and Key, 2015

What about anisotropy? It is quite common for physical properties of sediments to be different in the vertical and horizontal directions. For example, horizontal resistivity ρ_h is often smaller than vertical resistivity ρ_z .

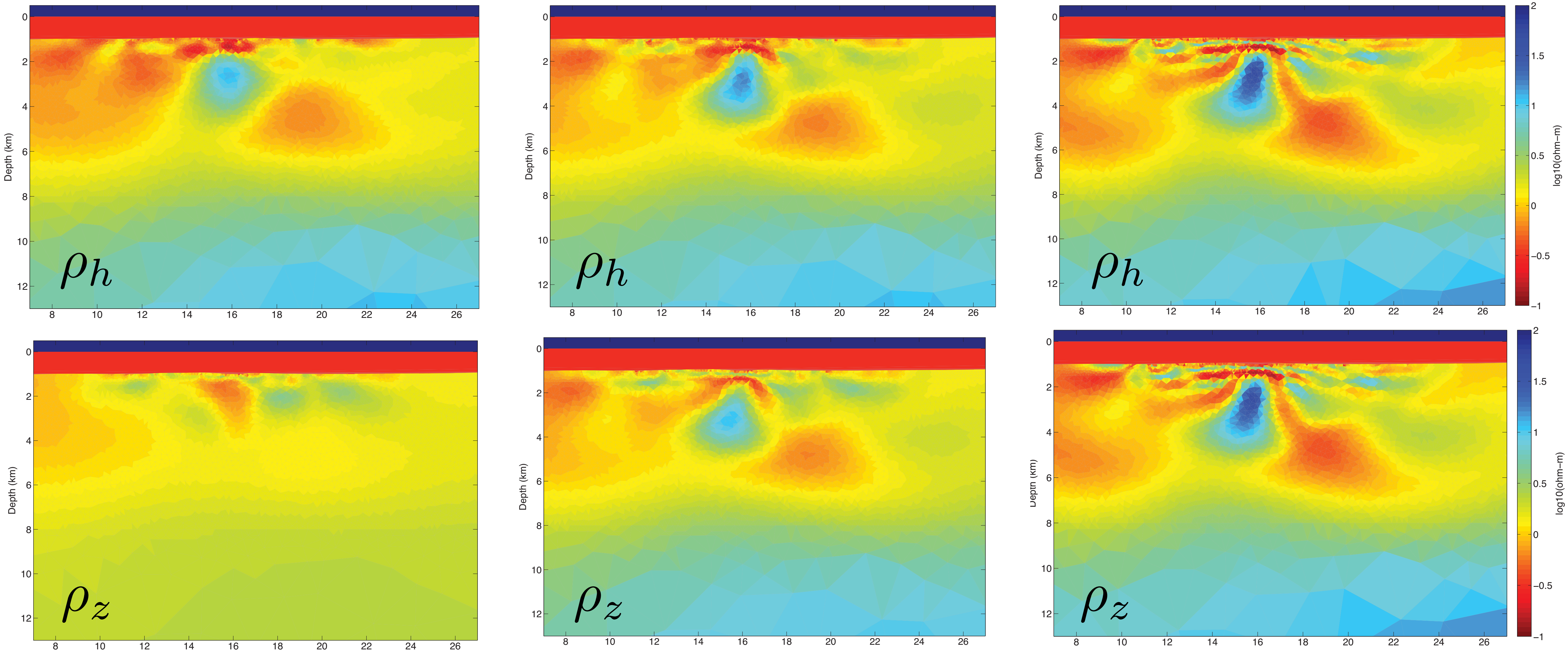
The problem is how to weight the penalty between the two models.



$$\mathbf{R}_1 = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & -1 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & -1 & 1 & 0 & \dots & 0 \\ & & \ddots & & & \ddots & \\ & & & & & & \\ & & & & & -1 & 1 \end{pmatrix}$$

$$\mathbf{R}_2 = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & -1 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & -1 & 1 & 0 & \dots & 0 \\ & & \ddots & & & \ddots & \\ & & & & & & \\ & & & & & -1 & 1 \end{pmatrix}$$

Joint Anisotropic Inversions versus penalty between rho-y and rho-z (all fitting to RMS 1.2):



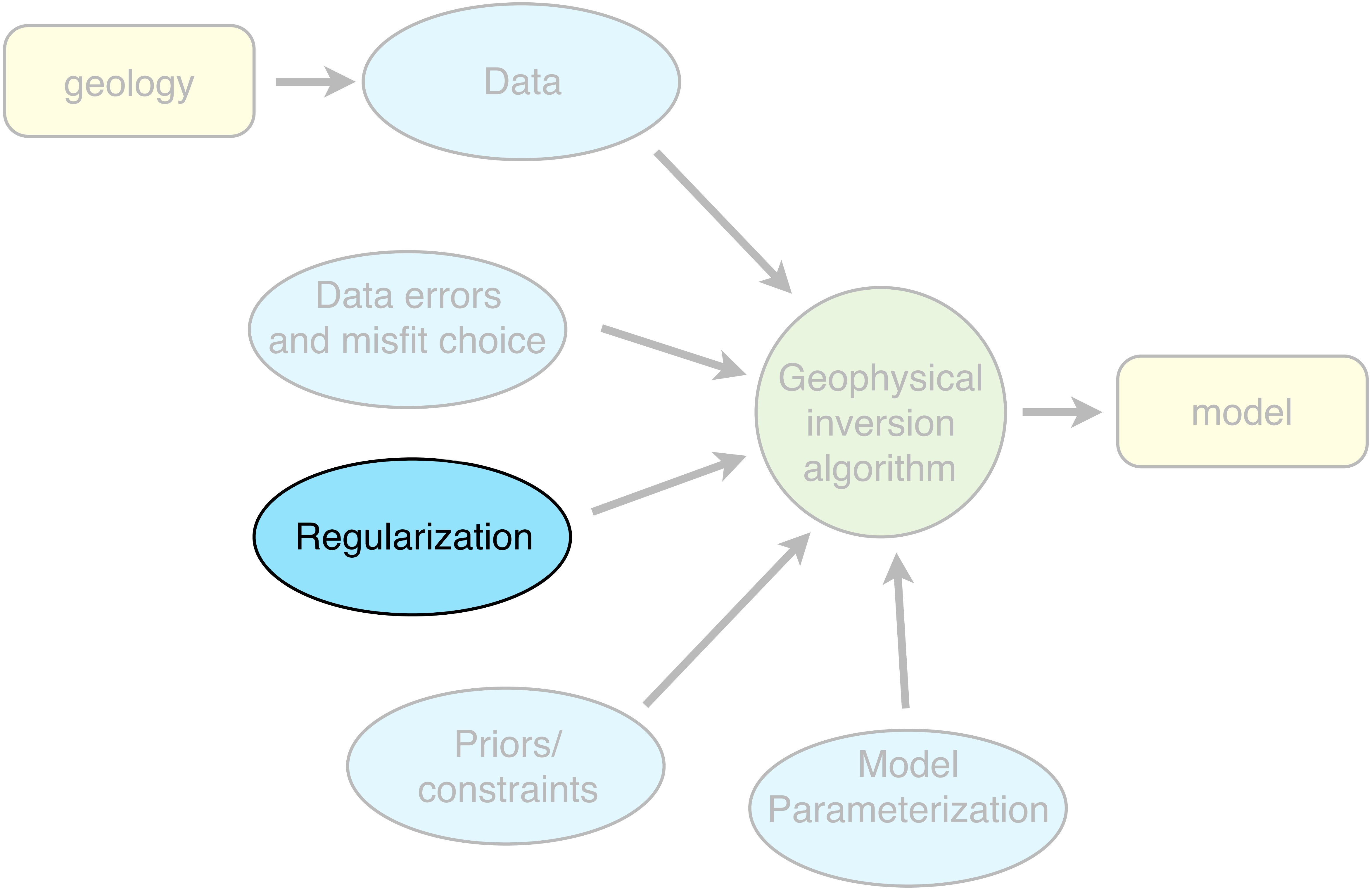
Low weight (0.1), models are independent.

Medium weight (1), models look sensible.

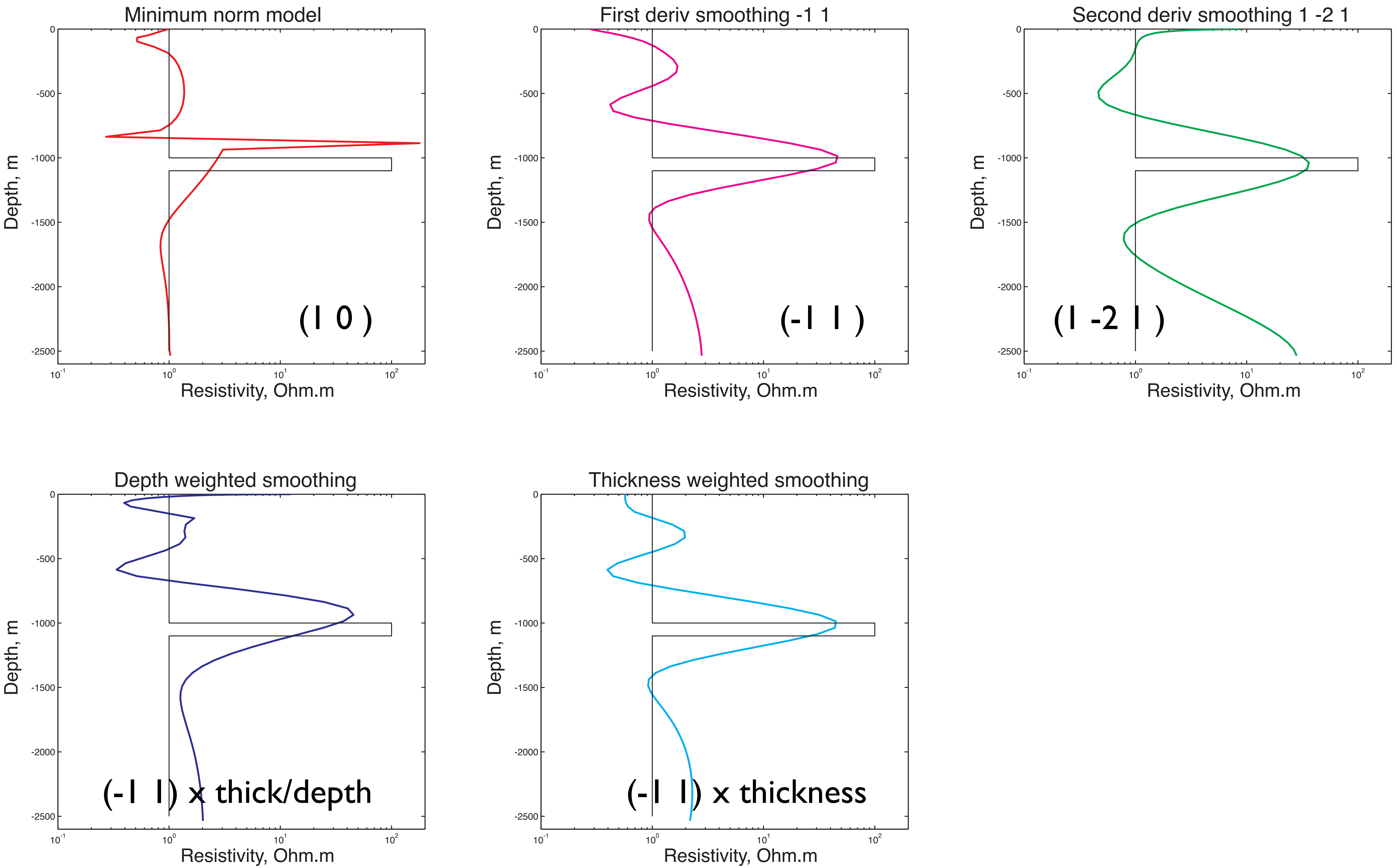
High weight (10), models are identical.

Constable, Orange, and Key, 2015

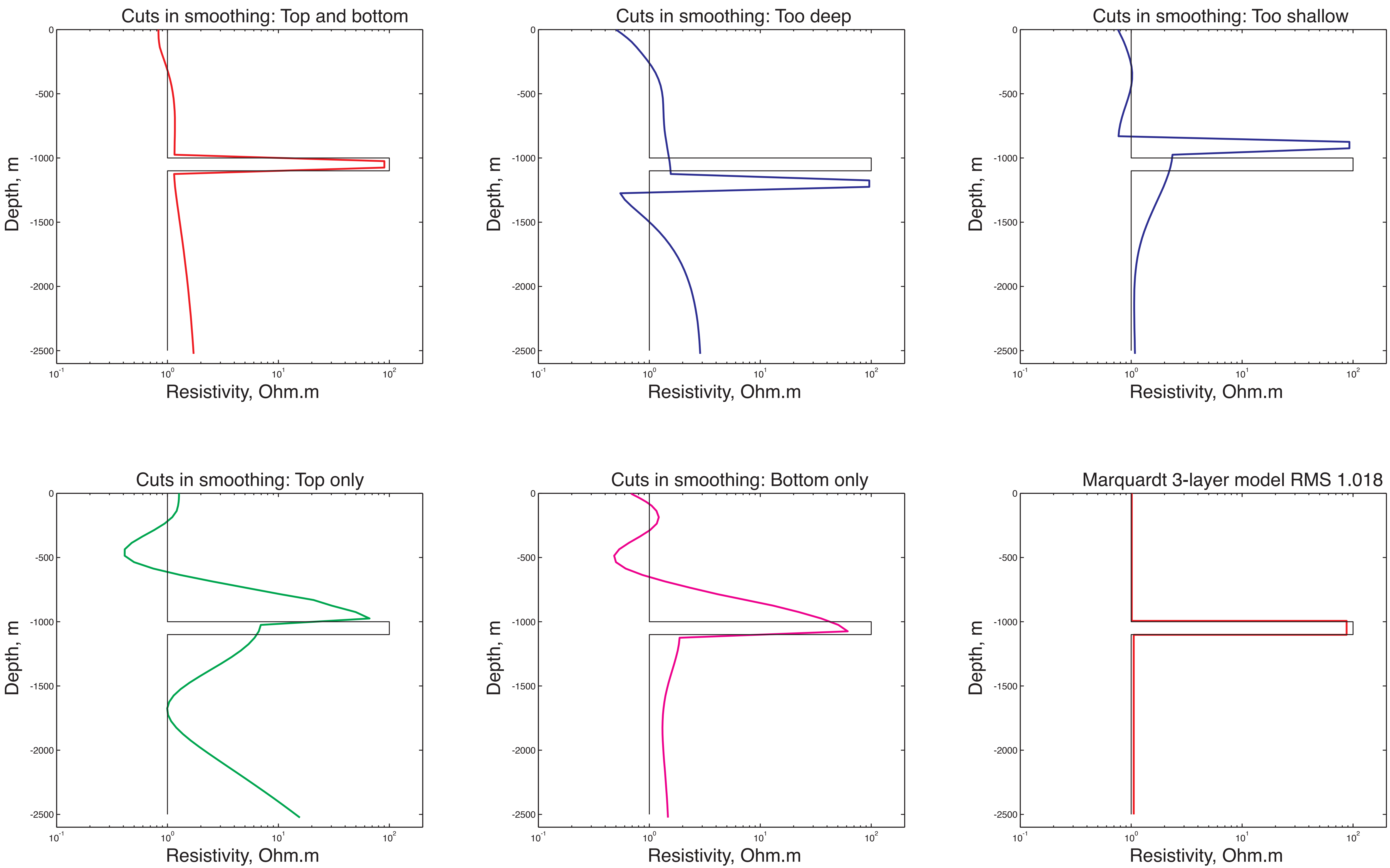
There are many ways to choose how to regularize the problem, and this matters too.



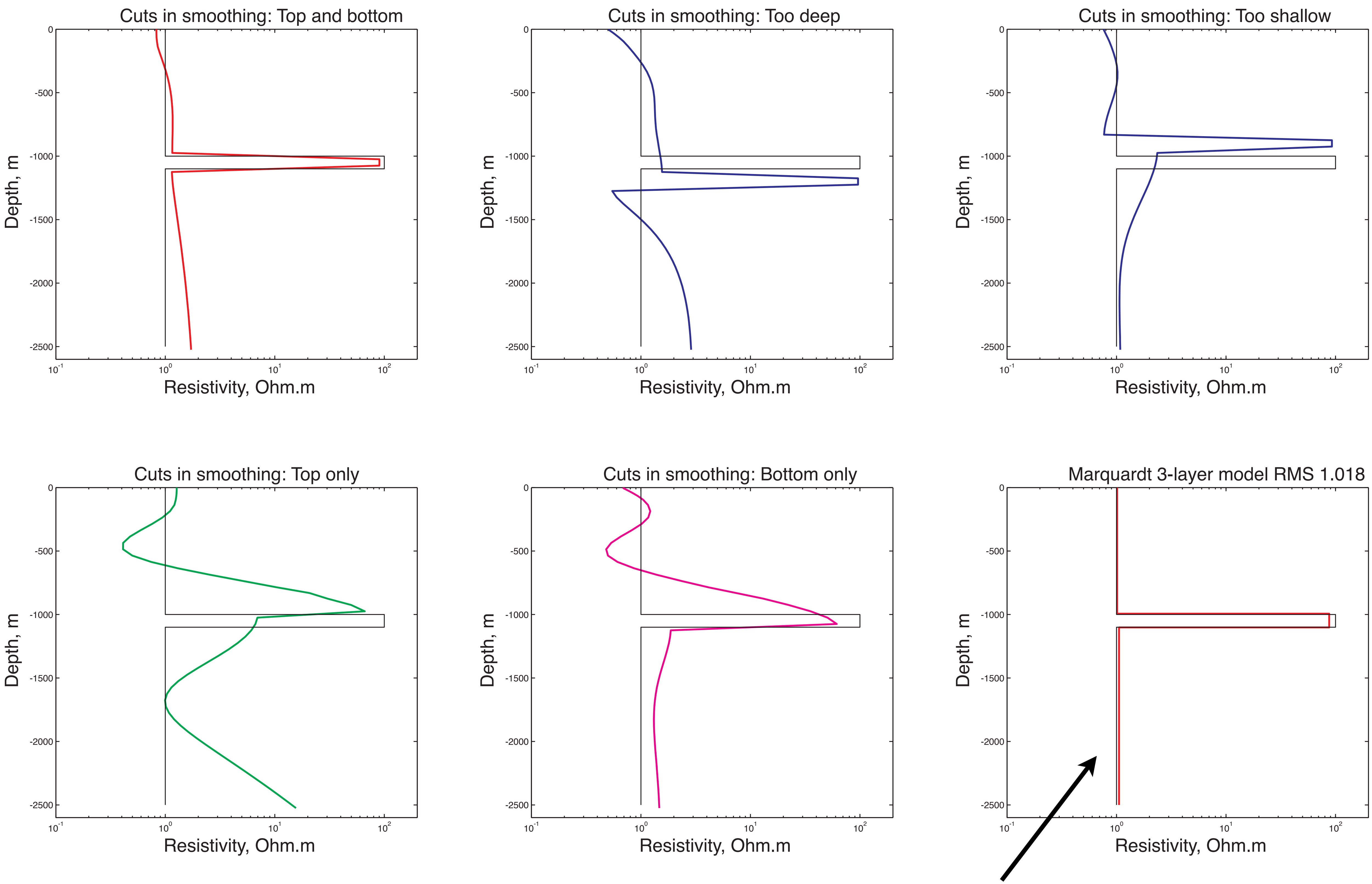
For a given misfit, the model depends on **R**



You can have fun with cuts (removing a row of $\mathbf{R}$):

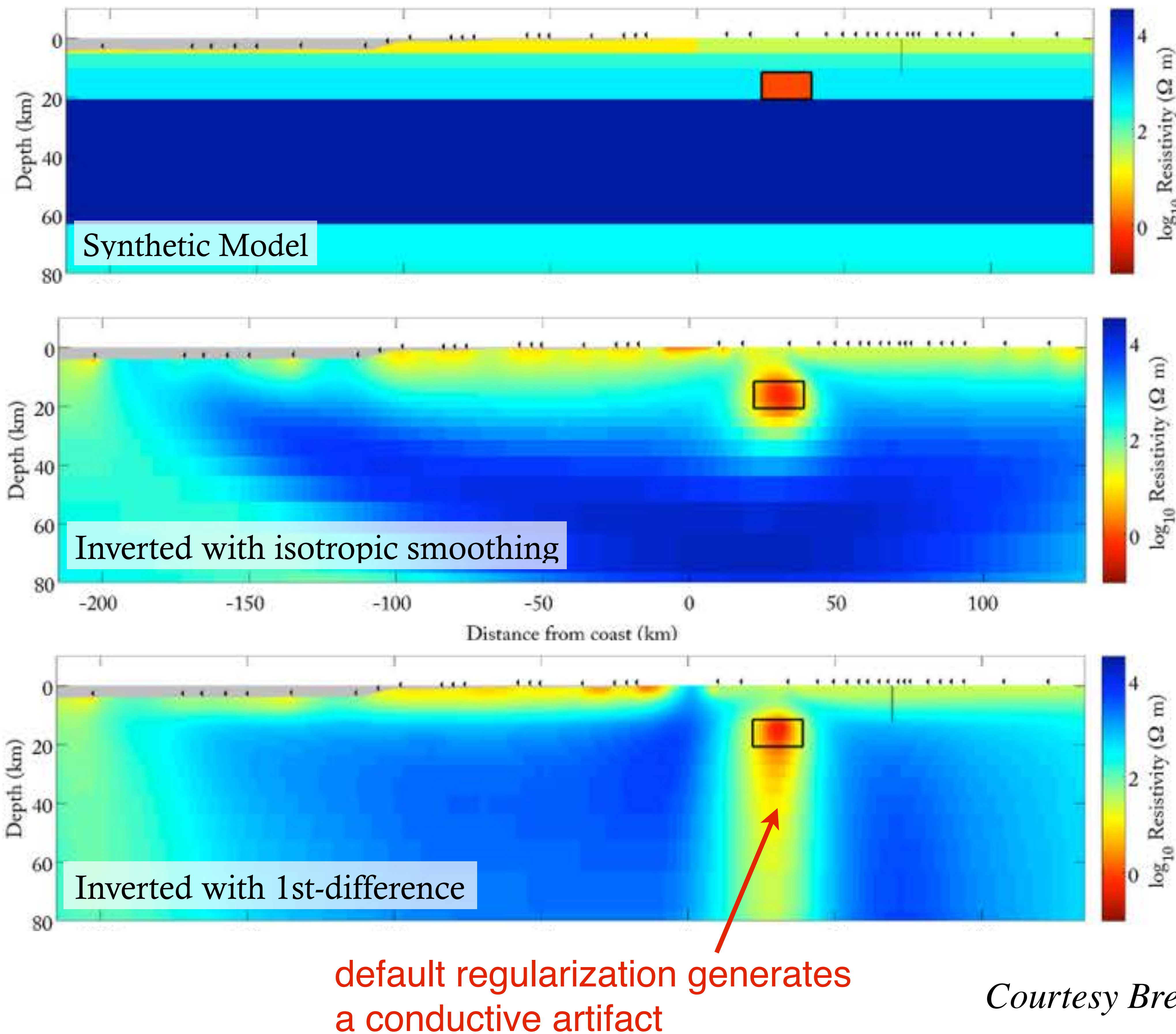


You can have fun with cuts (removing a row of $\mathbf{R}$):



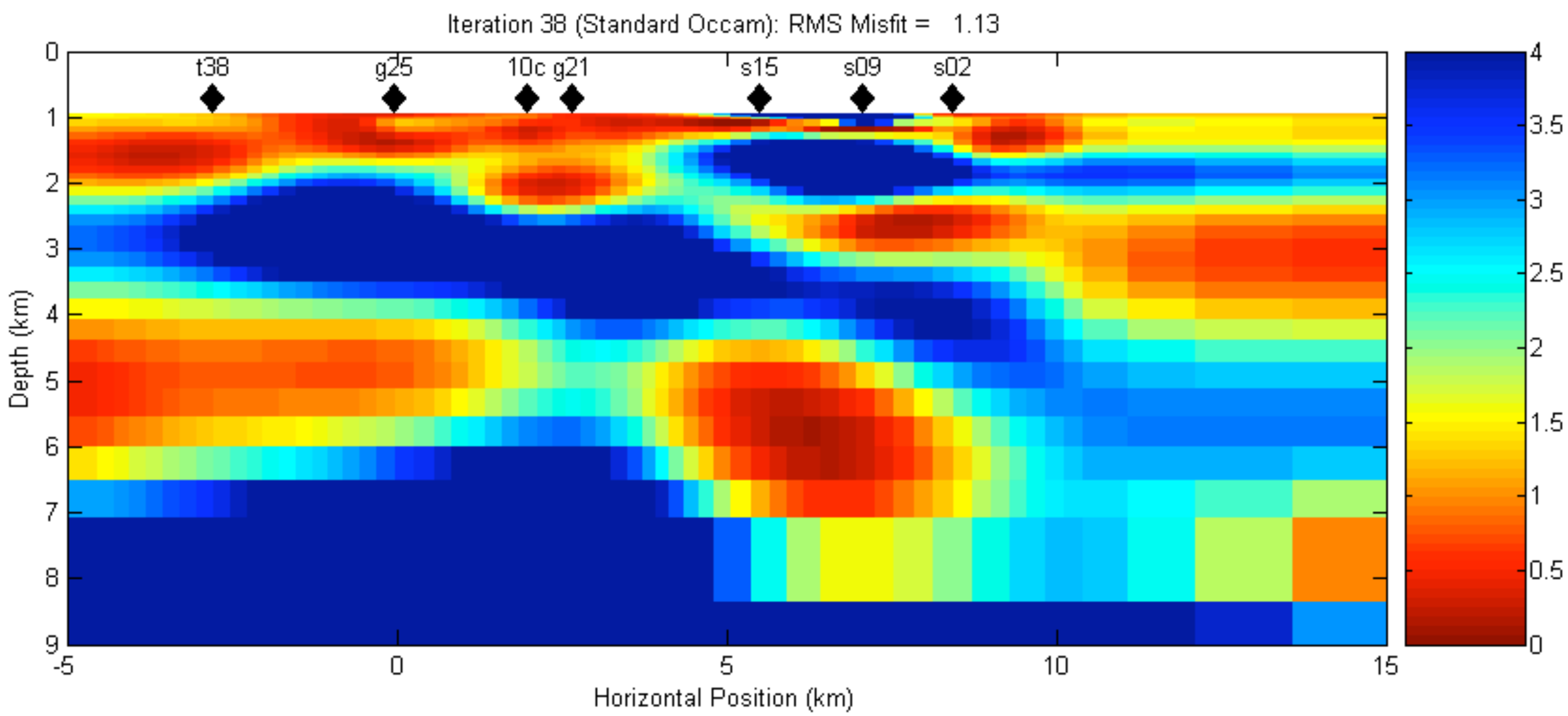
Sparsely parameterized model does well! (Why?)

In 2D:



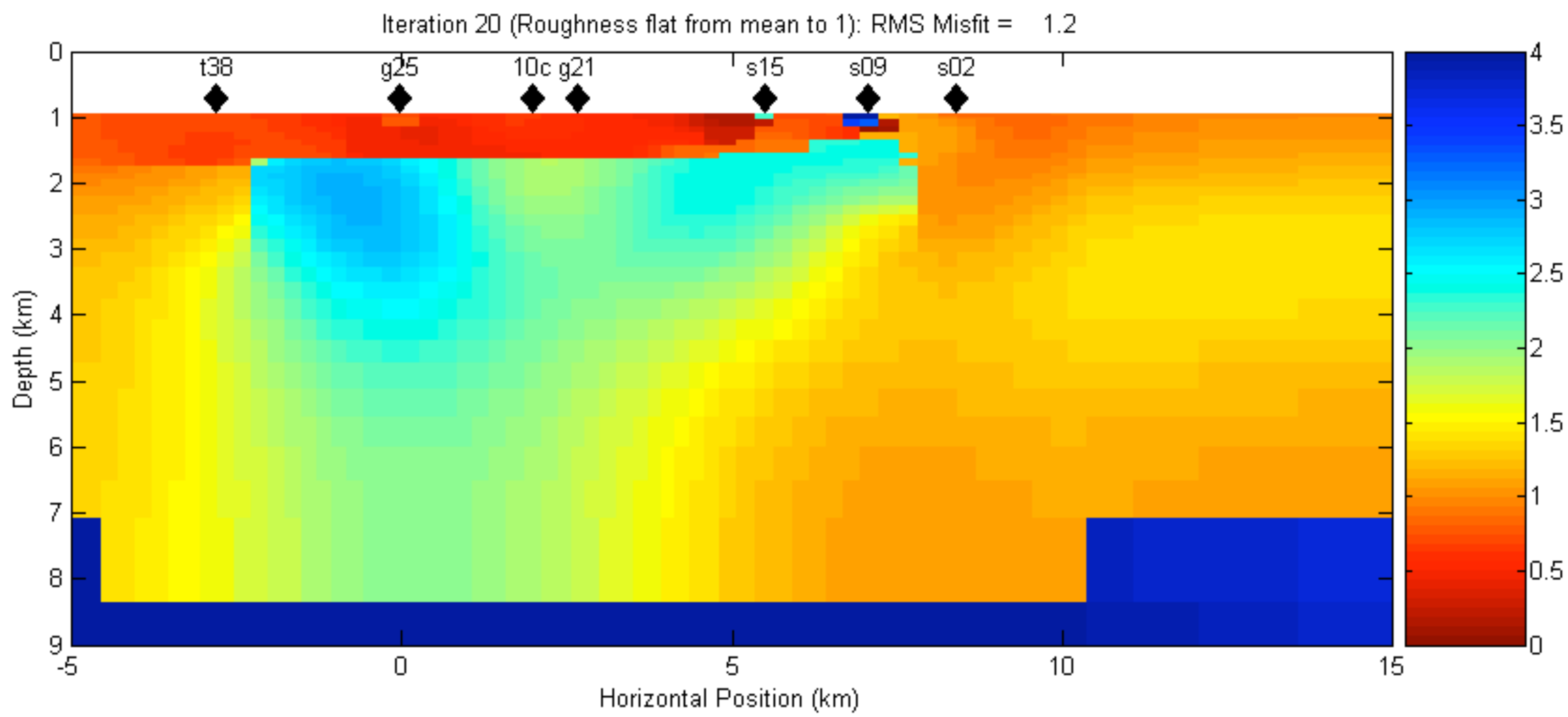
Courtesy Brent Wheelock.

Always remember that regularization has an input into the model solution. Both of these MT models fit the data equally well.



Default **R:**

$$
\mathbf{R}_1 = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & -1 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & -1 & 1 & 0 & \dots & 0 \\ & & \ddots & & & \ddots & \\ & & & & & -1 & 1 \end{pmatrix}
$$



A nonlinear and adaptive **R**, providing little penalty for big contrasts

Courtesy David Myer.

Can we place errors or uncertainties on regularized models?

No!

For sparse parameterizations, data errors are often projected onto model parameters through the Jacobian. This is a dangerous practice because

- it depends on the parameterization
- the Jacobian depends on the solution

Can we place errors or uncertainties on regularized models?

No!

For sparse parameterizations, data errors are often projected onto model parameters through the Jacobian. This is a dangerous practice because

- it depends on the parameterization
- the Jacobian depends on the solution

For regularized models, however, we compound this by generating huge amounts of covariance between the model parameters as part of the smoothness constraint. It is much more useful to think of regularized models as extremal solutions, and vary the regularization to ask the questions you may have.

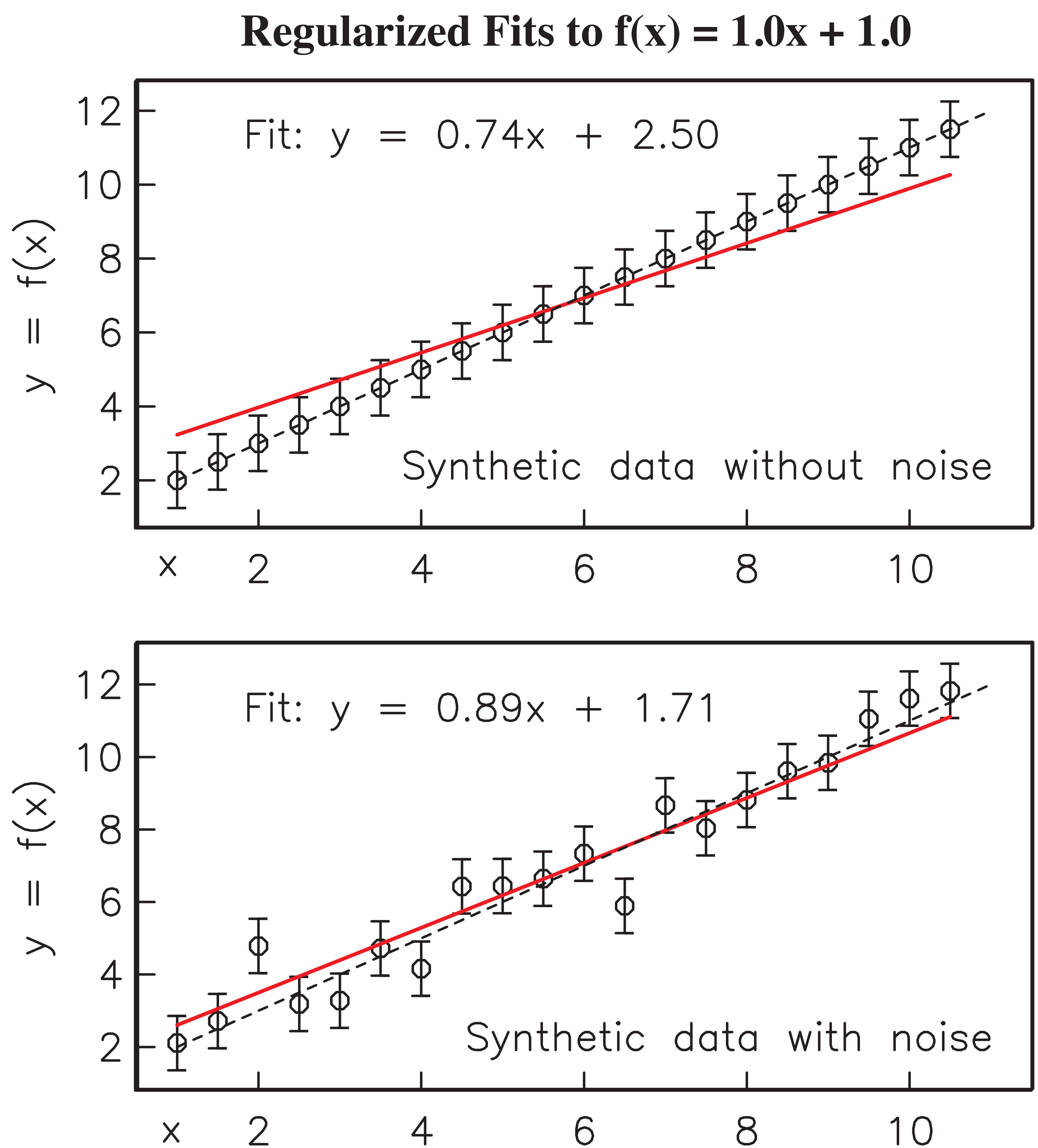
Stochastic methods provide a useful way to assess model uncertainty, but they are still restricted to simple, mostly 1D, models.

More on errors and regularization:

When creating synthetic data for inversion tests, *always* perturb the data with noise, don't just add error bars.

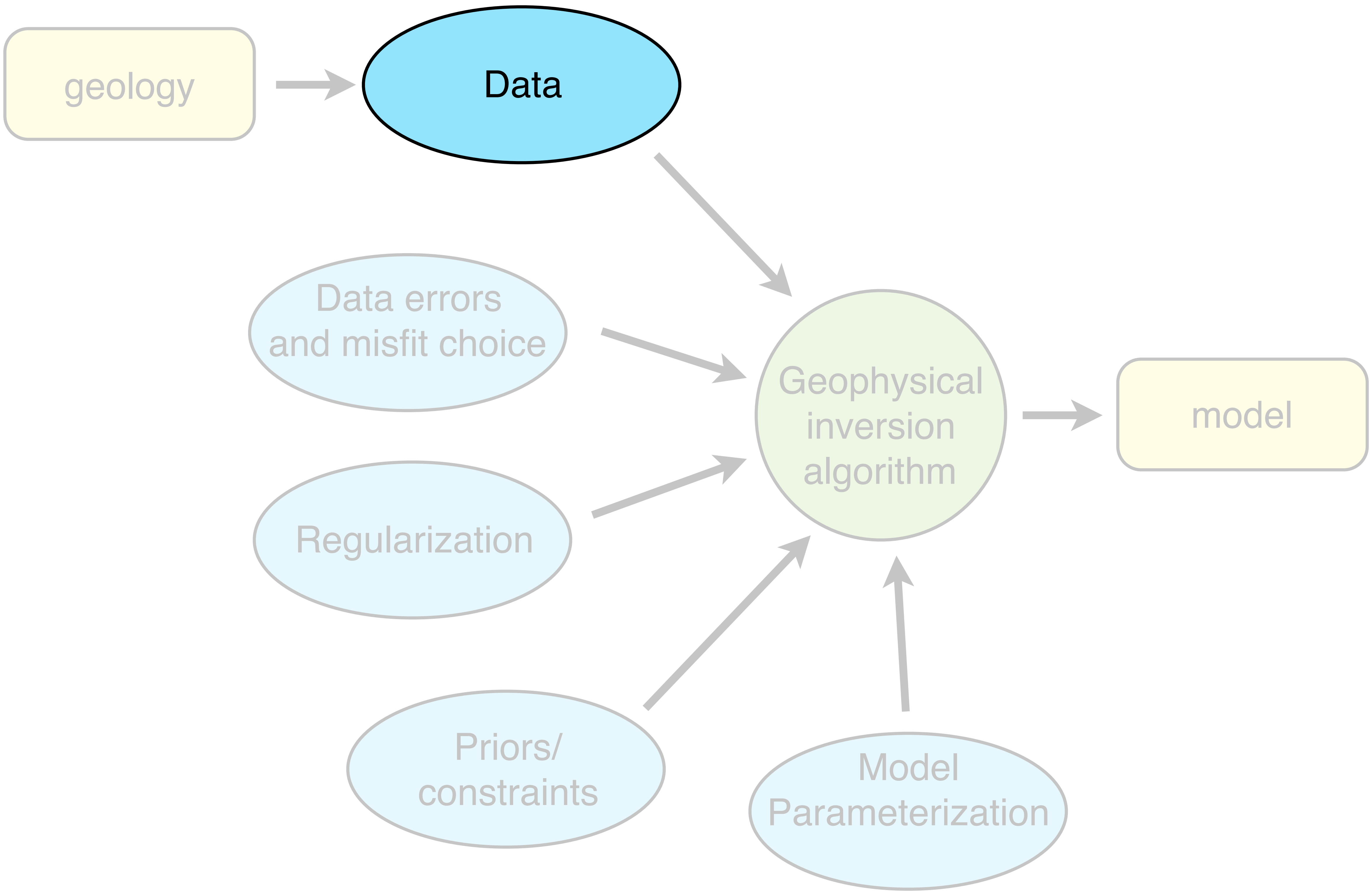
This is because regularized inversion will use its misfit budget to make the model smaller.

You actually get *better* models by adding noise.



Constable, 1991

It can even matter how you scale the data.



In EM, both MT apparent resistivities and CSEM amplitudes can vary by many orders of magnitude. This suggests that one should use error floors that are percentages.

One might also parameterize the data as logs. For small ϵ:

$$d' \pm 0.434\epsilon = \log_{10}(d \pm \epsilon d)$$

log data data → d'

linear data → d

fractional error → ϵ

where $$0.434 = 1/\ln(10)$$

In EM, both MT apparent resistivities and CSEM amplitudes can vary by many orders of magnitude. This suggests that one should use error floors that are percentages.

One might also parameterize the data as logs. For small ϵ:

$$d' \pm 0.434\epsilon = \log_{10}(d \pm \epsilon d)$$

log data data

linear data

fractional error

where

$$0.434 = 1/\ln(10)$$

It ought not to matter how you parameterize the data (so long as the errors are properly scaled and the appropriate chain rule is applied to the Jacobian):

$$\mathbf{m}_1 = \left[\mu \mathbf{R}^T\mathbf{R} + (\mathbf{WJ})^T\mathbf{WJ}\right]^{-1}(\mathbf{WJ})^T\mathbf{W}(\mathbf{d} - f(\mathbf{m_0}) + \mathbf{J}\mathbf{m}_0) \quad .$$

make data log

take the log of the forward model

use the chain rule to convert $\partial\mathbf{d}/\partial\mathbf{m}$ to $\partial\log(\mathbf{d})/\partial\mathbf{m}$

In EM, both MT apparent resistivities and CSEM amplitudes can vary by many orders of magnitude. This suggests that one should use error floors that are percentages.

One might also parameterize the data as logs. For small ϵ:

$$d' \pm 0.434\epsilon = \log_{10}(d \pm \epsilon d)$$

log data (↑ d') — linear data (↑ d) — fractional error (↑ ϵ)

where $$0.434 = 1/\ln(10)$$

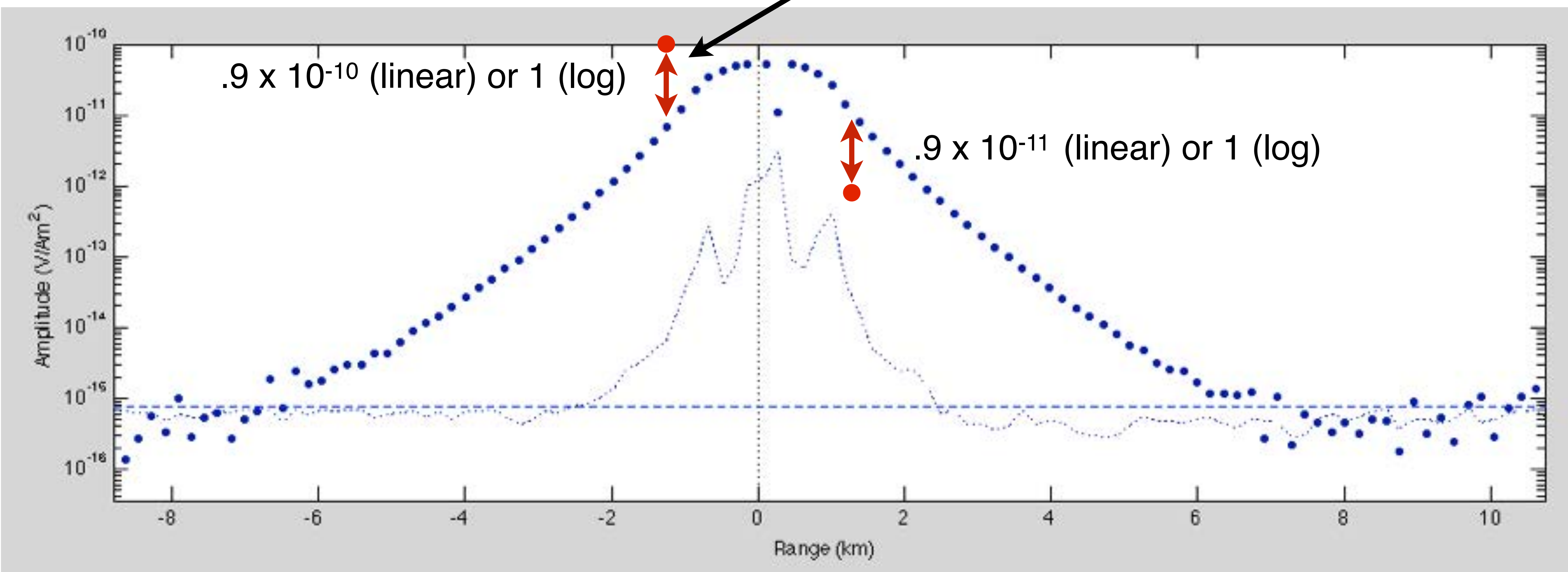
It ought not to matter how you parameterize the data (so long as the errors are properly scaled and the appropriate chain rule is applied to the Jacobian):

$$\mathbf{m}_1 = \left[\mu \mathbf{R}^T\mathbf{R} + (\mathbf{WJ})^T\mathbf{WJ}\right]^{-1}(\mathbf{WJ})^T\mathbf{W}(\mathbf{d} - f(\mathbf{m_0}) + \mathbf{J}\mathbf{m}_0) \quad .$$

But ...

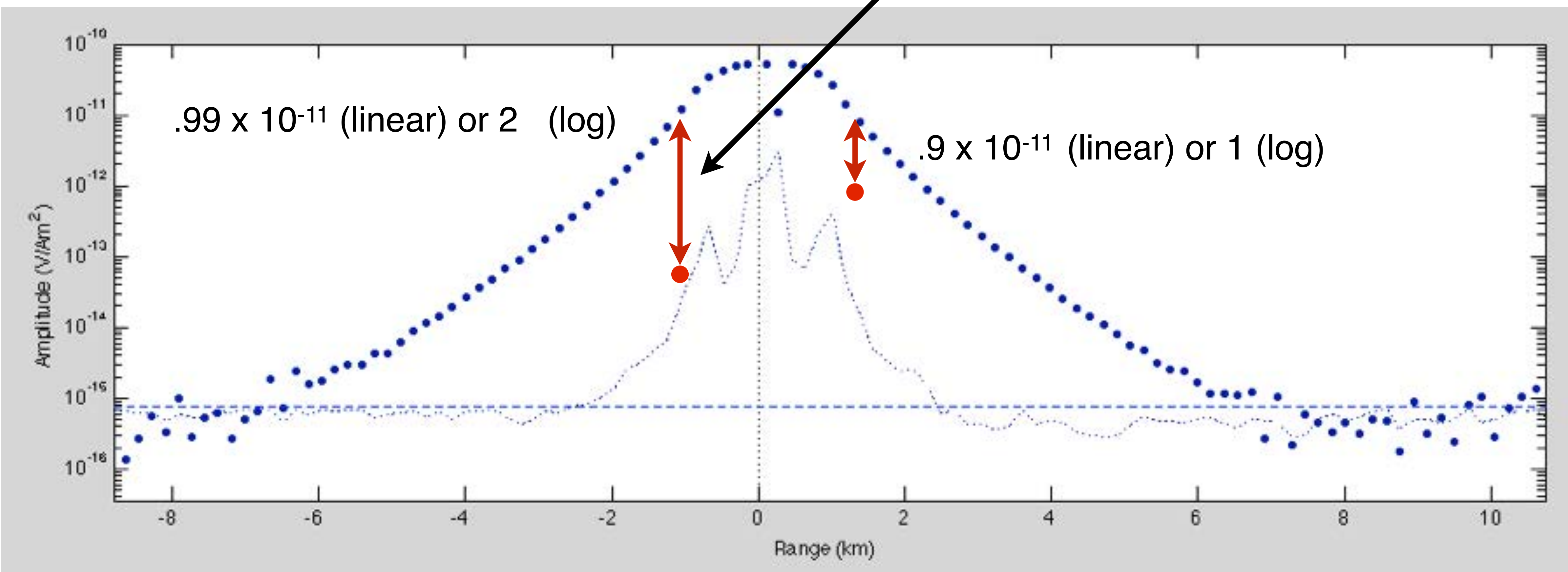
... it does. Consider misfits in marine CSEM data:

in the linear domain, this misfit is 10 times bigger than the other one



... it does. Consider misfits in marine CSEM data:

in the linear domain, this misfit is only 10% bigger than the other one



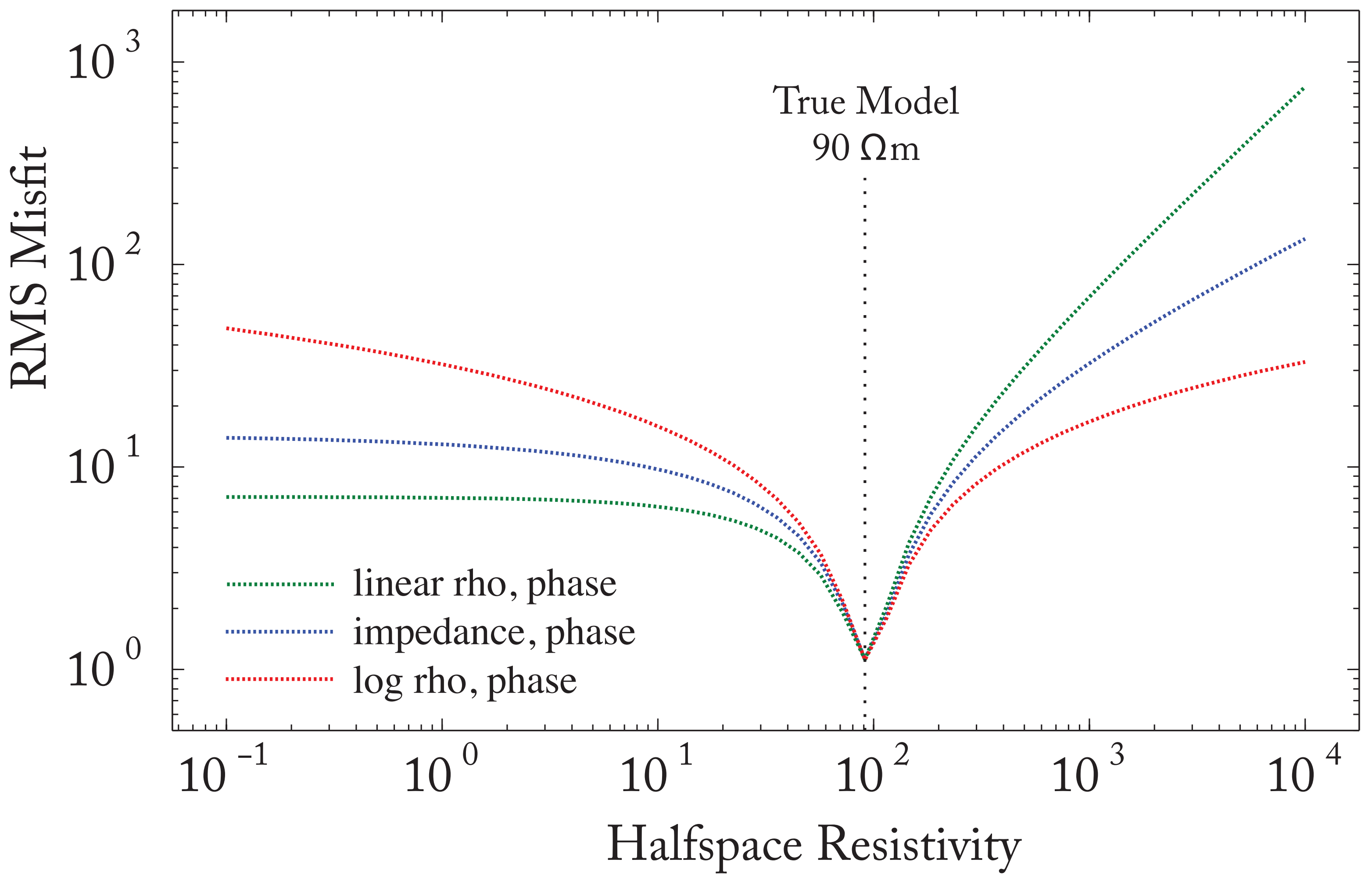
Here we consider MT data over a half-space, varying only half-space resistivity R.

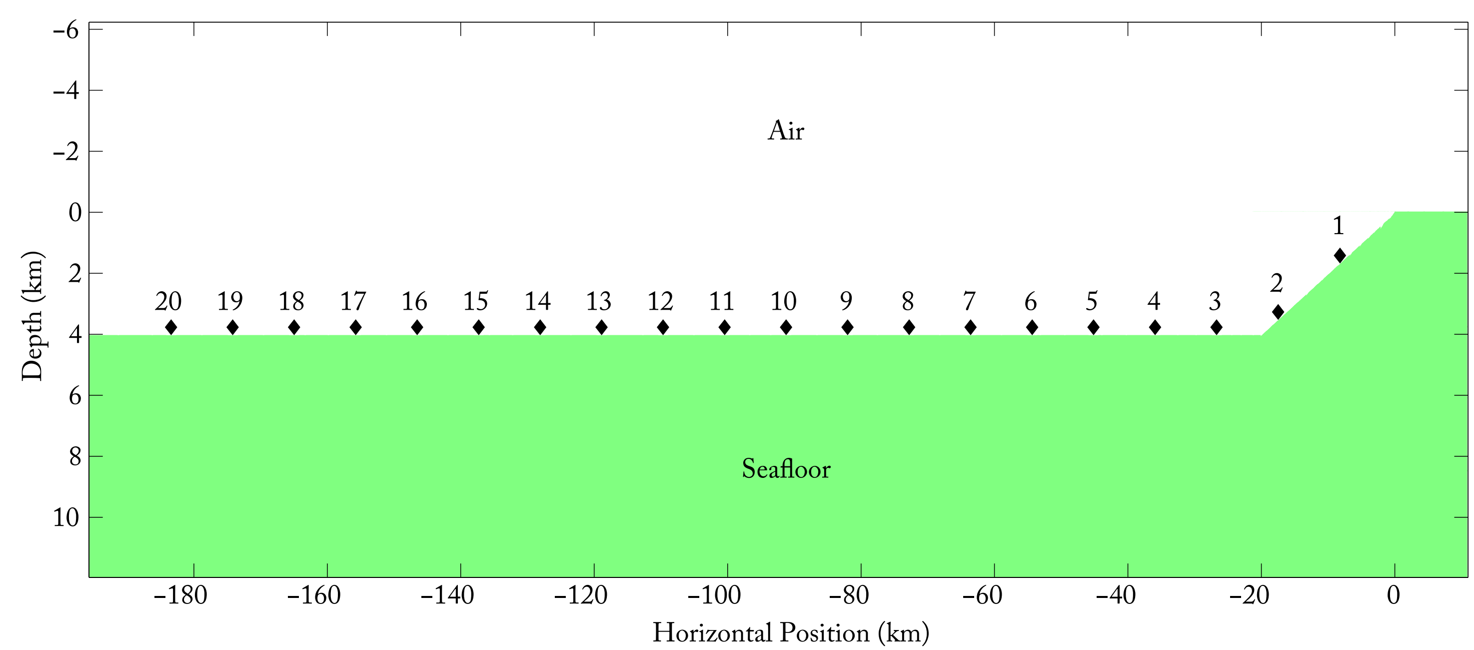
Recall that MT impedance (Z) is:

$$E_x = ZH_y$$

$$\rho = \frac{1}{2\pi f \mu} |Z|^2$$

For small R, misfit flattens for linear ρ





(modified from Wheelock et al., 2015)

Here we consider MT data over a half-space, varying only half-space resistivity R.

Recall that MT impedance (Z) is:

$$E_x = ZH_y$$

$$\rho = \frac{1}{2\pi f \mu} |Z|^2$$

For small R, misfit flattens for linear ρ

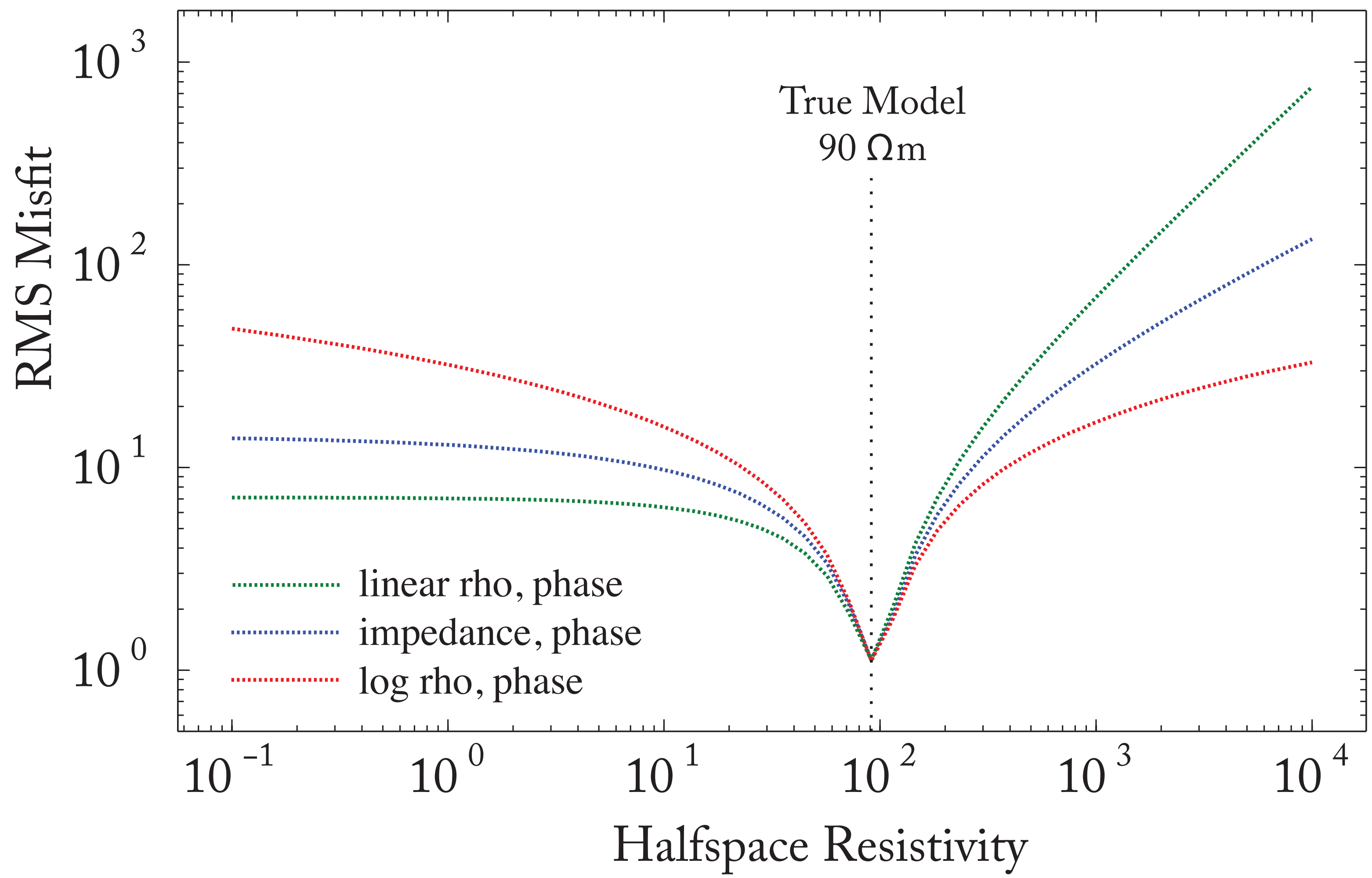
This is because

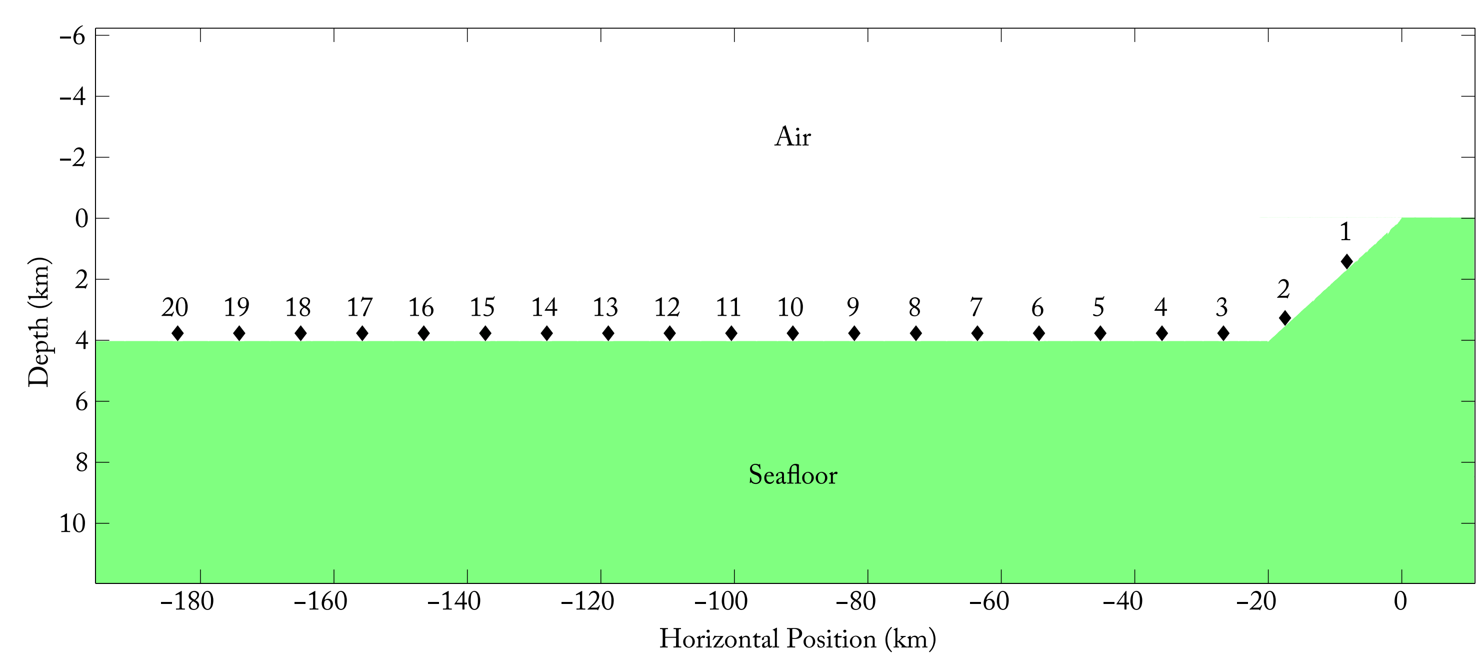
$$(\mathbf{d} - f(\mathbf{m})) \rightarrow \text{const.}$$

as $f(\mathbf{m}) \rightarrow 0$ but

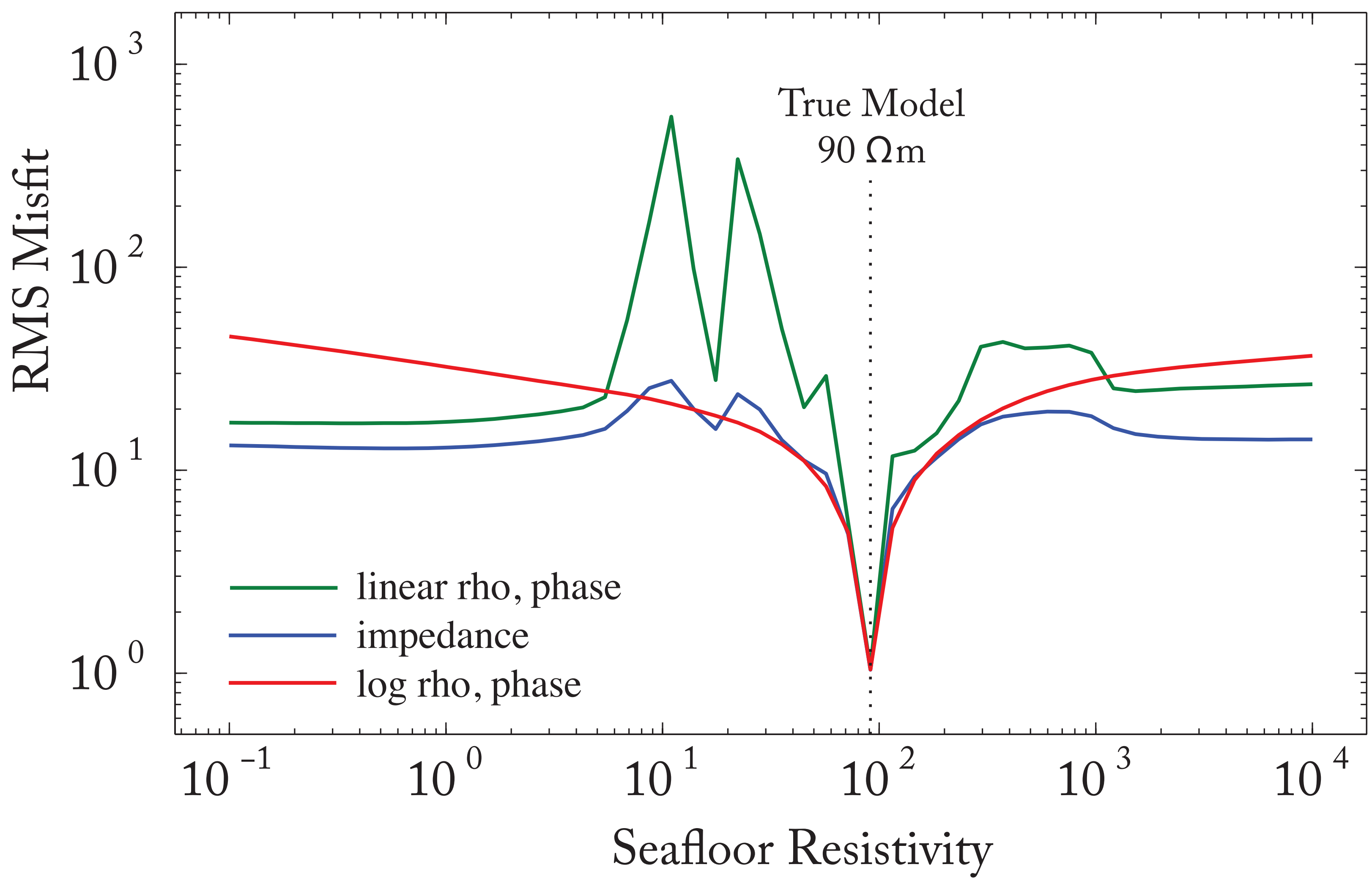
$$(\log \mathbf{d} - \log f(\mathbf{m}))$$

does not.



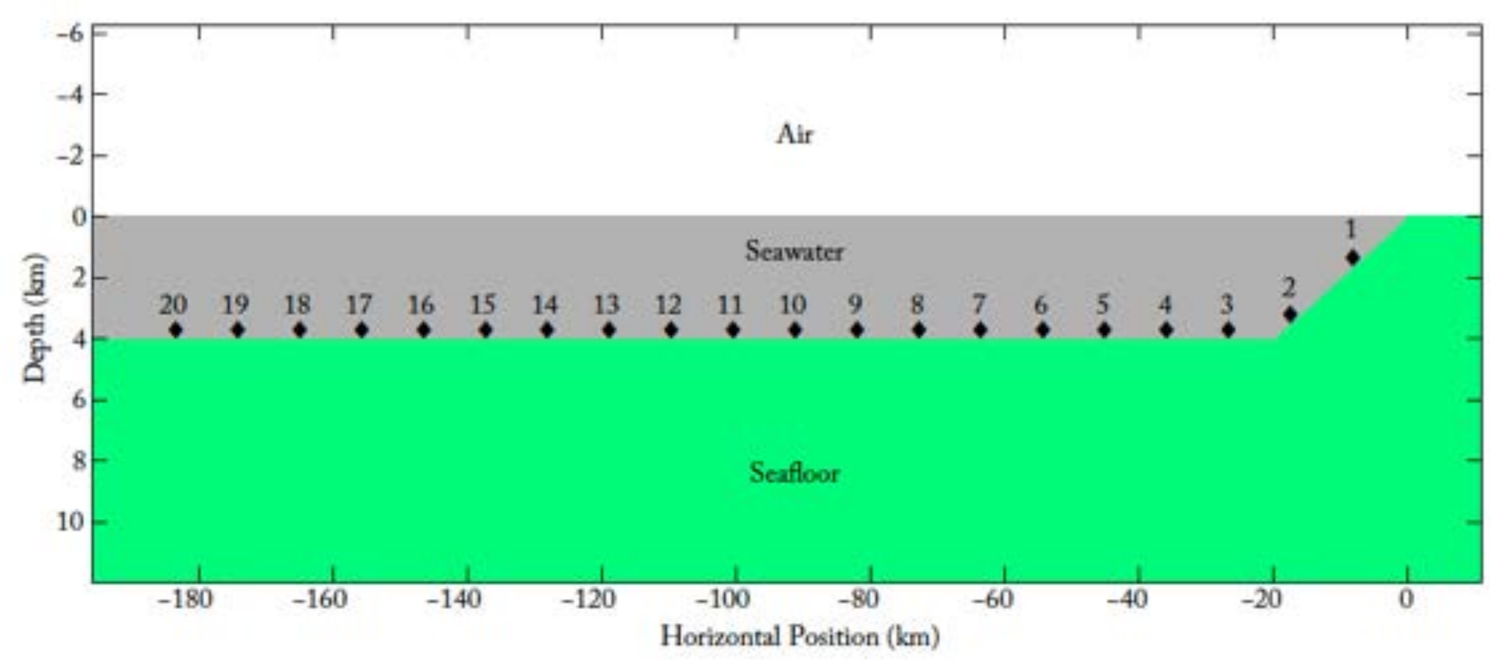


(modified from Wheelock et al., 2015)



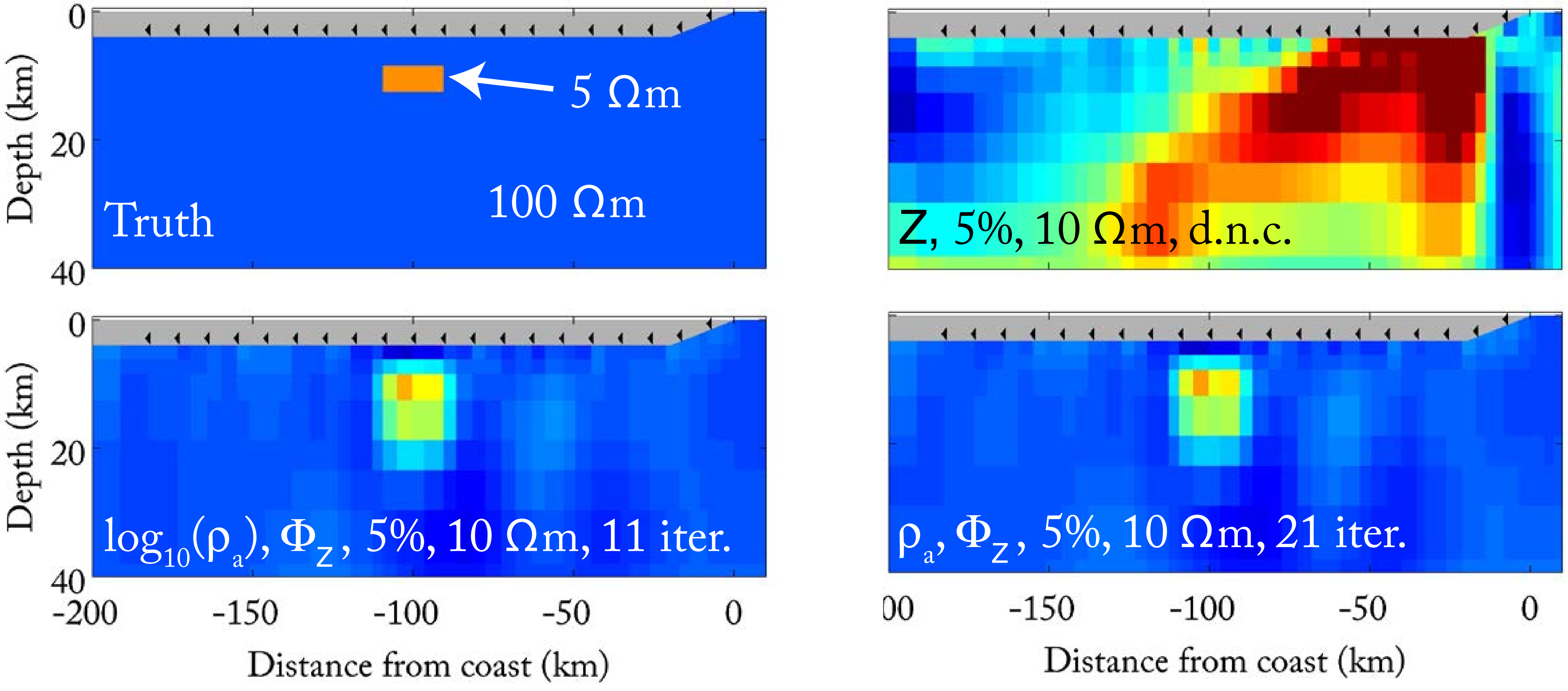
This effect can be even worse for marine MT data affected by bathymetry.

Local minima develop, and misfit flatlines at low R.



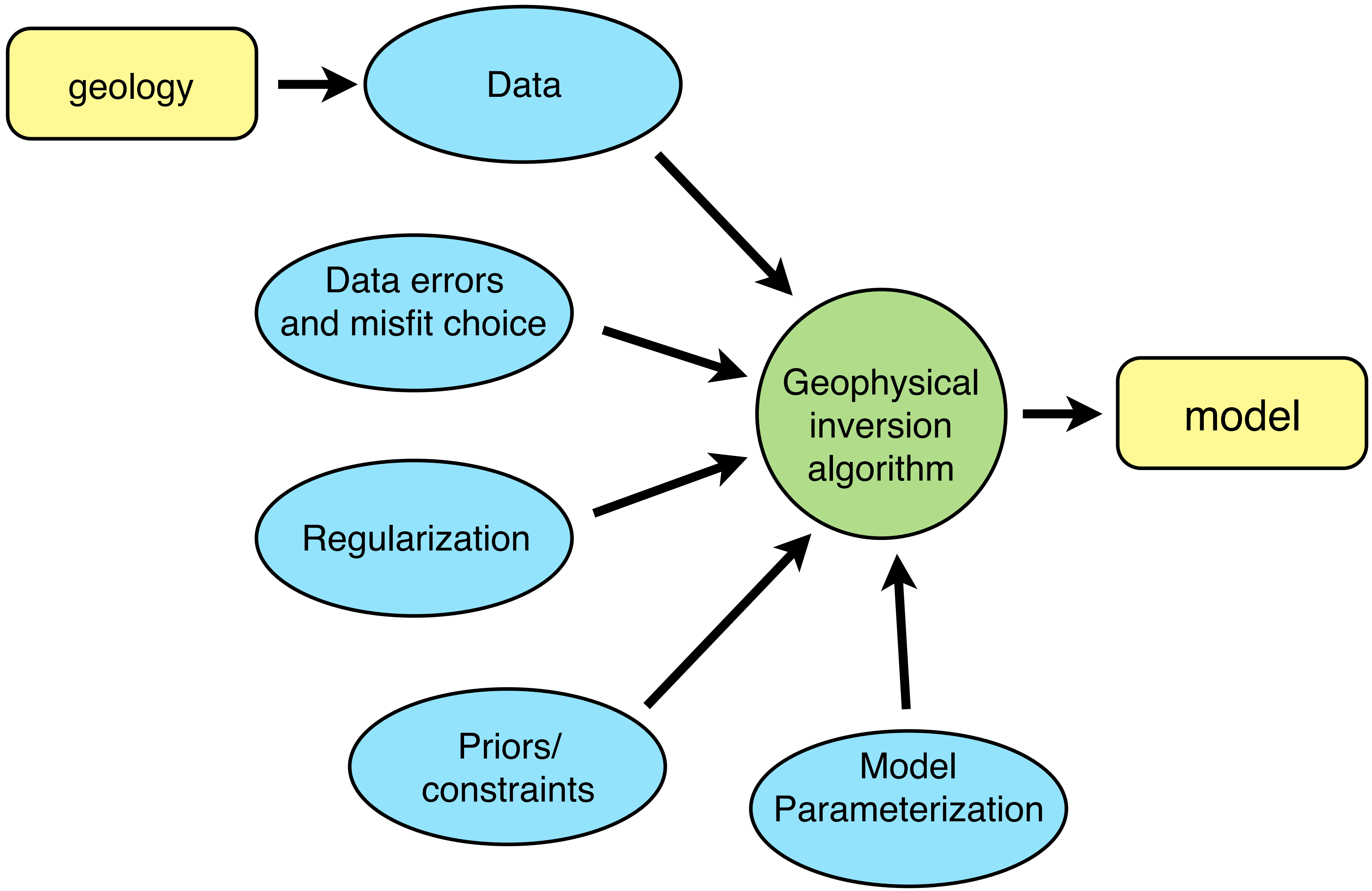
(modified from Wheelock et al., 2015)

Here is a 2D example. Linear apparent resistivity and phase converged, but log(resistivity) converges to the same model in half the iterations. MT impedance, Z, did not converge (d.n.c) at all.

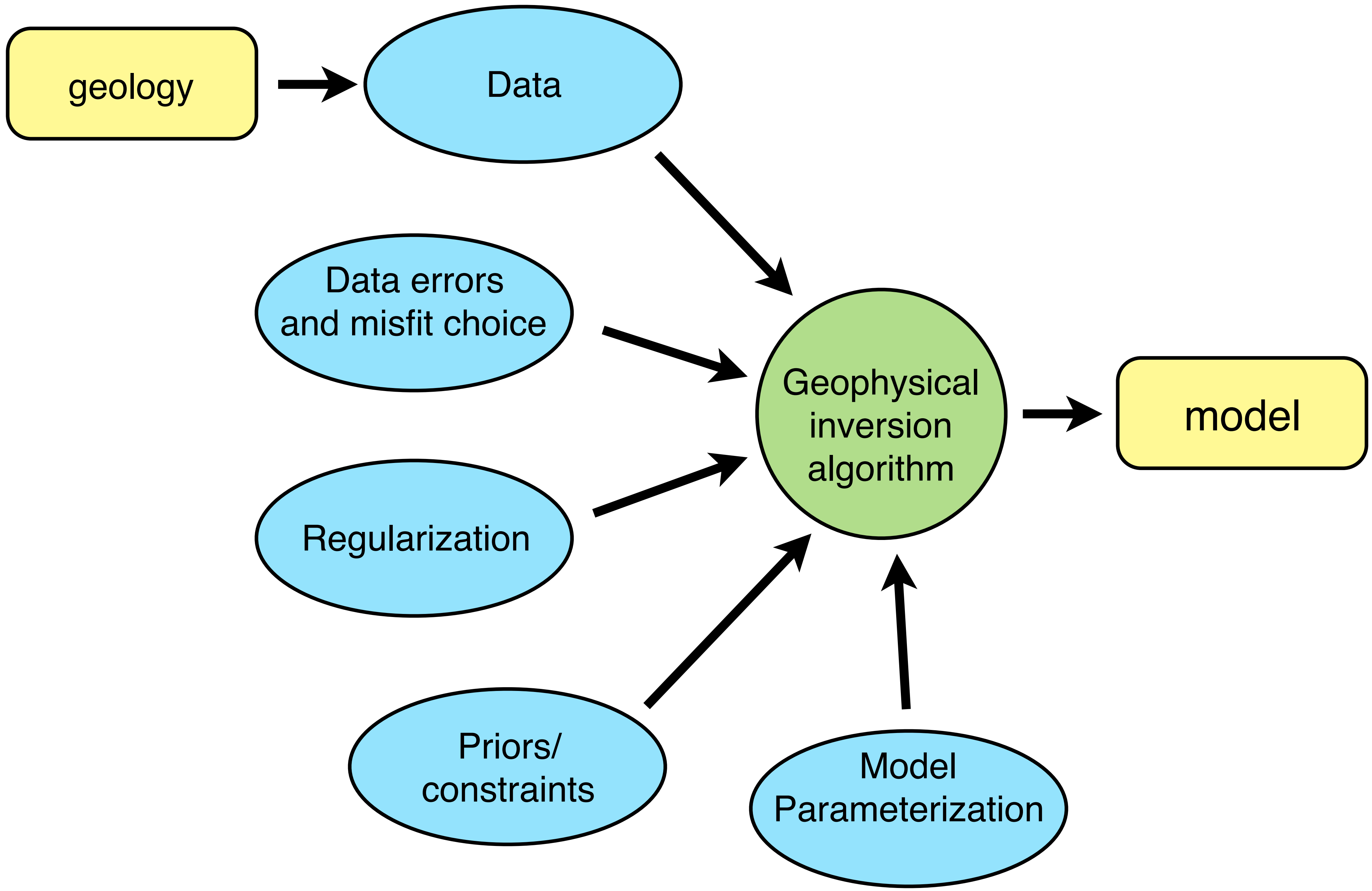


(modified from Wheelock et al., 2015)

So I hope I have convinced you that models from geophysical inversion depend on much more than the data alone:



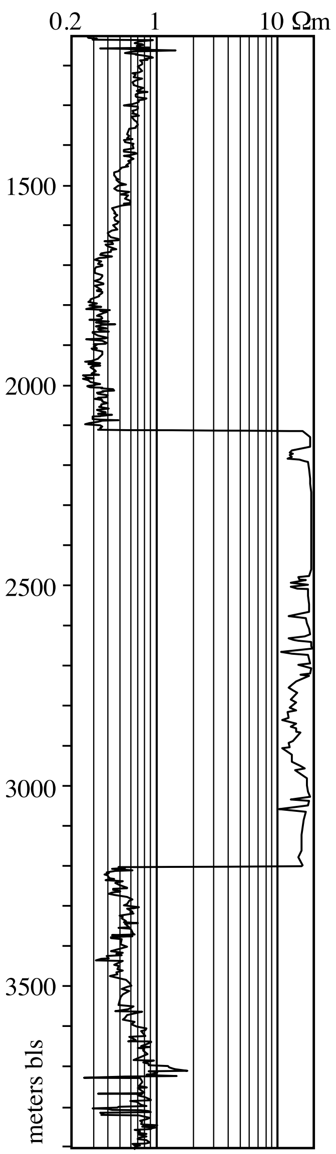
So I hope I have convinced you that models from geophysical inversion depend on much more than the data alone:

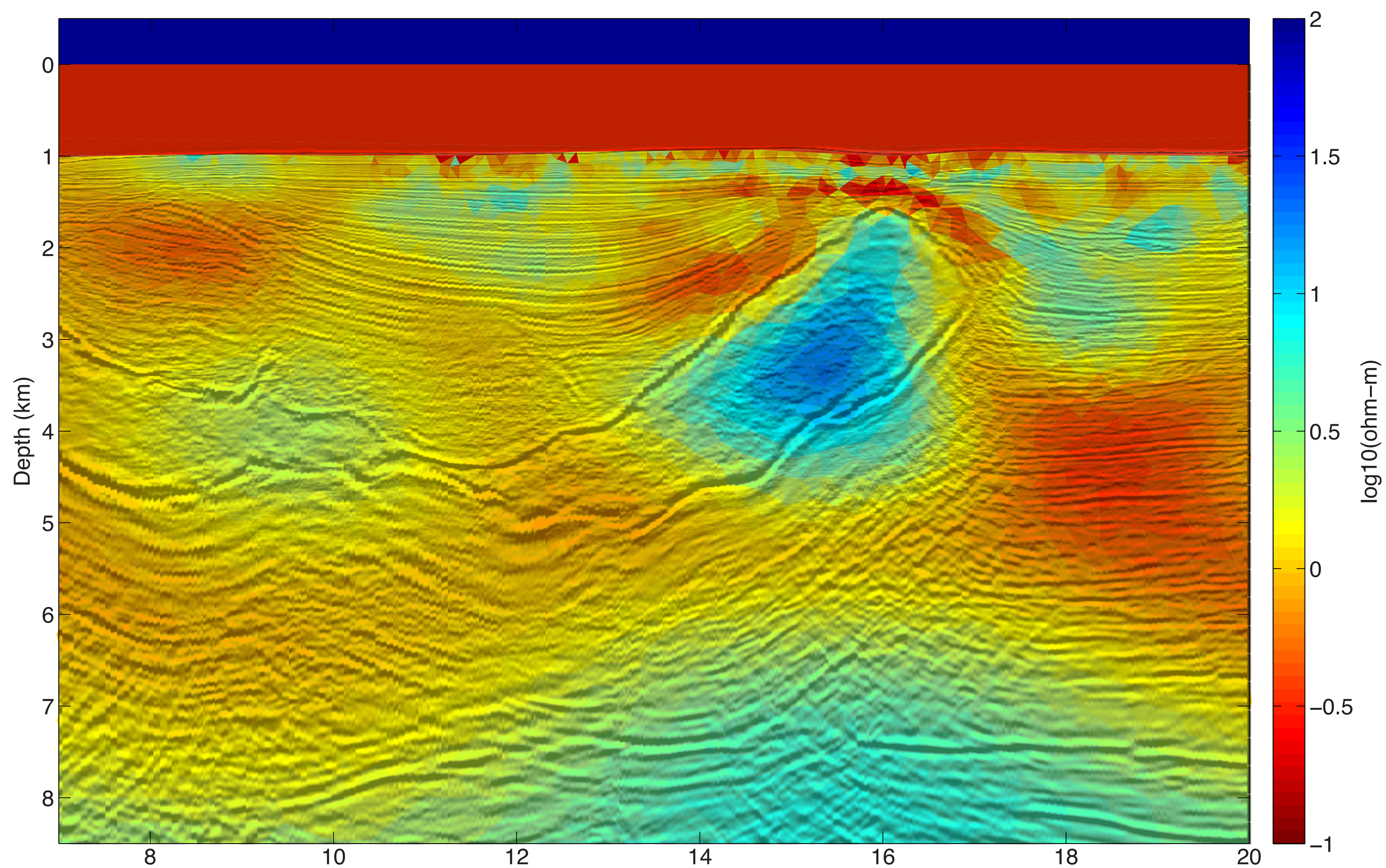


Does this mean that geophysical inversion is useless?

Not at all! There is plenty of evidence that geophysical inversion works very well. You just have to know what you are doing, how your code and algorithm work, and pay attention to the factors besides the data that determine the result.

And, you will probably need to run more than one inversion… many more.





SEG℠

SOCIETY OF EXPLORATION GEOPHYSICISTS

Connecting the world of applied geophysics