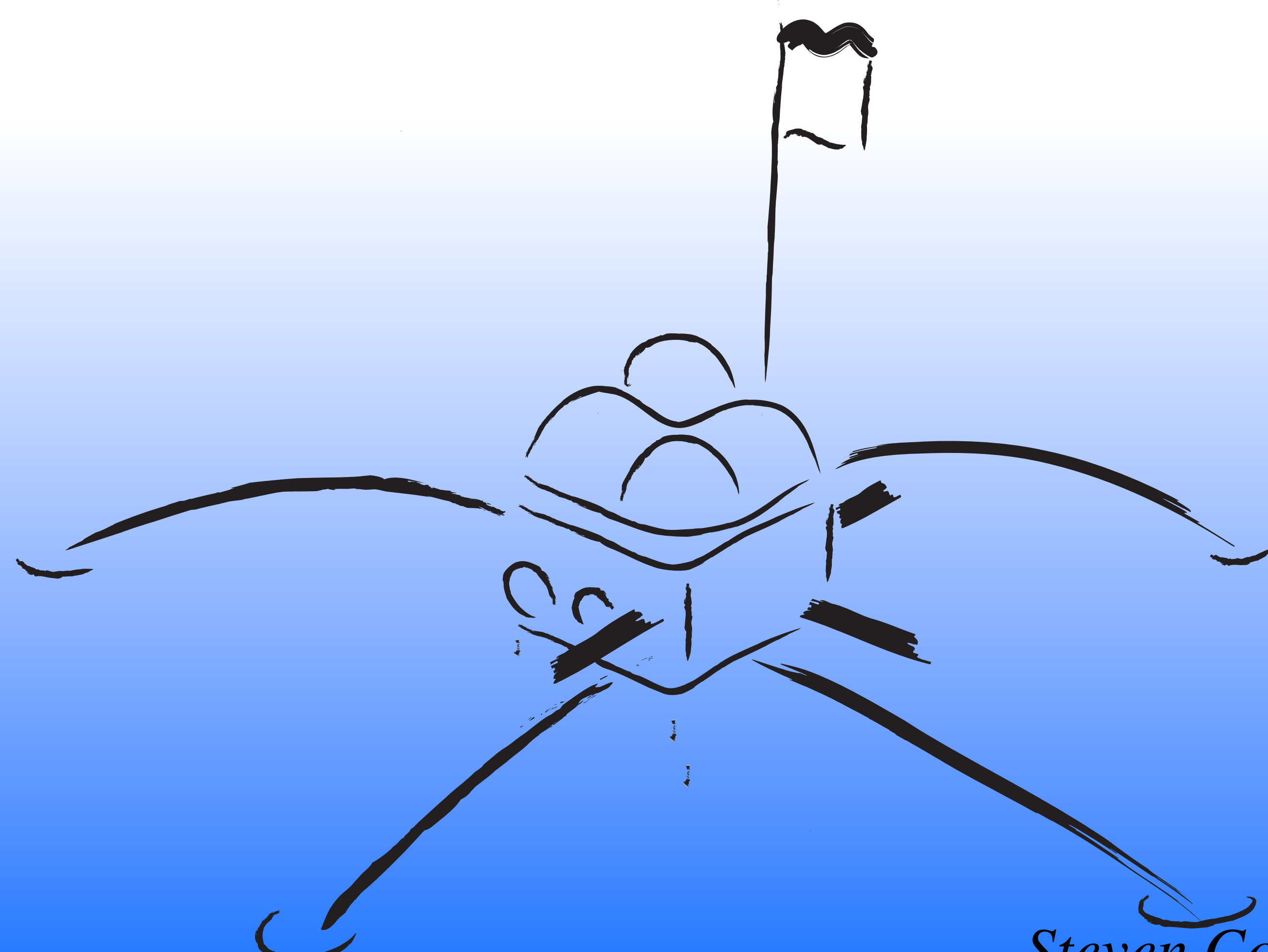
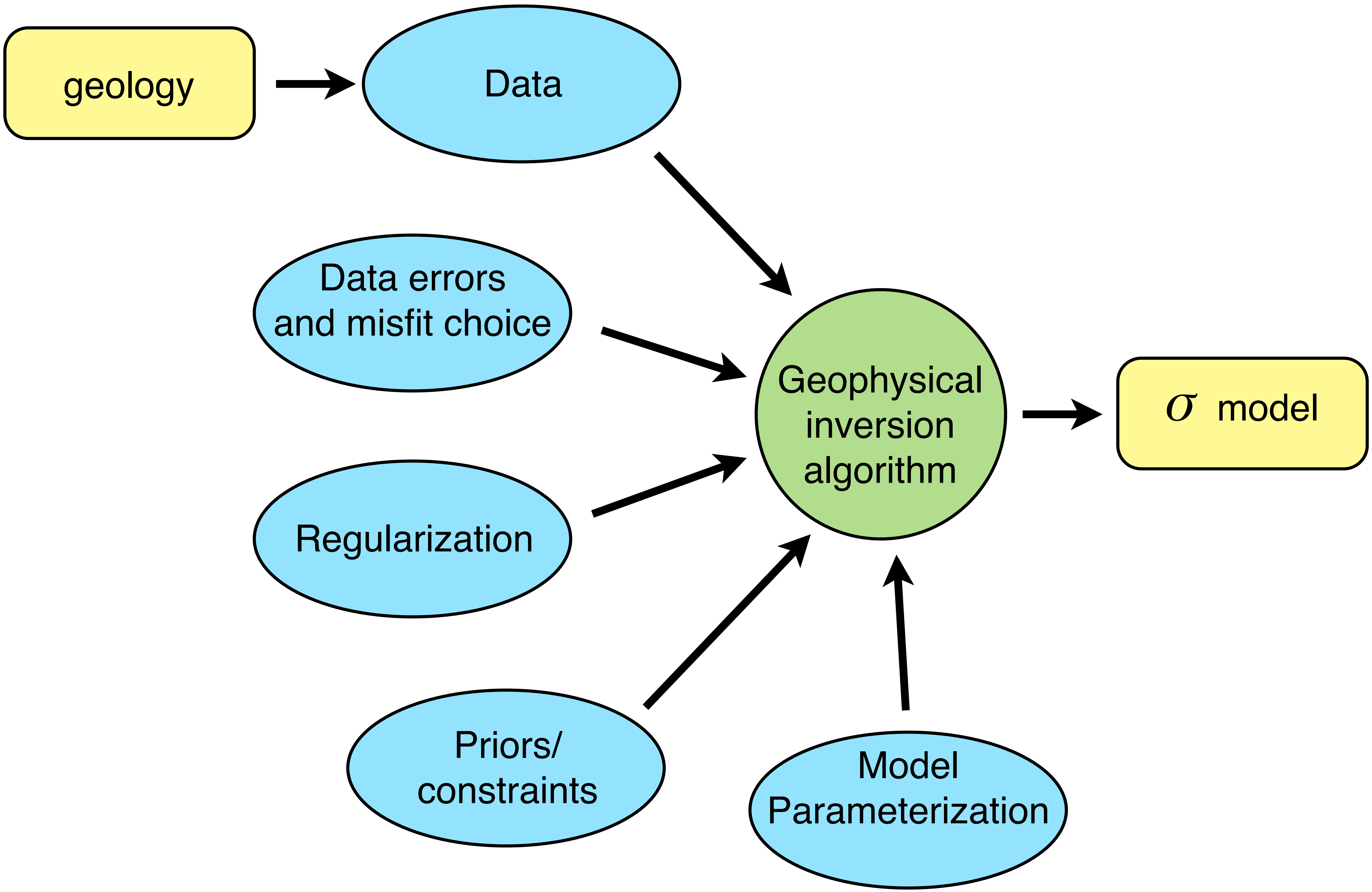
A Young Person's Guide to Regularized Inversion

Steven Constable
Scripps Institution of Oceanography
http://marineemlab.ucsd.edu

with thanks to Kerry Key, Arnold Orange, David Myer, and Brent Wheelock

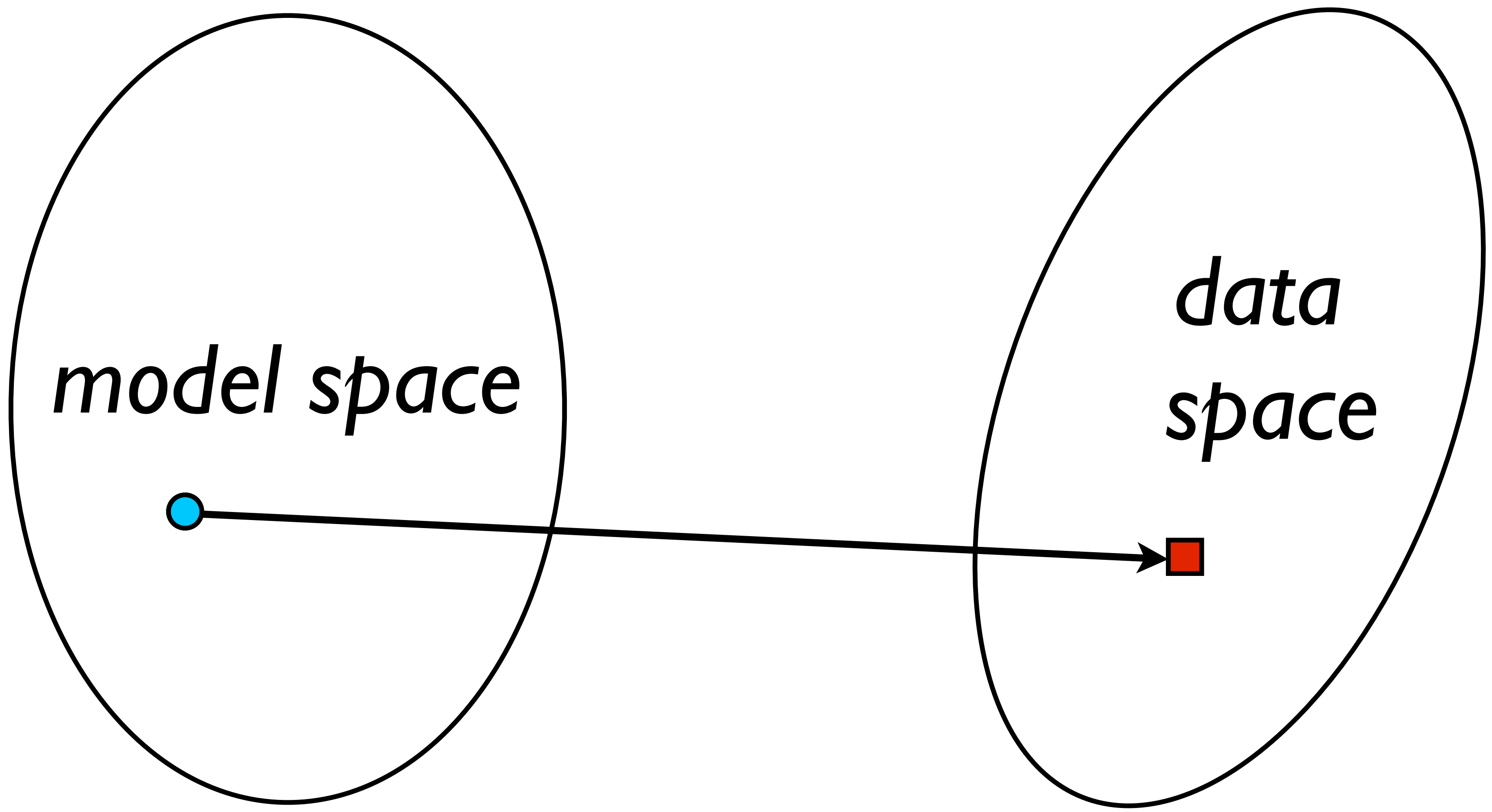


Models from geophysical inversion depend on much more than the data inputs:



Steven Constable, 2013 ASEG workshop

Forward modeling:



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

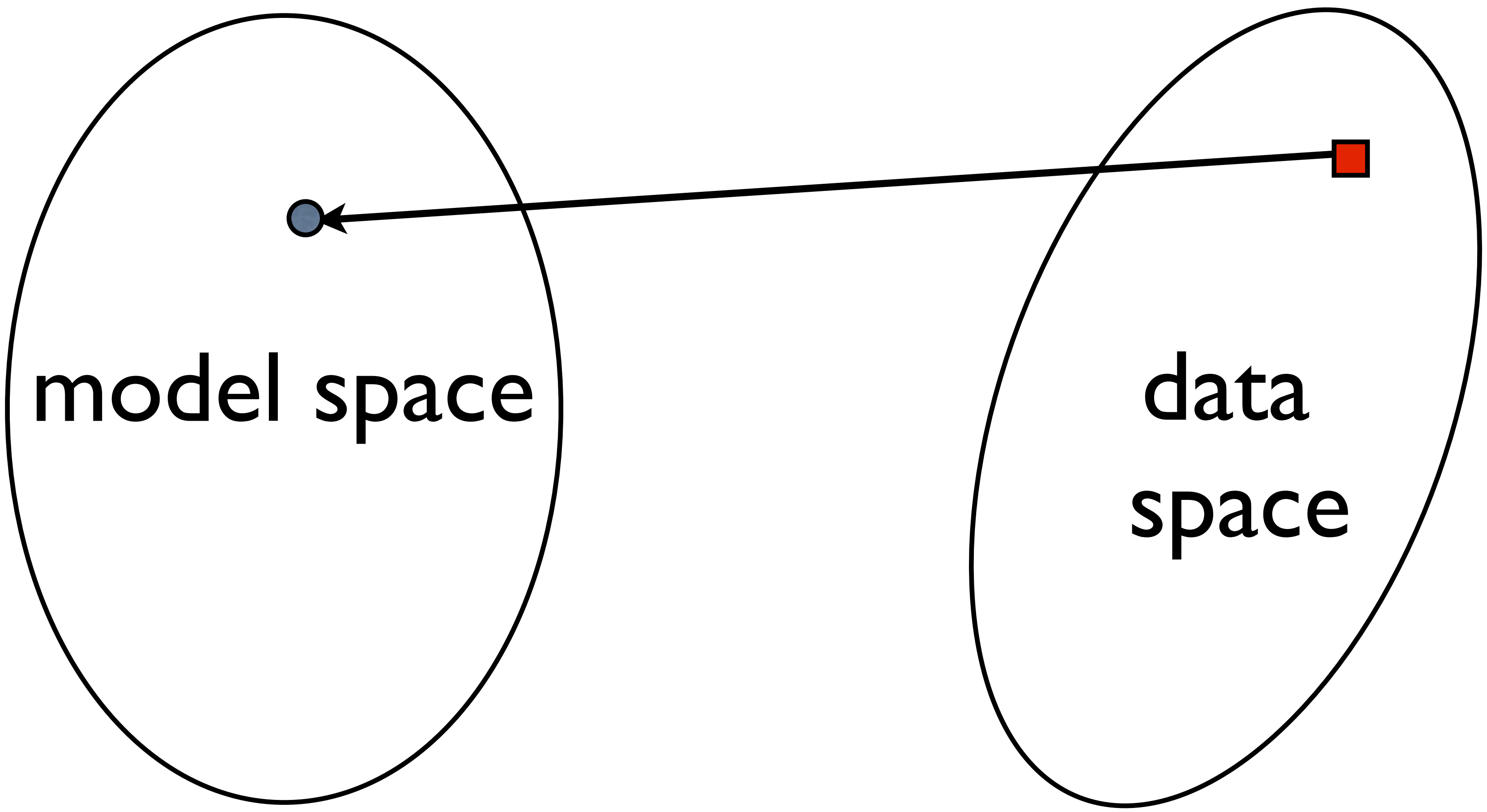
$$\mathbf{m} = (m_1, m_2, ..., m_N)$$ Model parameters (σ of layers, blocks, ...)

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

$$\hat{\mathbf{d}} = (\hat{d}_1, \hat{d}_2, \hat{d}_3, ..., \hat{d}_M)$$ Predicted data (amps., phases, ...)

Steven Constable, 2013 ASEG workshop

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 $\mathbf{m}$

Steven Constable, 2013 ASEG workshop

There are several approaches to inversion:

Stochastic

- Monte Carlo
- Genetic Algorithms
- Bayesian Searches
- Simulated annealing, etc.

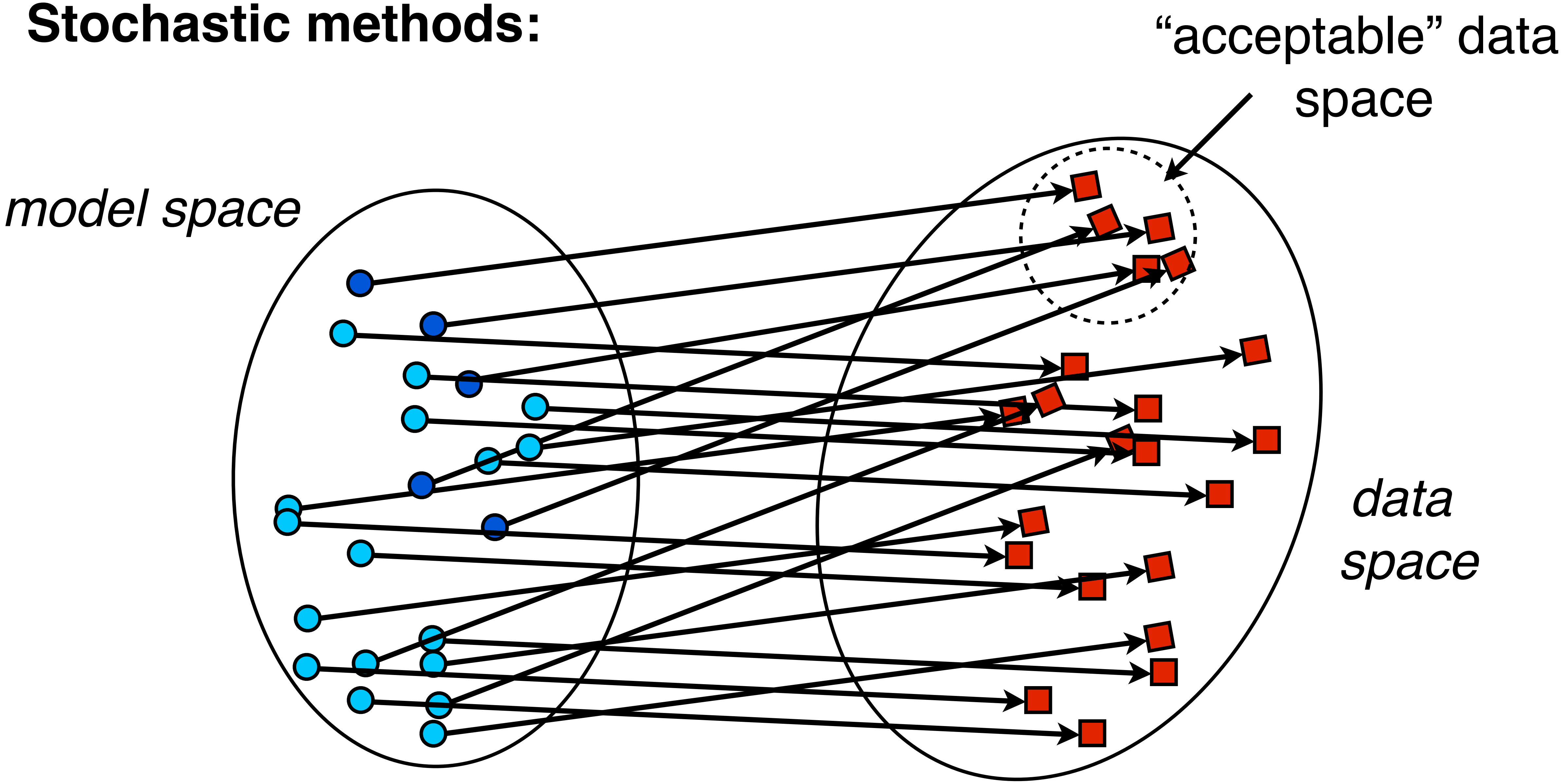
Deterministic

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

Analytical

- D+ (1D MT)
- Bilayer (1D resistivity)

Stochastic methods:



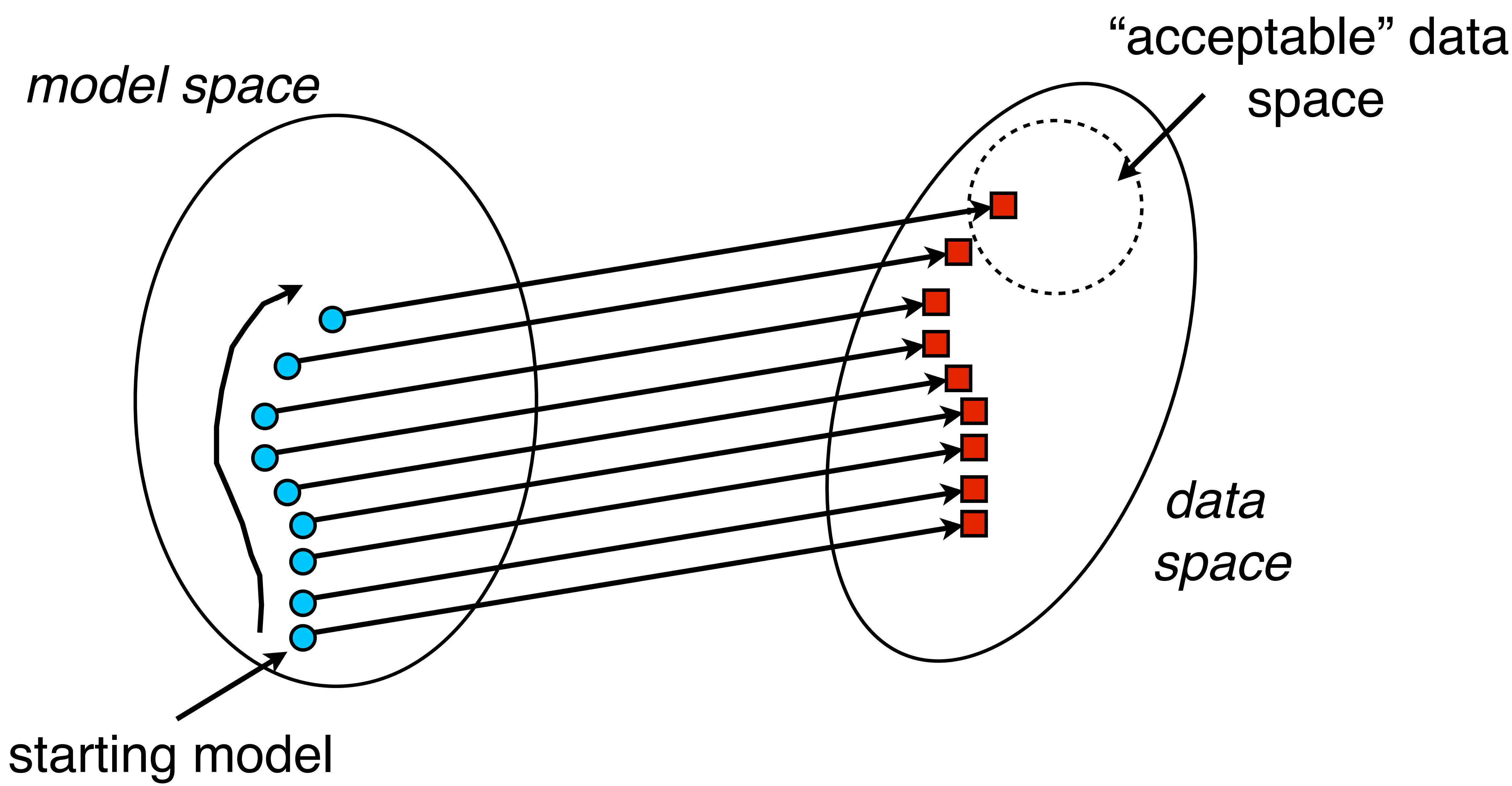
A useful approach, largely restricted to 1D (because millions of models required), with most of the subtlety in model generation methods.

The advantages are that (i) only forward calculations are made and (ii) some probabilities 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.

Steven Constable, 2013 ASEG workshop

Deterministic

Newton Algorithms
Steepest descent
Conjugate Gradients
Quadratic (and Linear) Programming



Steven Constable, 2013 ASEG workshop

Analytical

D+ (1D MT)

Bilayer (1D 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):

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

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

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

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

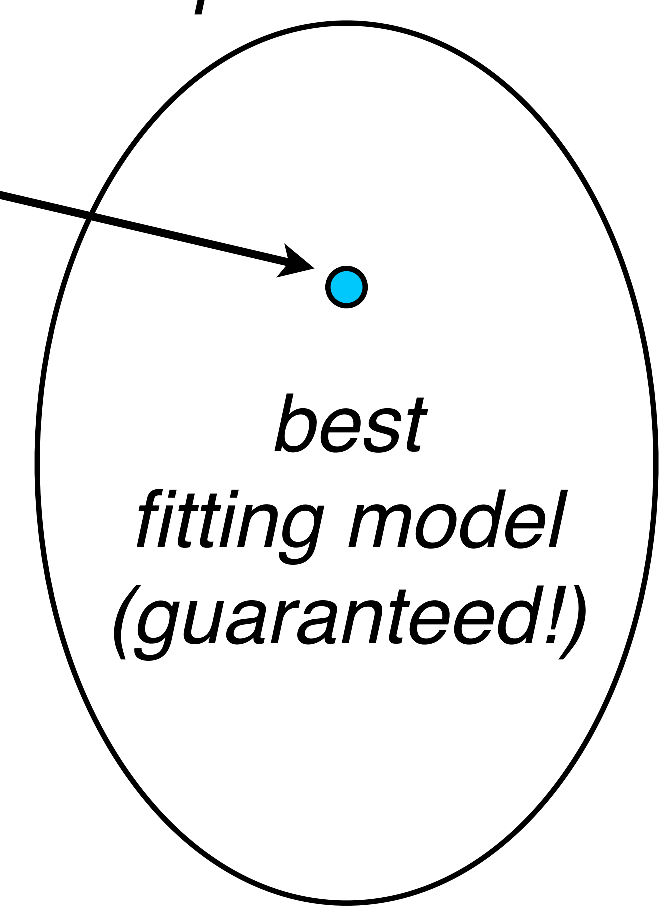
We could solve (48) by recurring upwards in the familiar way, starting with $E(H) = 0 = C_{K+1}$, to get the value of $E(0)$ and hence of $C_1 = c(\omega)$. But now we do something different: we substitute repeatedly from the top, and we get a magnificent **continued fraction** for the admittance:

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

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:

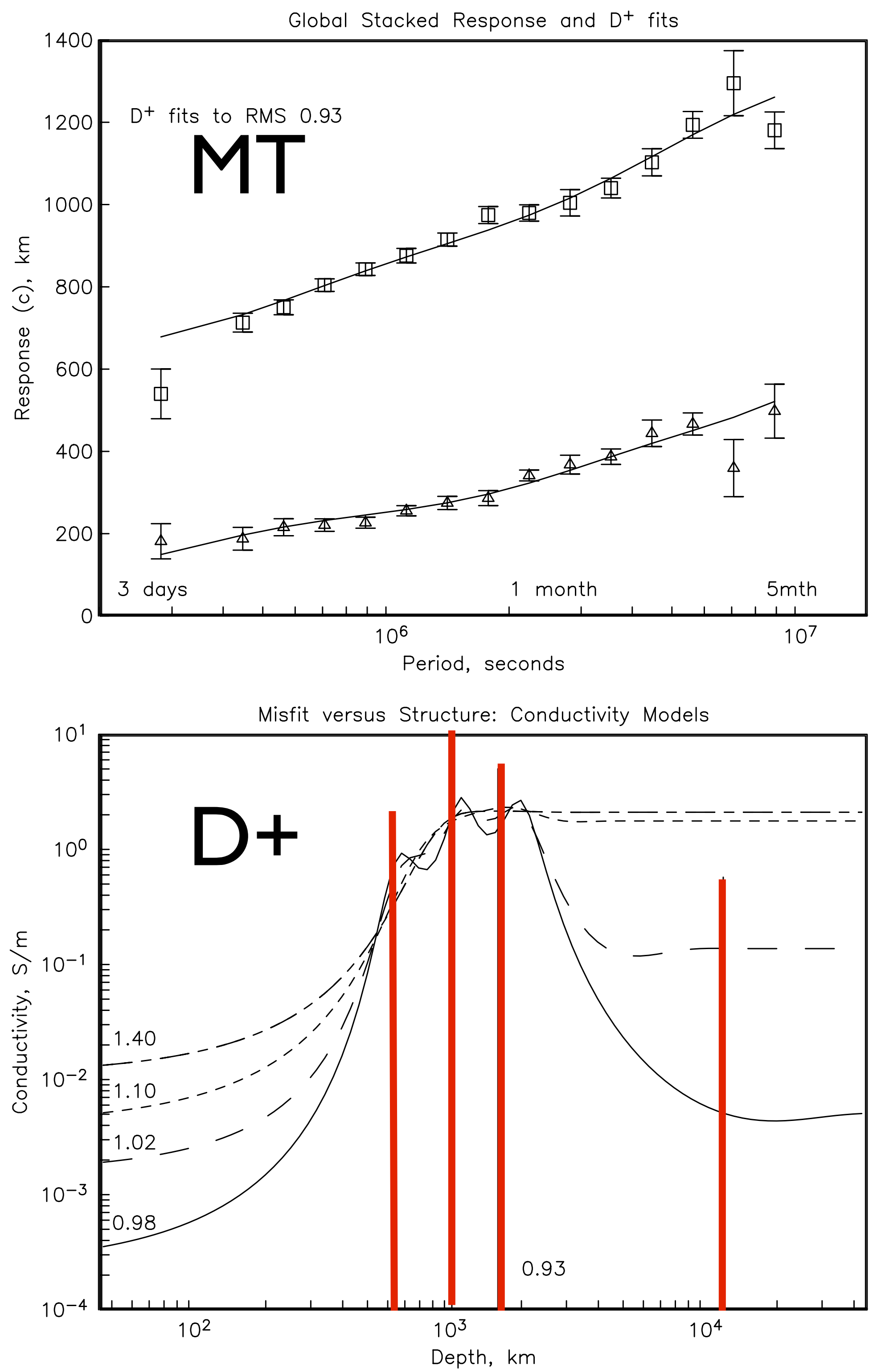
$$c(\omega) = z_1 + \sum_{k=1}^{K} \frac{\alpha_k}{\lambda_k + i\omega} . \tag{59}$$

model space

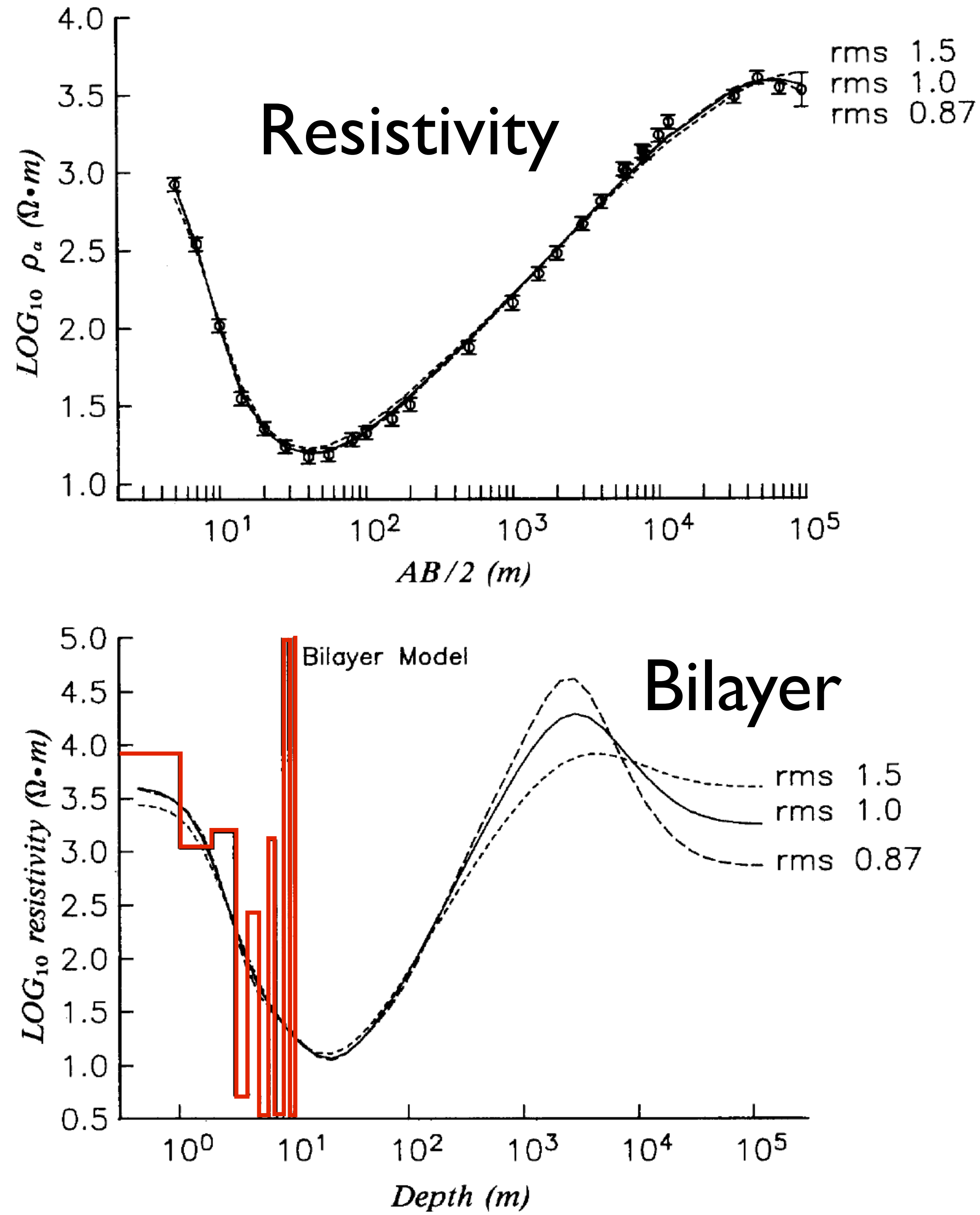


Steven Constable, 2013 ASEG workshop

These solutions are guaranteed best fitting but pathological.



MT: Delta functions of σ



Resistivity: Arbitrarily thin surface layers

Steven Constable, 2013 ASEG workshop

Existence and Uniqueness:

Infinite exact data

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

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

An infinite number of solutions fit the data exactly

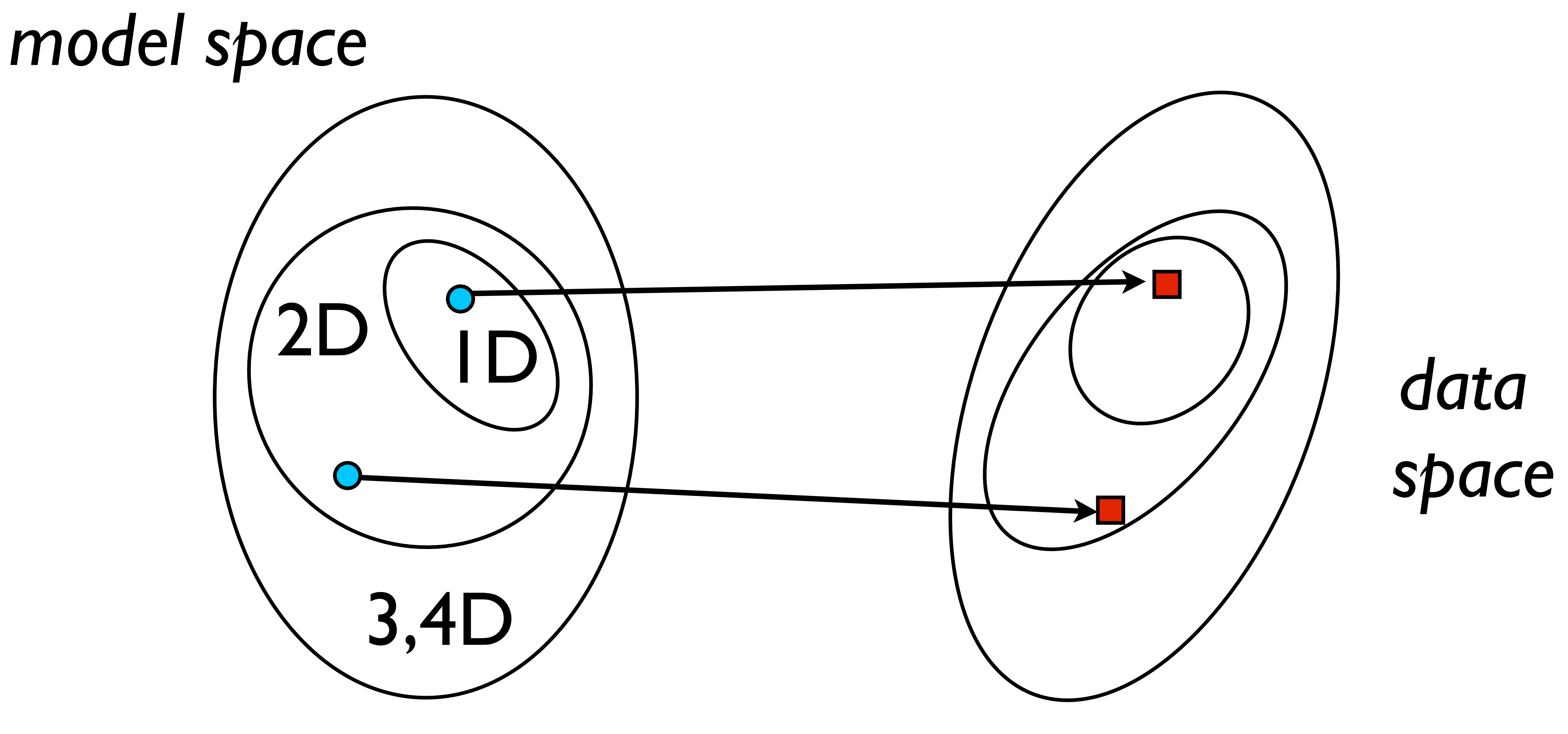
Finite noisy data for a nonlinear problem

Either zero or an infinite number of solutions fit OK

As Sven Treitel puts it, there is no such thing as being a little bit non-unique.

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.



Steven Constable, 2013 ASEG workshop

Sven Treitel 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 σ versus anisotropy)?

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

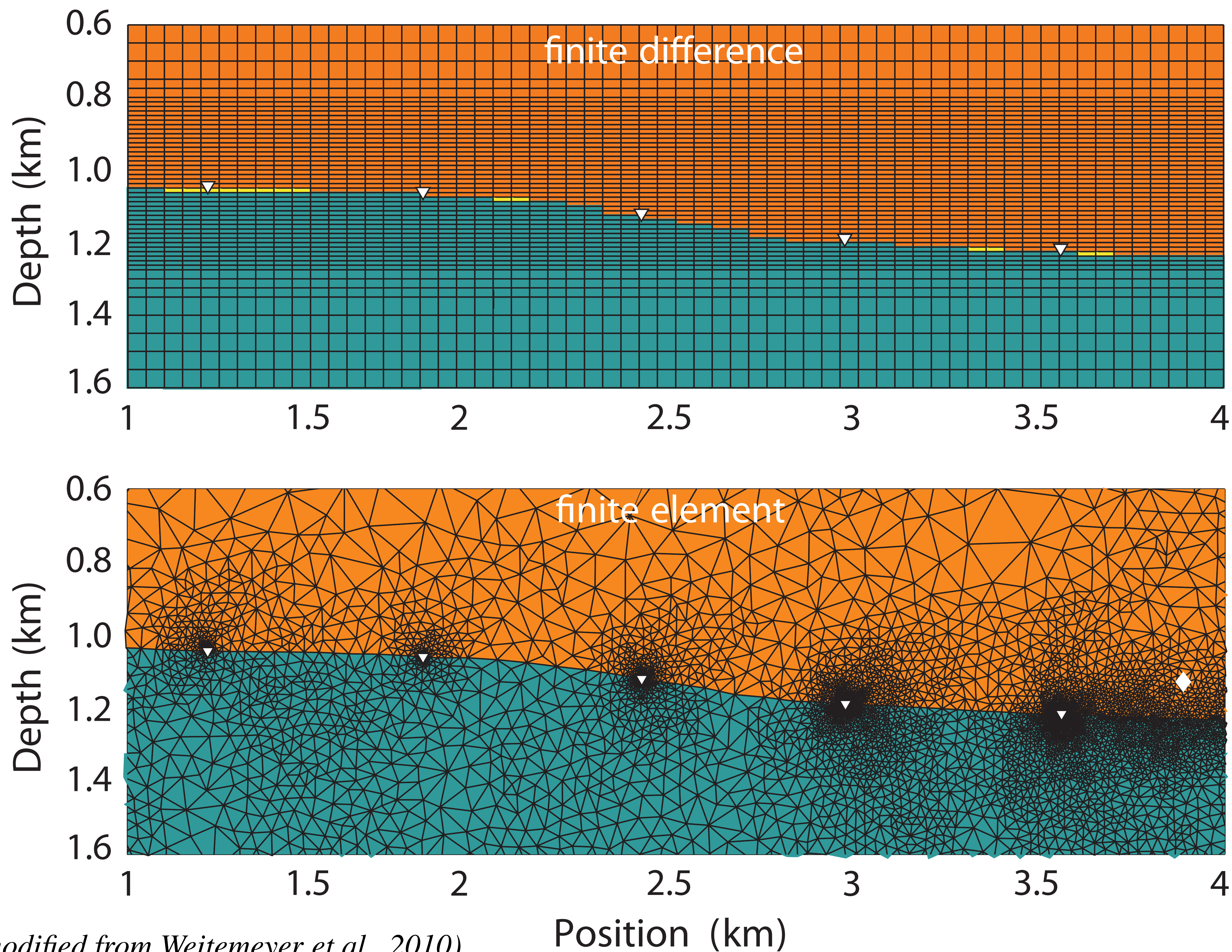
Is the dimensionality of model space large enough? (1D, 2D, 3D, 4D)

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.

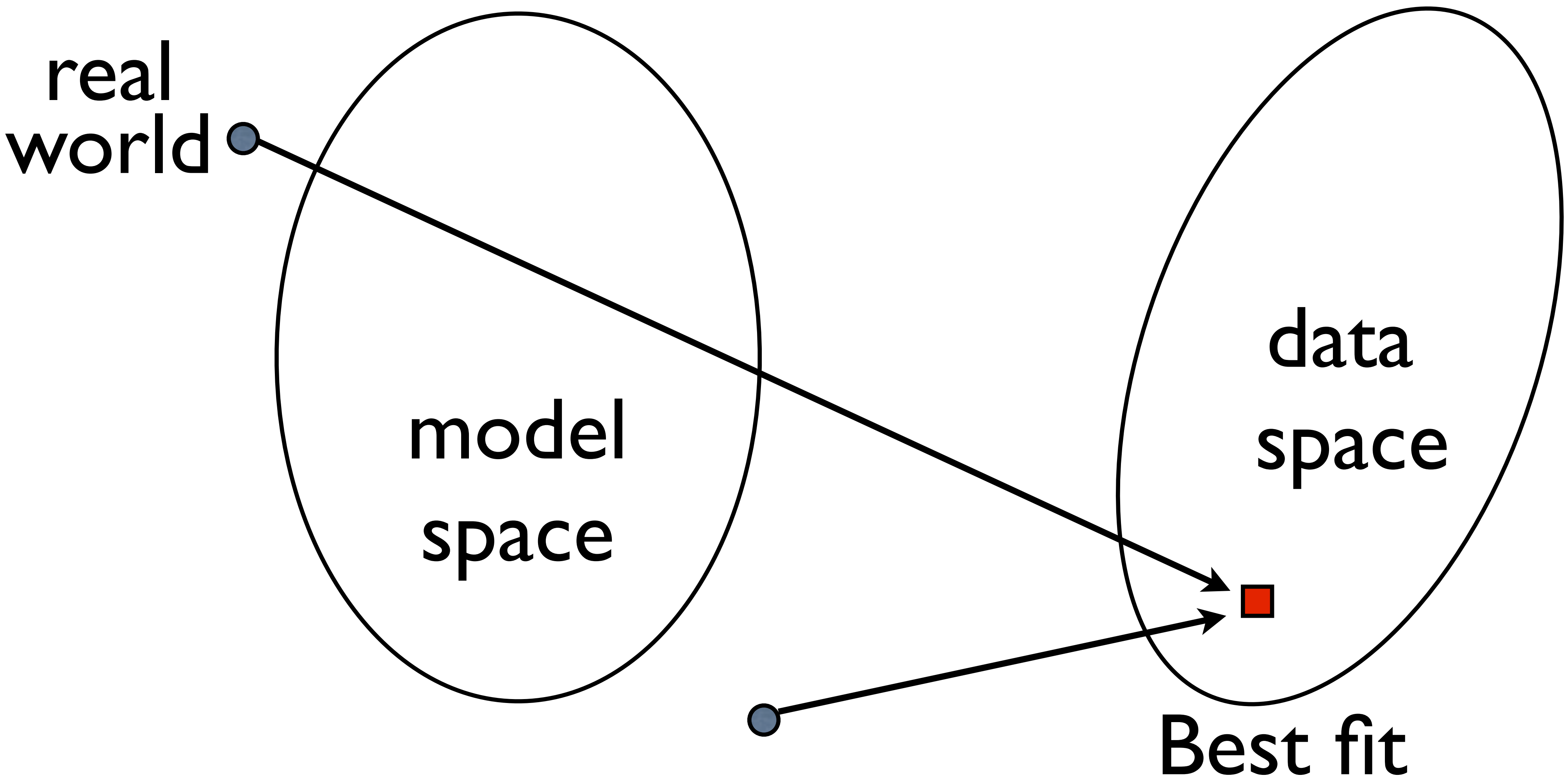
Steven Constable, 2013 ASEG workshop

An example where FE forward calculations were used to validate a FD inversion mesh of bathymetry:



(modified from Weitemeyer et al., 2010)

Steven Constable, 2013 ASEG workshop



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.

Steven Constable, 2013 ASEG workshop

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

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

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

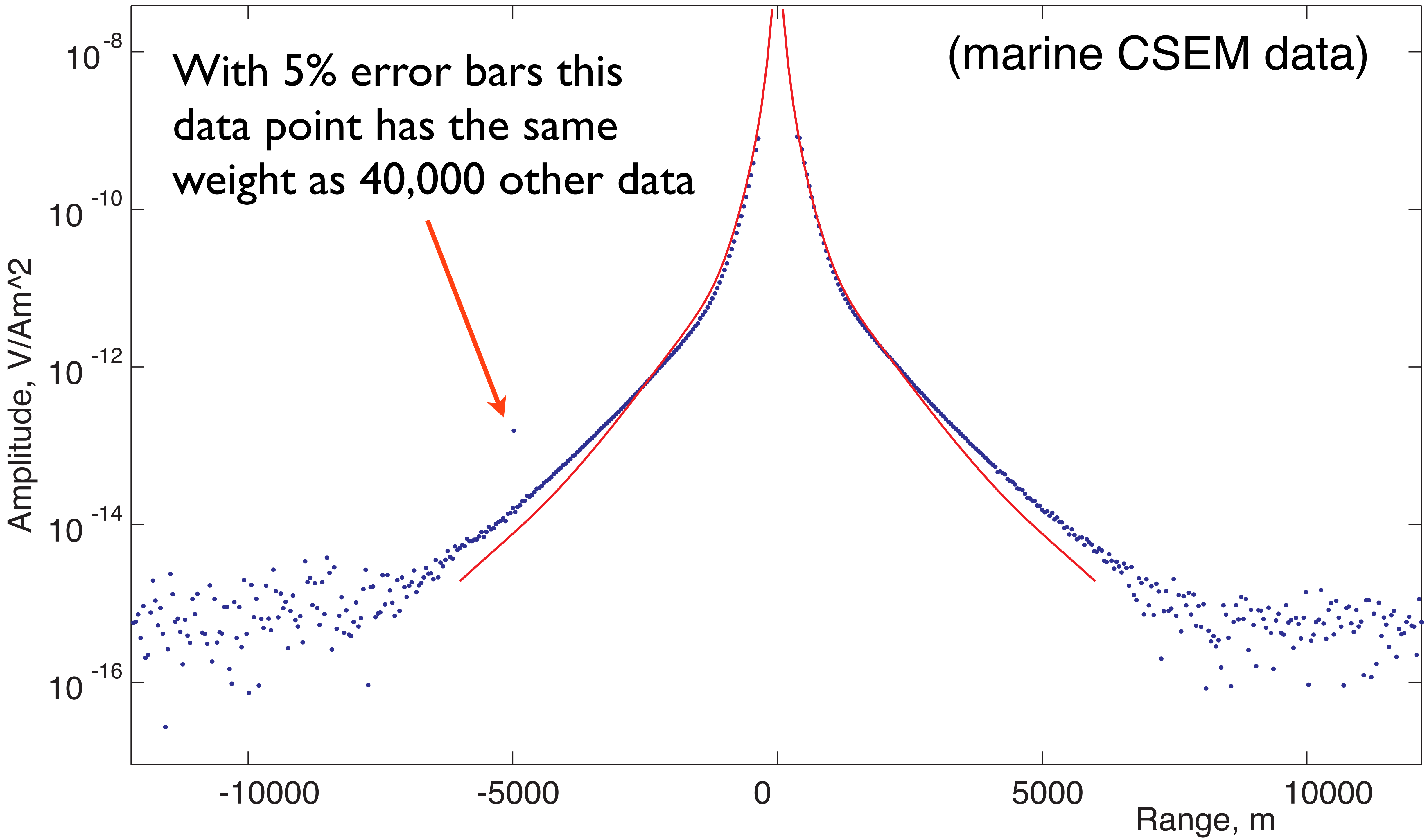
$$\mathbf{W} = \mathrm{diag}(1/\sigma_1, 1/\sigma_2, ..., 1/\sigma_M) \quad .$$

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

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

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. But...

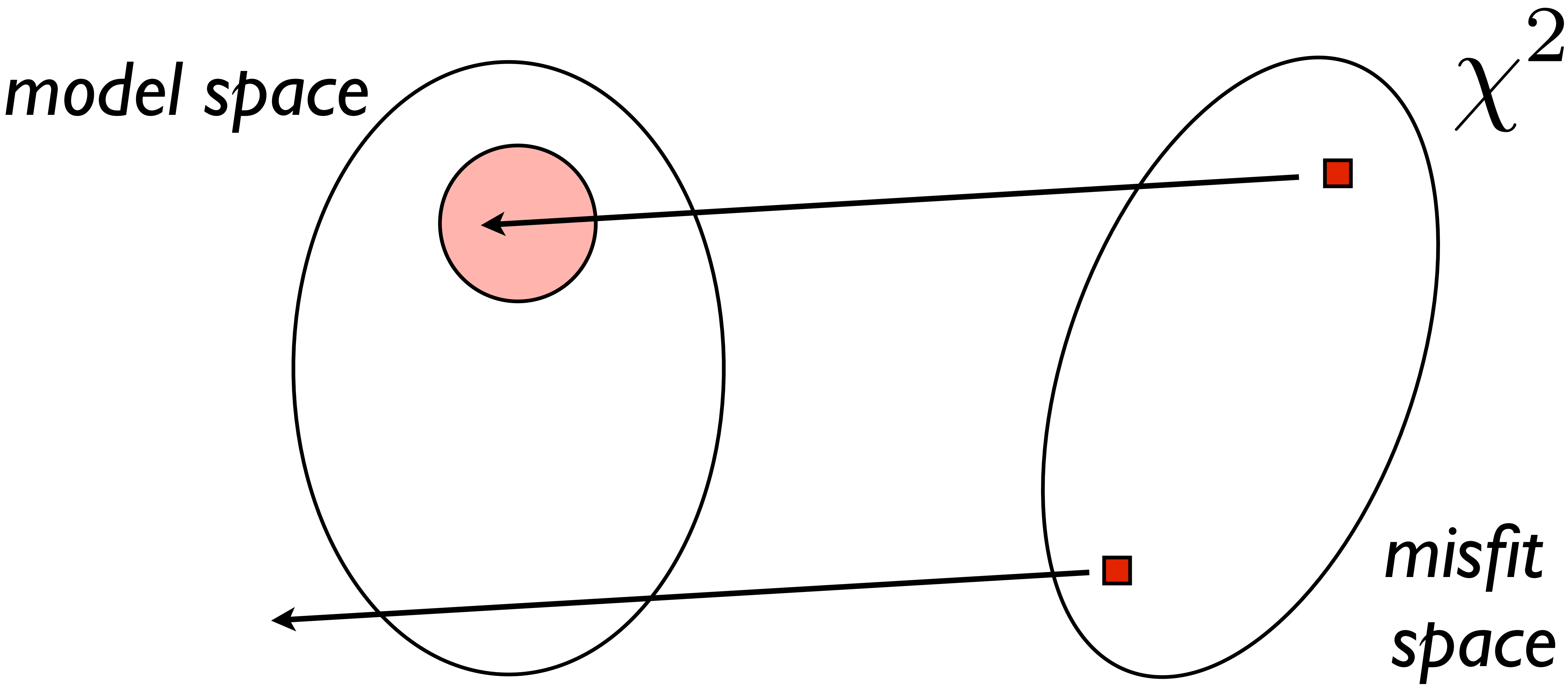
... sum-squared misfit measures are unforgiving of outliers:



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

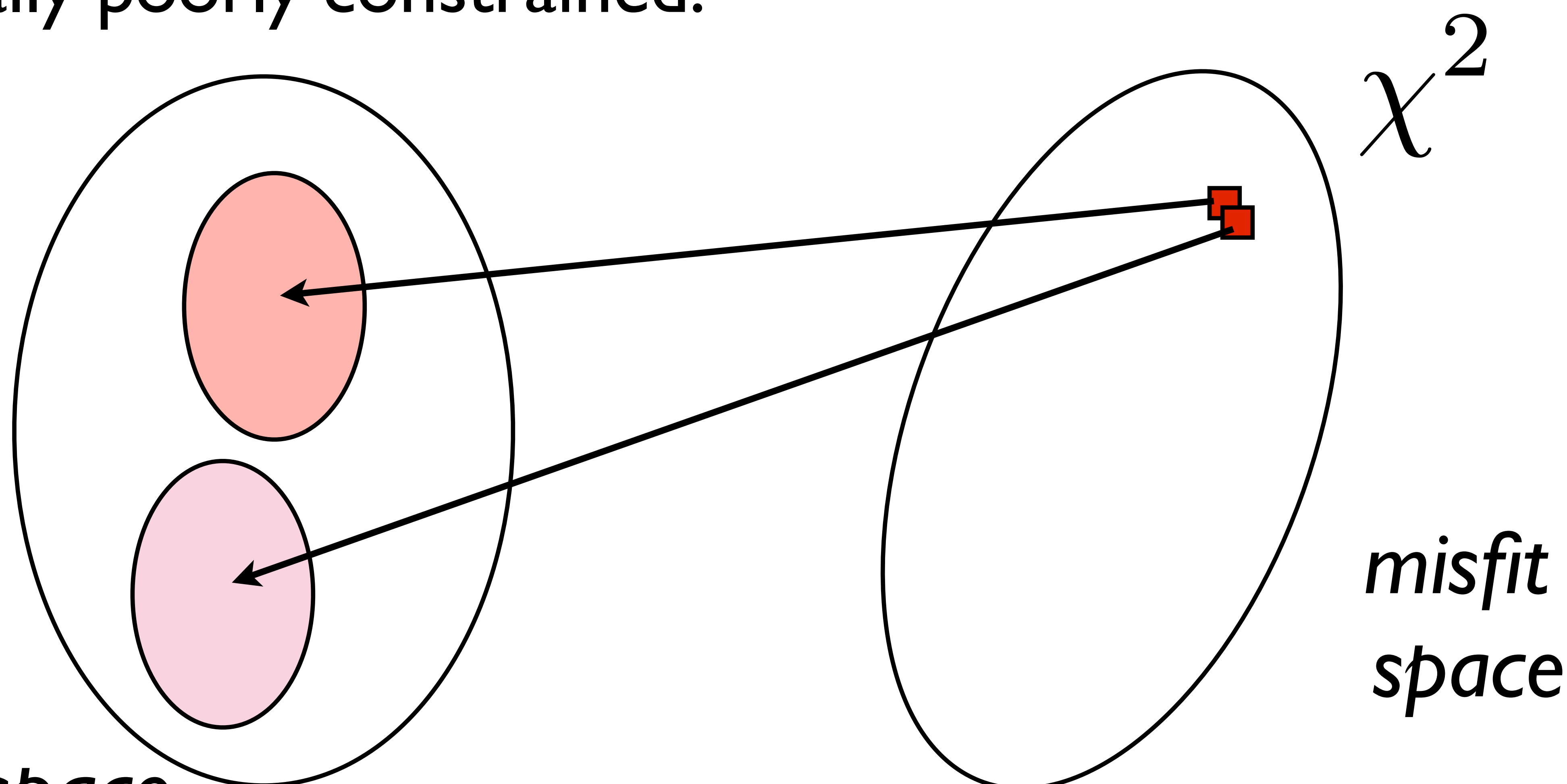
Steven Constable, 2013 ASEG workshop

Geophysical inversion is non-unique:



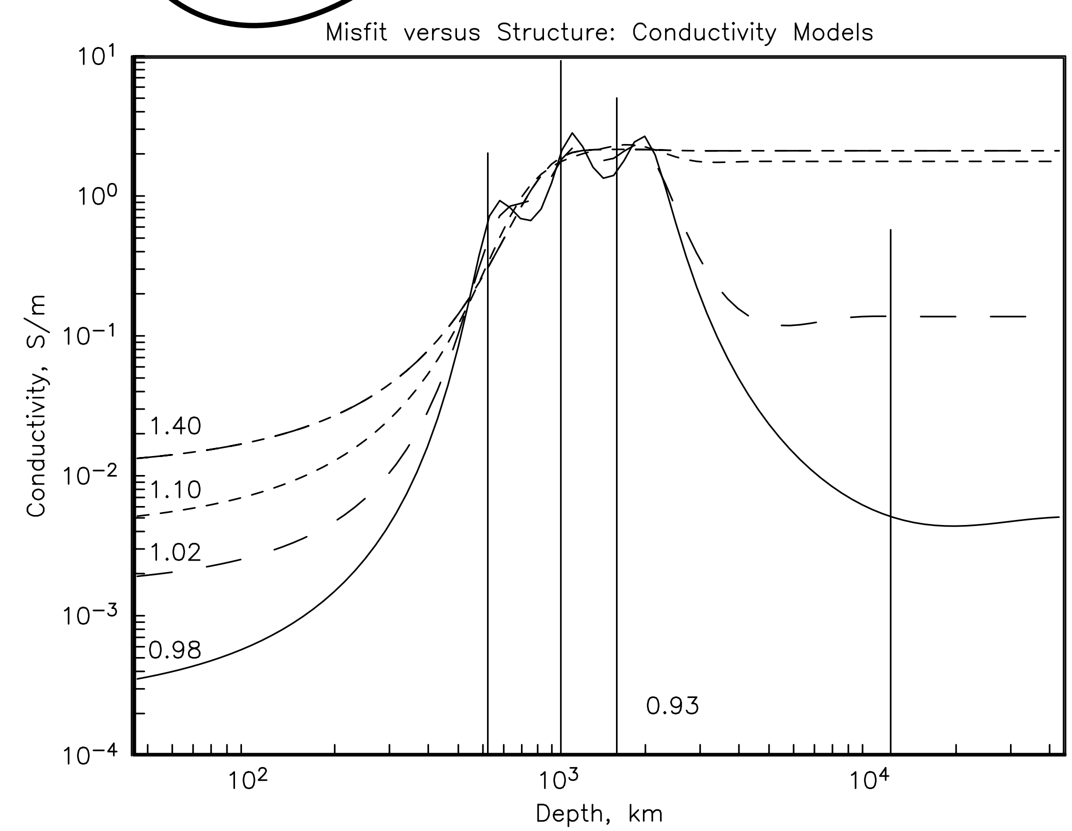
A single misfit will map into an infinite number of models (or none at all!).

and 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).



Steven Constable, 2013 ASEG workshop

Speaking of 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 theoretician.

This talk will just consider moderately large N.

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

$$\beta = (\mathbf{W}\mathbf{J})^T \mathbf{W}(\mathbf{d} - f(\mathbf{m_0}))$$
$$\alpha = (\mathbf{W}\mathbf{J})^T \mathbf{W}\mathbf{J} \quad .$$

So, given a starting model $\mathbf{m}_0$ we can find an update $\Delta\mathbf{m}$ and iterate until we converge. (This is Gauss-Newton.)

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

But this only works for small N (it isn't even defined for $N > M$). If N is big then the solutions become unstable, oscillatory, and generally useless (they are probably trying to converge to D+ type solutions).

Note also that we are solving for a model update - Bob Parker calls this a "creeping algorithm".

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

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

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:

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

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 unique solution, and manufactures models with useful properties.

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

$$U = ||\mathbf{R}\mathbf{m}_1||^2 + \mu^{-1}\left(||\mathbf{W}\mathbf{d} - \mathbf{W}f(\mathbf{m}_1)||^2\right)$$

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.

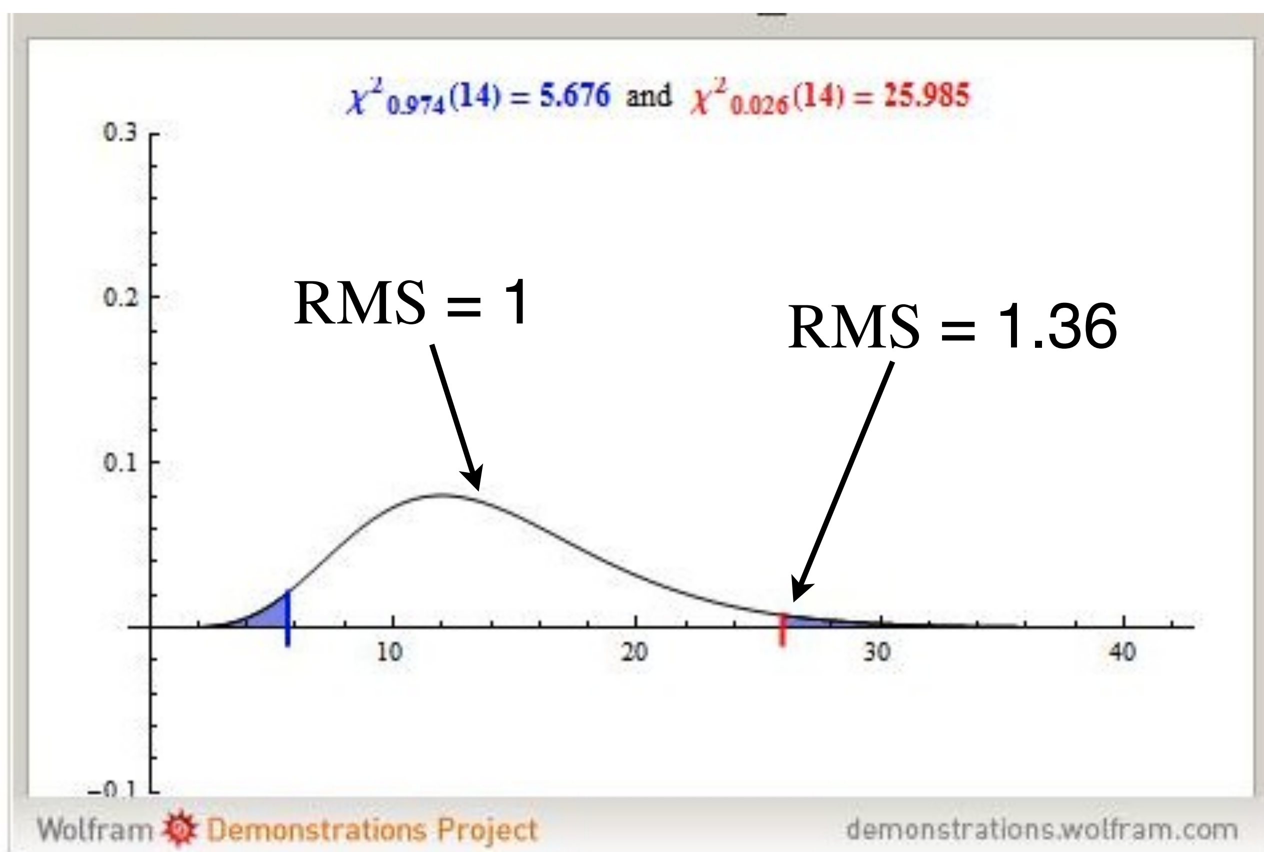
One approach is to choose μ and minimize U by least squares. There are various sets of machinery to do this (Newton, quasi-Newton, conjugate gradients, etc.). With many of these methods μ must be chosen by trial and error, increasing the computational burden and introducing some subjectivity. For some reason, many of the conjugate gradient algorithms add a penalty against the starting model, which adds a bias to the inversion.

Picking μ *a priori* is simply choosing how rough your model is compared to the data misfit. But, we've no idea how rough our model should be. However, we ought to have a decent idea of how well our data can be fit.

For zero-mean, Gaussian, independent errors, the sum-square misfit

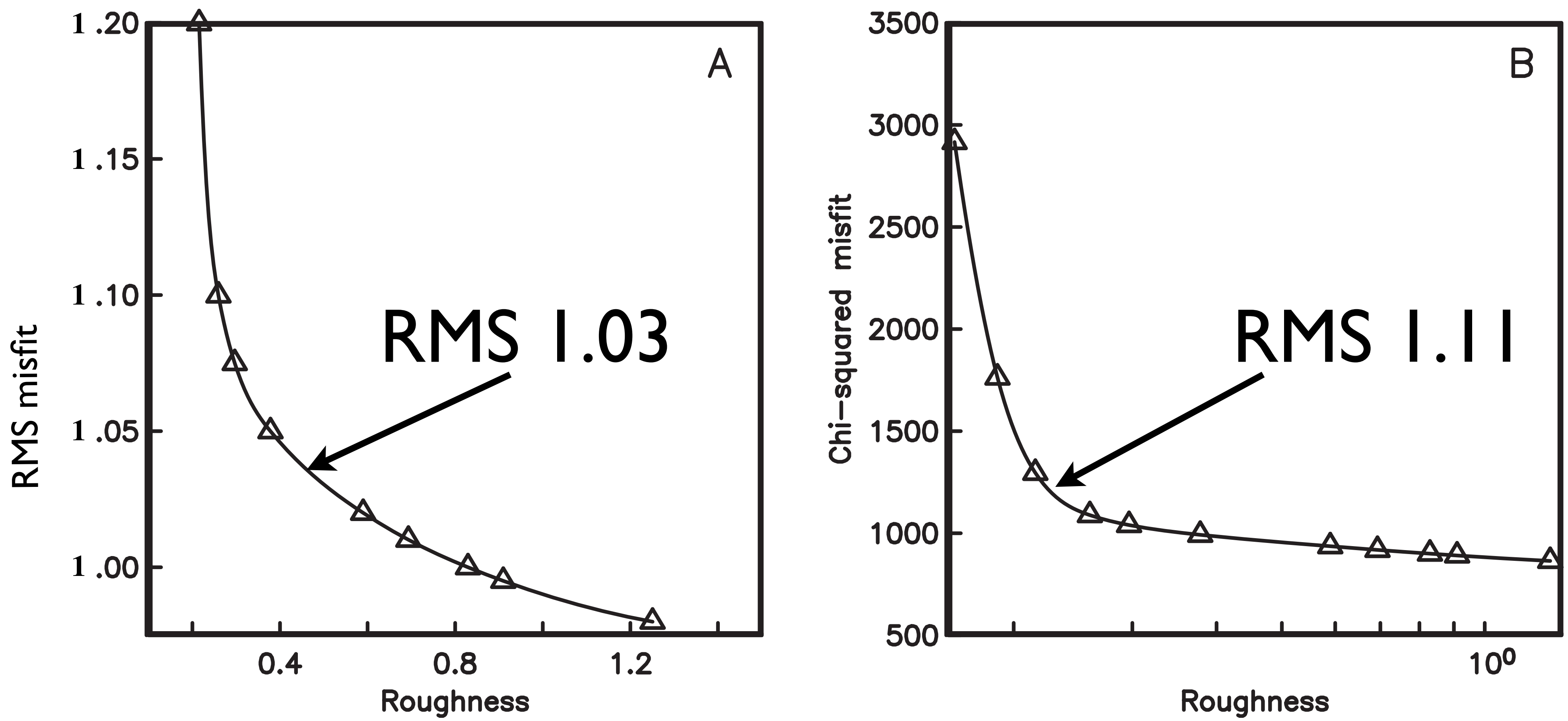
$$\chi^2 = ||\mathbf{Wd} - \mathbf{W\hat{d}}||^2$$

is chi-squared distributed with N degrees of freedom. The expectation value is just M, which corresponds to an RMS of one, and so this could be a reasonable target misfit. Or, one could look up the 95% (or other) confidence interval for chi-squared M.



Steven Constable, 2013 ASEG workshop

Beware of trade-off ("L") curves:



(they are not as objective as proponents say...)

Steven Constable, 2013 ASEG workshop

The Occam approach is to introduce some acceptable fit to the data (χ_*^2) and minimize:

$$U = ||\mathbf{R}\mathbf{m}_1||^2 + \mu^{-1}\big(||\mathbf{W}\mathbf{d} - \mathbf{W}f(\mathbf{m}_1)||^2 - \chi_*^2\big)$$

Linearizing:

$$U = ||\mathbf{R}\mathbf{m}_1||^2 + \mu^{-1}\Big(||\mathbf{W}\mathbf{d} - \mathbf{W}\big(f(\mathbf{m}_0) + \mathbf{J}(\mathbf{m}_1 - \mathbf{m}_0)\big)||^2 - \chi_*^2\Big)$$

After differentiation and setting to zero we get

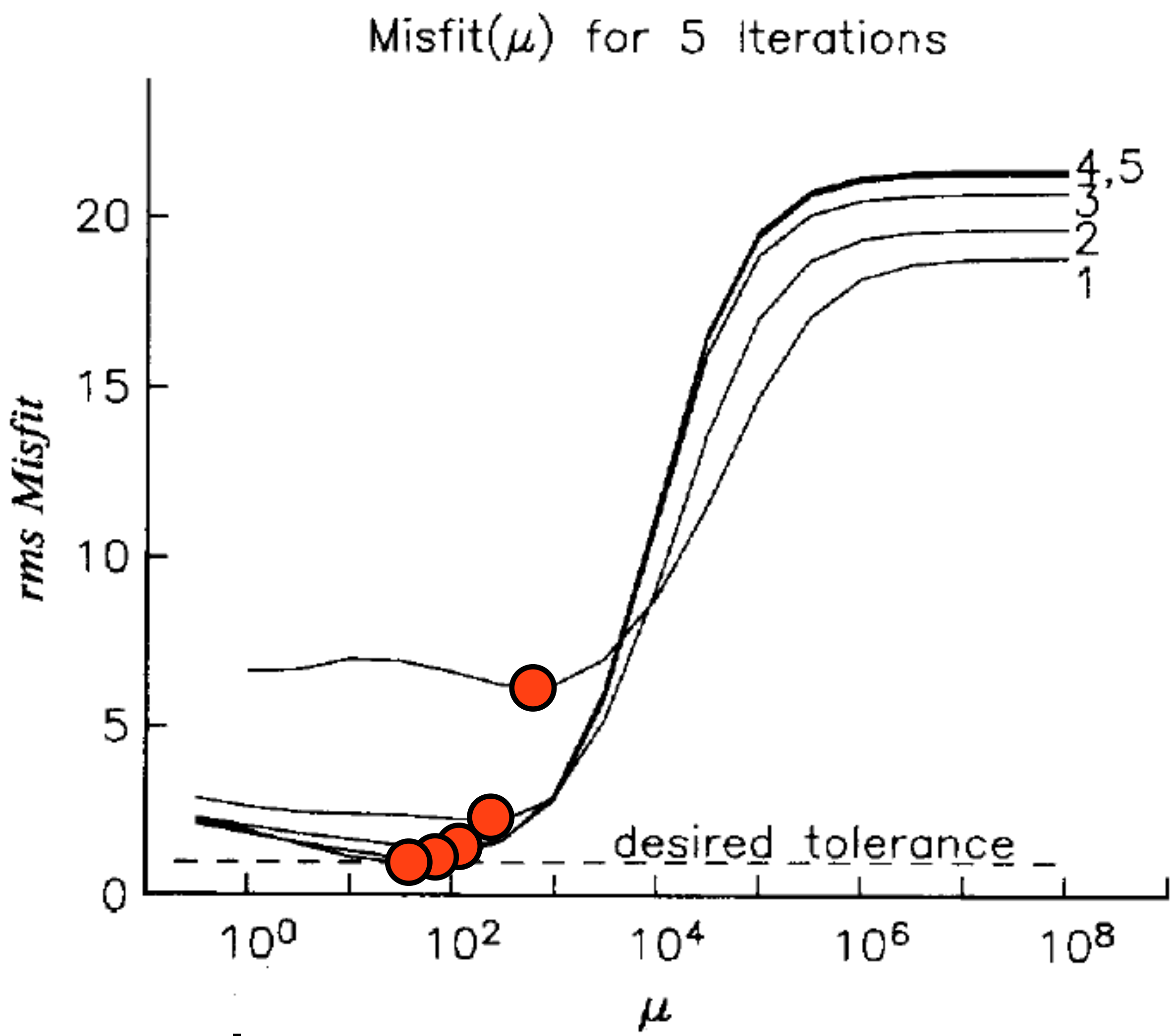
$$\mathbf{m}_1 = \big[\mu\mathbf{R}^T\mathbf{R} + (\mathbf{W}\mathbf{J})^T\mathbf{W}\mathbf{J}\big]^{-1}(\mathbf{W}\mathbf{J})^T\mathbf{W}(\mathbf{d} - f(\mathbf{m_0}) + \mathbf{J}\mathbf{m}_0)$$

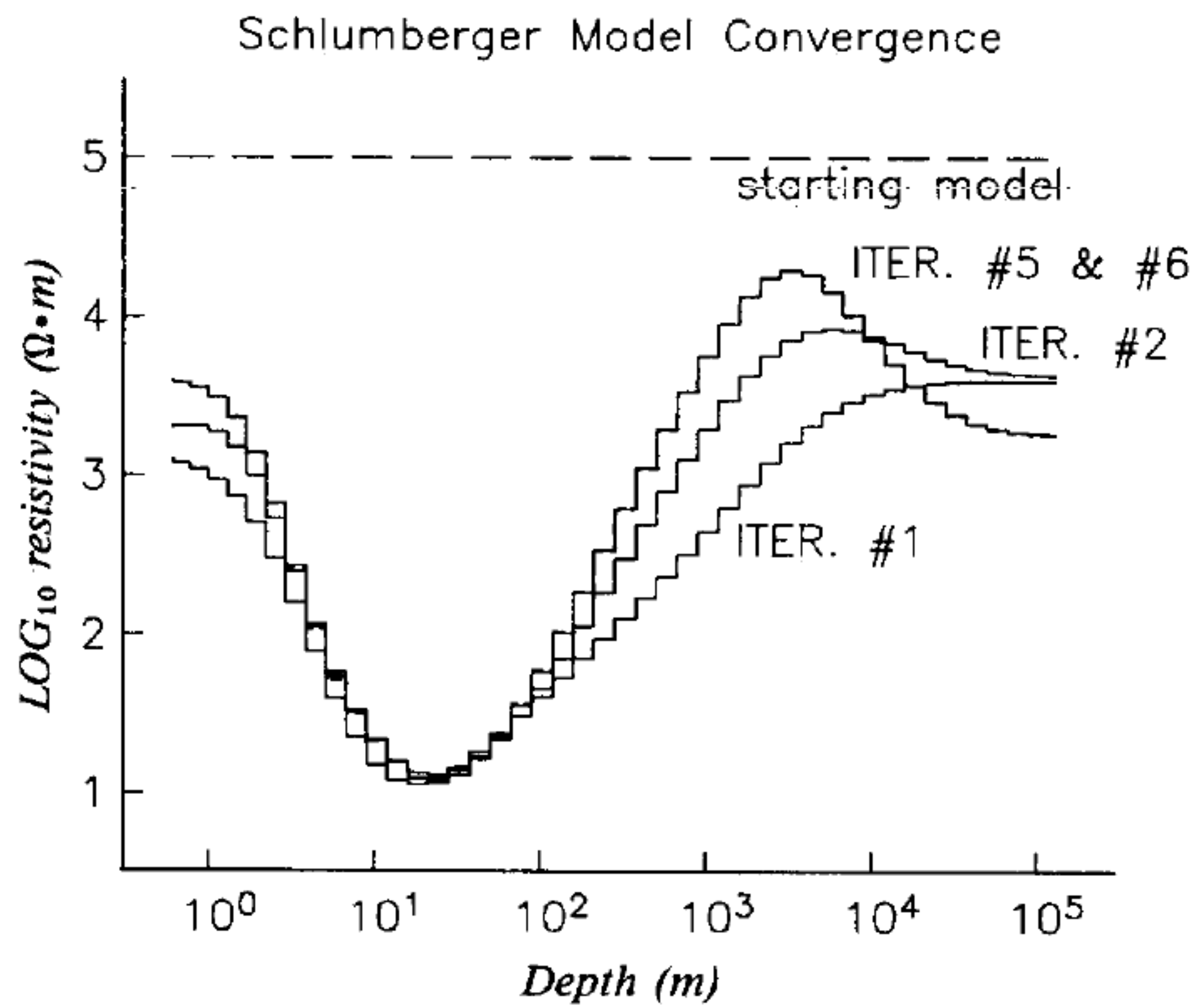
The only thing we need is to find the right value for μ .

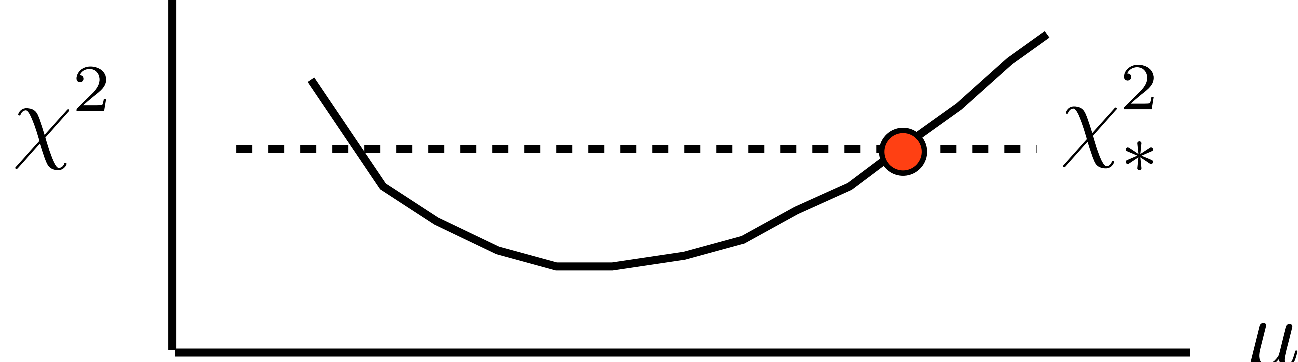
(Note we are solving for the next model directly; Bob Parker calls this a "leaping" algorithm.)

Steven Constable, 2013 ASEG workshop

Occam does this by carrying out a line search to find the ideal μ. Before χ^2_* is reached, we minimize χ^2. After χ^2_* is reached we choose the μ which gives us exactly χ^2_*.

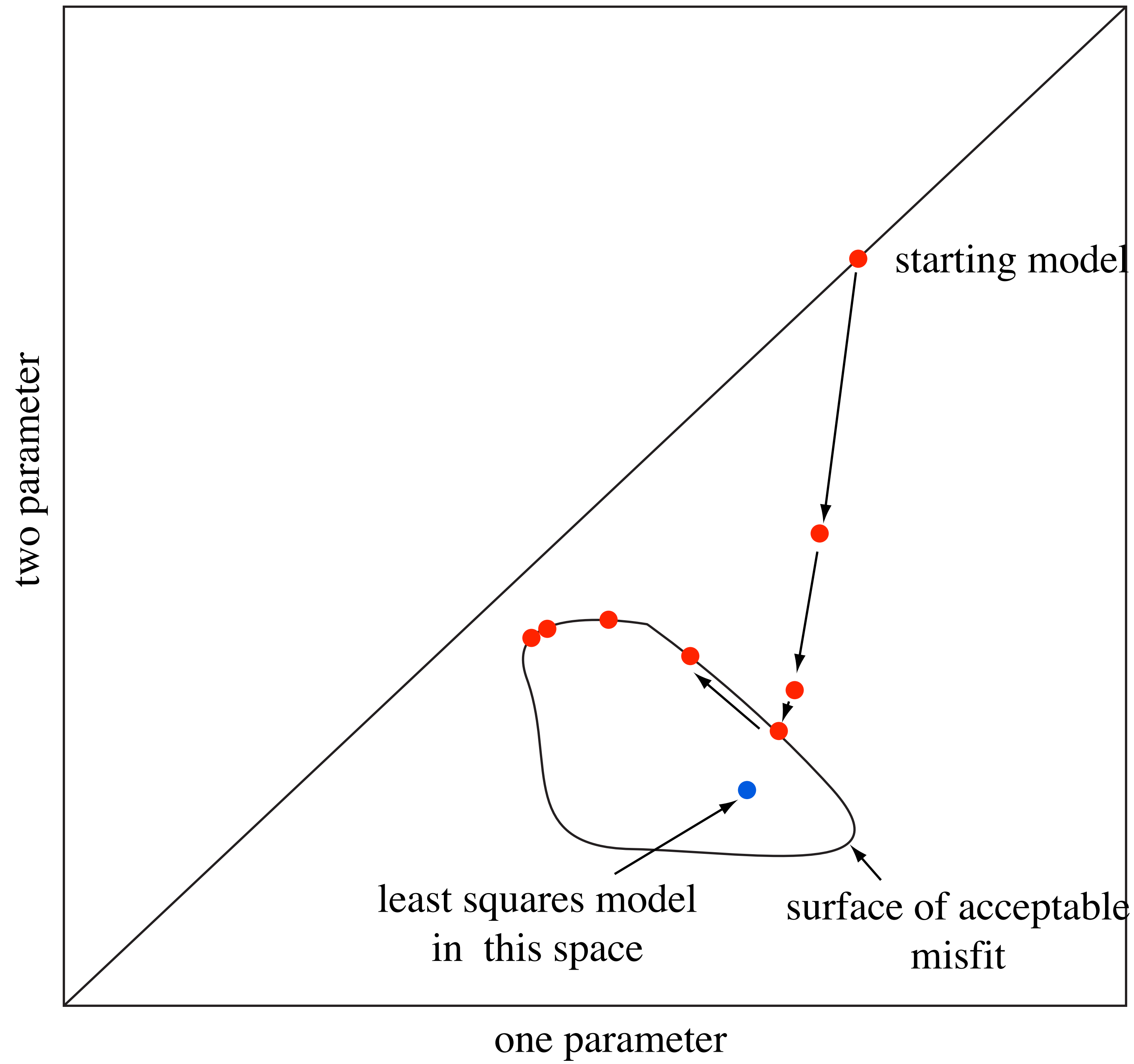






Steven Constable, 2013 ASEG workshop

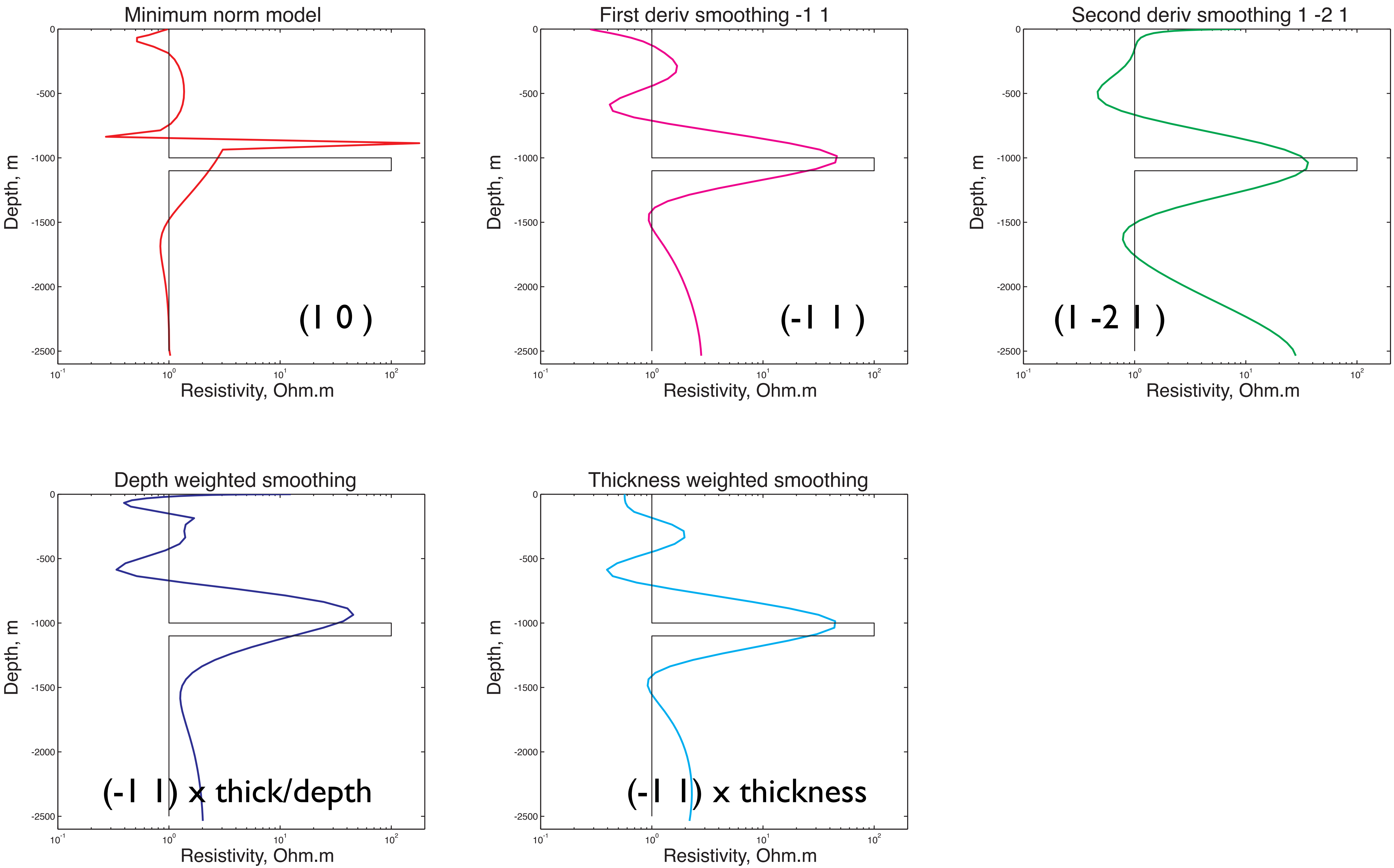
We illustrate this with a toy 2-parameter problem:



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

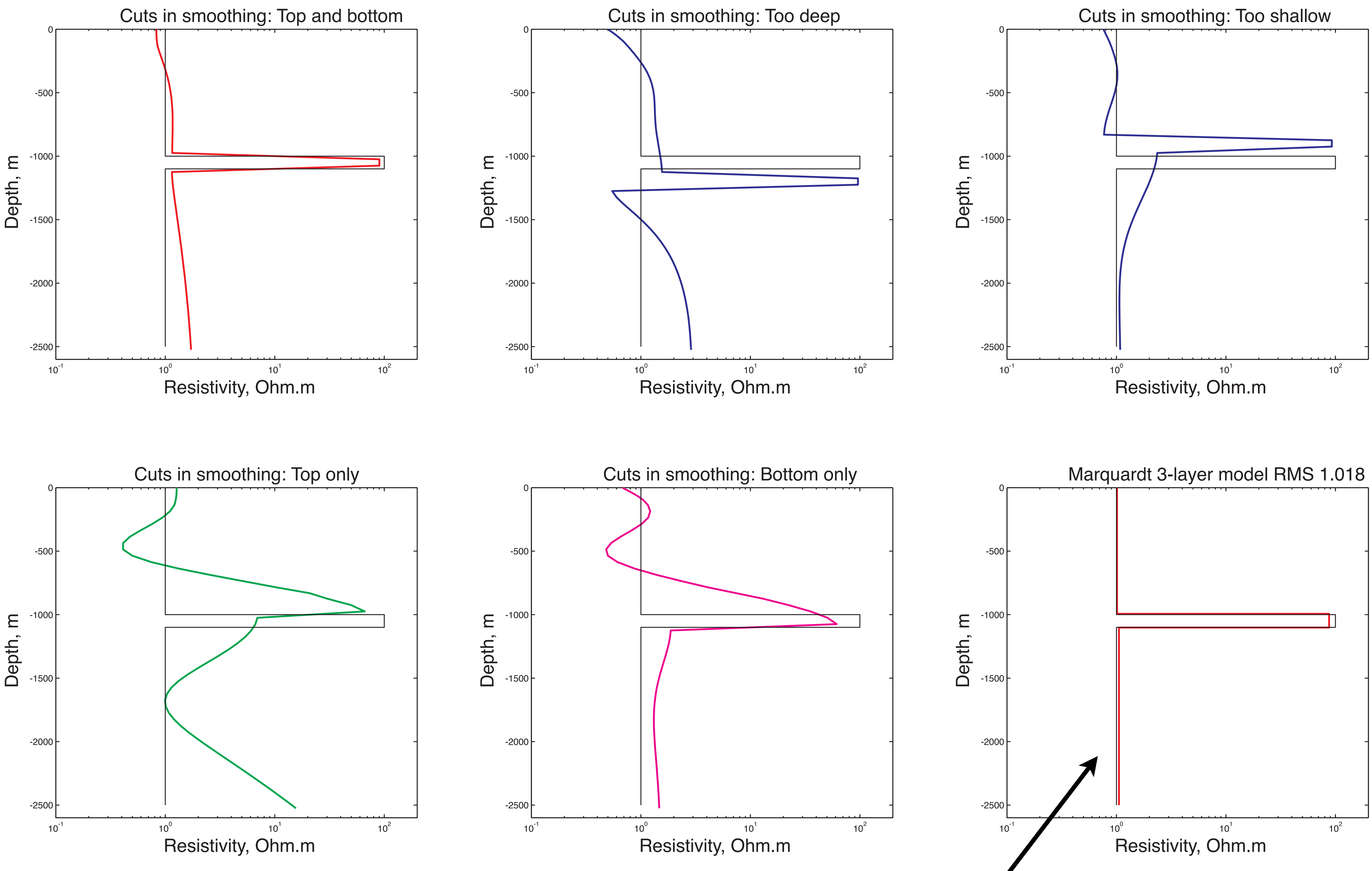
Steven Constable, 2013 ASEG workshop

For a given misfit, the model now depends on (**R**)



Steven Constable, 2013 ASEG workshop

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



Sparsely parameterized model does well! (Why?)

Steven Constable, 2013 ASEG workshop

Some words here about the derivative matrix:

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

For numerical efficiency, the derivatives are often calculated using an analytical expression. Often the chain rule has to be used to relate a derivative of *Re(E)* and *Im(E)* wrt σ (say) to */E/* wrt ρ.

If your derivatives are wrong, you are going in the wrong direction. It is a good idea to check analytical derivatives with a forward

$$\frac{\partial f(x_i, \mathbf{m}_0)}{\partial m_j} \approx \frac{f(x_i, \mathbf{m}_0 + \Delta\mathbf{m}) - f(x_i, \mathbf{m}_0)}{\Delta\mathbf{m}}$$

or central difference:

$$\frac{\partial f(x_i, \mathbf{m}_0)}{\partial m_j} \approx \frac{f(x_i, \mathbf{m}_0 + \Delta\mathbf{m}) - f(x_i, \mathbf{m}_0 - \Delta\mathbf{m})}{2\Delta\mathbf{m}}$$

The central difference is quadratically more accurate than the forward difference, but takes twice as long to compute.

How to choose χ^2_*? If our errors are well estimated and well behaved, we have the statistical guidelines of expectation value or cumulative probability.

Errors come from

- statistical processing errors (spectral estimation for MT and stacking for CSEM)

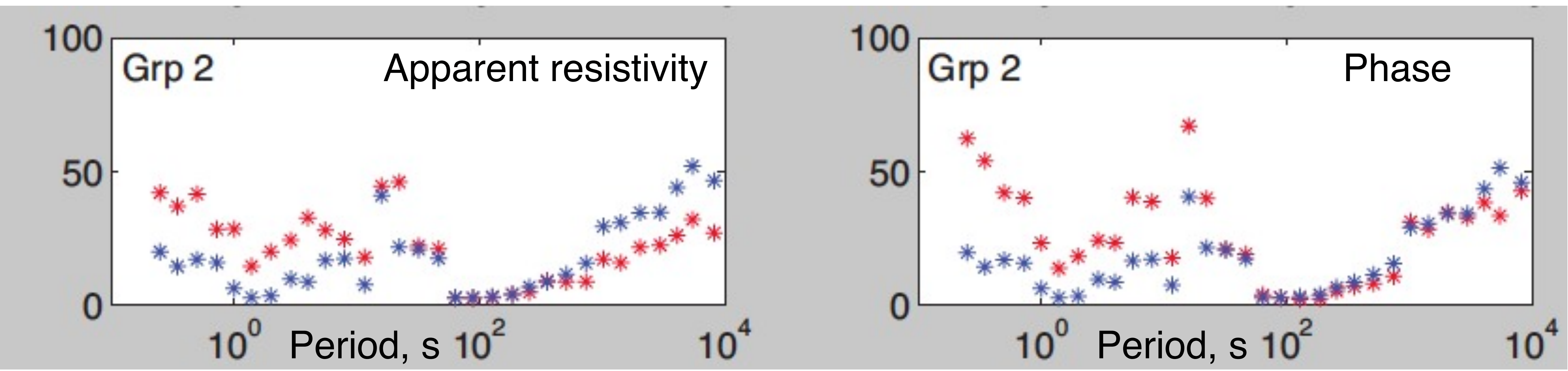
- systematic errors such and navigation errors and instrument calibrations, and

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

Instrument noise should be captured by processing errors, but some error models assume stationarity (i.e. noise statistics don’t vary with time). In practice, we only have a good handle on processing errors - everything else is lumped into a noise floor.

How good are your processing errors? Sometimes it is hard to tell. Here we took 28 days of MT data and processed up 4-day chunks and estimated a variance.

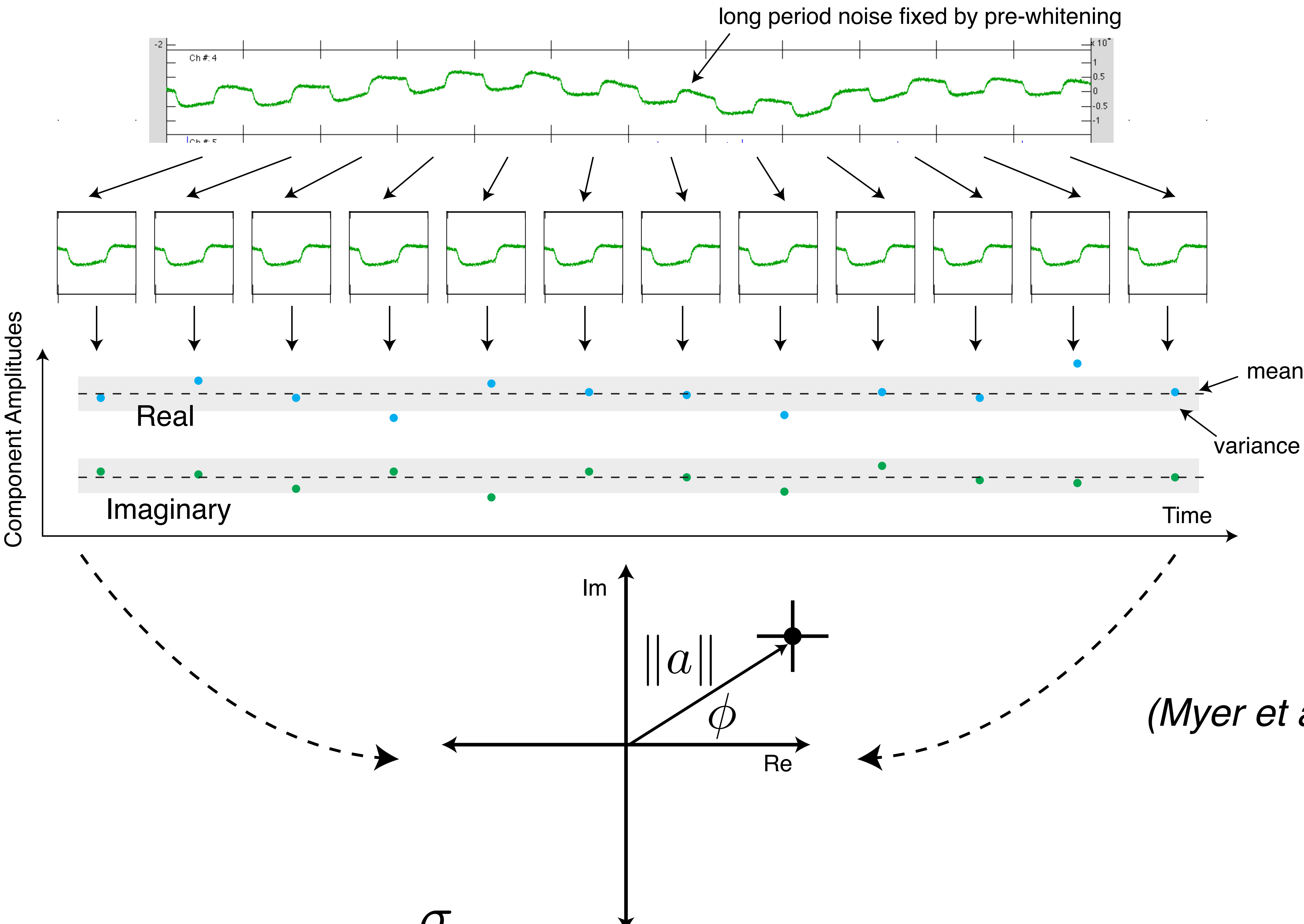
blue = processing errors
red = errors from data variance



Processing errors underestimate variance at short periods, overestimate it at long periods. This work is ongoing.

(courtesy Dallas Sherman)

For stacked, frequency domain data, we can do this over every stack frame:

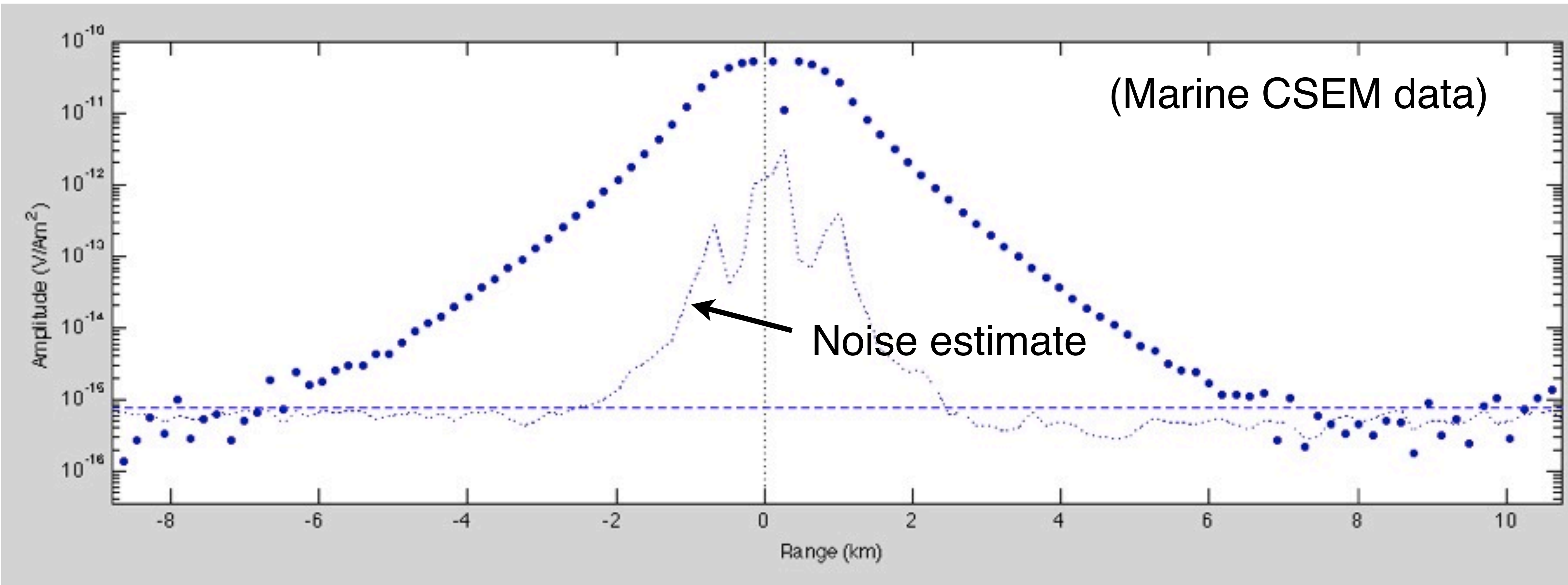


(Myer et al., 2011)

$$\sigma_{\text{mean}} = \frac{\sigma}{\sqrt{N}}$$ or equivalently $$\text{SNR} \to \sqrt{\text{time}}$$

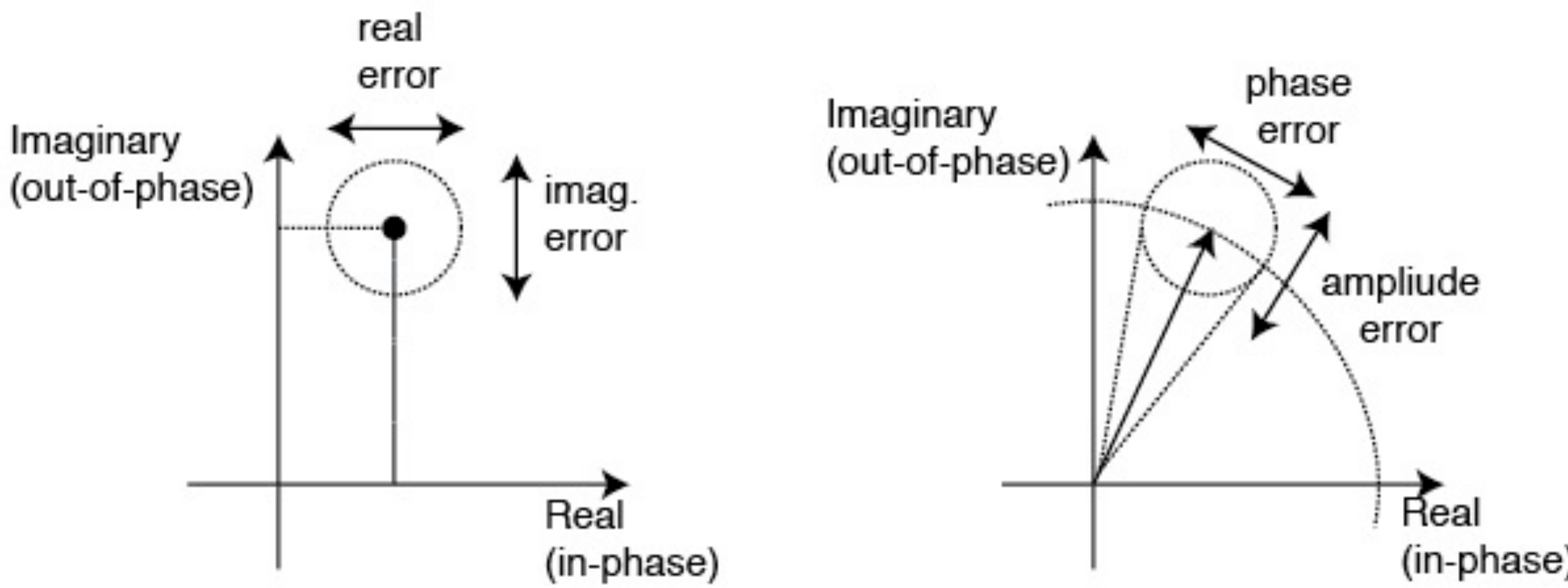
Steven Constable, 2013 ASEG workshop

This approach captures both non-stationarity and bias from variations during the stack frame.



Steven Constable, 2013 ASEG workshop

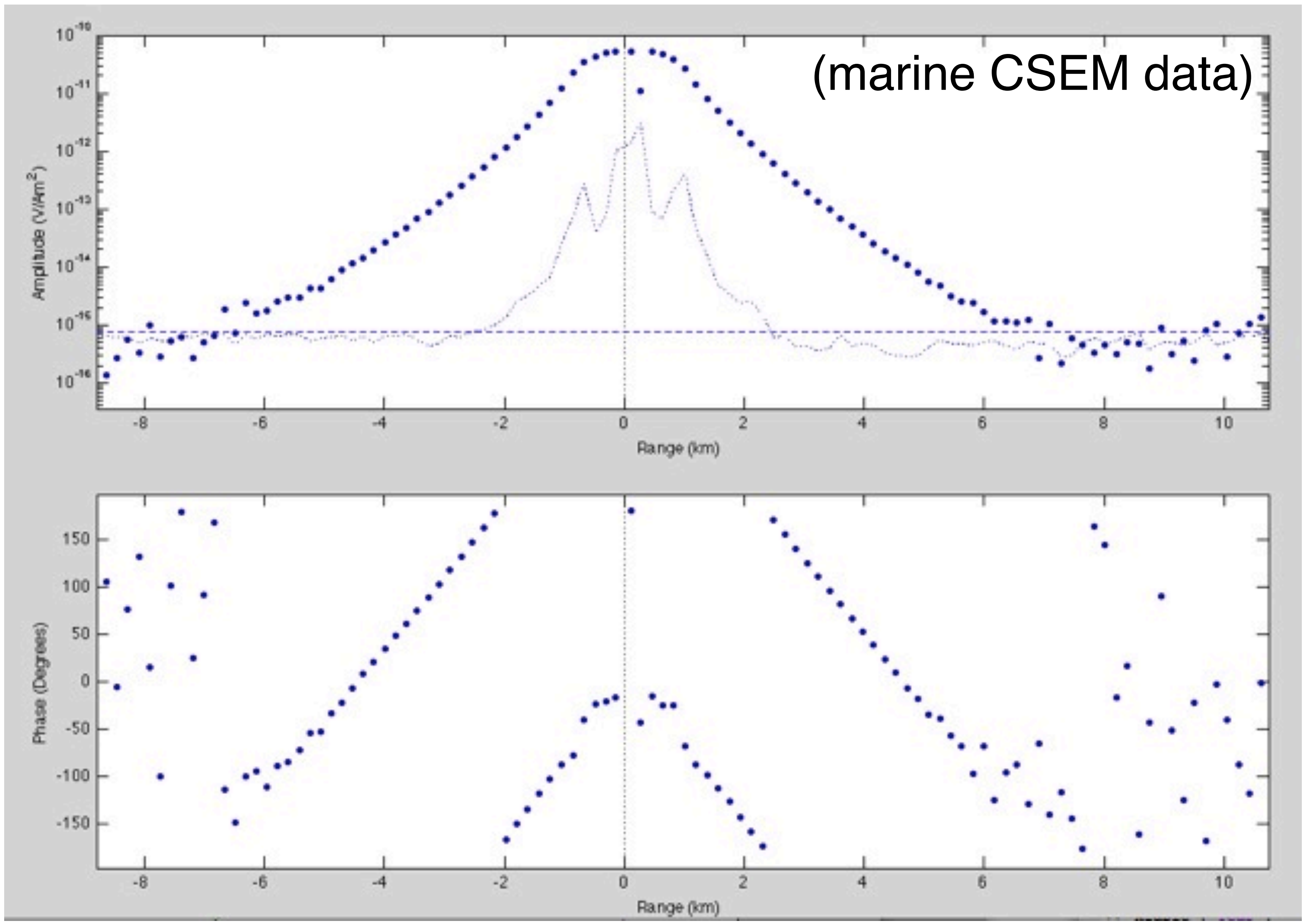
For frequency domain data, processing errors in *Re* and *Im* (in-phase and quadrature) components should be equal and Gaussian.



This defines a relationship between amplitude and phase: 10% = 5.7° (or 10% and 2.8° for MT).

(Note that you can't use relative (percent) errors for *Re* and *Im* because they can go through zero as phase goes through 0°, 90°, 180° ...)

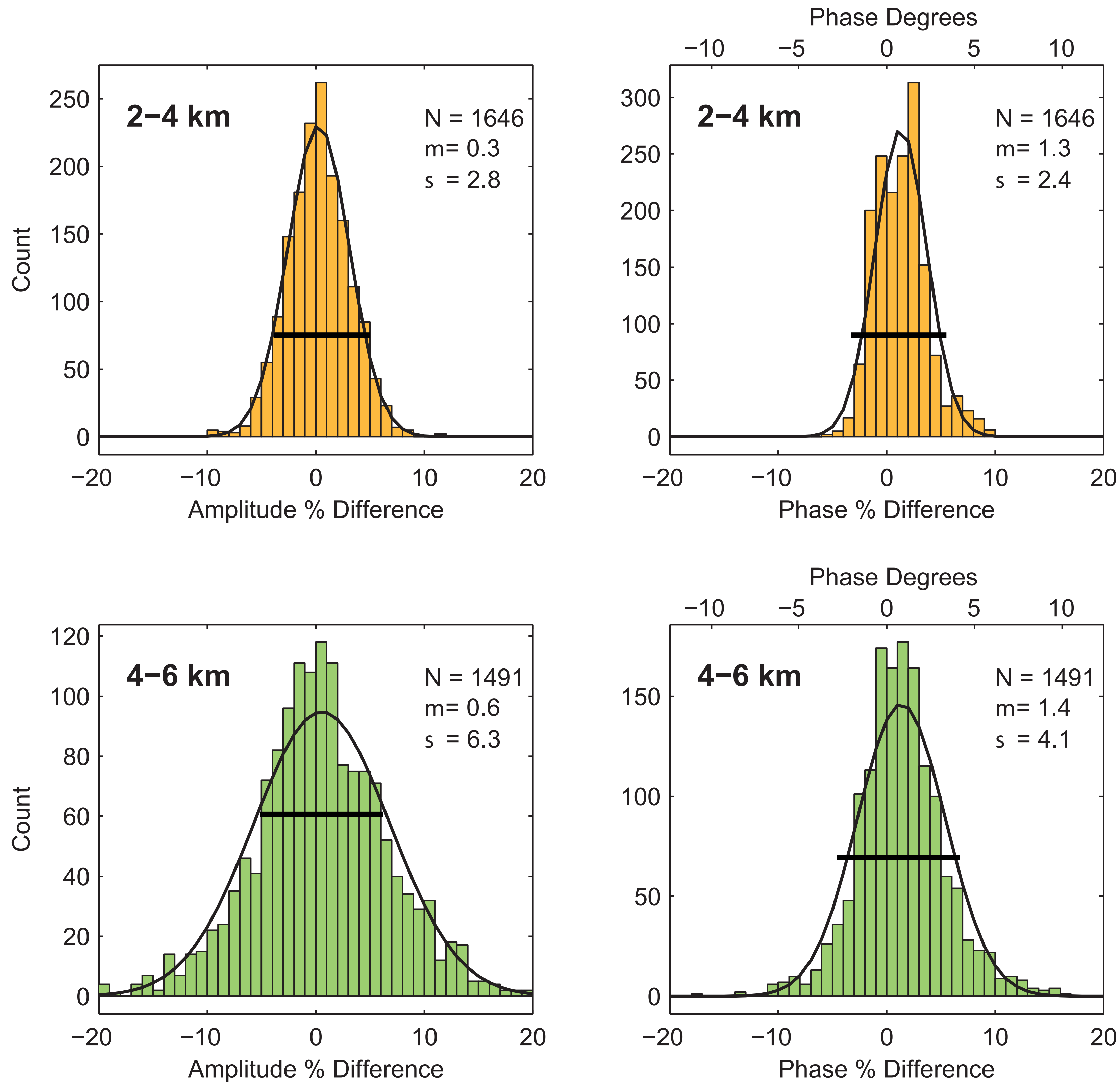
CSEM phase can exceed 360° or jump quadrant (so can marine MT data). This can be troublesome in inversions, so sometimes it is better to invert *Re* and *Im* components (but don't use percentage errors!).



However, there may be several drawbacks to inverting *Re* and *Im*.

Steven Constable, 2013 ASEG workshop

Phase errors can be smaller than they should be from processing analysis.



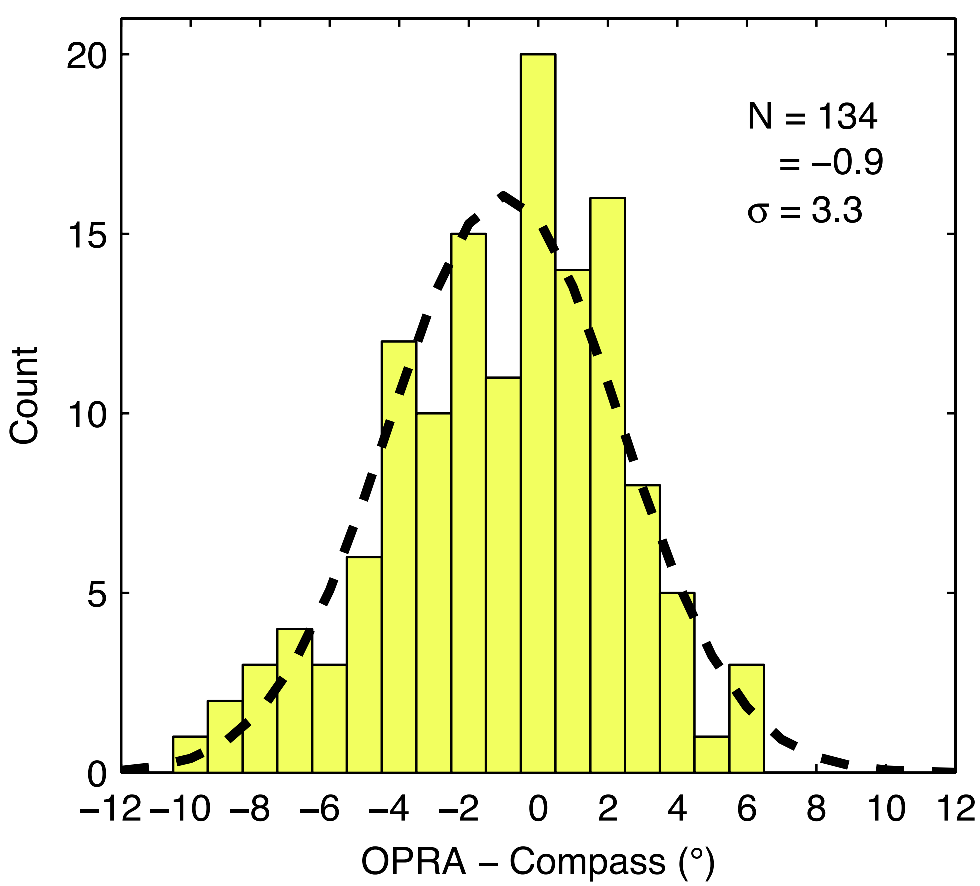
(modified from Myer et al., 2012)

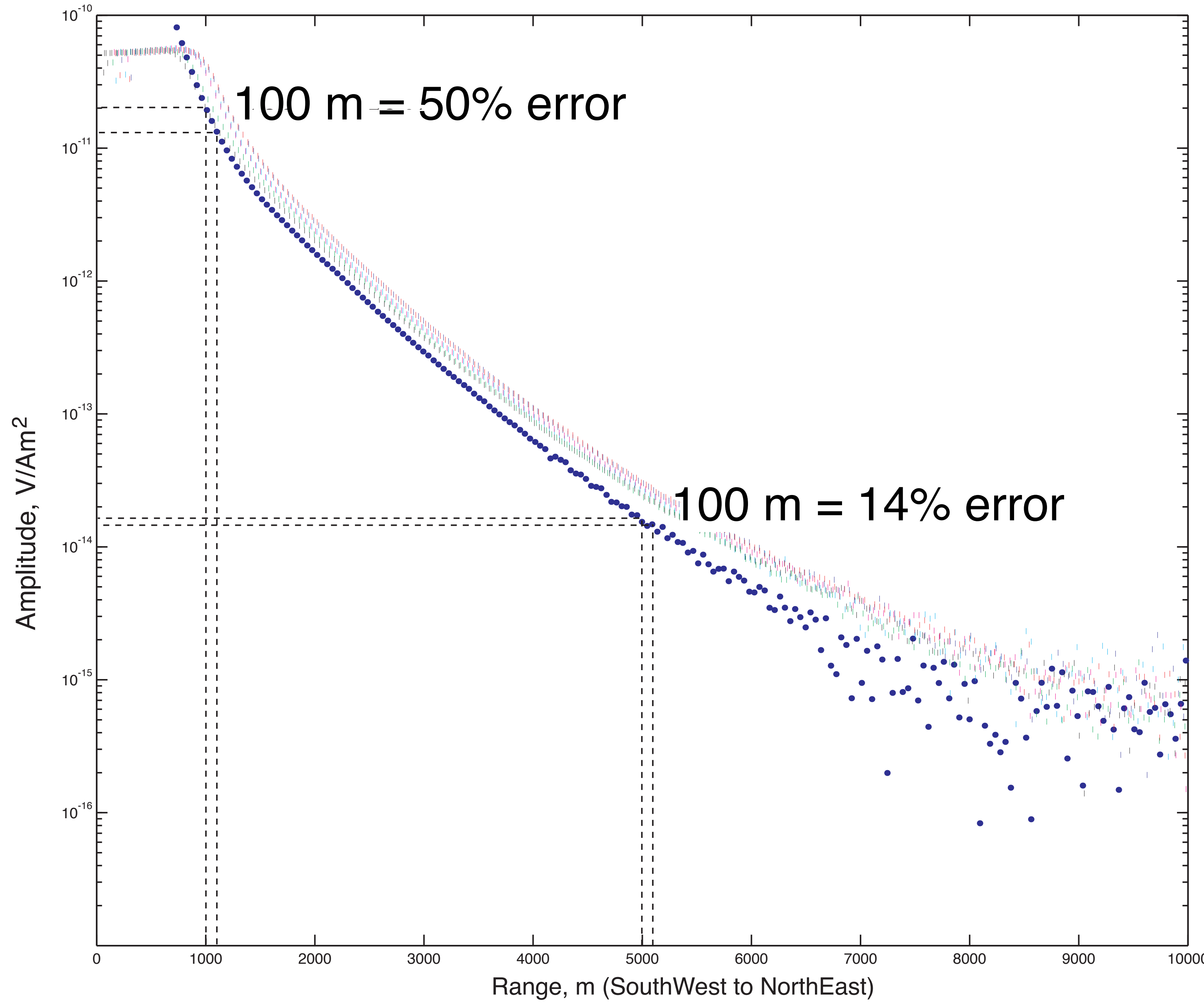
(differences taken from actual repeat tows for marine CSEM)

Steven Constable, 2013 ASEG workshop

Trouble is, processing errors are only relevant near the noise floor. For marine CSEM navigation is probably the biggest error, and can have a bigger effect on amplitude than phase.

Receiver orientation is also a critical parameter in data quality.



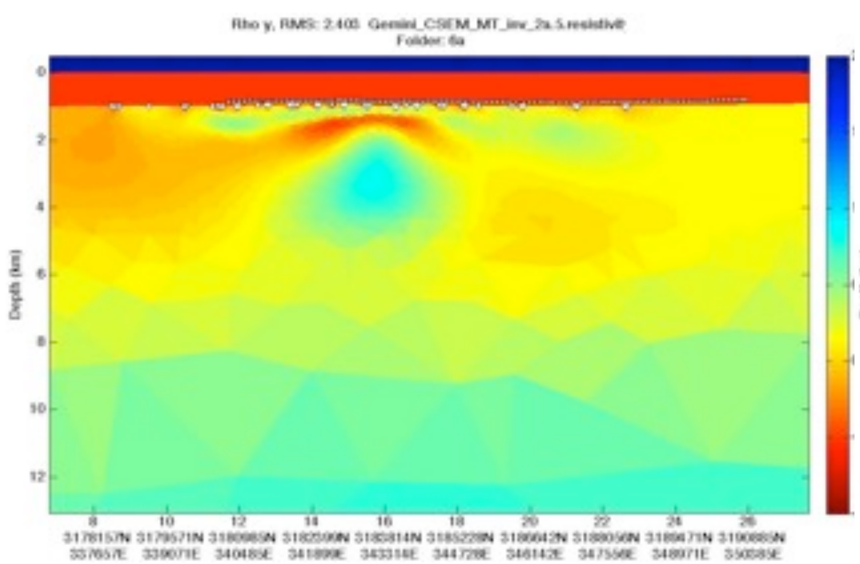


Understanding your data error structure can take a lot of work.

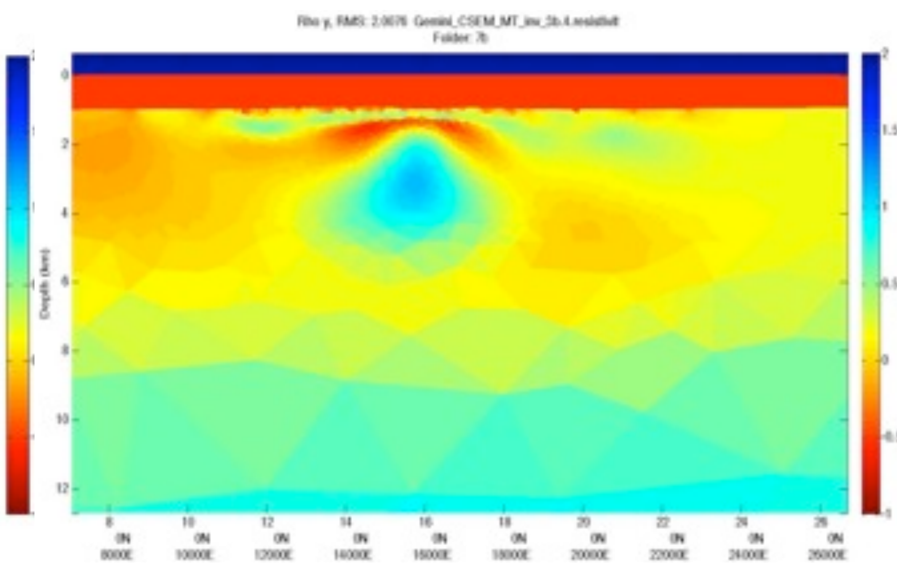
Steven Constable, 2013 ASEG workshop

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

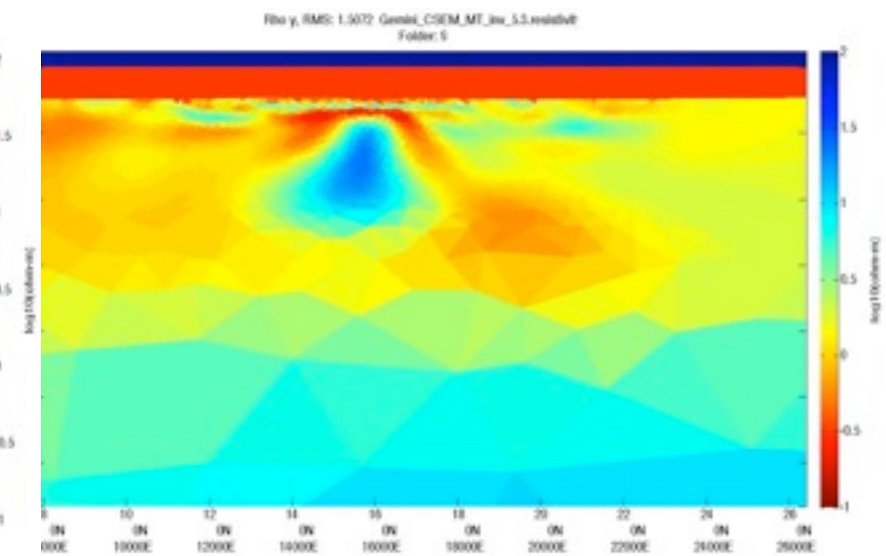
Joint 2D inversion of marine CSEM (3 frequencies, no phase) and MT (Gemini salt prospect, Gulf of Mexico):



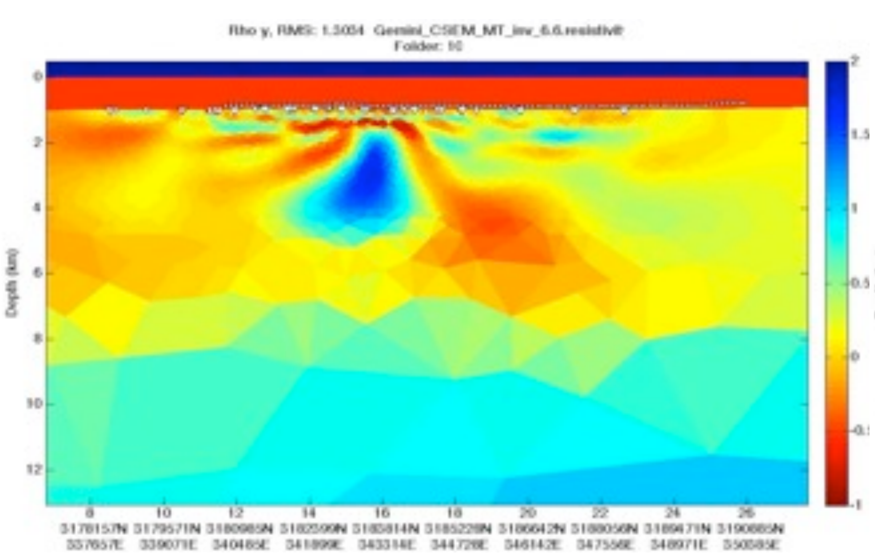
Target misfit 2.4



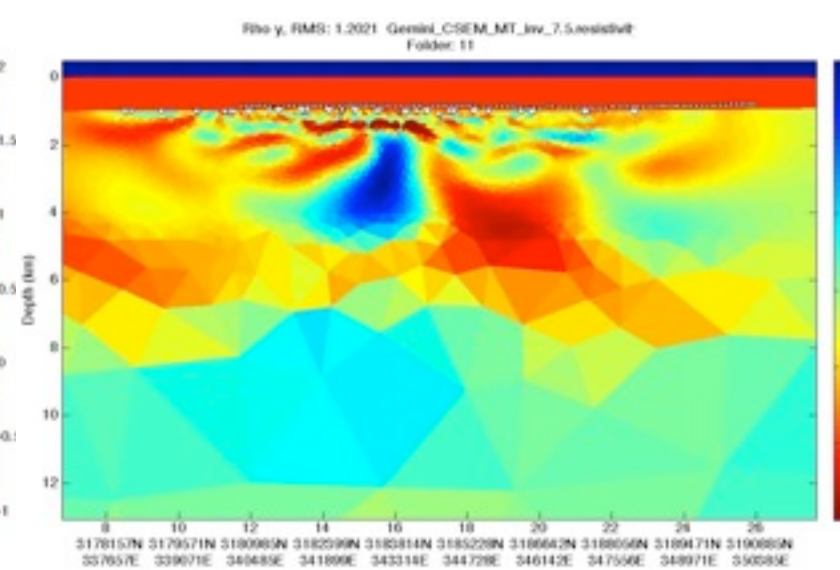
Target misfit 2.0



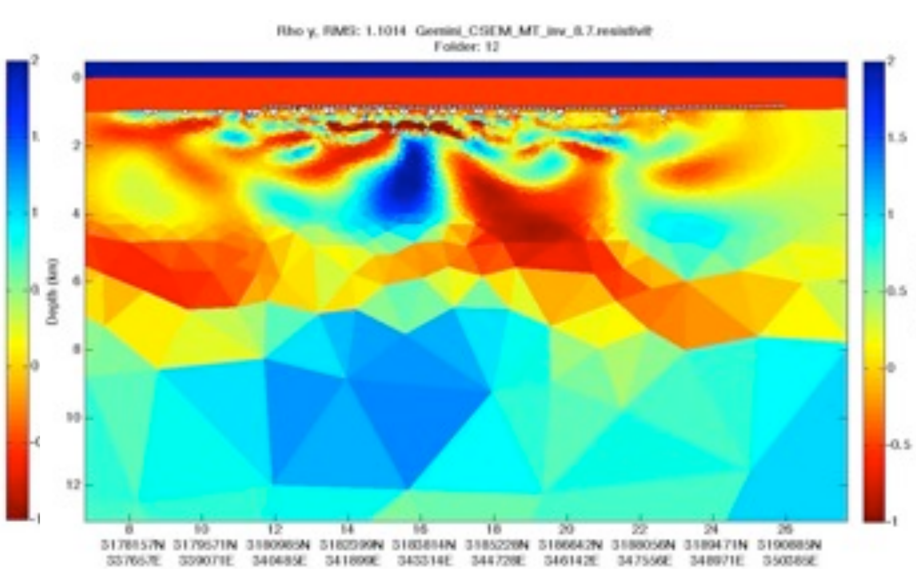
Target misfit 1.5



Target misfit 1.3



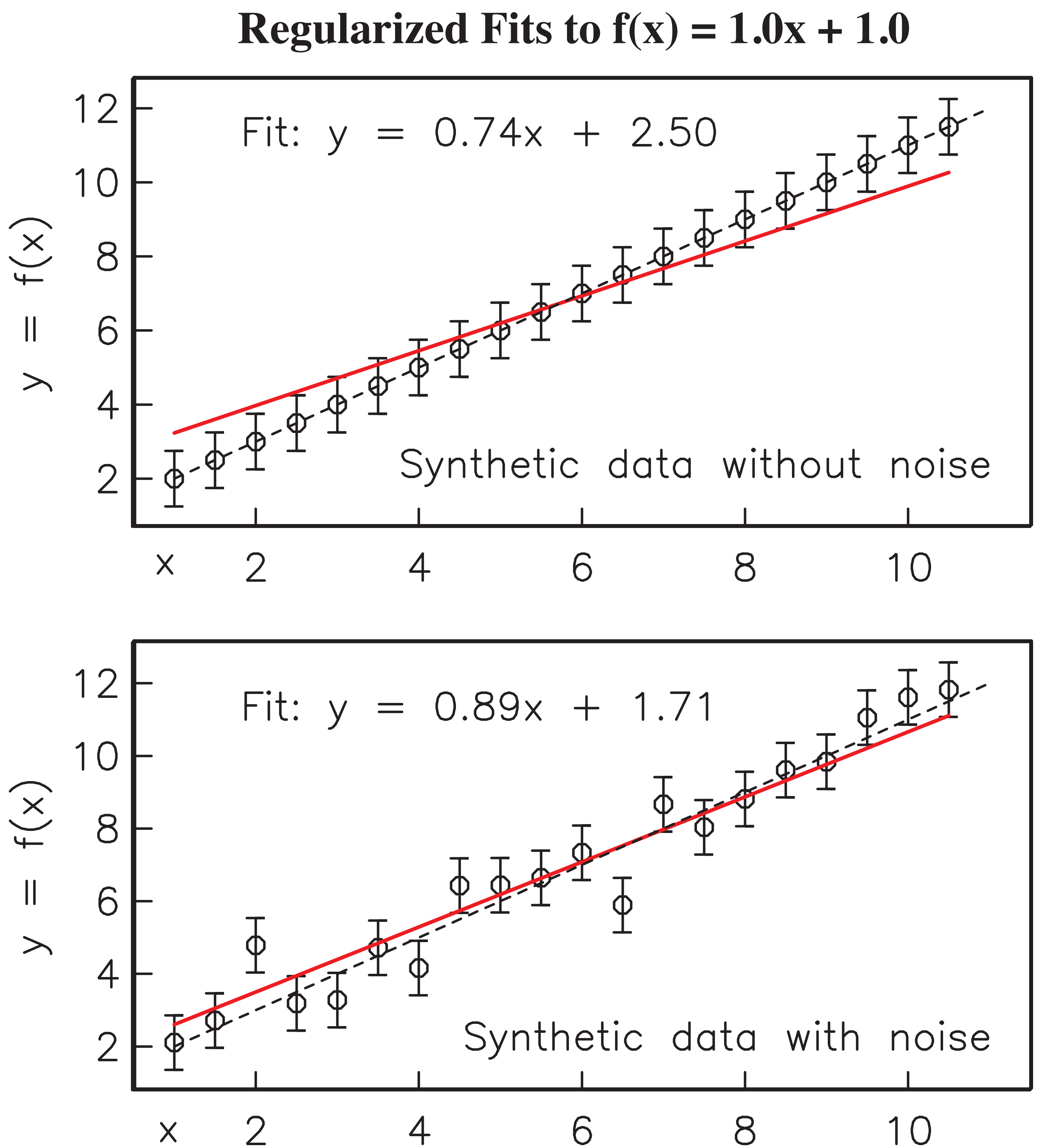
Target misfit 1.2



Target misfit 1.1

(courtesy Arnold Orange)

Noise matters even with synthetic model studies. Here slope is penalized and data are fit to RMS = 1:



Error bars with no noise: Misfit is used to minimize slope and the result is heavily biased.

Error bars with noise: Misfit budget used for errors, fit to model is better.

(from Constable, 1991)

Steven Constable, 2013 ASEG workshop

Your inversion scheme must not allow negative resistivities (or conductivities) to develop in the model. The easiest way to do this is by inverting for $\log(\boldsymbol{m})$. (There are other ways, e.g. NNLS.)

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:

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

where

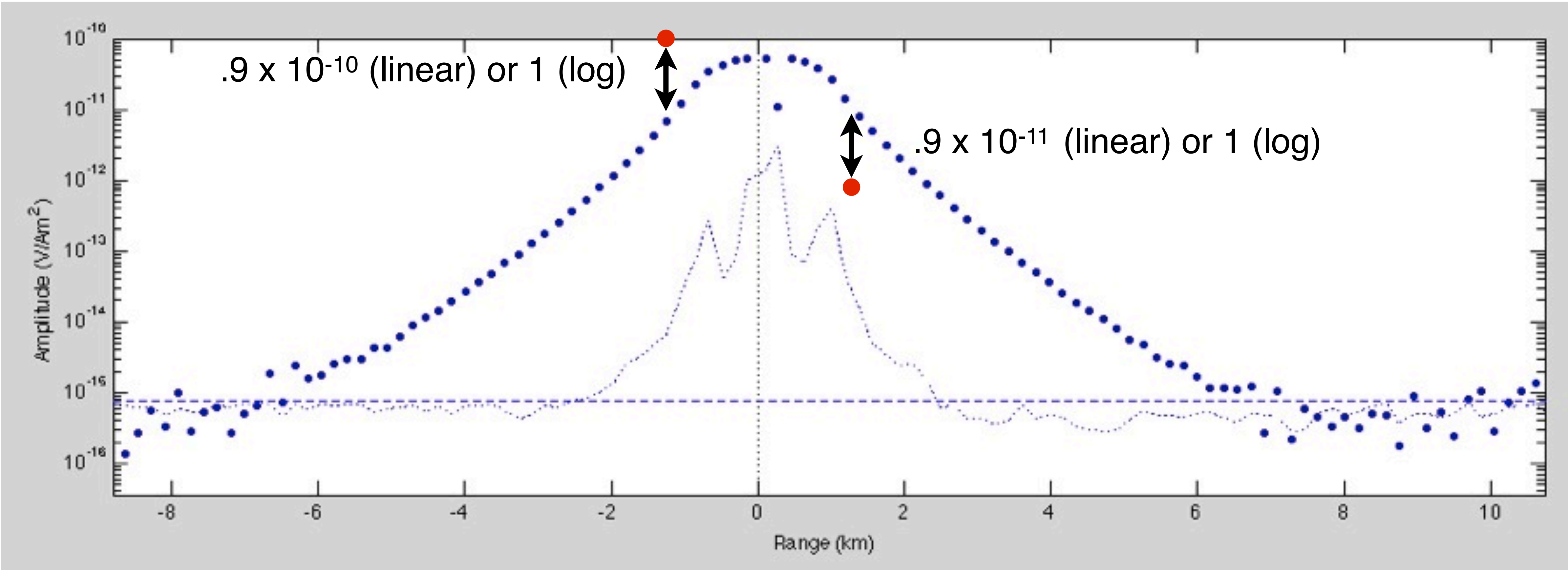
$$0.434 = 1/\ln(10)$$

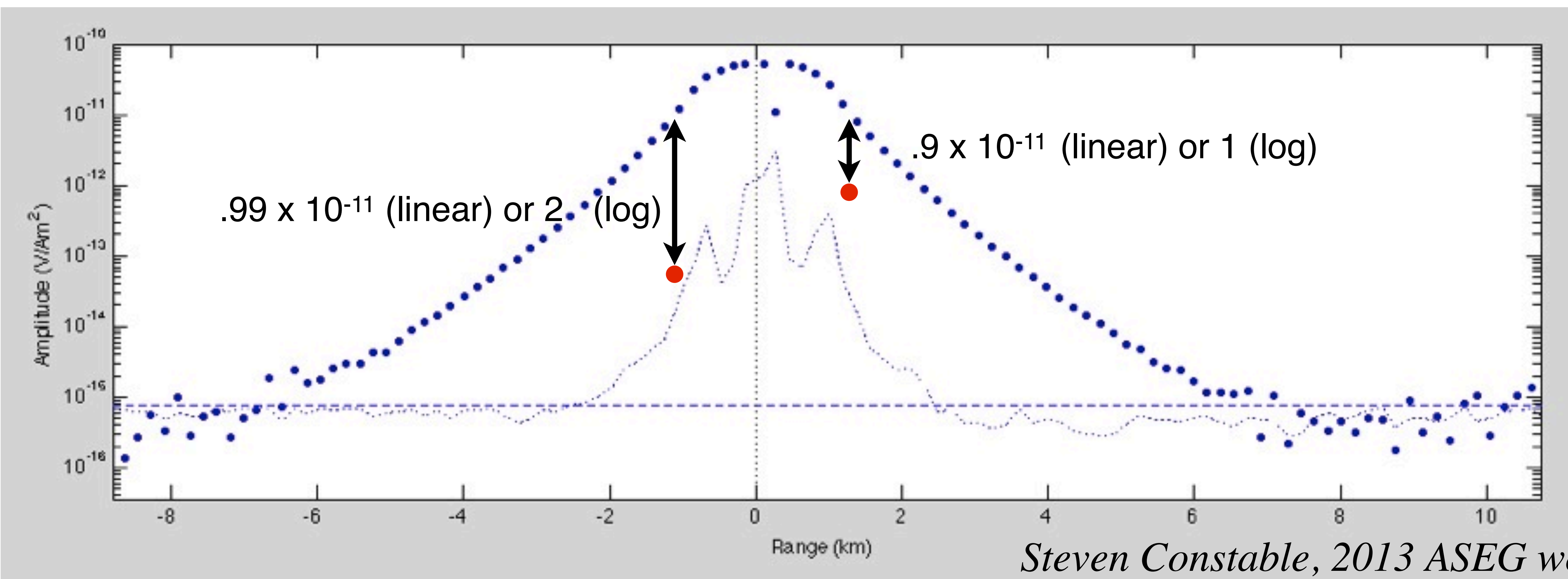
It ought not to matter how you parameterize the data (so long as the errors are properly scaled and the appropriate chain rule is applied to the Jacobian).

But...

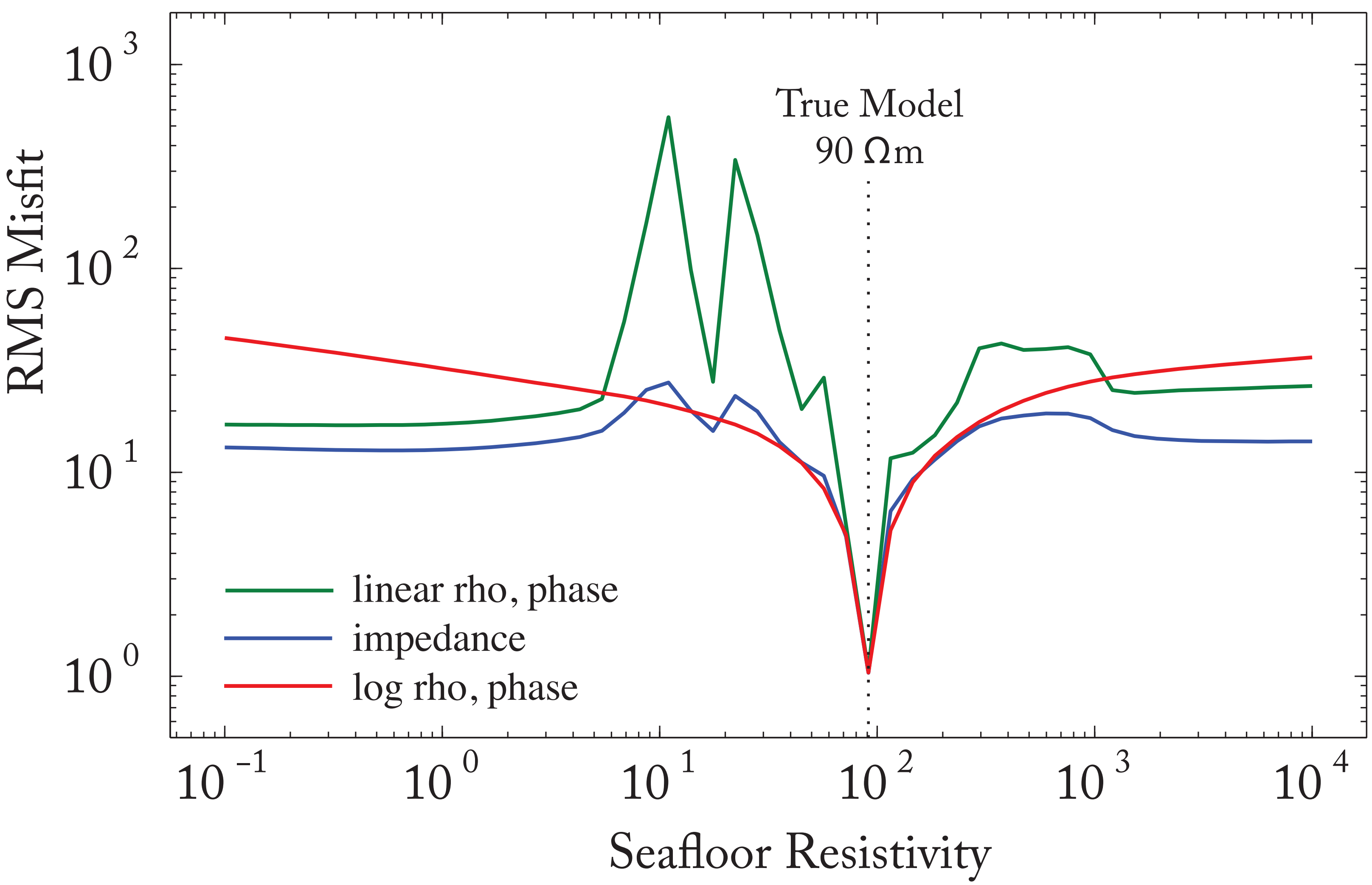
Steven Constable, 2013 ASEG workshop

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





Steven Constable, 2013 ASEG workshop



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

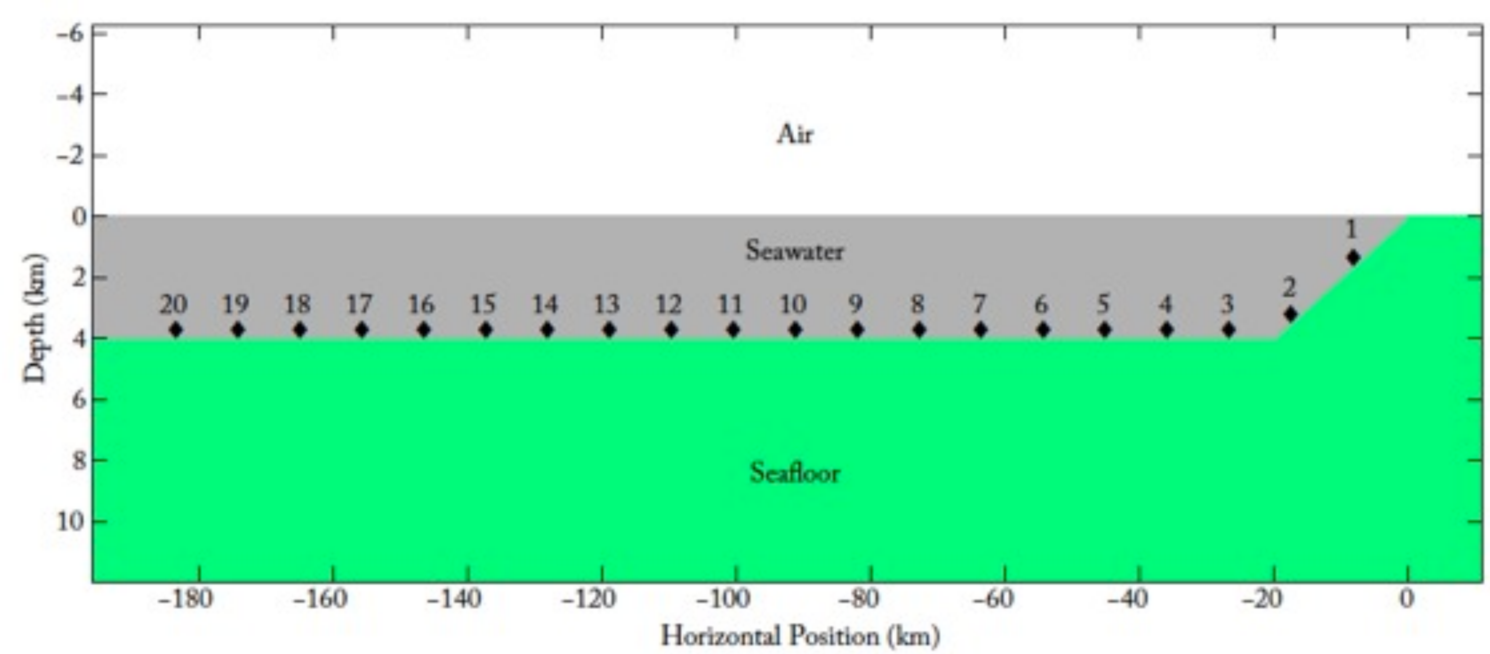
Local minima develop, and misfit flatlines at low ρ .

(modified from Wheelock, 2012)

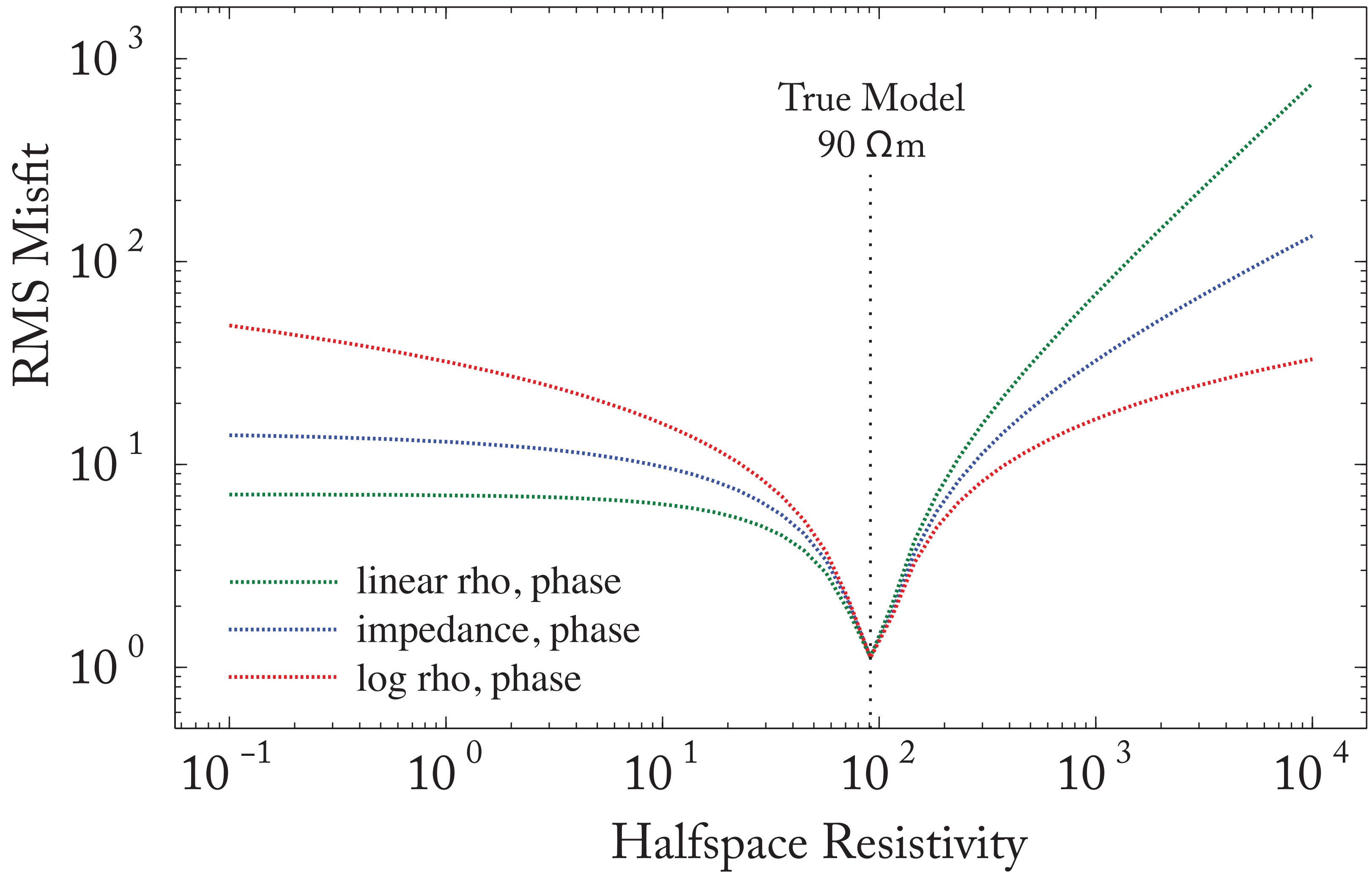
Recall that impedance (Z) is:

$$E_x = ZH_y$$

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



Steven Constable, 2013 ASEG workshop



The local minima are a product of bathymetry, but flatlining still occurs at low resistivities even without seawater.

(modified from Wheelock, 2012)

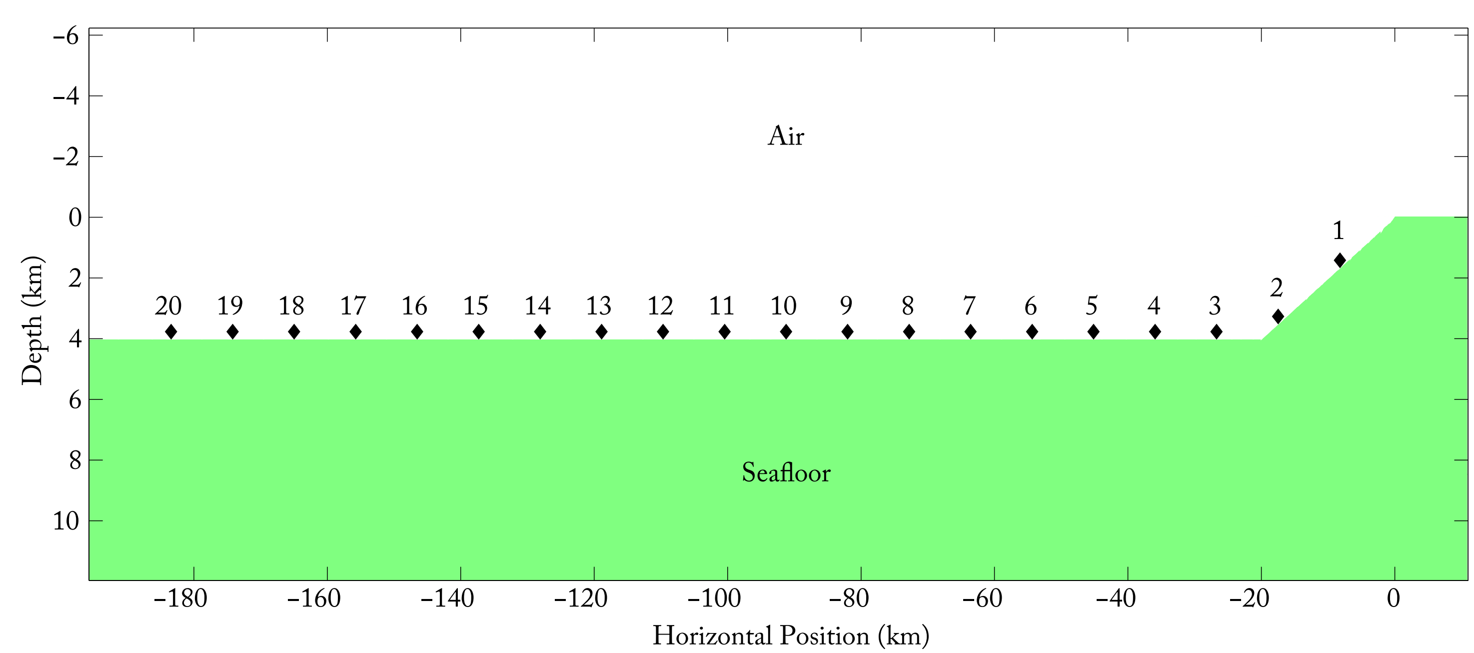
This is because

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

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

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

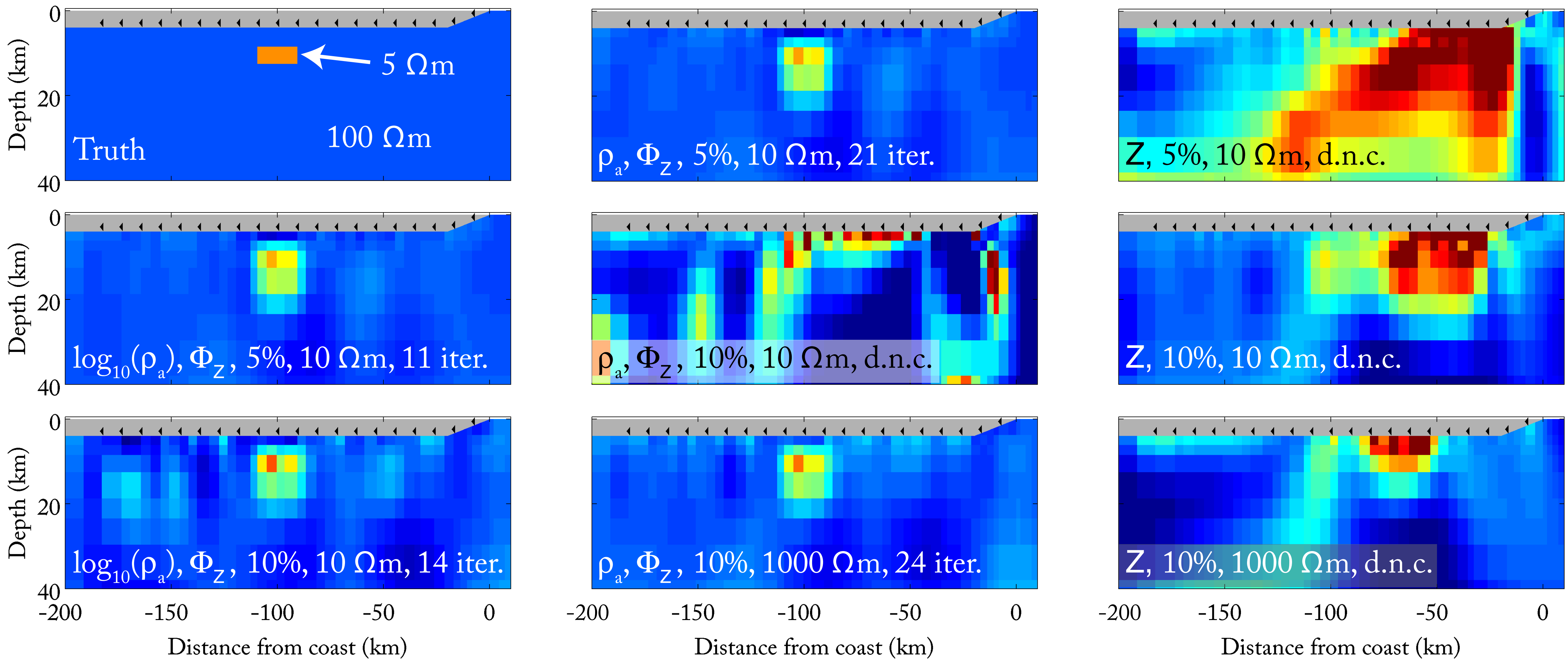
does not.



Steven Constable, 2013 ASEG workshop

Brent showed that not only do $\log \rho$ models converge more reliably, they converge faster.

(Wheelock, 2012)

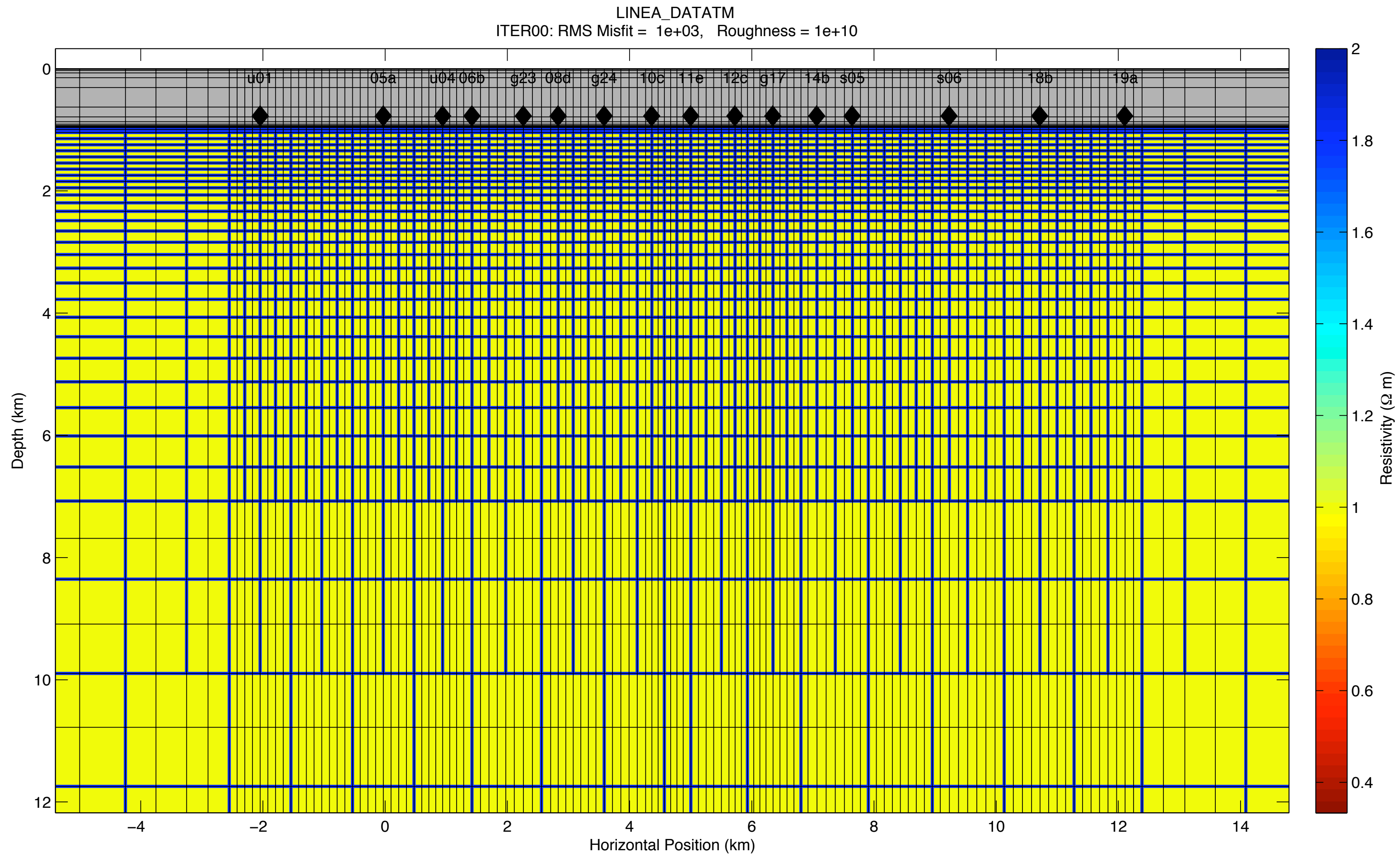


"I don't do marine EM - so why should I care?"

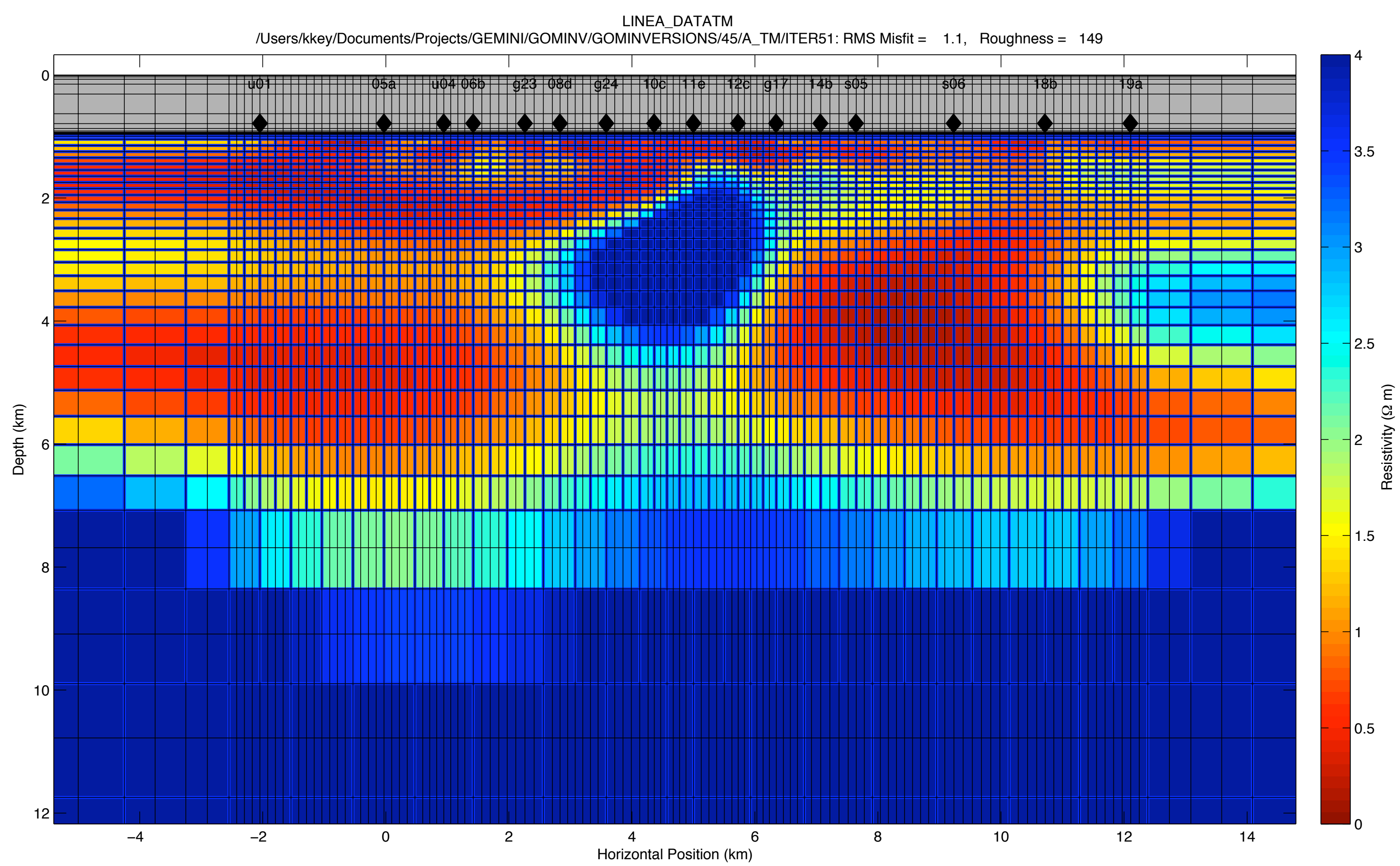
Well... With the advent of 3D MT inversion, more people are inverting the diagonal elements of Z, and not converting them to apparent resistivity, let alone $\log \rho$.

Steven Constable, 2013 ASEG workshop

For rectangular meshes we use a “dual grid” to subsample the computational mesh and allow model blocks to grow with depth. This also keeps the inversion matrices smaller.



However, the aspect ratio of the model blocks can get quite large - this can distort the roughness measure unless appropriate weights are used in **R**.

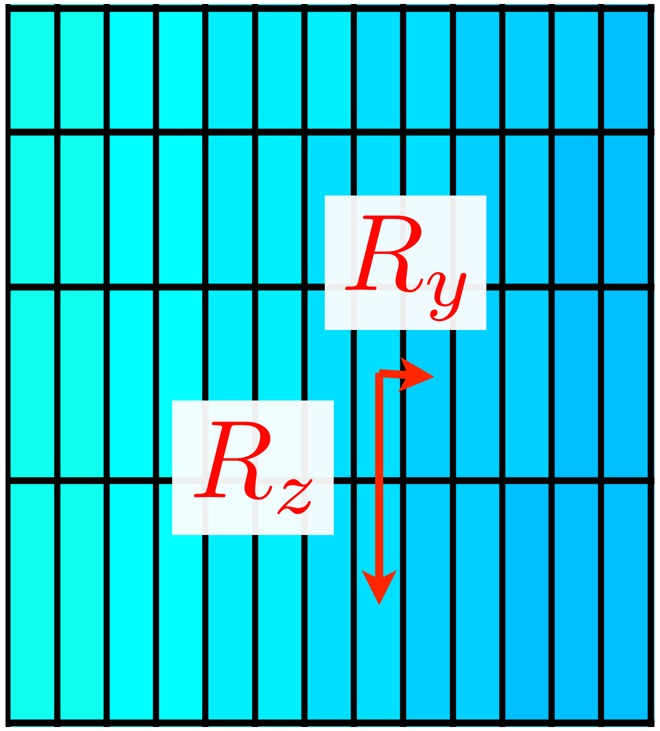


Steven Constable, 2013 ASEG workshop

Choosing R

Striving for Isotropic Smoothing

model mesh



Old Penalty

$$||R\mathbf{m}||^2 = \underbrace{(m_2 - m_1)^2}_{R_y} + \underbrace{(m_3 - m_2)^2}_{R_z} \ldots$$

New Penalty

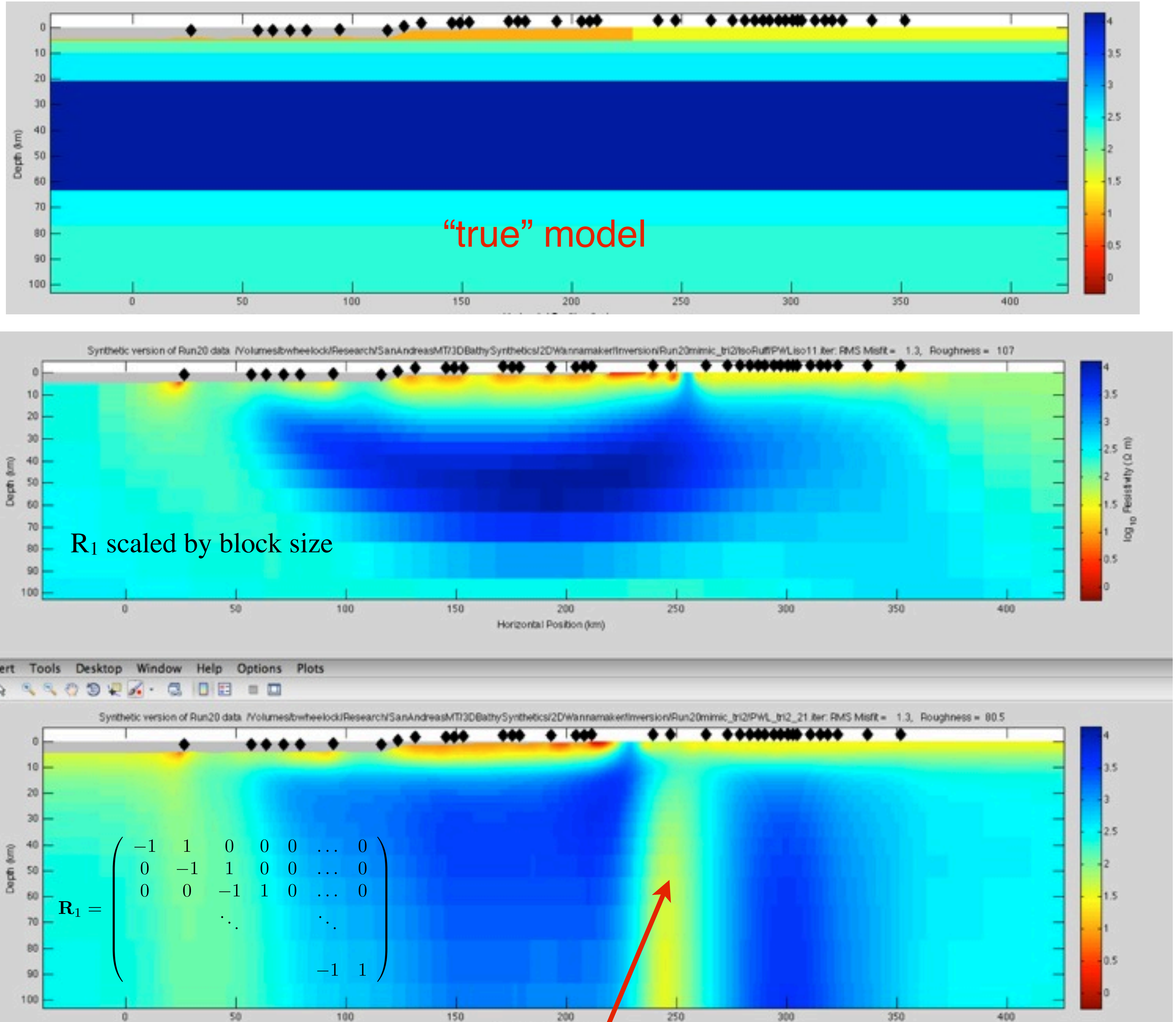
continuous form: $$R_\Omega[m] = \int_\Omega ||\nabla m||^2 \, d^2 s$$

discrete form: $$||R\mathbf{m}||^2 = \underbrace{(m_2 - m_1)^2}_{R_y} \frac{\Delta z}{\Delta y} + \underbrace{(m_3 - m_2)^2}_{R_z} \frac{\Delta y}{\Delta z} \ldots$$

Courtesy Brent Wheelock.

Steven Constable, 2013 ASEG workshop

A subtle and insidious example:

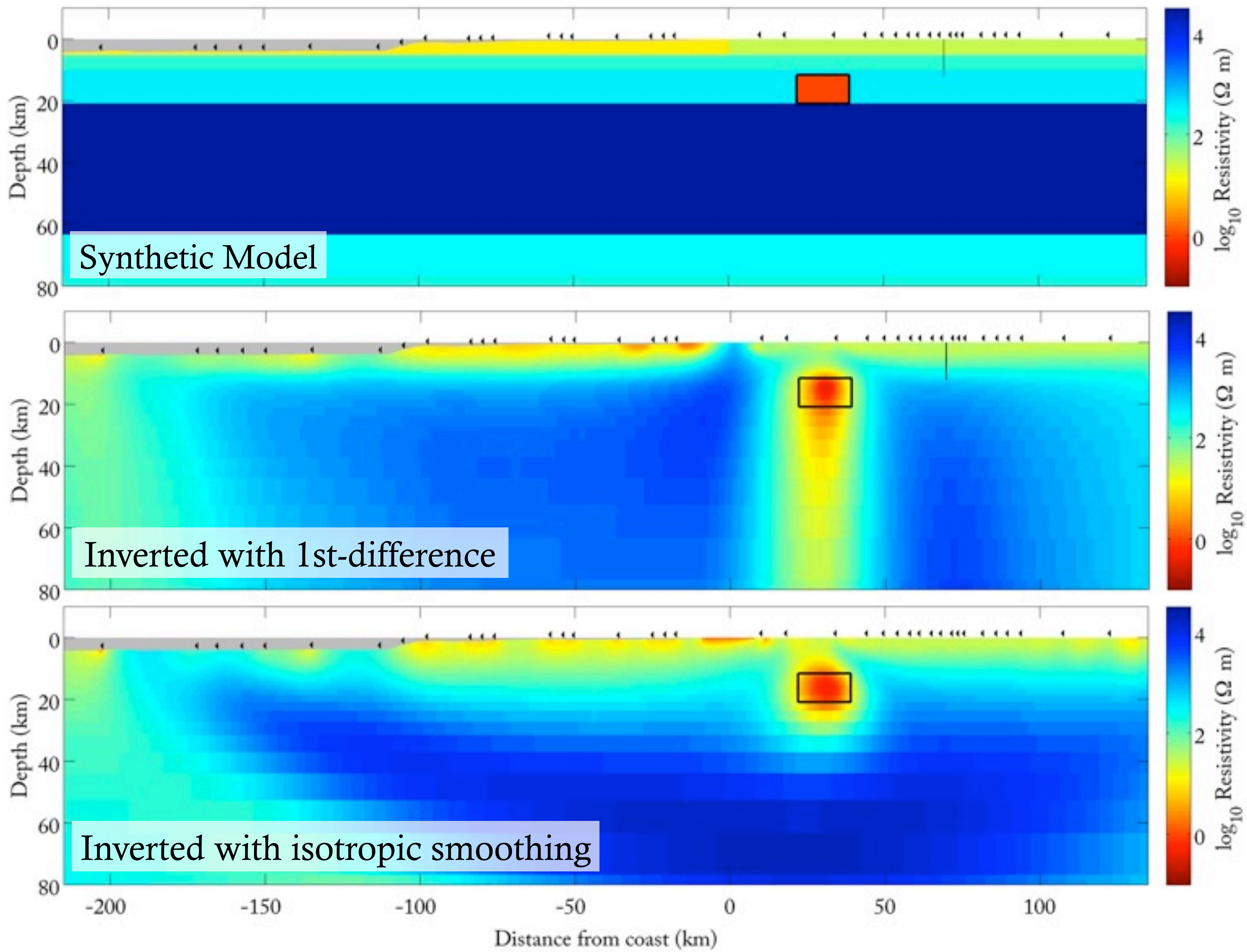


Courtesy Brent Wheelock.

default regularization generates a conductive artifact

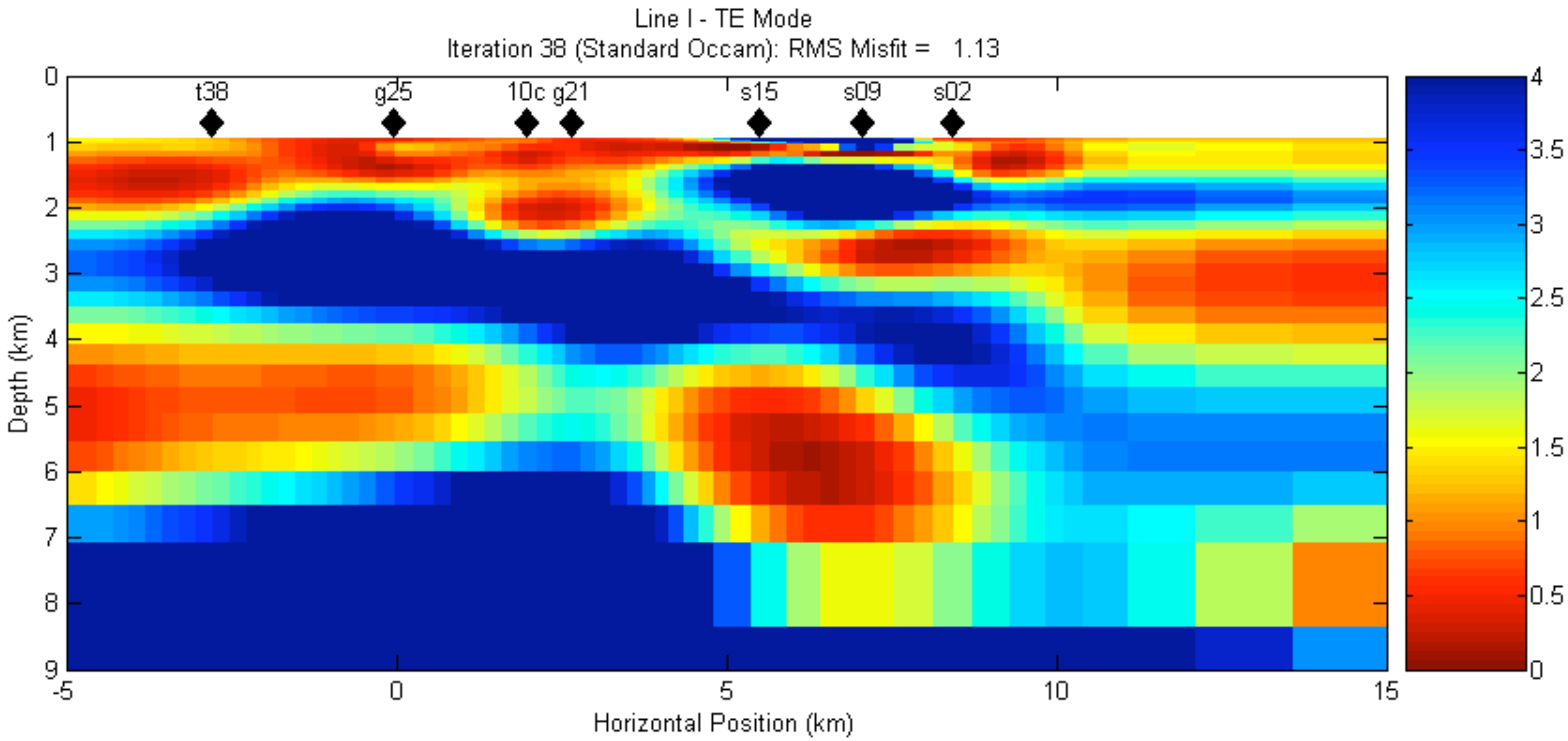
Steven Constable, 2013 ASEG workshop

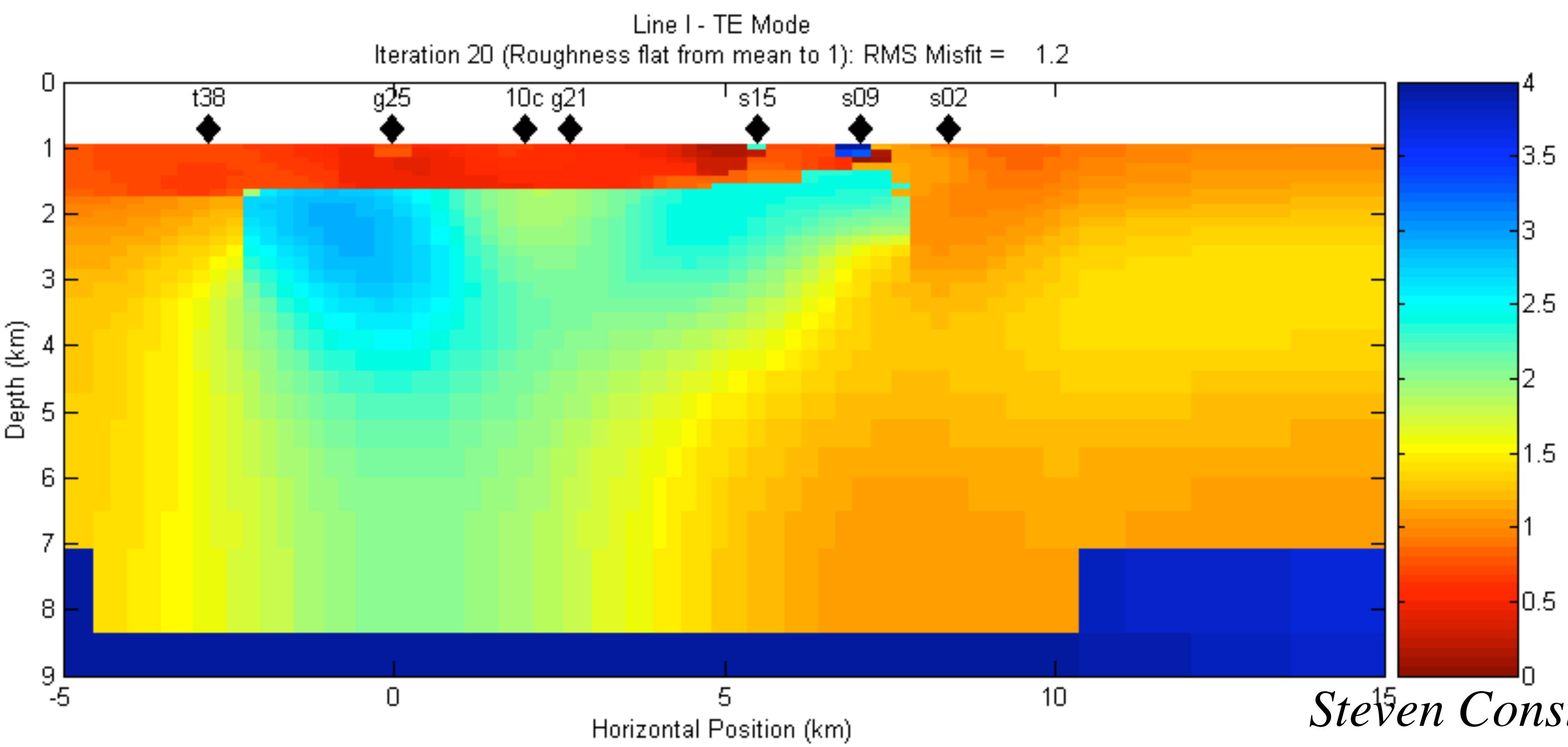
Another example:



Steven Constable, 2013 ASEG workshop

Always remember that regularization has an input into the model solution.





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

These two models both fit data to RMS 1.1-1.2

R is nonlinear and adaptive, providing little penalty for big contrasts

Courtesy David Myer.

Steven Constable, 2013 ASEG workshop

Can we place errors or uncertainties on regularized models?

No.

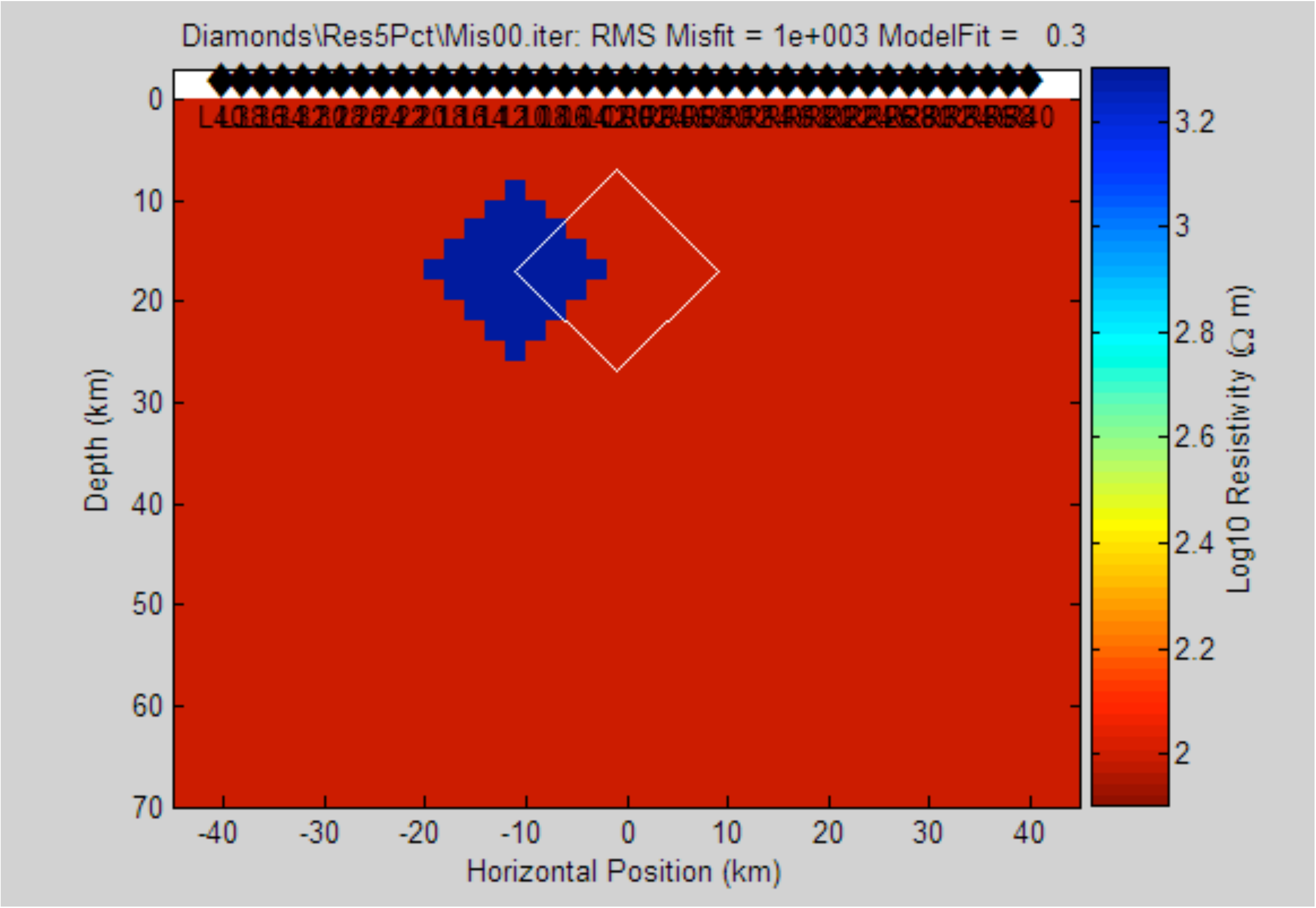
For sparse parameterizations, data errors can be projected onto model parameters through the Jacobian. This is a dangerous practice because

- it depends on the parameterization

- the Jacobian depends on the solution

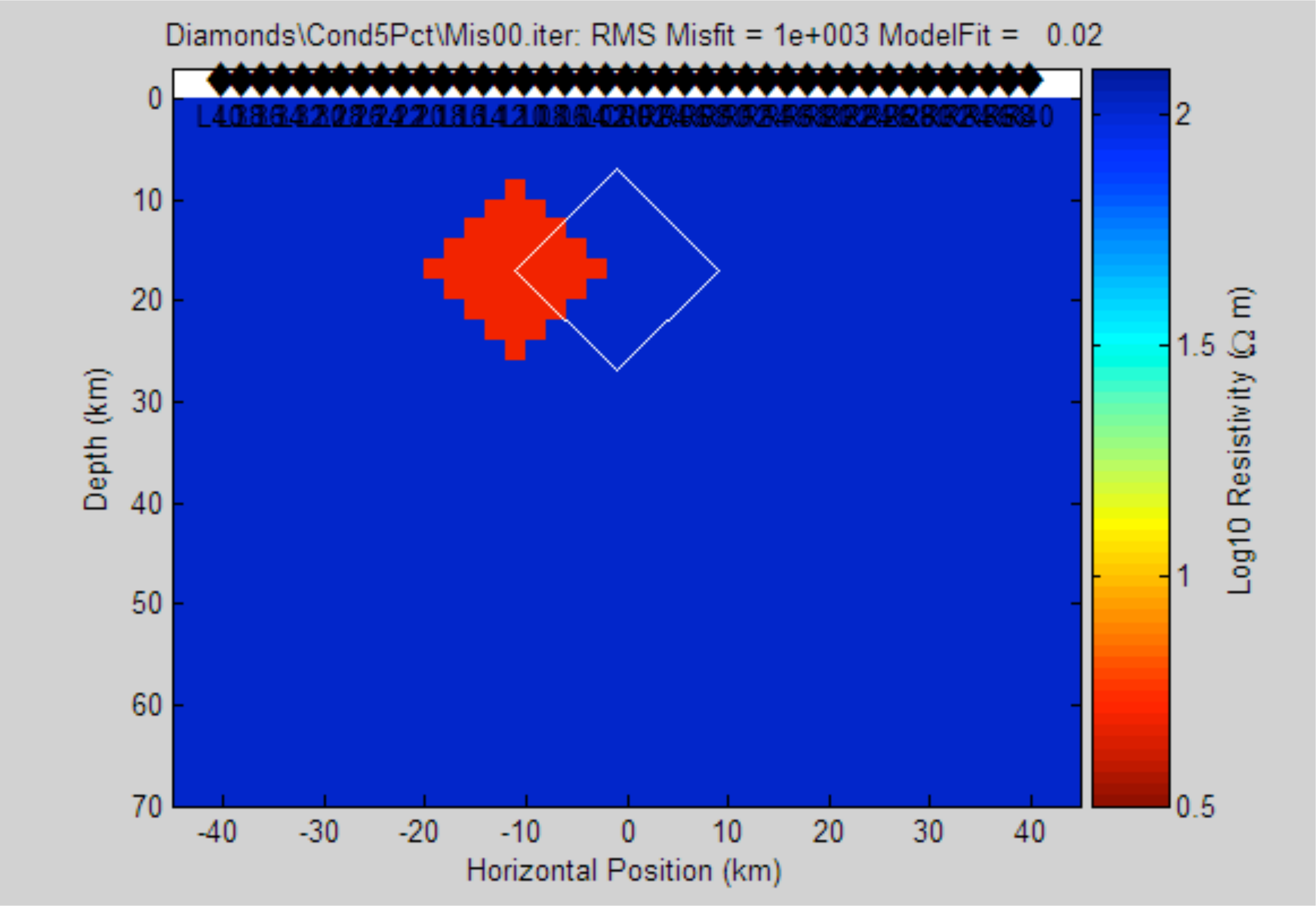
For regularized models, however, we compound this by generating huge amounts of covariance between the model parameters as part of the smoothness constraint. It is much more useful to think of regularized models as extremal solutions, and vary the regularization to ask the questions you may have.

Because **J** depends on **m**, it is best start inversions from a half-space



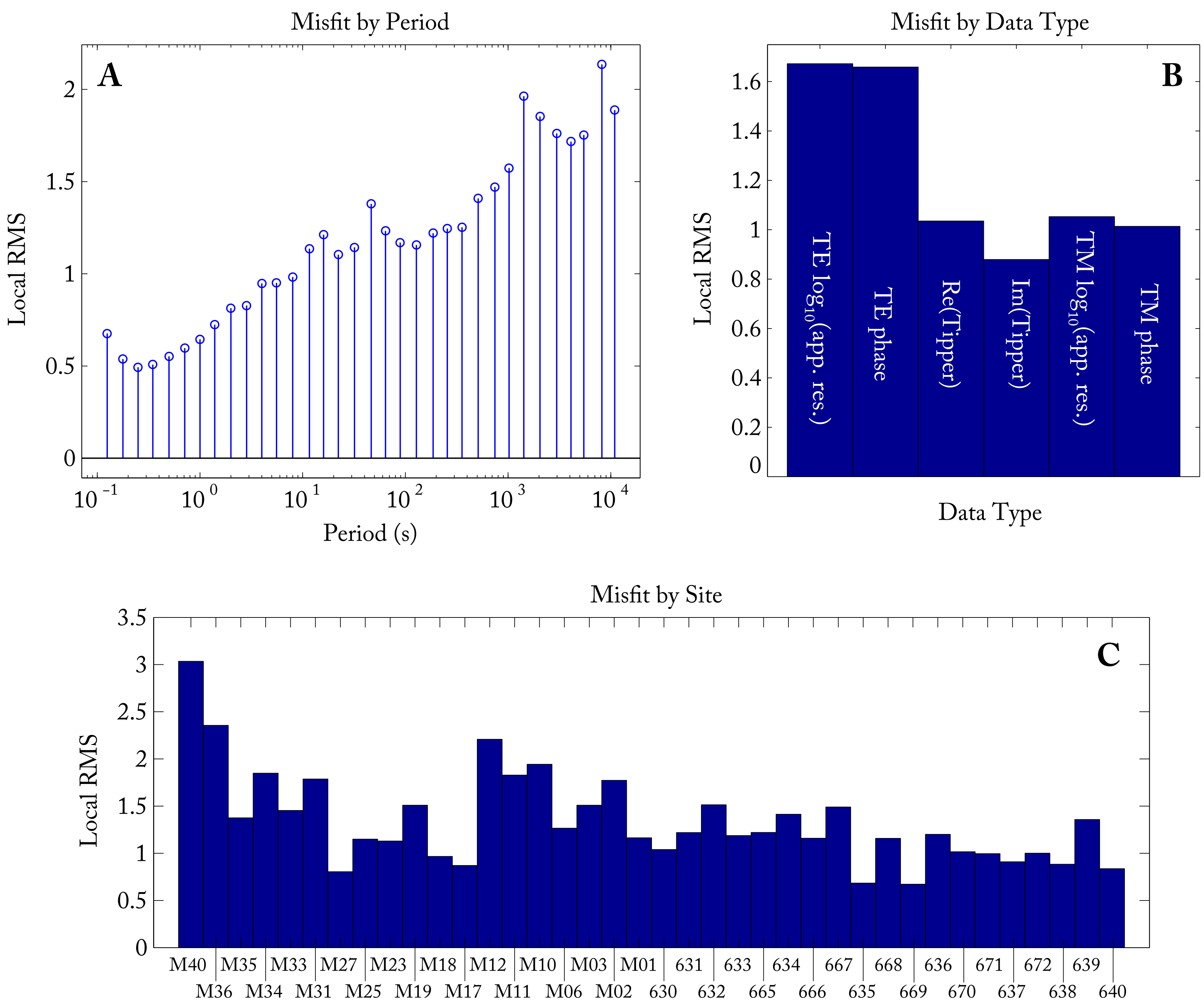
MT: misaligned starting resistor - no harm done

Courtesy David Myer.



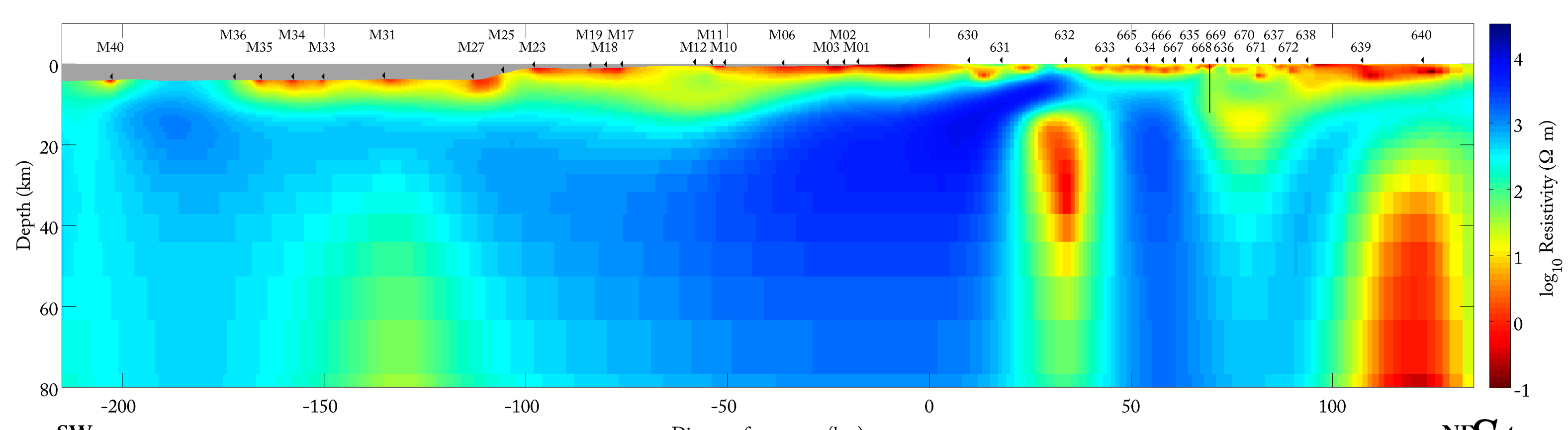
misaligned starting conductor - forever trapped by **J**

Courtesy David Myer.



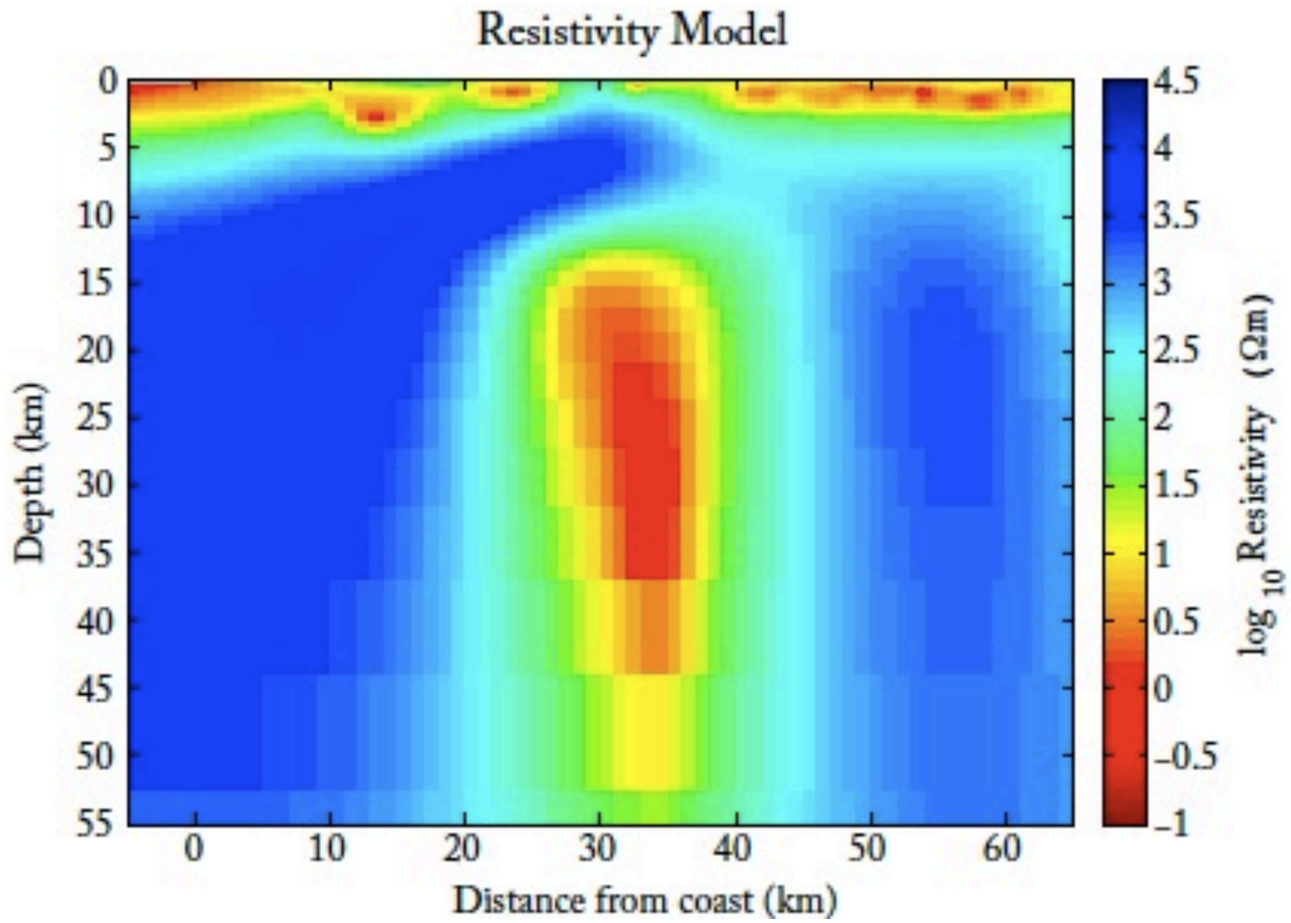
PhD thesis, Brent Wheelock, 2012.

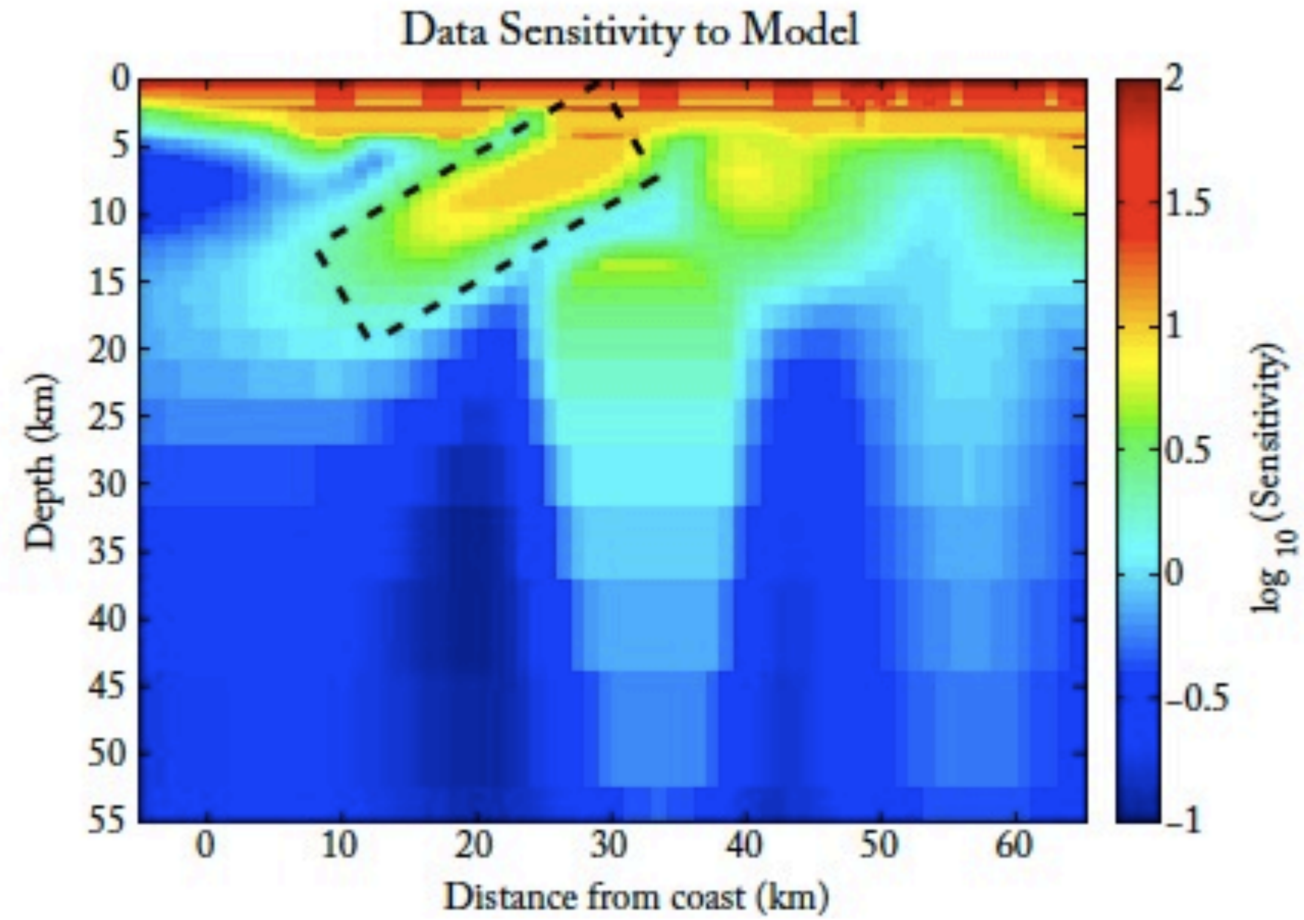
It is a good idea to examine residuals for outliers, and also how the error is partitioned; ideally it should be random across all data, but in practice very rarely is.



Steven Constable, 2013 ASEG workshop

Projecting the Jacobian onto the model can provides some insights into sensitivity (c.f. uncertainty):

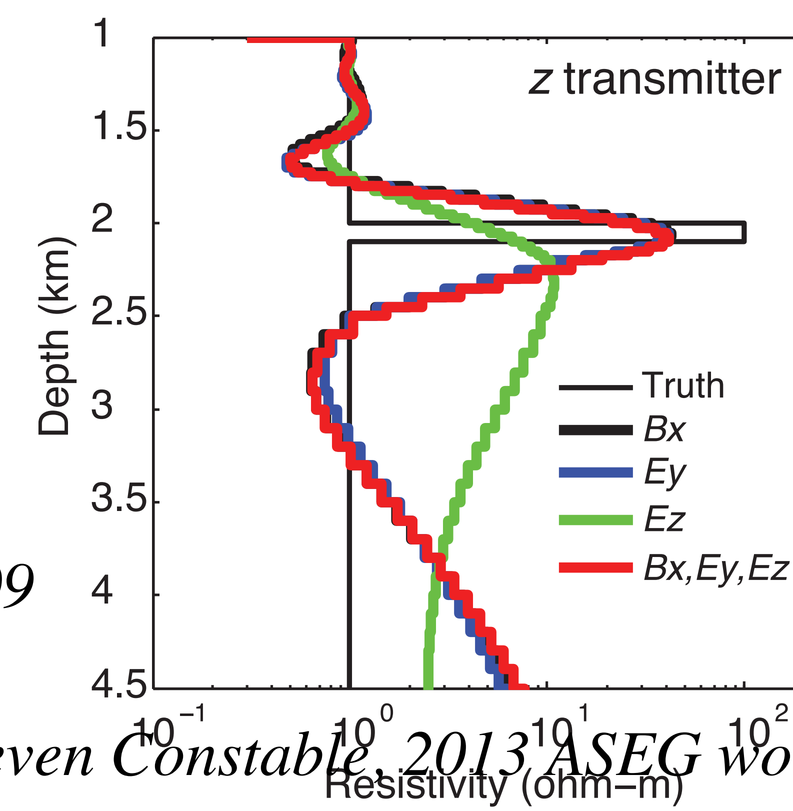




PhD thesis, Brent Wheelock, 2012.

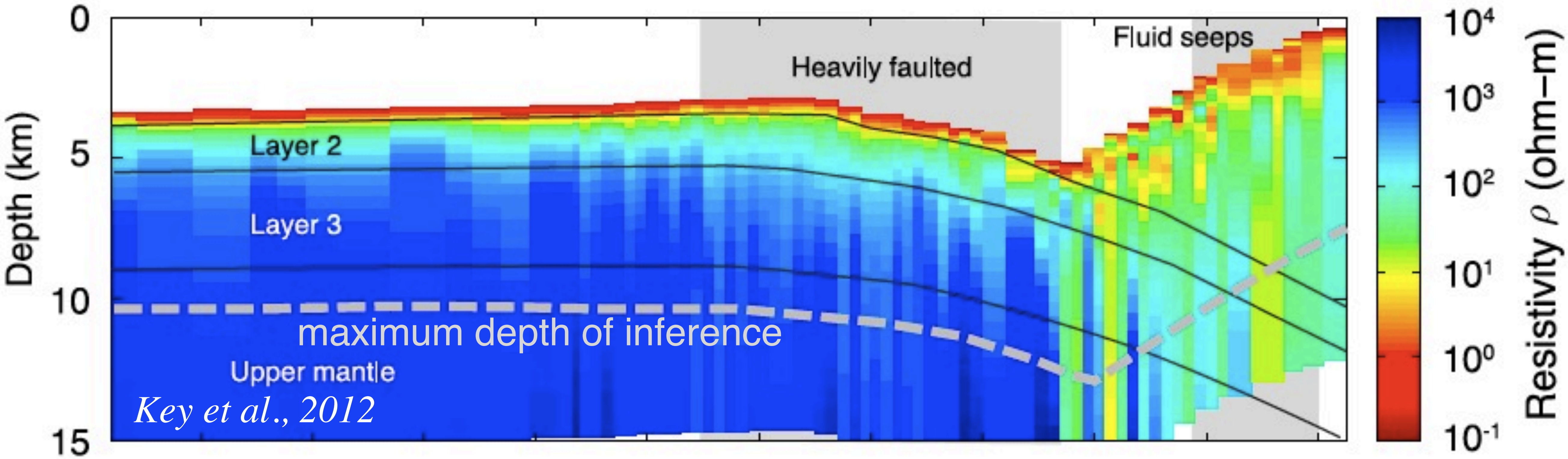
But don’t confuse sensitivity with resolution.

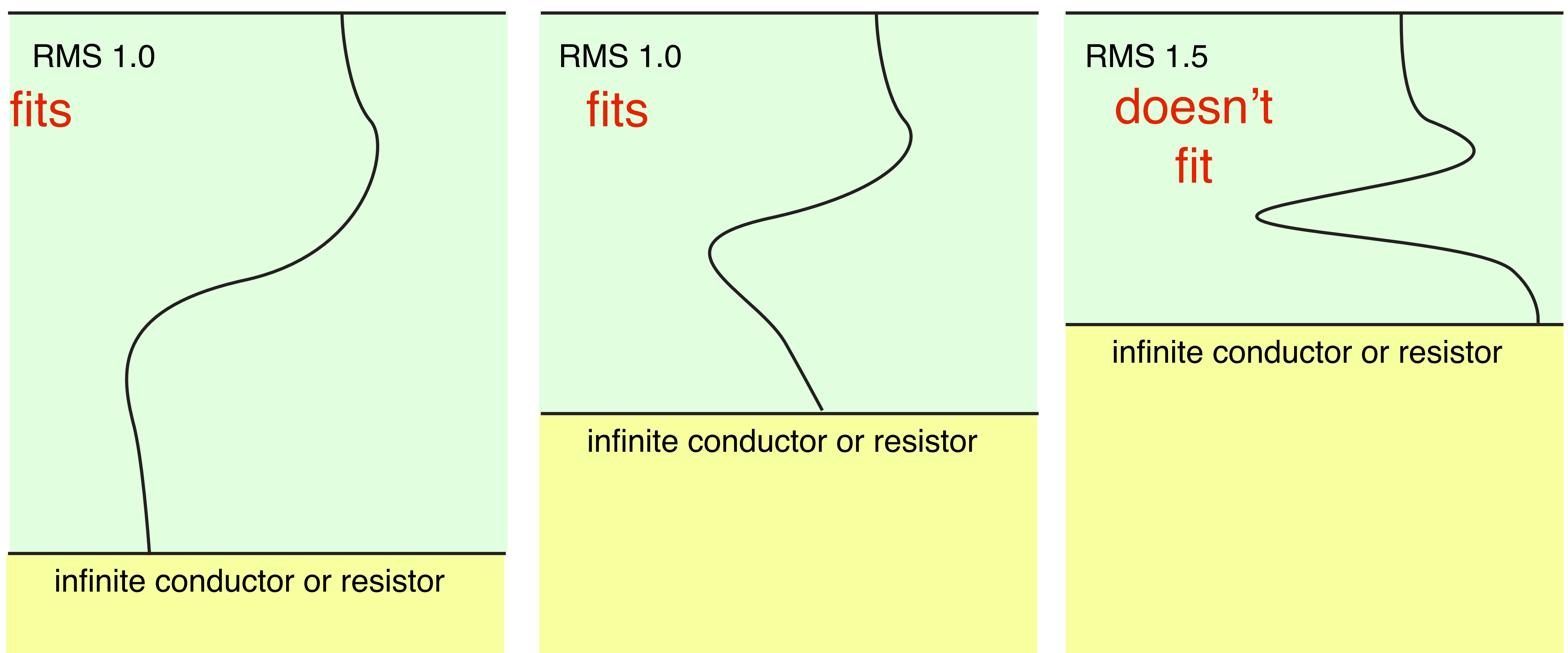
Key, 2009



Steven Constable, 2013 ASEG workshop

Sensitivity and resolution: Maximum depth of inference, *a la* Parker.

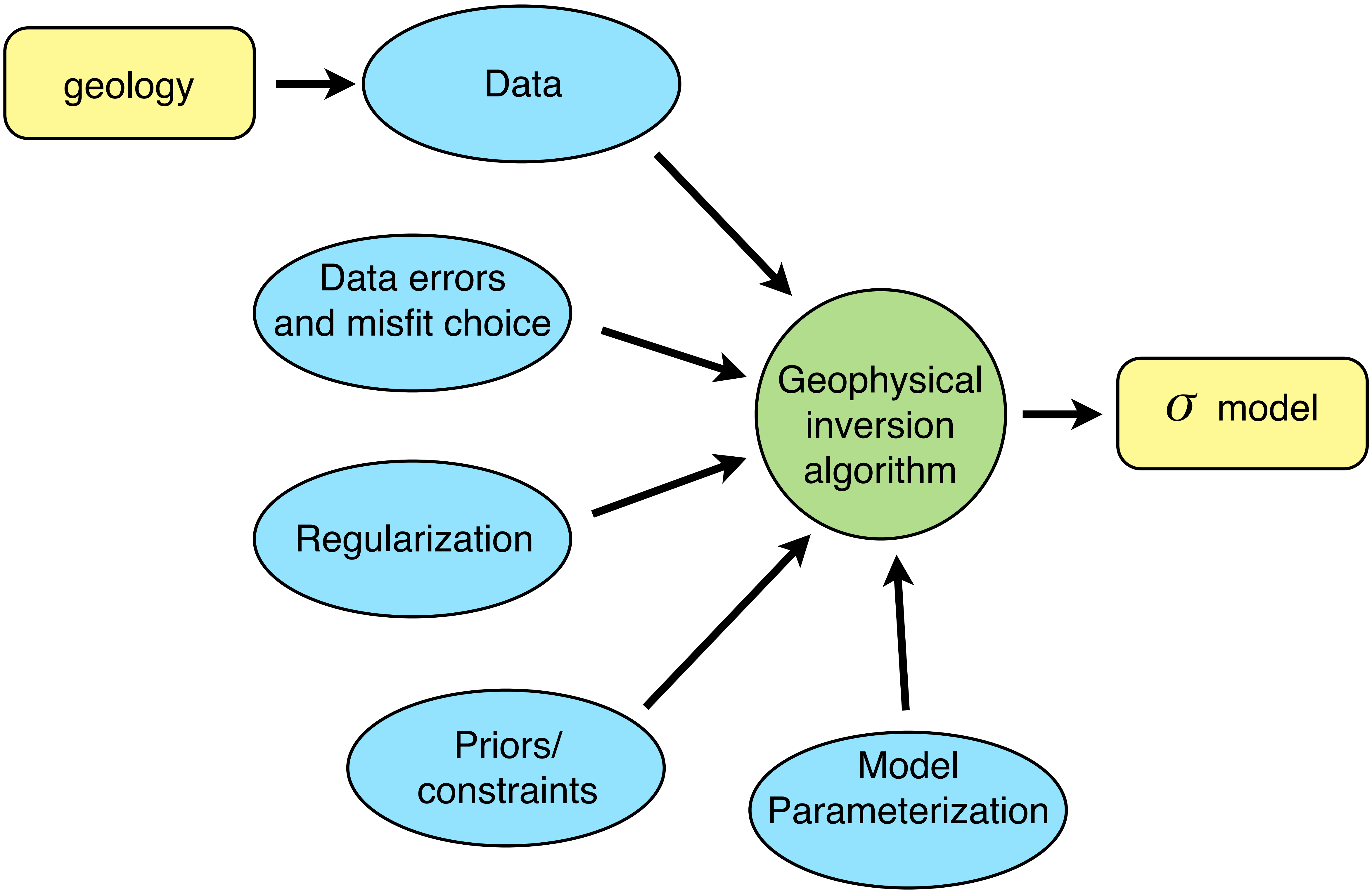




drive a high conductor or resistor up until an inversion cannot fit the data

Steven Constable, 2013 ASEG workshop

So I will finish as I started: models from geophysical inversion depend on much more than the data alone:



Steven Constable, 2013 ASEG workshop