Inverse Theory and Seismic Tomography

CIDER 2018

VED LEKIC

JULY 13, 2018

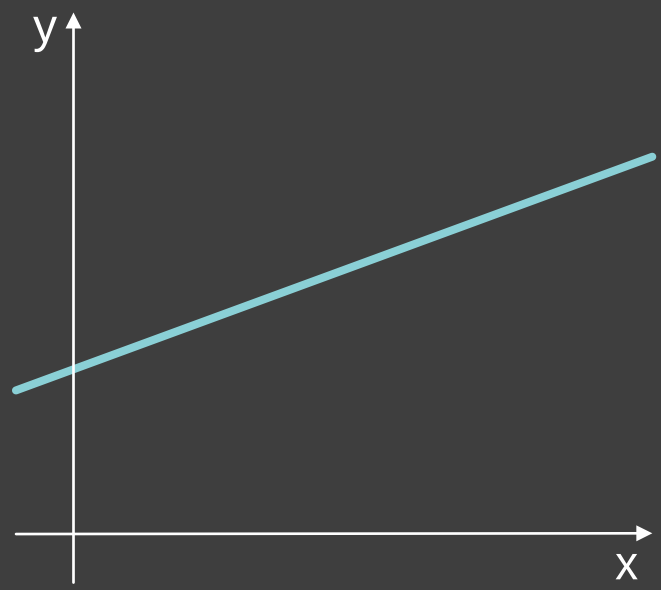
What is Inverse Theory?

- A combination of approaches for determining and evaluating physical models from observed data
- An approach to calculate data from known model must be possible → **forward problem**
- Ingredients:
  - **Physics** (usually) to define forward problem
  - **Linear algebra** (usually) to map between model and data spaces
  - **Probability and statistics** (sometimes) to quantify uncertainty of our physical model determinations

Types of forward problems

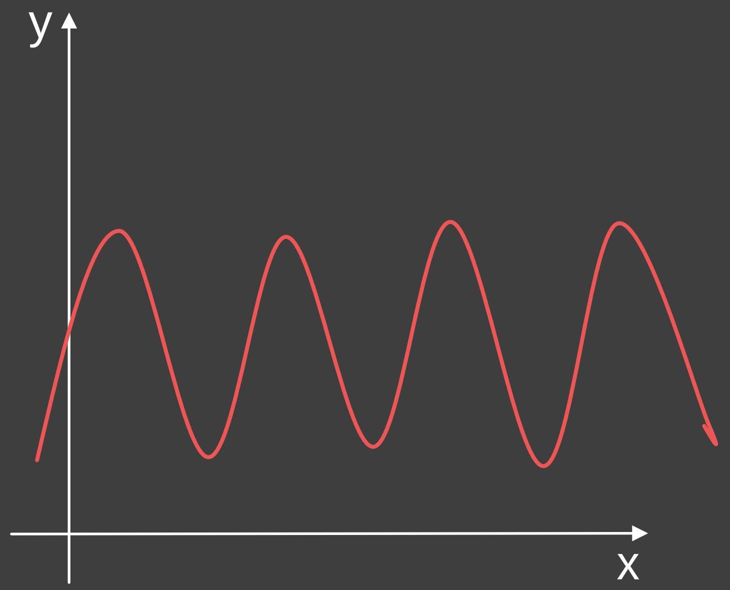
Linear

- $y(x) = m_1 x + m_2$
- e.g. predict gravity given density of an object at known location



Non-linear

- $y(x) = \sin(x m_1) + m_1 m_2$
- e.g. predict seismic waveform given 3D velocity structure of the Earth



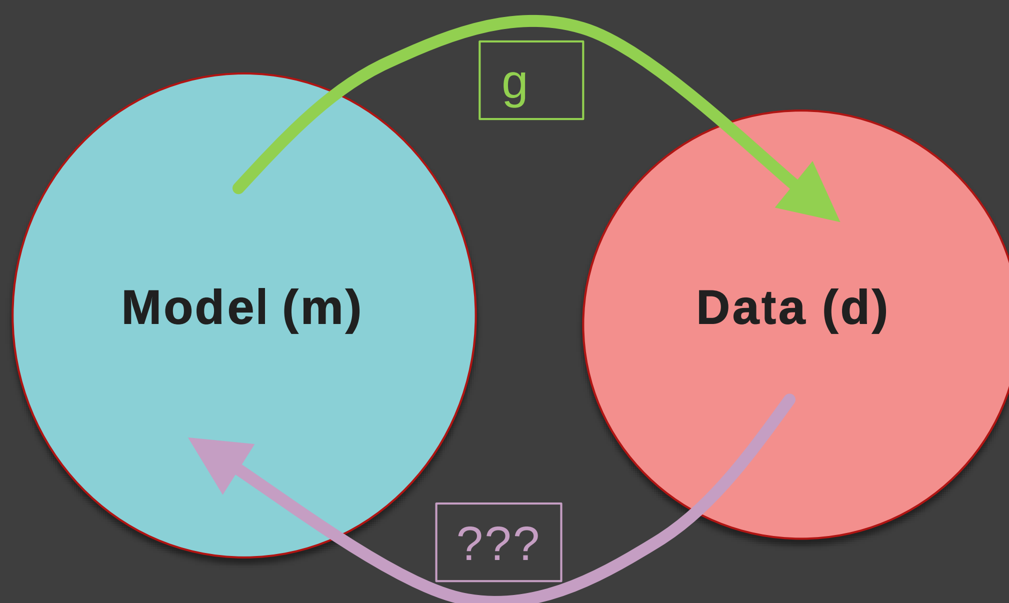
Typical formulation

- Observations are sampled in time and/or space and lumped into a data vector (**d**)
- Physical quantities of interest are expressed as a finite set of parameters that are lumped into a model vector (**m**)

$$\mathbf{d} = g(\mathbf{m})$$

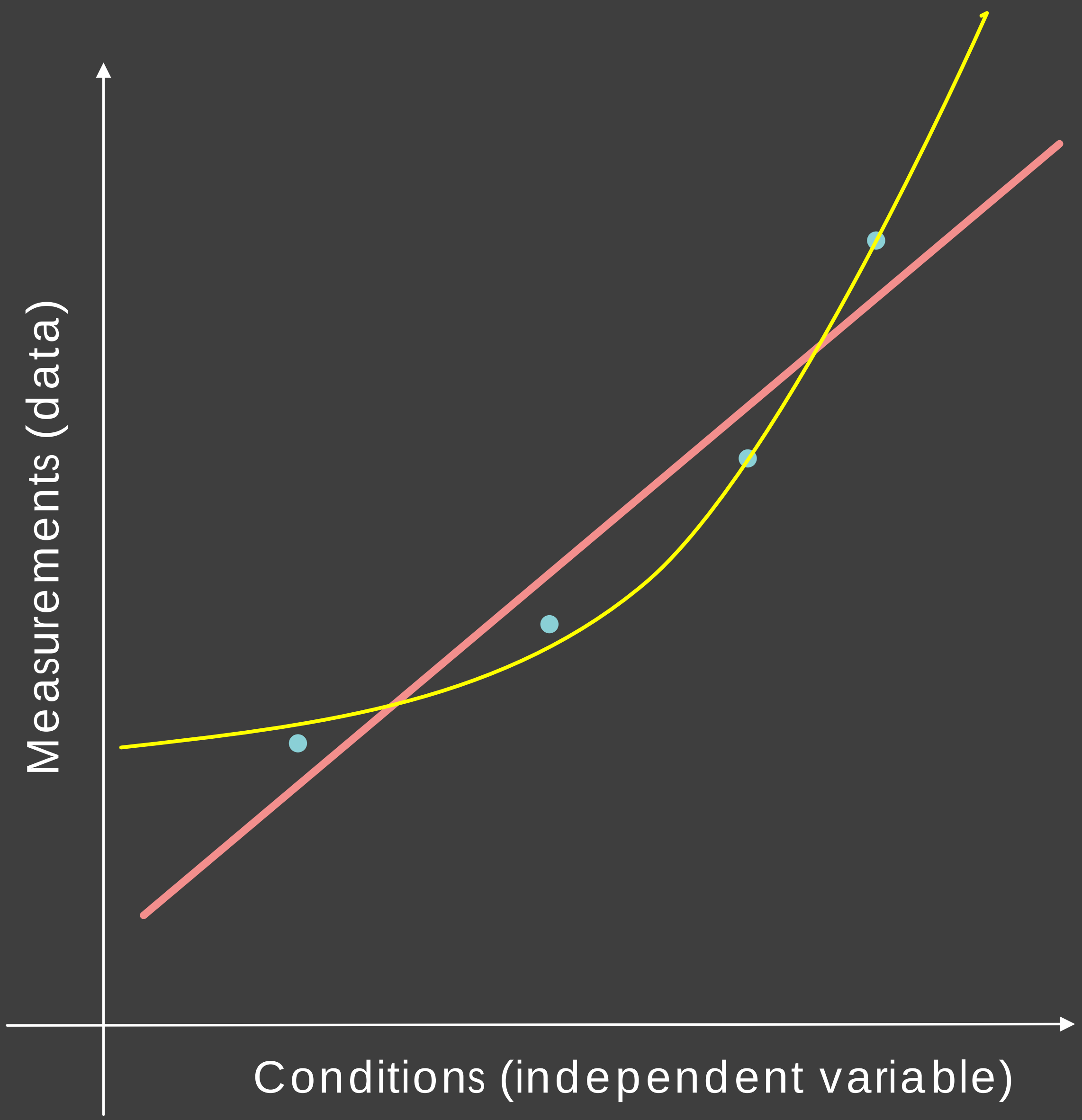
Data vector

Model vector



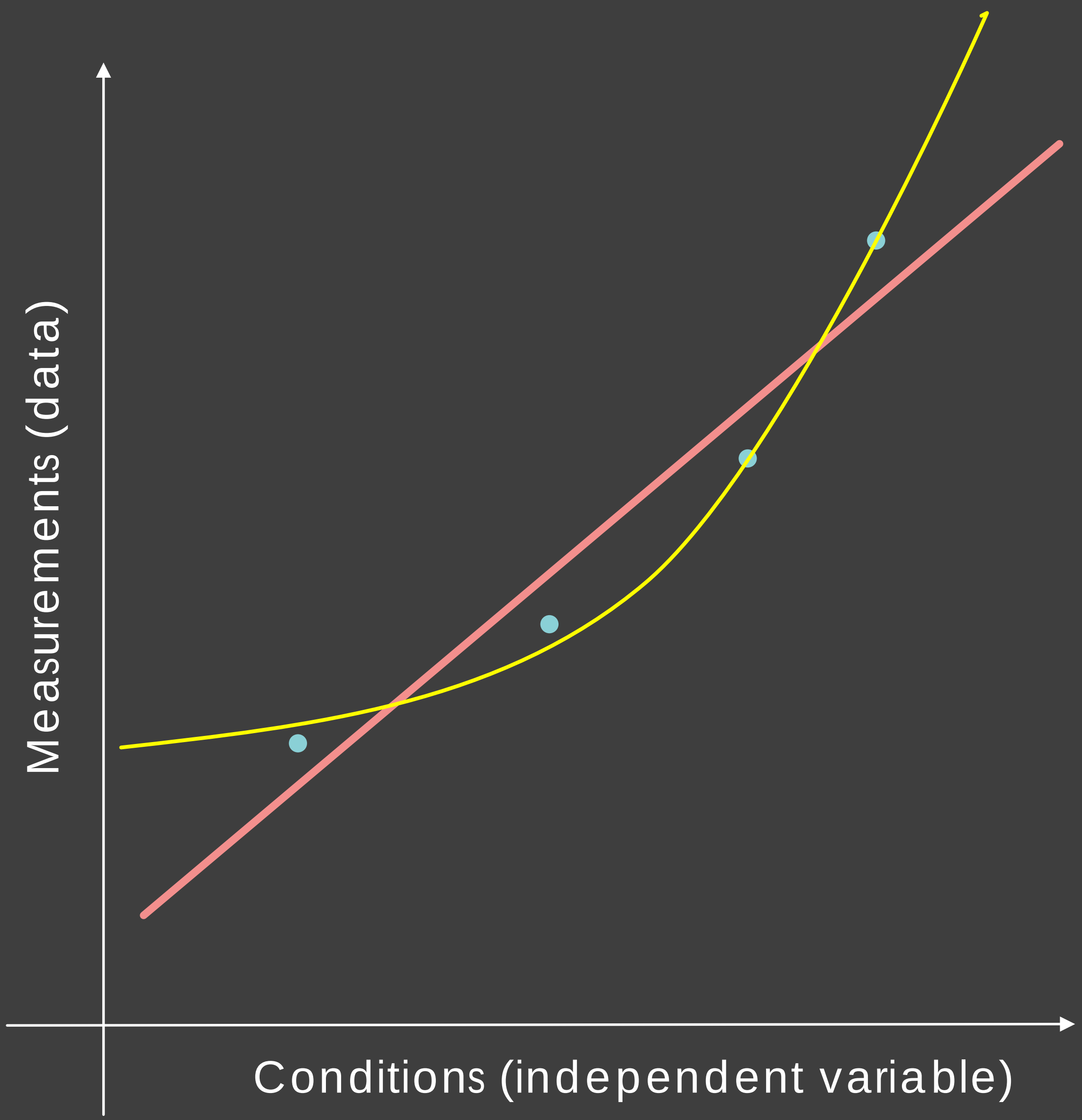
Parameterization

- What quantities (parameters) should we consider?
- Consider pairs of quantities, such as measurements (y) made at conditions (x)



Parameterization

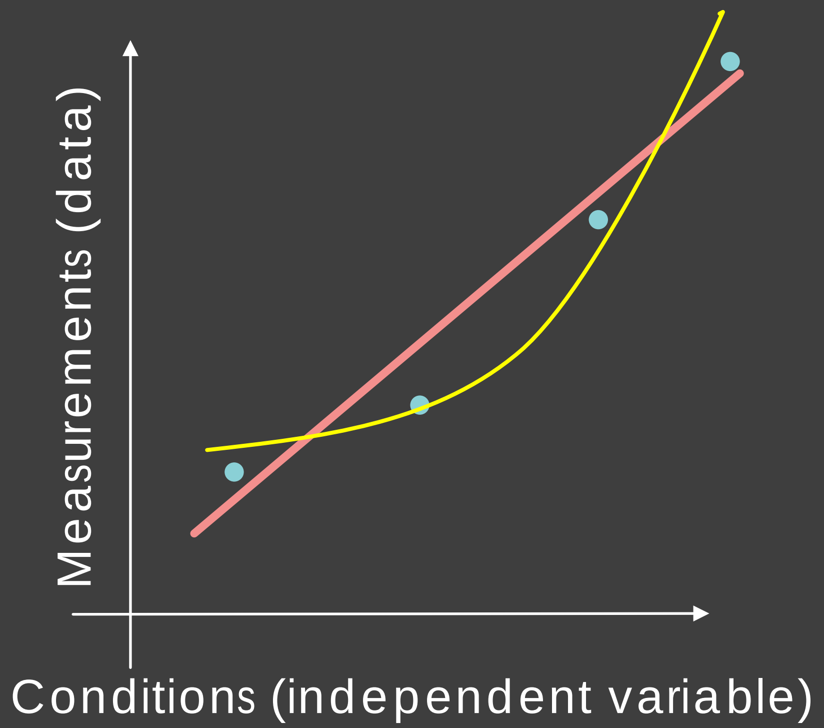
- Physical model may provide insight into how to parameterize a problem
- Sometimes we seek “simplest” model (fewest parameters) that can fit data



Law of parsimony – a useful heuristic

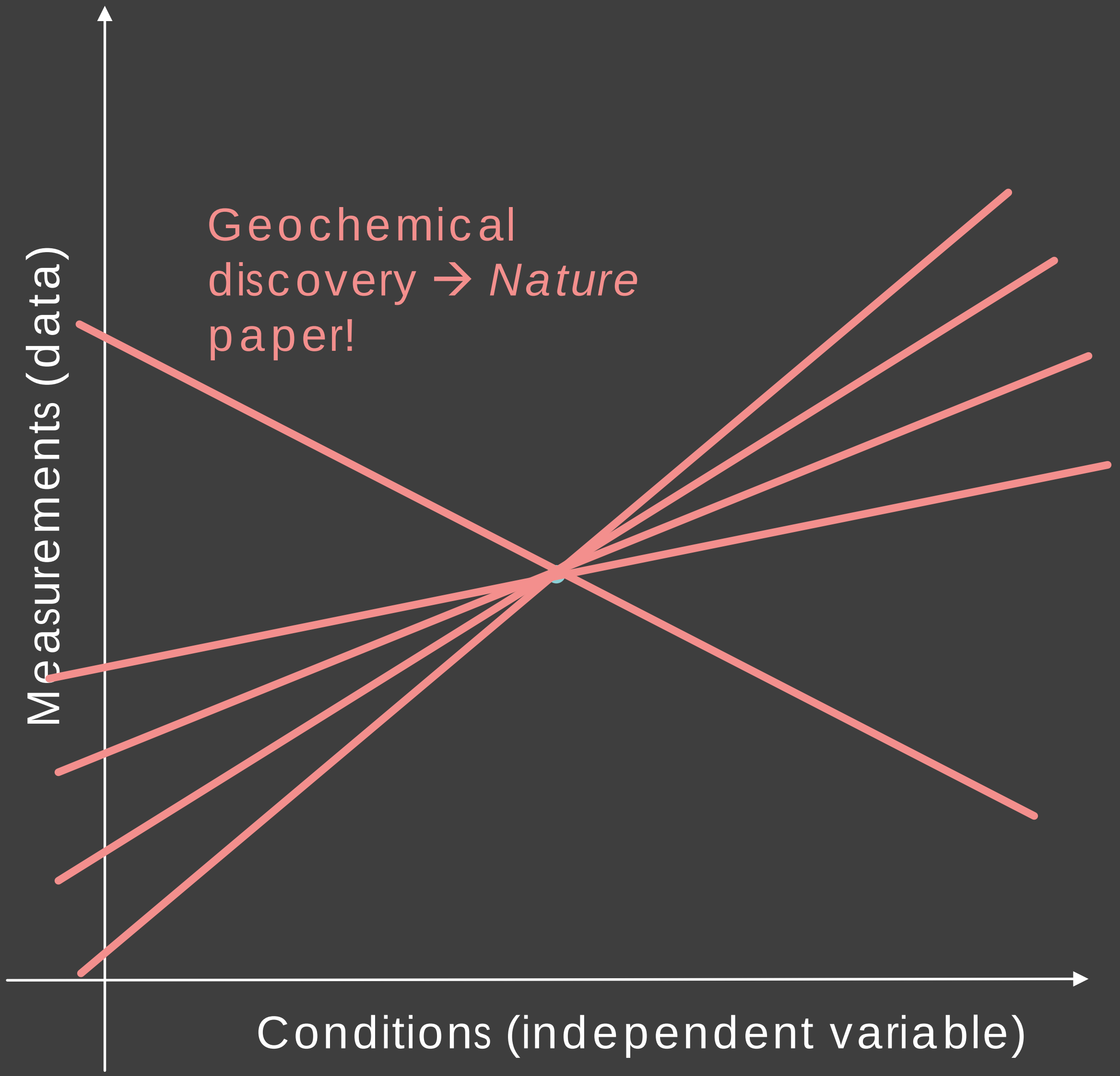
- Preference for “simpler” models is often justified by invoking “Occam’s Razor”.
  - e.g. “It is vain to do more with what you can do with less.”
    – some 13th century monk
- Karl Popper argues that simplest models are preferable because they can be more easily falsified
- *"I remember my friend Johnny von Neumann used to say, with four parameters I can fit an elephant, and with five I can make him wiggle his trunk."*

  -- Enrico Fermi, Nature (2004).



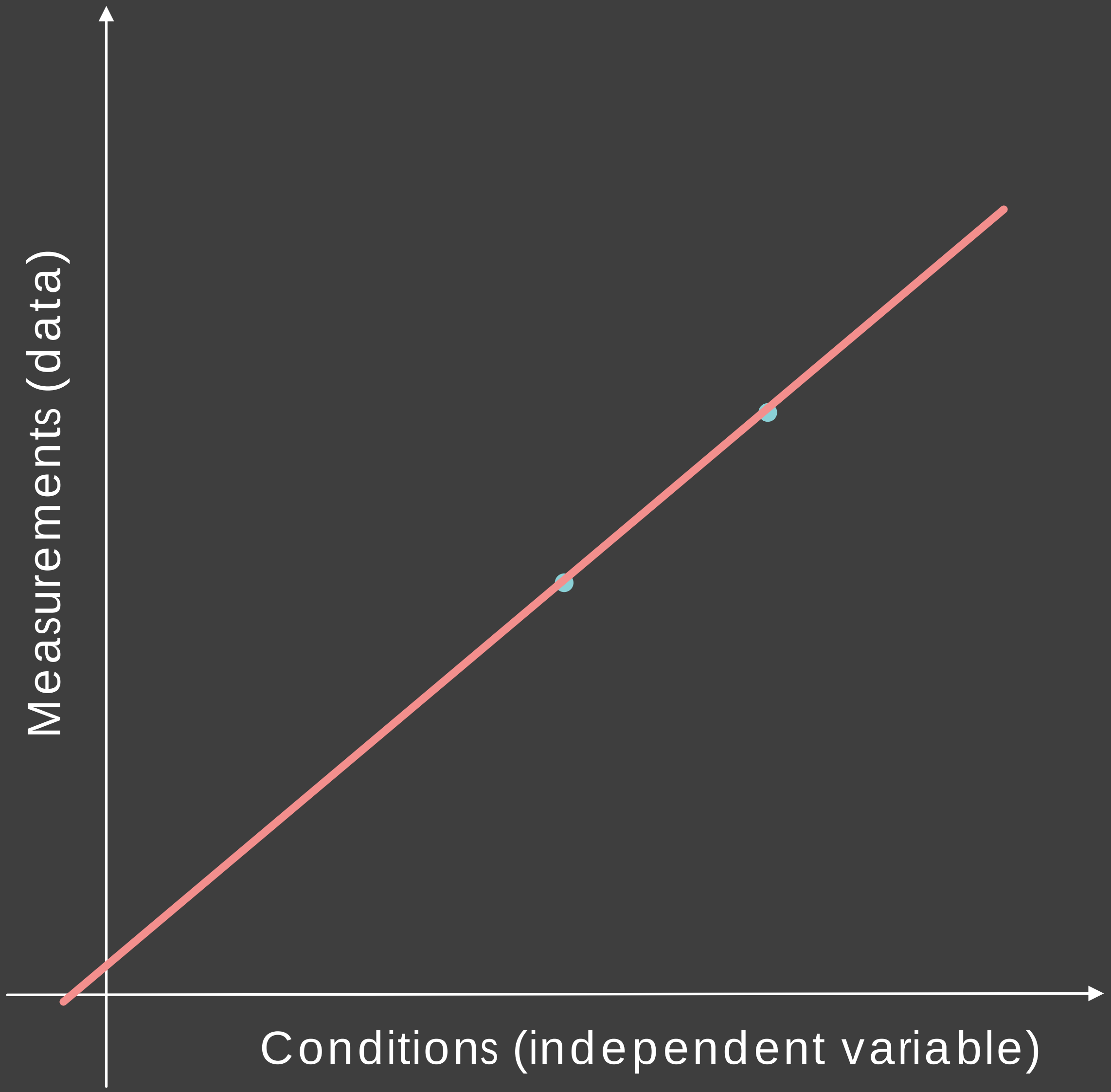
Under-determined problem

- With 1 data point, we cannot uniquely determine slope and y-intercept
- Complete trade-off between m_1 and m_2!



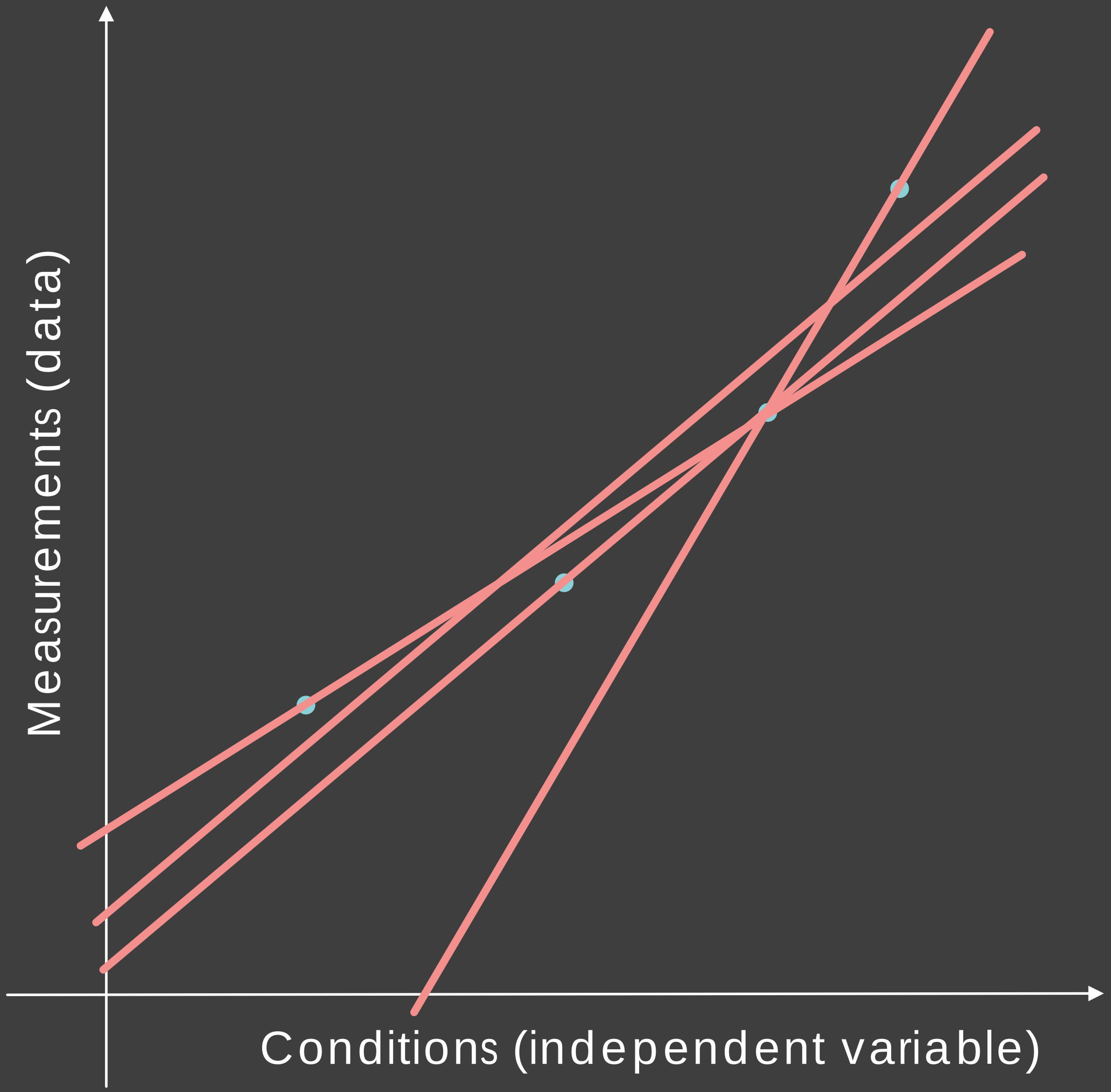
Even-determined problem

- When we have 2 data points, we can uniquely determine m_1 and m_2



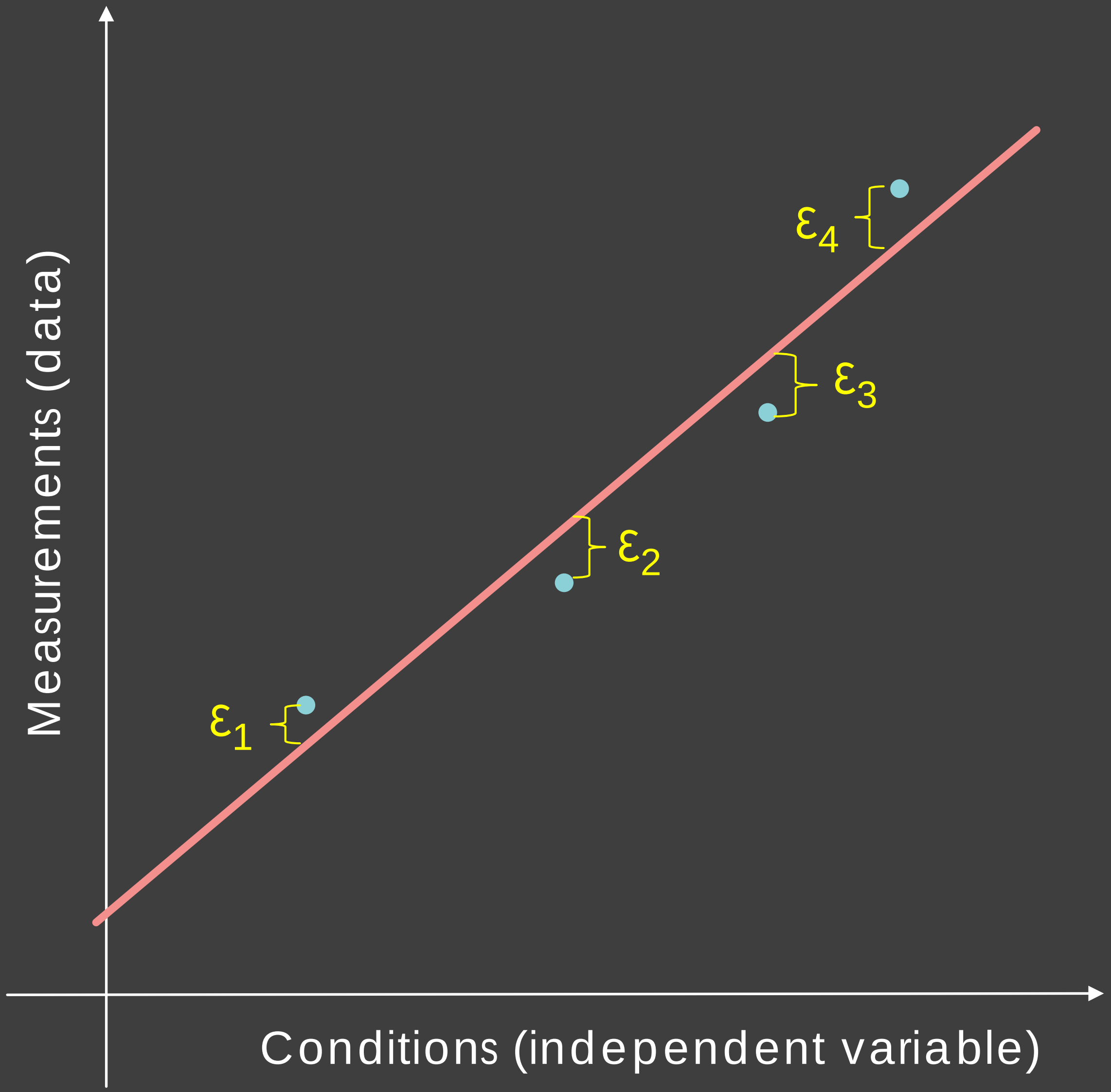
Over-determined problem

- With >2 points, the line is not guaranteed to pass through all points.
- What do we choose?
- What's the **BEST** line … objectively speaking?



The "objective" function

- We introduce an objective (a.k.a. cost, misfit) function that is minimum when our line provides the "best" fit
- i.e. we minimize ε_j's



Choices of objective functions

- **L-1**
  - Sum of absolute values of residuals
  - Tends to ignore outliers
  - "LAR","LAD","LAE","LAV"

$$\varphi(m) = \sum_{j=1}^{N_d} \left| d_j - g_j(m) \right|$$

- **L-2**
  - Sum of squared residuals
  - Tries to fit outliers
  - "Least-squares"

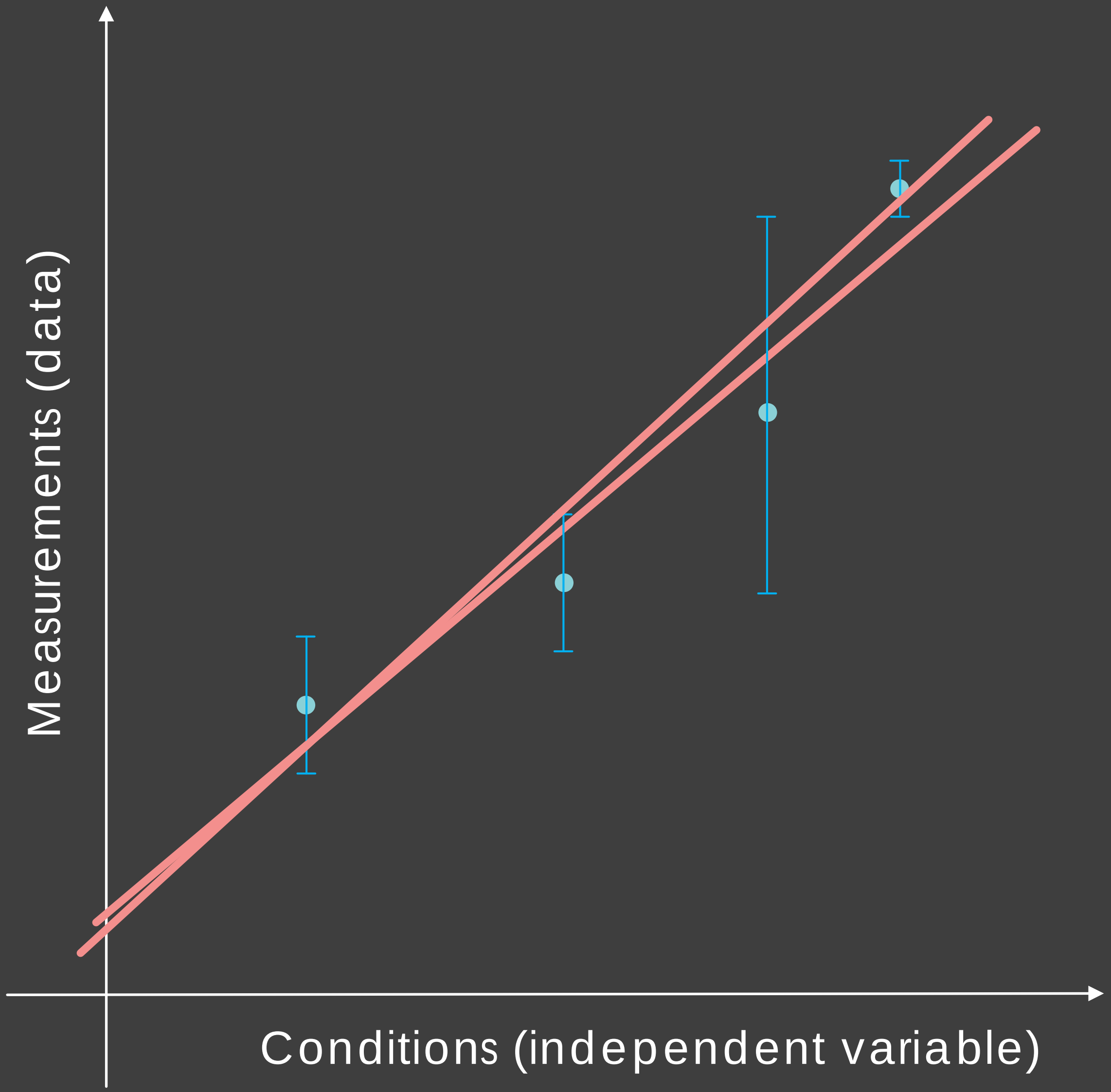
$$\varphi(m) = \sum_{j=1}^{N_d} \left(d_j - g_j(m) \right)^2$$

- **L-∞**
  - Maximum residual
  - Fit is "never terrible" but usually not good, either
  - "Minimax"

$$\varphi(m) = max \left| d_j - g_j(m) \right|$$

What about data uncertainty?

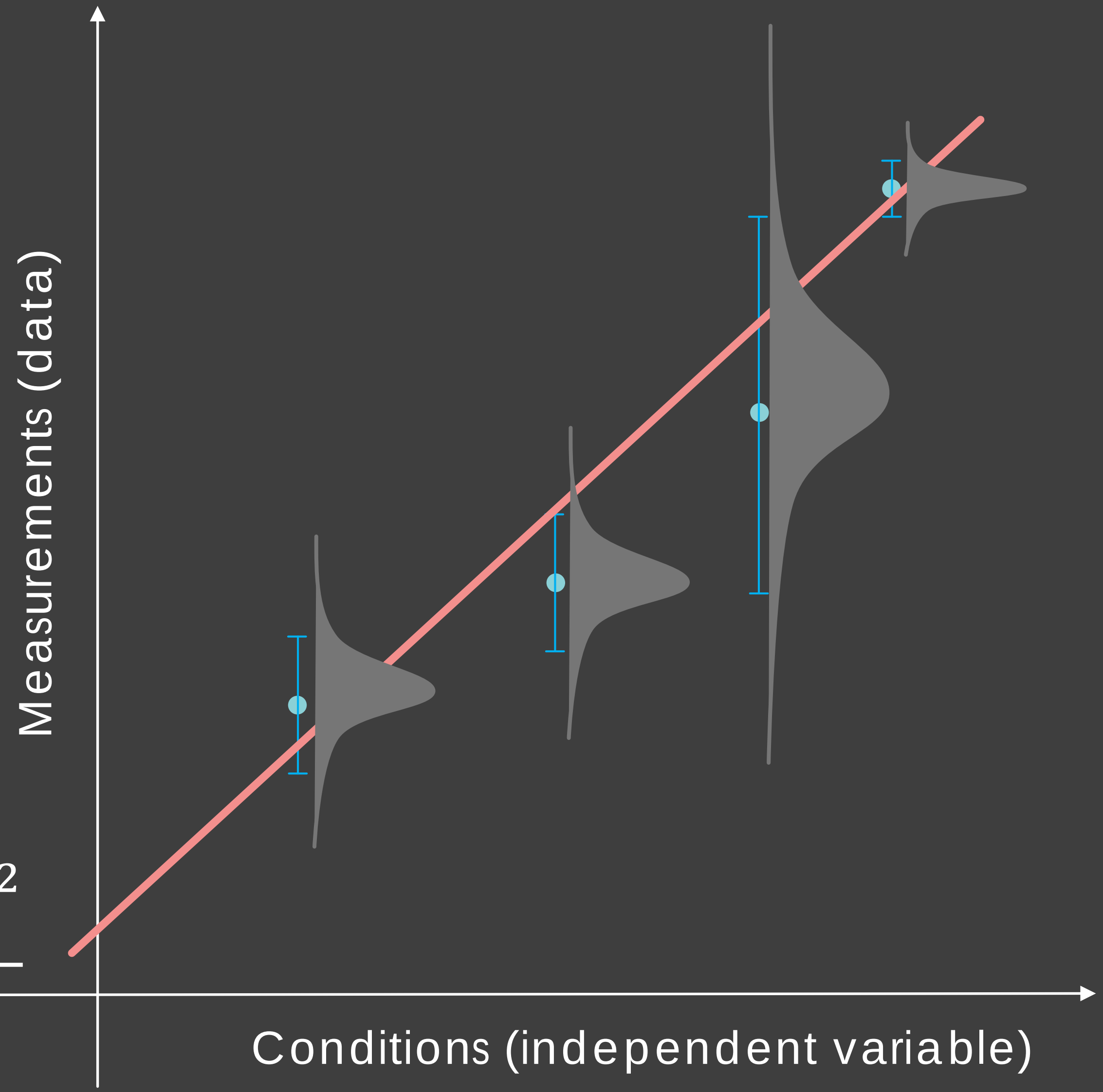
- If error of one measurement is expected to be larger than that of another, then the line need not pass close to it, nor should the objective function count it as much as more accurate measurements



What about data uncertainty?

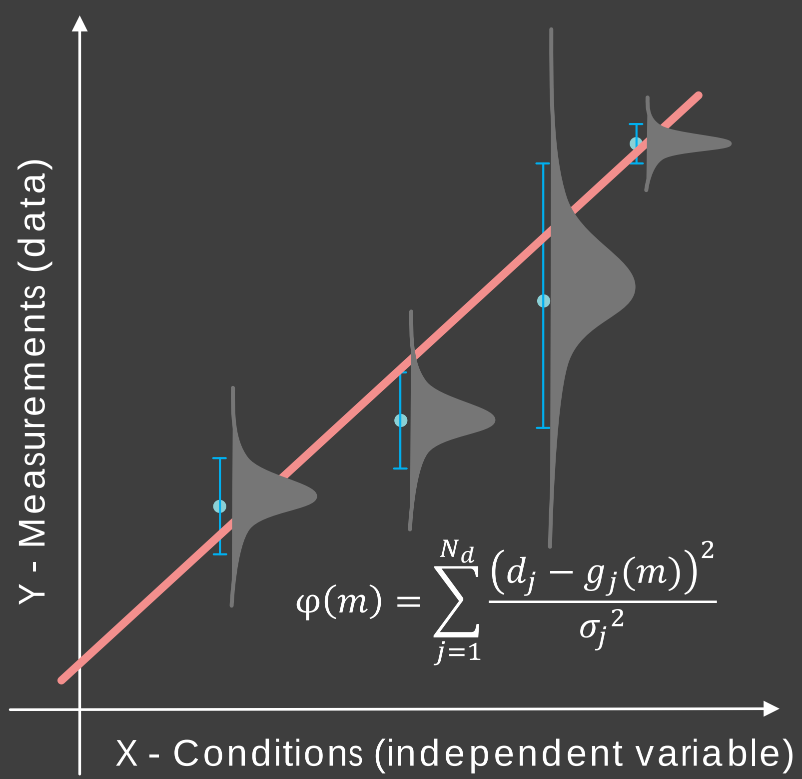
- If data error is normally distributed with variance σ_j^2 then the correct objective function is **L-2**
- Each squared residual gets weighted by variance:

$$\varphi(m) = \sum_{j=1}^{N_d} \frac{\left(d_j - g_j(m)\right)^2}{\sigma_j^2}$$



Fitting a line to some points

- Linear problem:
- $y_j = m_1 x_j + m_2$



- Matrix notation:

$$C_D^{-1/2} d = C_D^{-1/2} G m$$

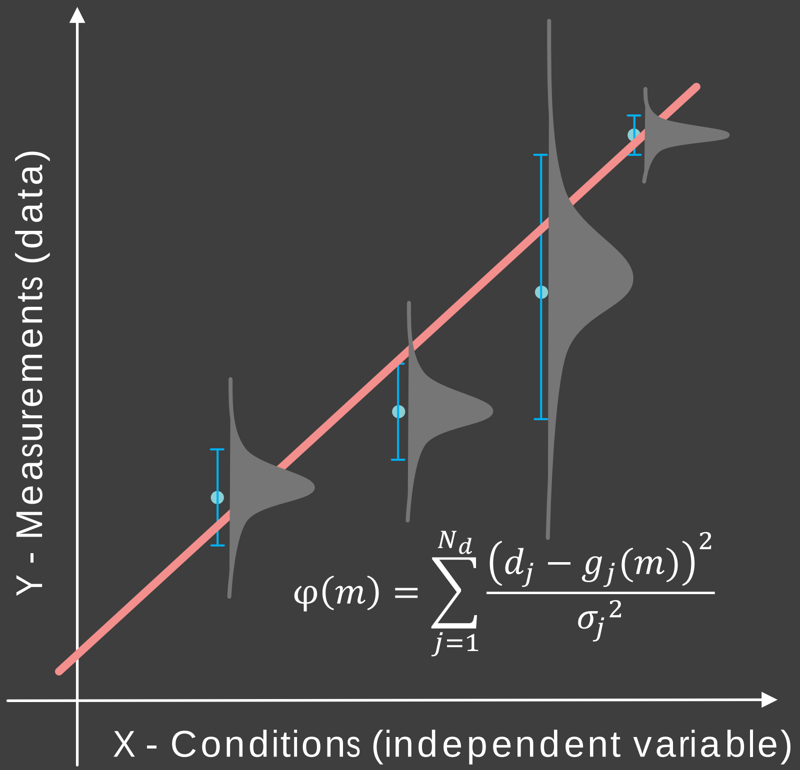
$$\begin{bmatrix} 1/\sigma_1 & 0 & 0 & 0 \\ 0 & 1/\sigma_2 & 0 & 0 \\ 0 & 0 & 1/\sigma_3 & 0 \\ 0 & 0 & 0 & 1/\sigma_4 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} =$$

$$\begin{bmatrix} 1/\sigma_1 & 0 & 0 & 0 \\ 0 & 1/\sigma_2 & 0 & 0 \\ 0 & 0 & 1/\sigma_3 & 0 \\ 0 & 0 & 0 & 1/\sigma_4 \end{bmatrix} \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \\ x_4 & 1 \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \end{bmatrix}$$

$$\varphi = (d - Gm)^T C_D^{-1} (d - Gm)$$

Fitting a line to some points

- Linear problem:
- $y_j = m_1 x_j + m_2$



- Matrix notation:

$$C_D^{-1/2} d = C_D^{-1/2} G m$$

$$\varphi = (d - Gm)^T C_D^{-1} (d - Gm)$$

- Minimize φ with respect to m

$$\frac{\partial \varphi}{\partial m} = 0$$

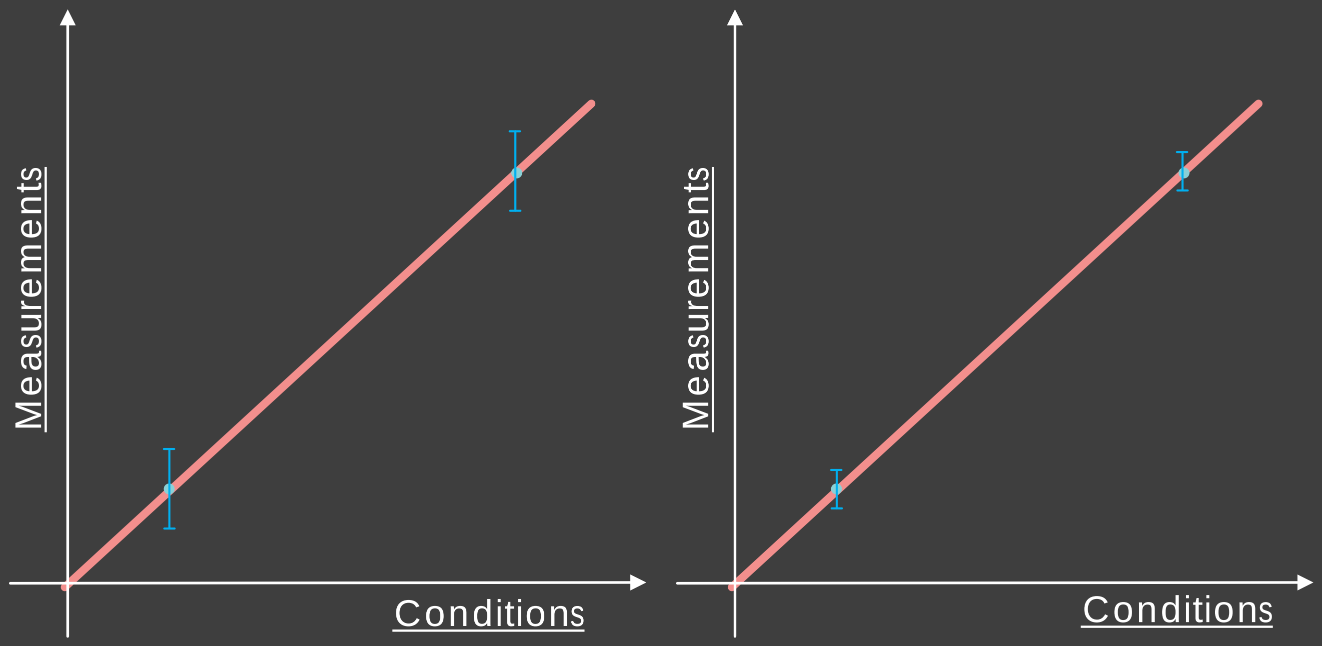
- Condition is satisfied when:

$$m_{est} = (G^T C_D^{-1} G)^{-1}\ G^T C_D^{-1} d$$

- "Least squares" estimate of model parameters: m_{est}

How confident are we in m_{est}?

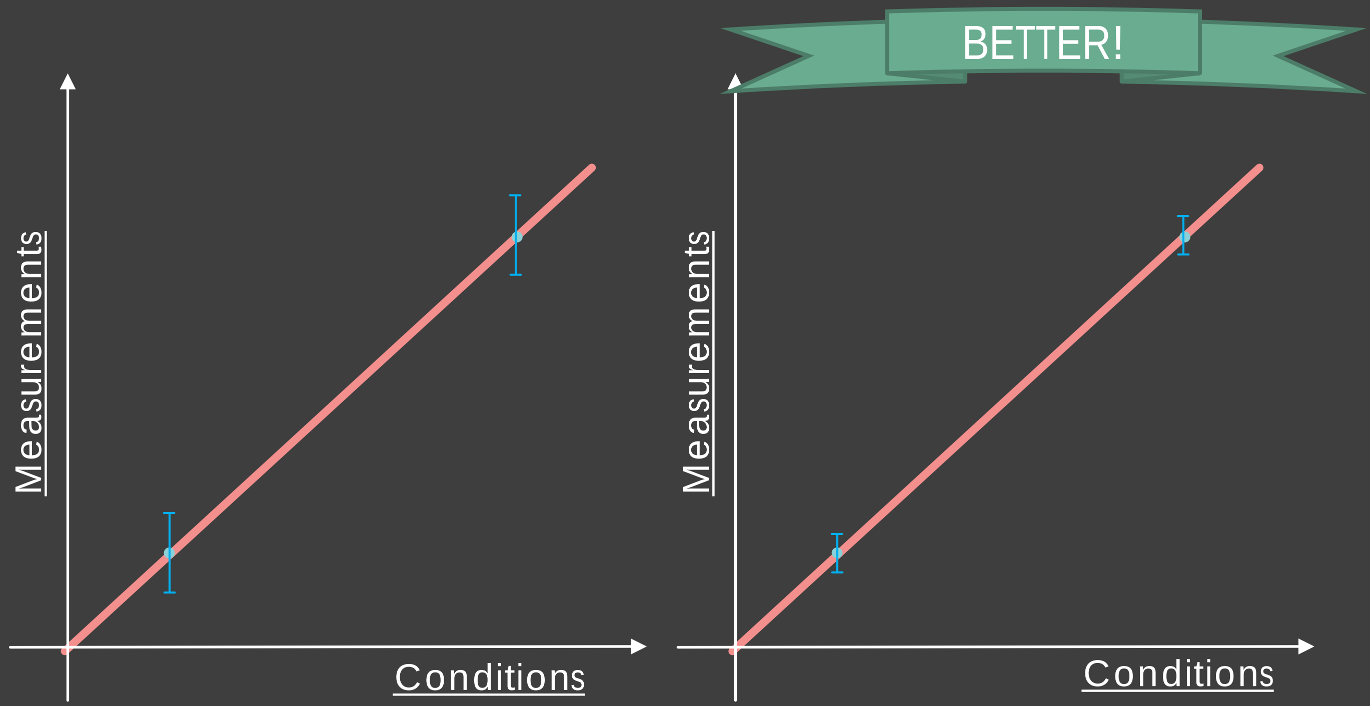
- We can evaluate the uncertainty of model parameters using the posterior model covariance matrix: $\tilde{C}_M$
- But first, let's build some intuition…



Which scenario – left or right – will yield a less uncertain estimate of slope?

How confident are we in m_{est}?

- We can evaluate the uncertainty of model parameters using the posterior model covariance matrix: $\tilde{C}_M$
- But first, let's build some intuition…

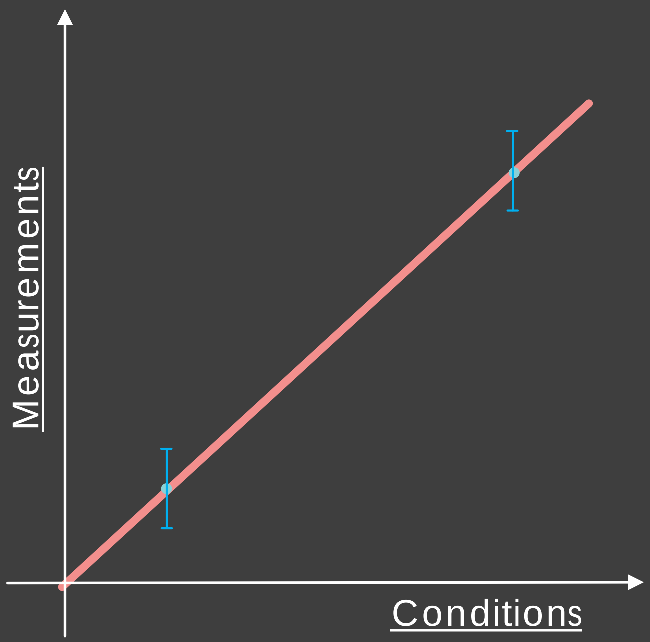


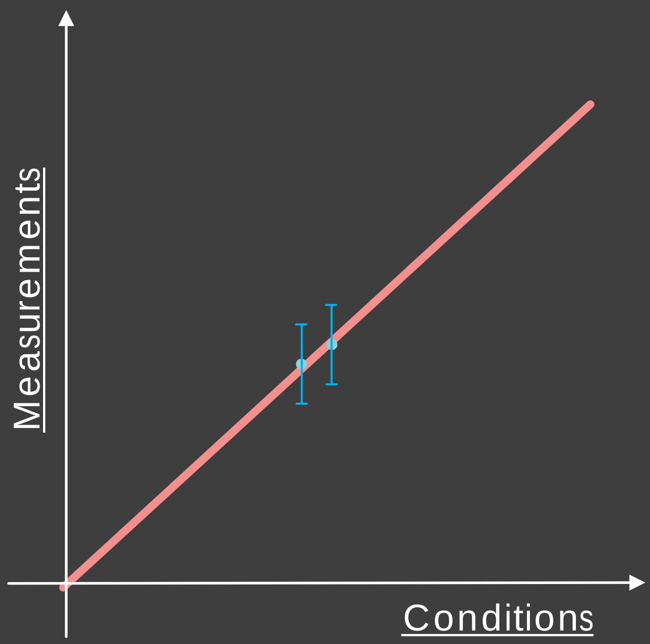
Objective function is more affected by slope variations when data uncertainty is small

→ more accurate data tends to yield more reliable estimates of parameters

How confident are we in m_{est}?

- We can evaluate the uncertainty of model parameters using the posterior model covariance matrix: $\tilde{C}_M$
- But first, let's build some intuition…

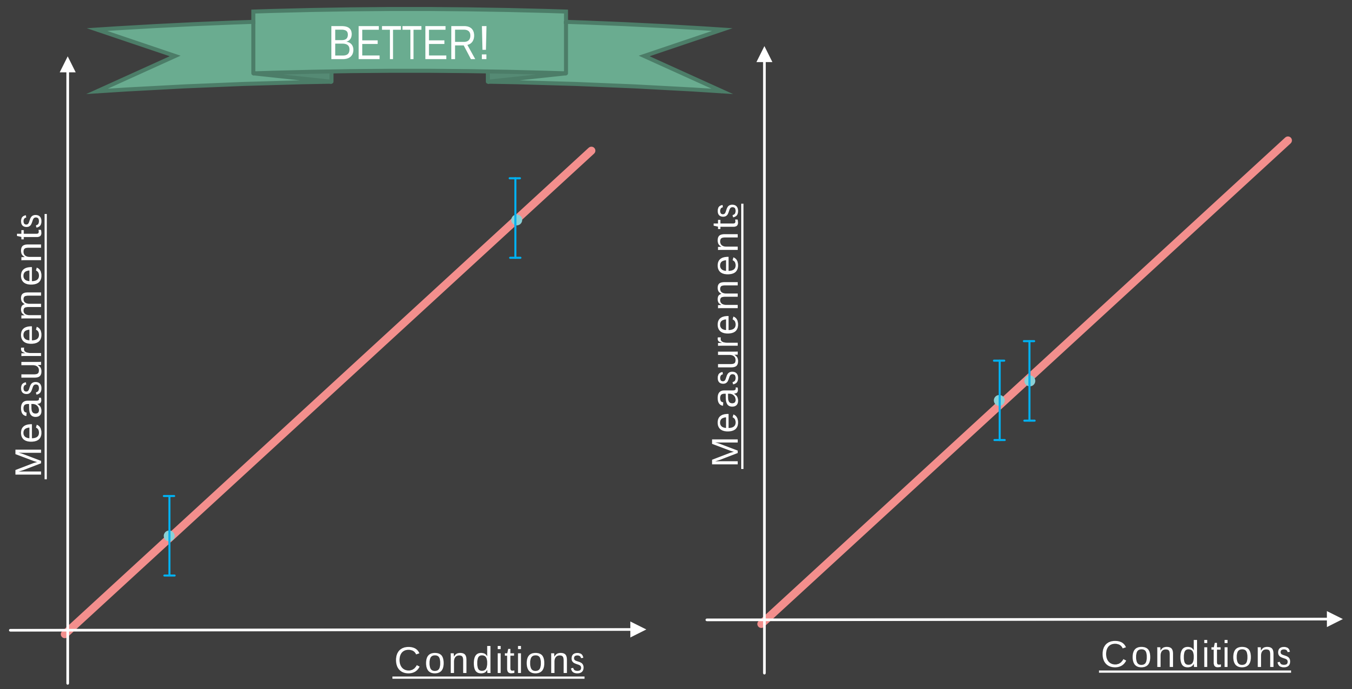




Which scenario – left or right – will yield a less uncertain estimate of slope?

How confident are we in m_{est}?

- We can evaluate the uncertainty of model parameters using the posterior model covariance matrix: $\tilde{C}_M$
- But first, let's build some intuition…



Predictions of widely spaced (in x) points are more sensitive to slope variations

→ estimates of parameters to which you are more sensitive tend to be better!

How confident are we in m_{est}?

- We can evaluate the uncertainty of model parameters using the posterior model covariance matrix: $\tilde{C}_M$

$$m_{est} = (G^T C_D^{-1} G)^{-1} G^T C_D^{-1} d$$

$$\tilde{C}_M = (G^T C_D^{-1} G)^{-1}$$

$$\tilde{C}_M = \begin{bmatrix} \sigma_{slope}^2 & \sigma_{slope}\sigma_{int}\rho \\ \sigma_{slope}\sigma_{int}\rho & \sigma_{int}^2 \end{bmatrix}$$

Variance of slope estimate

Correlation between estimates of slope and y-intercept → how much parameters **trade-off**!

Variance of y-intercept estimate

Limitation of $\tilde{C}_M$

- Posterior model covariance matrix only describes the generalized Gaussian distribution around the optimal model

$$m_{est} = (G^T C_D^{-1} G)^{-1} \, G^T C_D^{-1} d$$

$$\tilde{C}_M = (G^T C_D^{-1} G)^{-1}$$

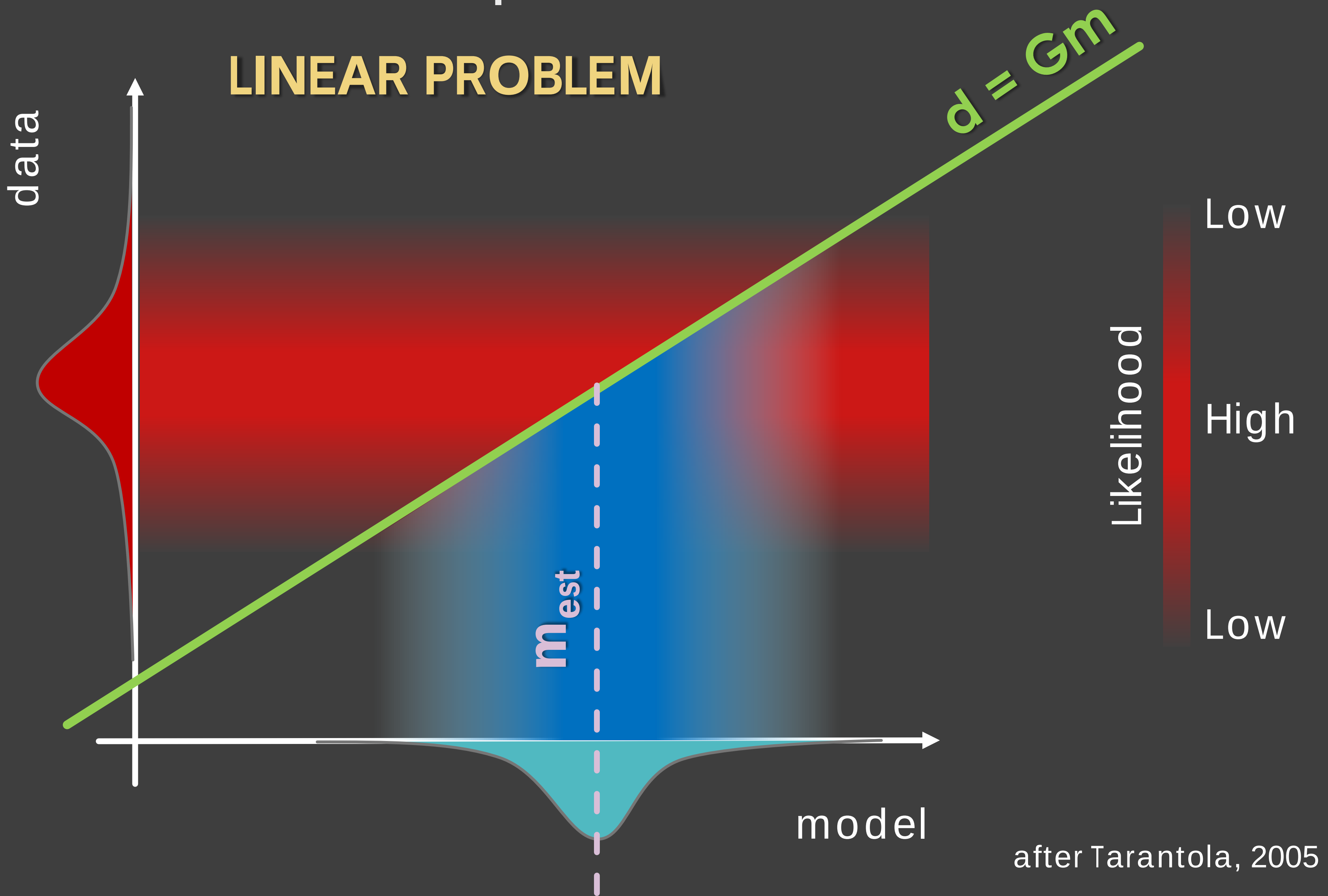
$$\tilde{C}_M = \begin{bmatrix} \sigma_{slope}^2 & \sigma_{slope}\sigma_{int}\rho \\ \sigma_{slope}\sigma_{int}\rho & \sigma_{int}^2 \end{bmatrix}$$

Variance of slope estimate

Correlation between estimates of slope and y-intercept → how much parameters **trade-off**!

Variance of y-intercept estimate

Schematic representation

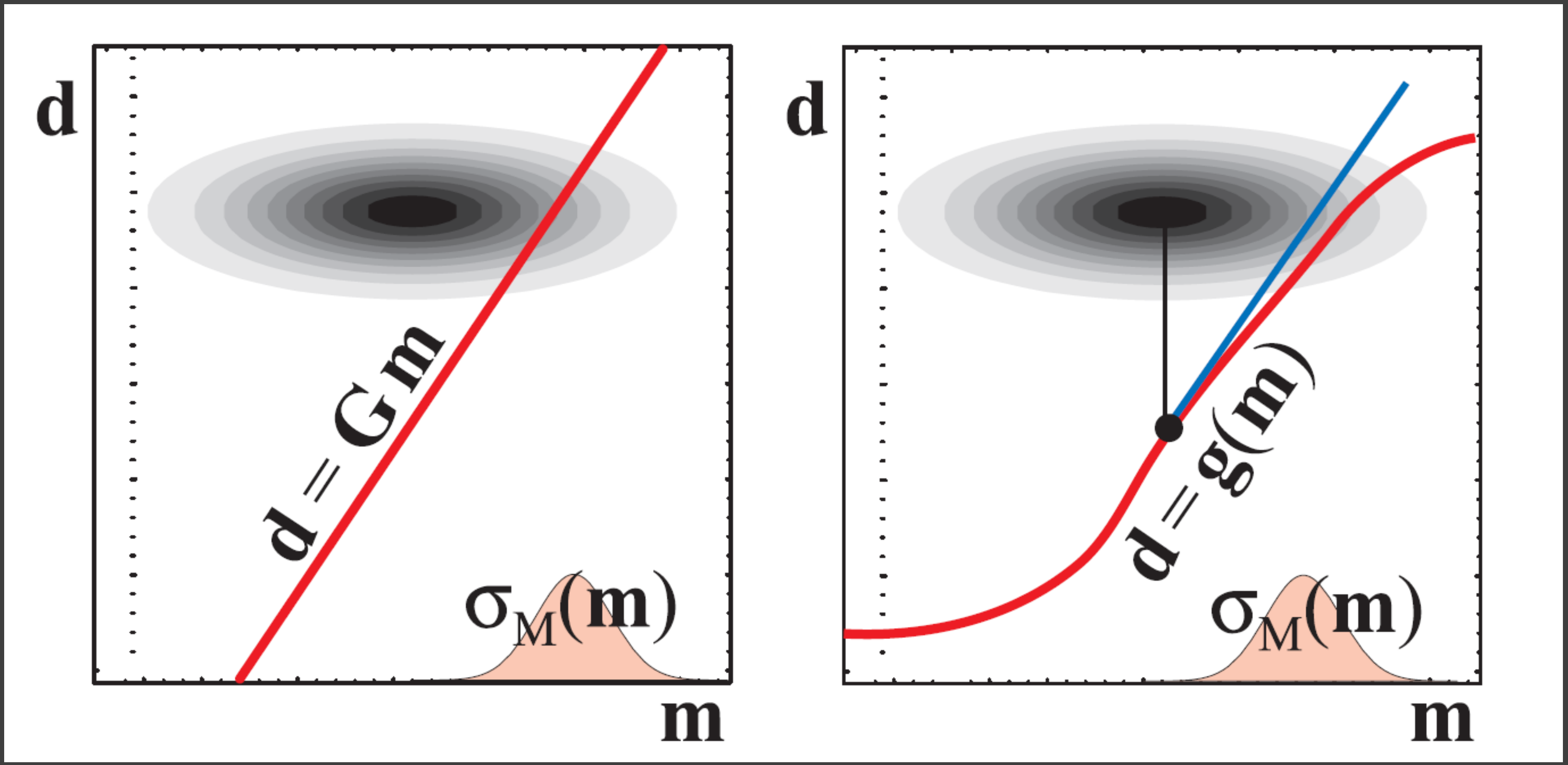


after Tarantola, 2005

Types of inverse problems

Linear – least squares works and $\tilde{C}_M$ is accurate

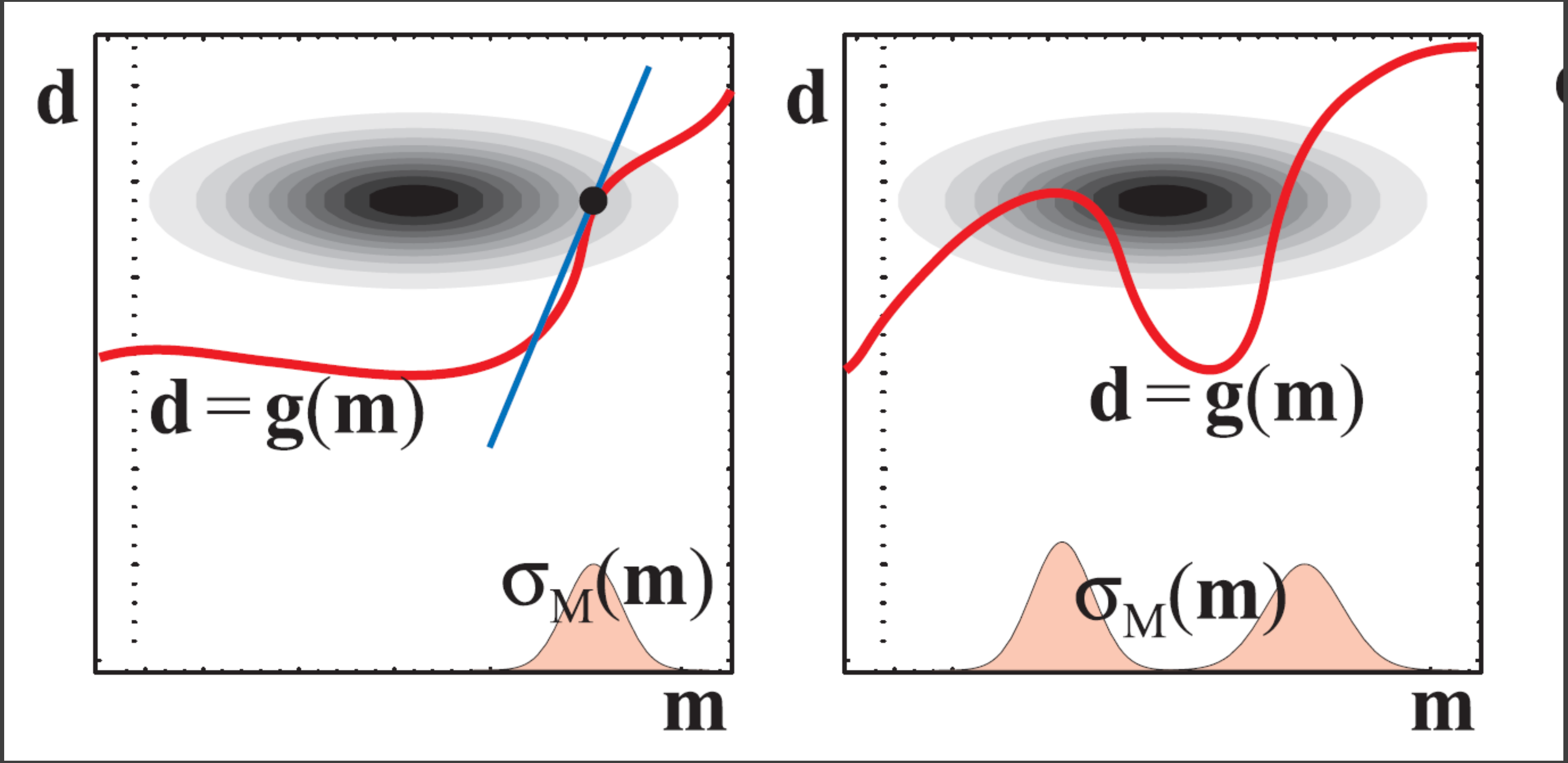
Linearizable around starting model $\tilde{C}_M$ is probably OK



Tarantola, 2005

Linearizable around most likely model
Must iterate!
$\tilde{C}_M$ might be OK

Non-linear → multi-modal posterior on **m,**
$\tilde{C}_M$ is woefully invalid!
Don't use least-squares! **SAMPLING!**☺

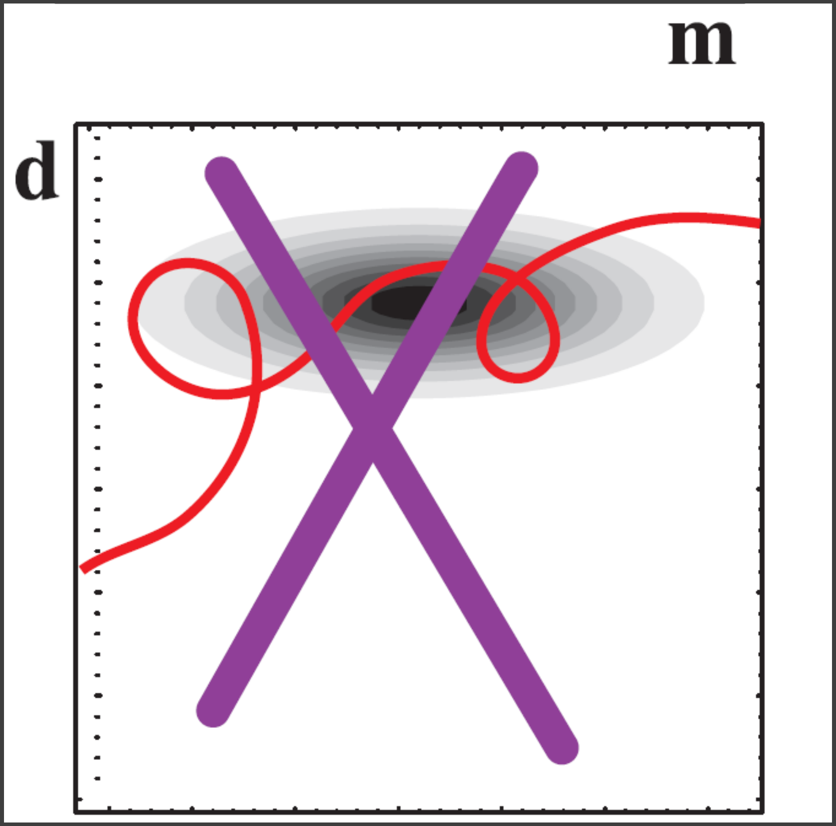


Tarantola, 2005

Trumpian / Brexistential

nothing will work! *Good luck*!

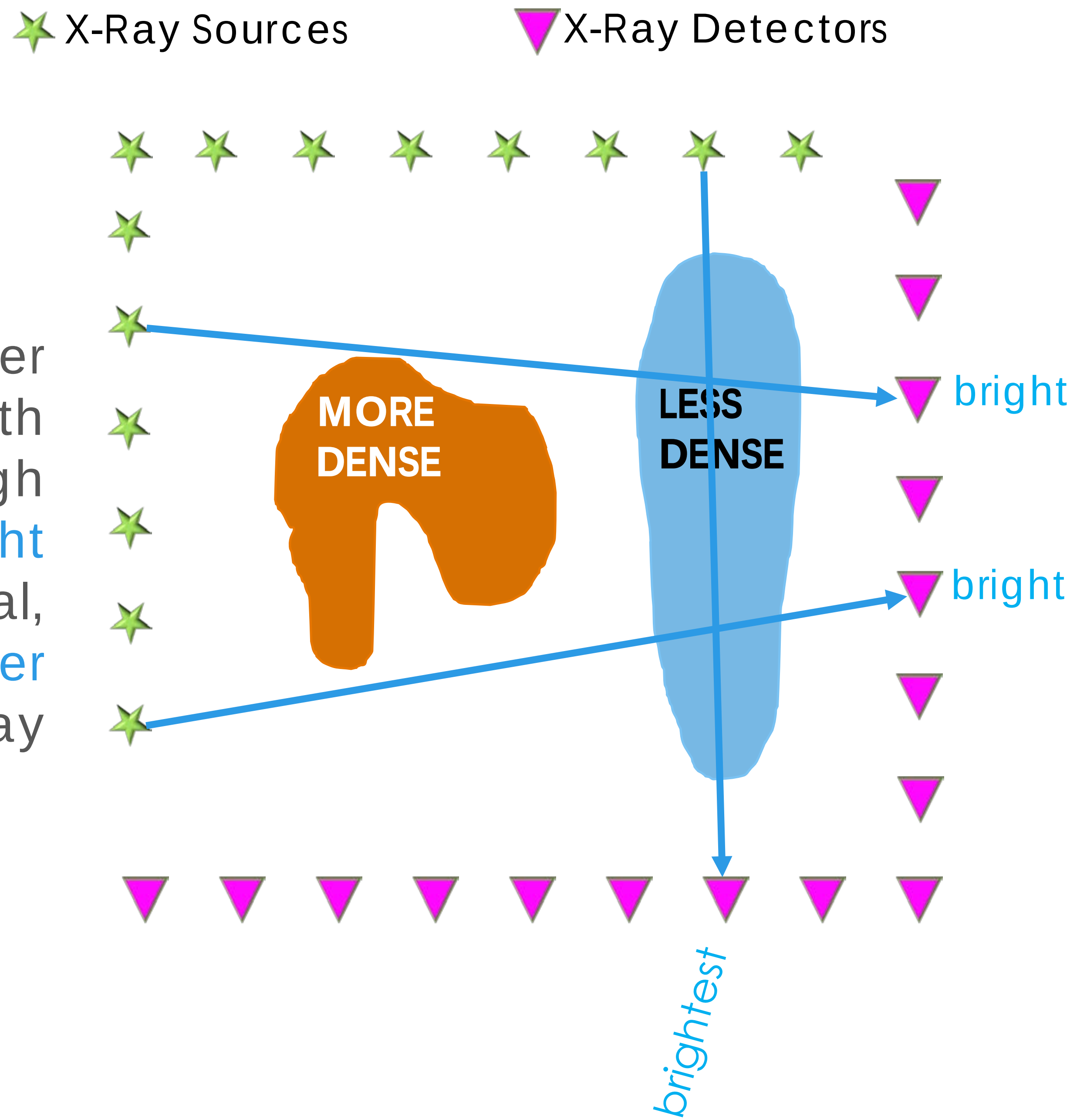
#resist



Tarantola, 2005

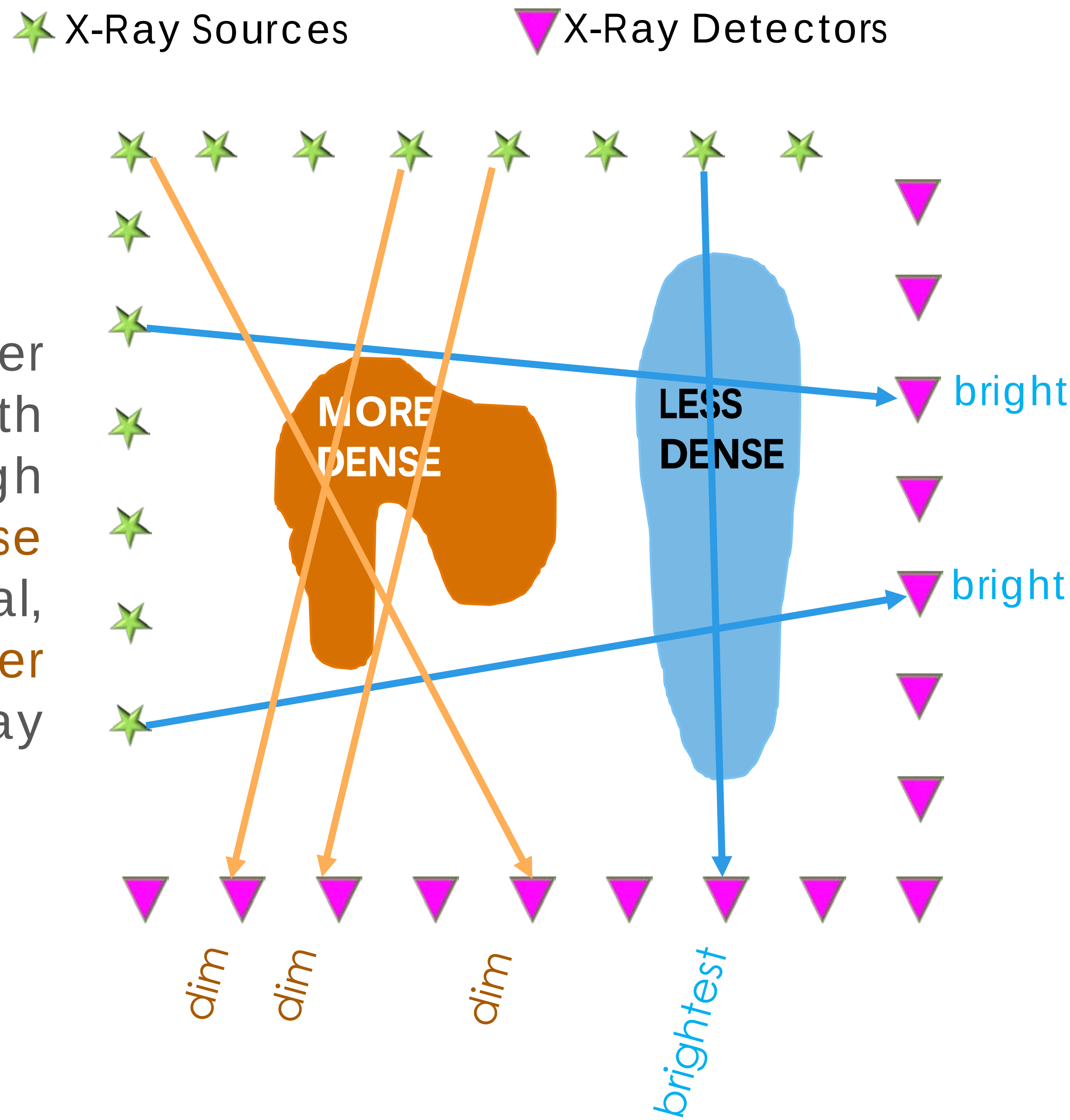
How does it work?

The longer the path through light material, the brighter the x-ray



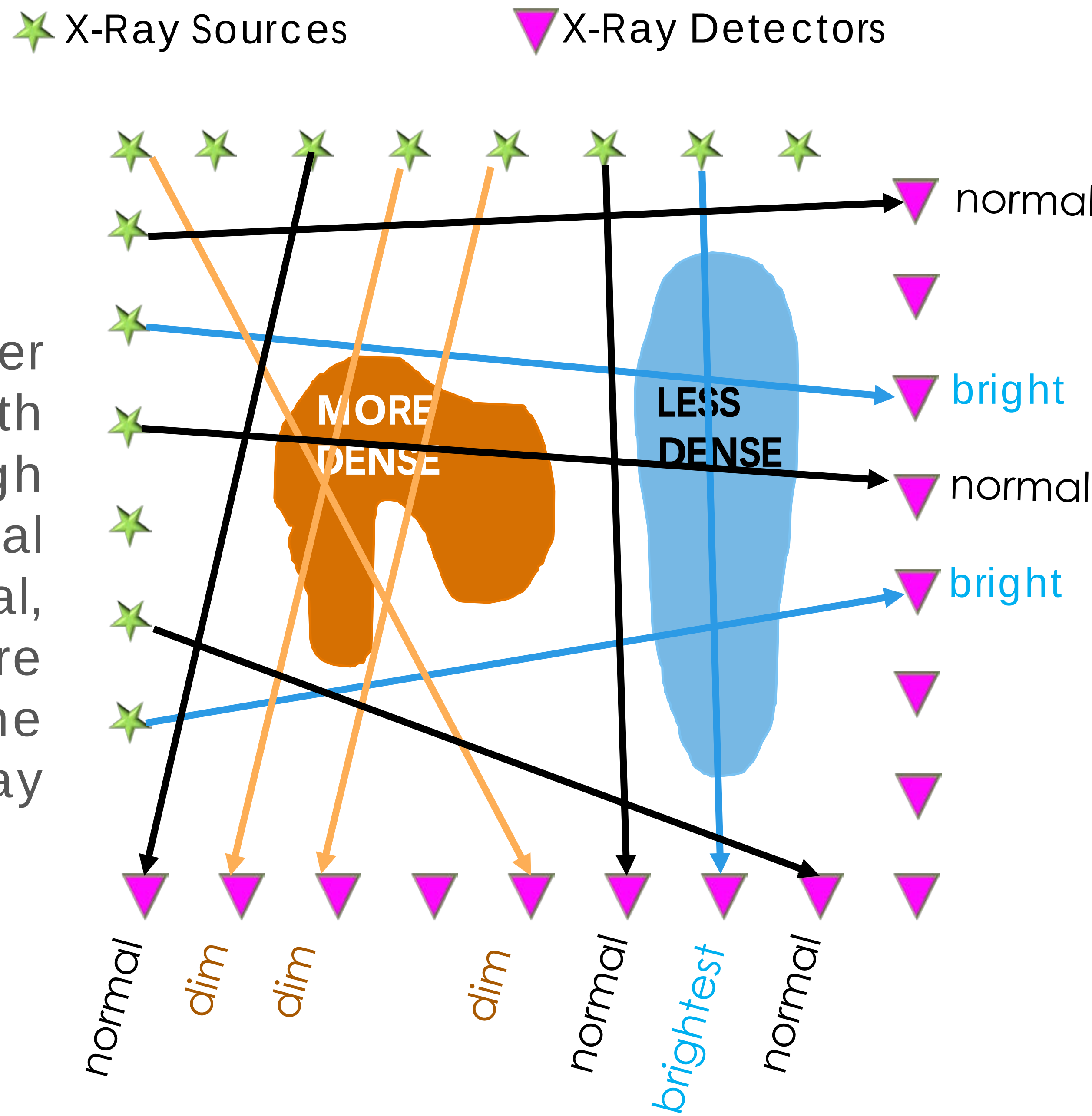
How does it work?

The longer the path through dense material, the dimmer the x-ray



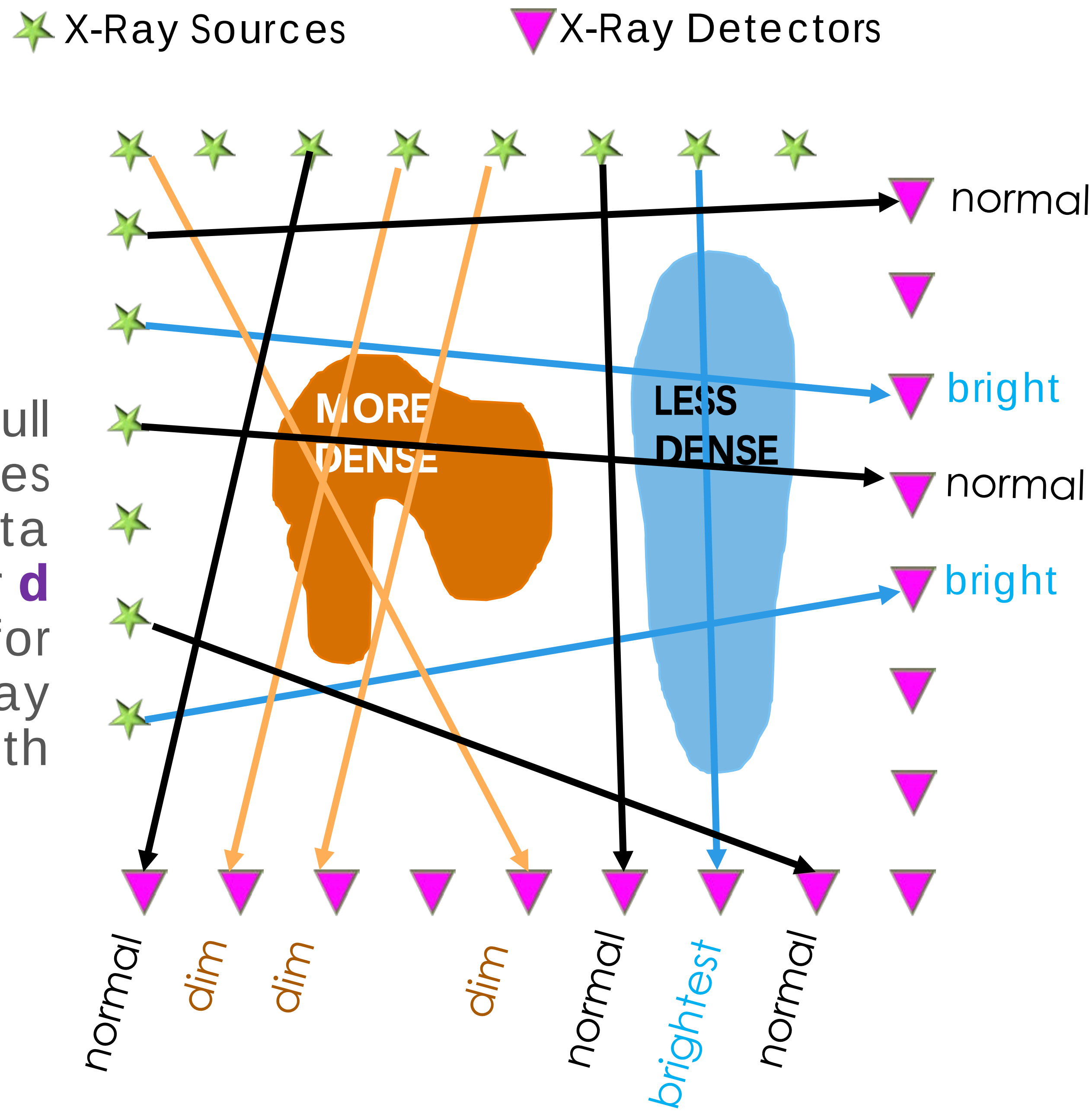
How does it work?

The longer the path through normal material, the more normal the x-ray



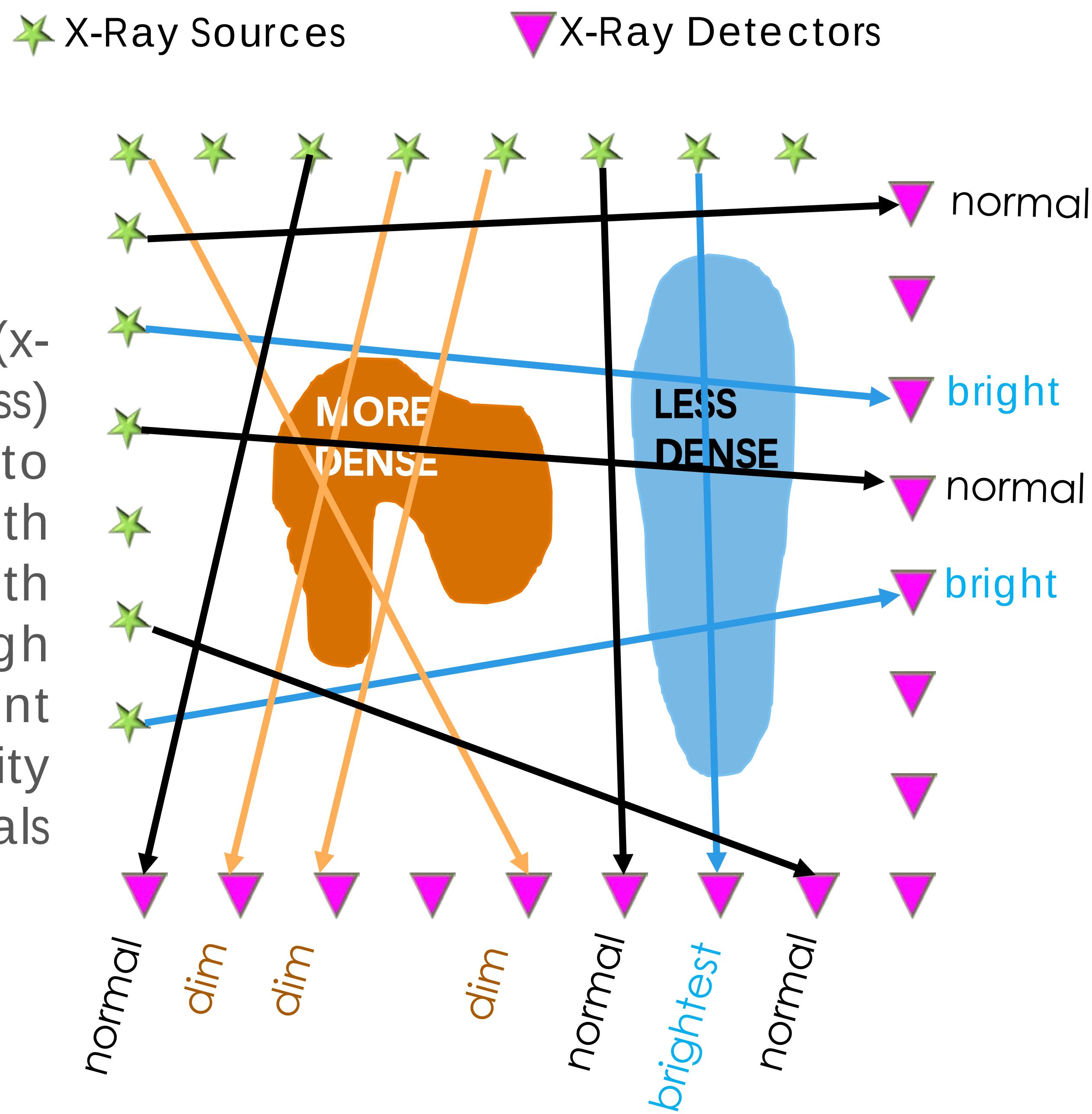
How does it work?

Pull brightnesses into data vector **d** one for each x-ray path



How does it work?

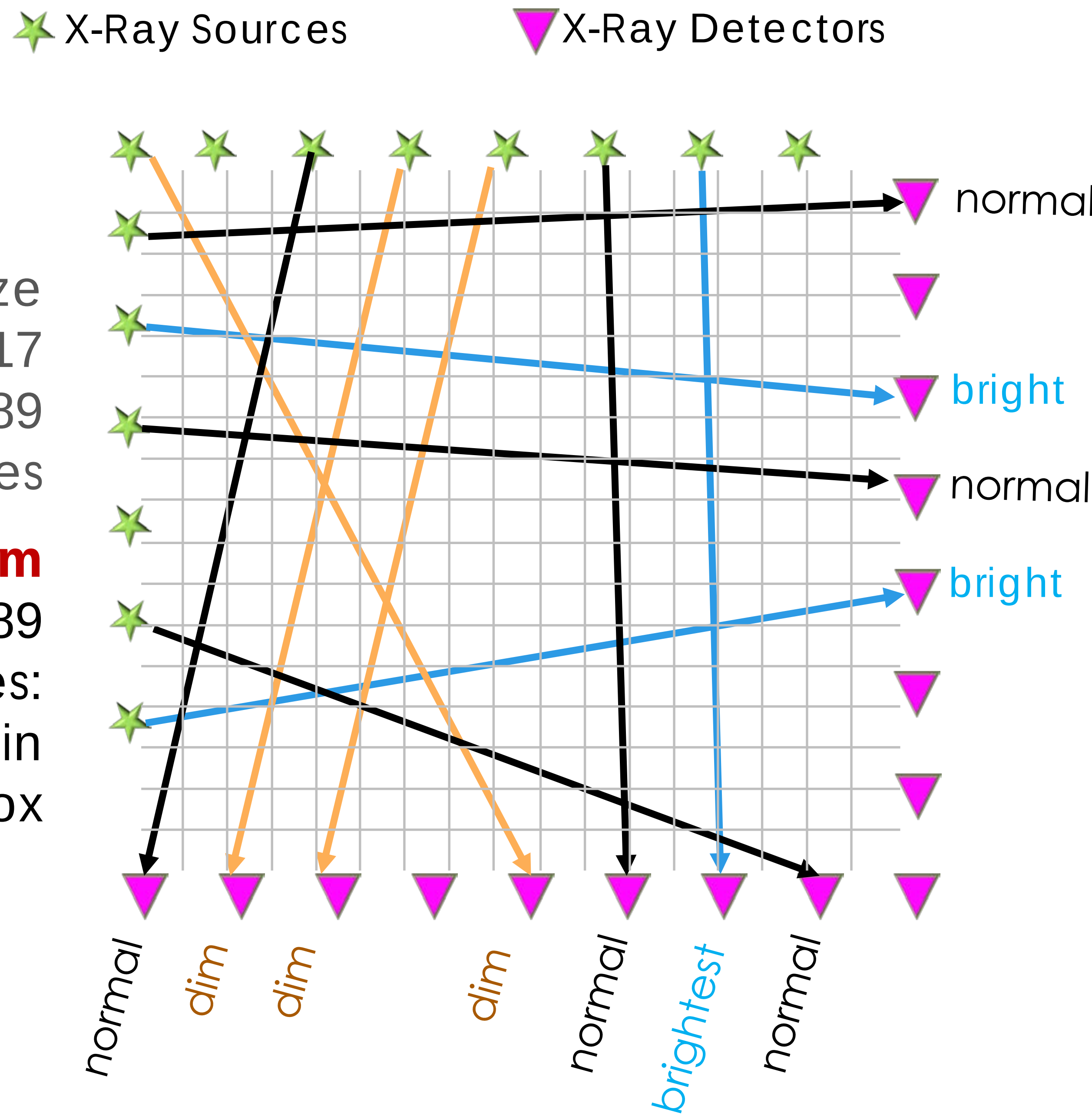
The data (x-ray brightness) is related to the path length through different density materials



How does it work?

Parameterize into 17x17 grid = 289 boxes

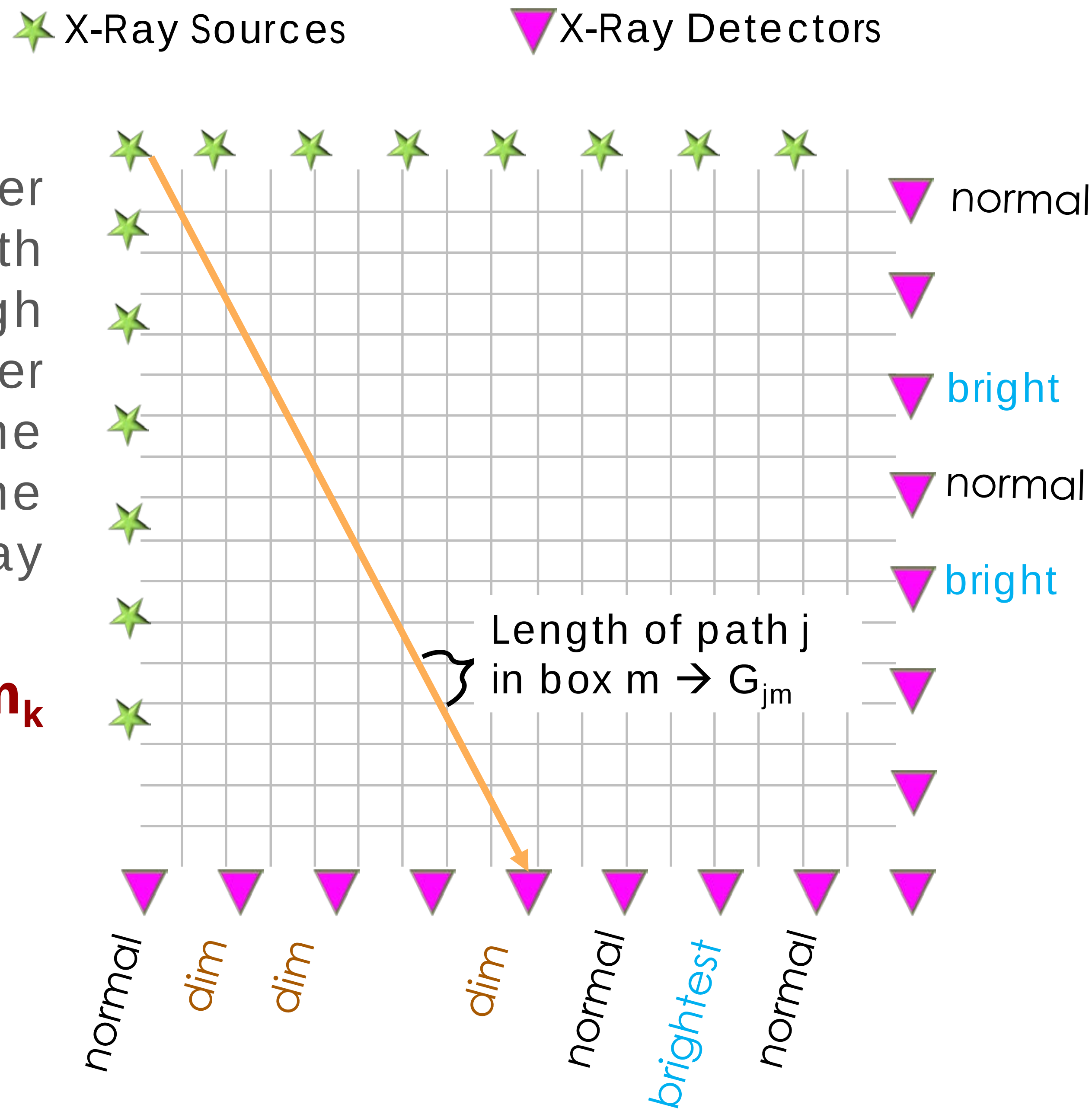
Model **m** has 289 entries: density in each box



How does it work?

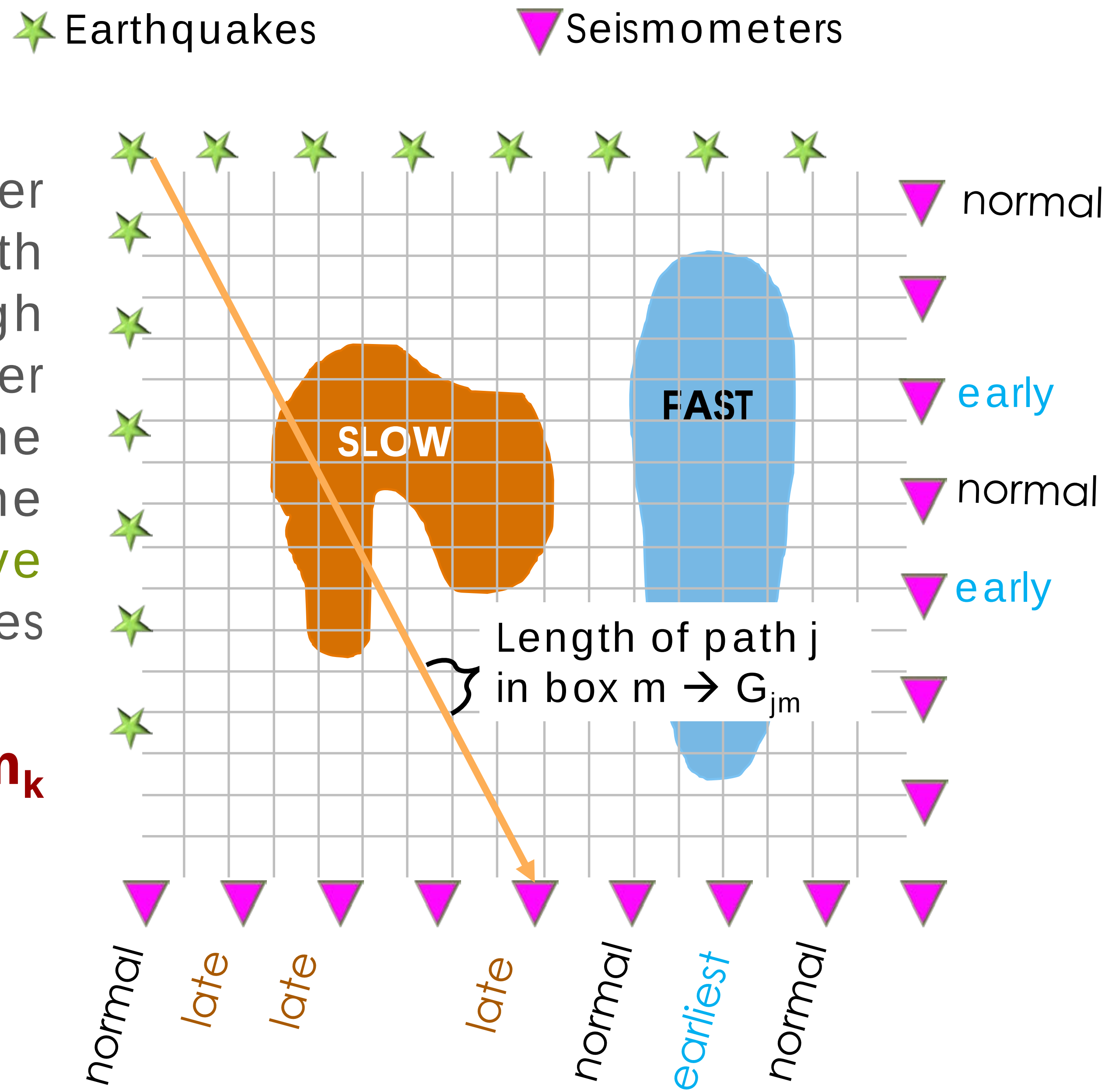
The longer the path through denser boxes, the dimmer the x-ray

$d_j = G_{jk} m_k$



How does it work?

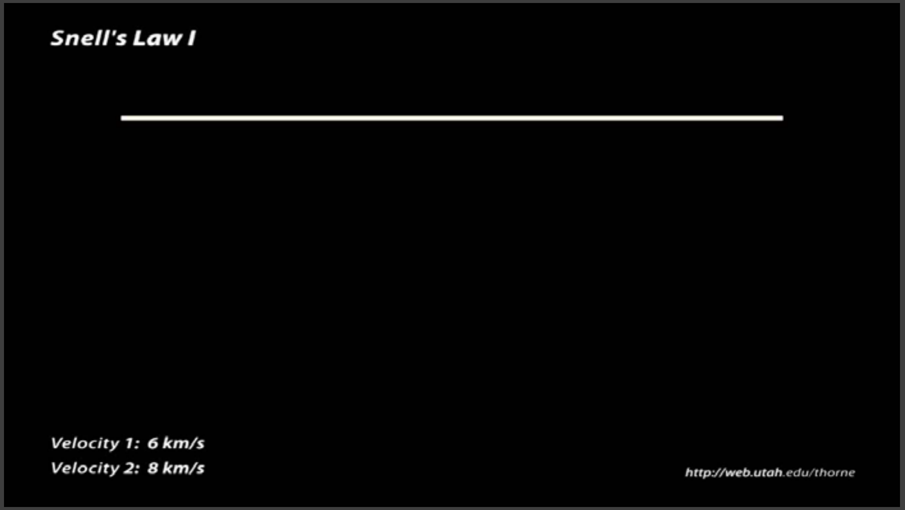
The longer the path through slower boxes, the later the seismic wave arrives



$$d_j = G_{jk} m_k$$

Linearize from starting model

- For wavefield to remain continuous along interface between regions with different velocities, waves will change direction in addition to their speed.



- $d_j = G_{jk} m_k$ where Gjk is the path distance spent by wave j within box k (or total sensitivity of wave j in box k)
  → **But path depends on slowness!!!**
- So we have to linearize: either from starting model, or if needed, iteratively until we find optimal model
  - Write $d_j = T_j - T_j^0$ and $m_k = 1/V_k - 1/V_k^0$
  - G_{jk} must be computed in model $m_k^0 = 1/V_k^0$ or even at each iteration

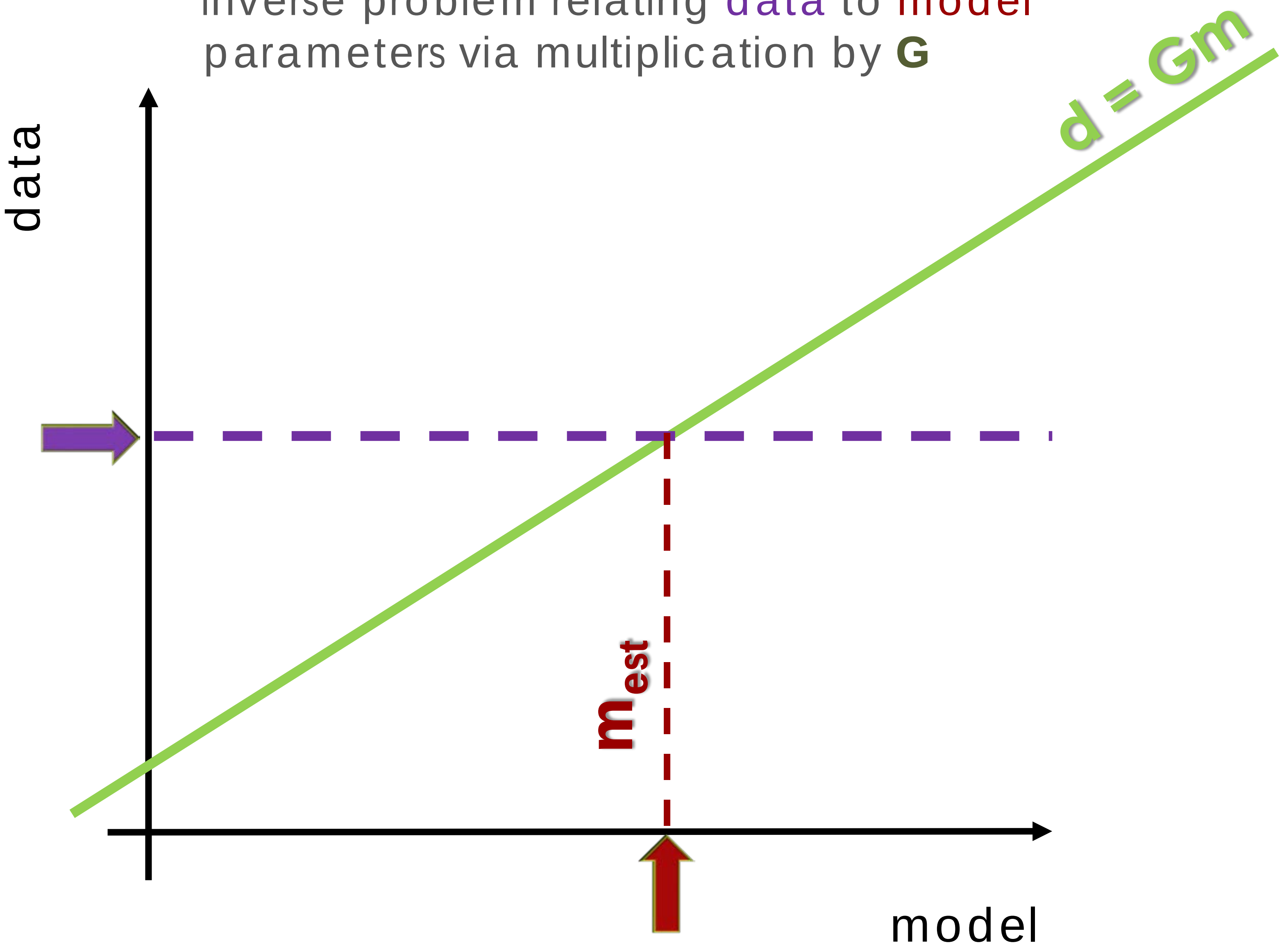
Basic Tomographic Problem

$$\mathbf{d} = \mathbf{g}(\mathbf{m})$$
$$\mathbf{d} = \mathbf{G}\mathbf{m}$$

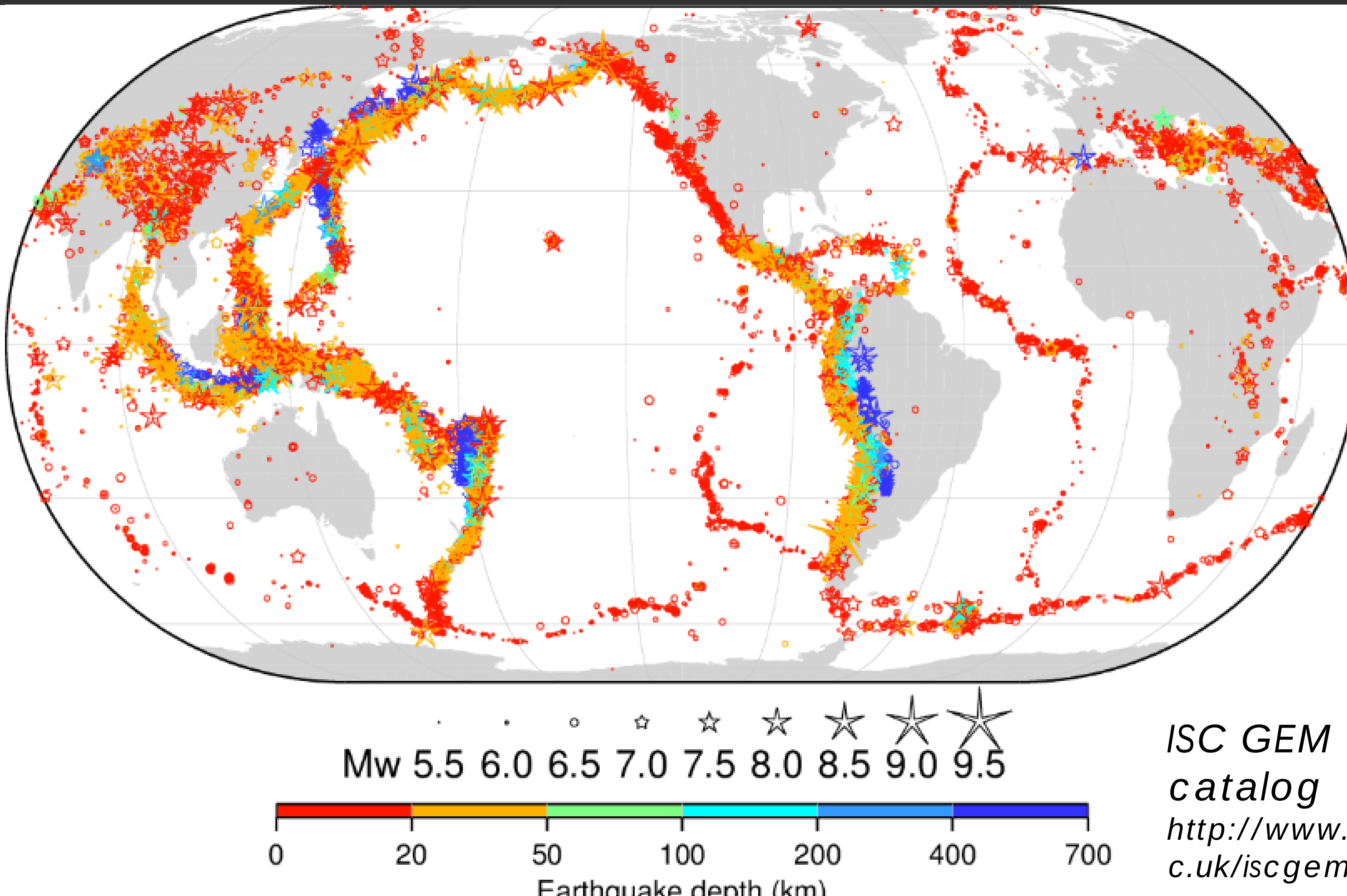
- **d** = seismic data (e.g. travel-times, waveforms, dispersion measurements)
- **m** = model describing spatial variations of seismic velocities or density (e.g. in blocks, splines, spherical harmonics)
- **g** = function describing how data d depend on model parameters $\mathbf{m}_k$, for linear problems $\mathbf{g}(\mathbf{m}) \rightarrow \mathbf{G}\mathbf{m}$

Linear Inverse Problem

- We can schematically represent a linear inverse problem relating data to model parameters via multiplication by **G**



Distribution of Earthquakes

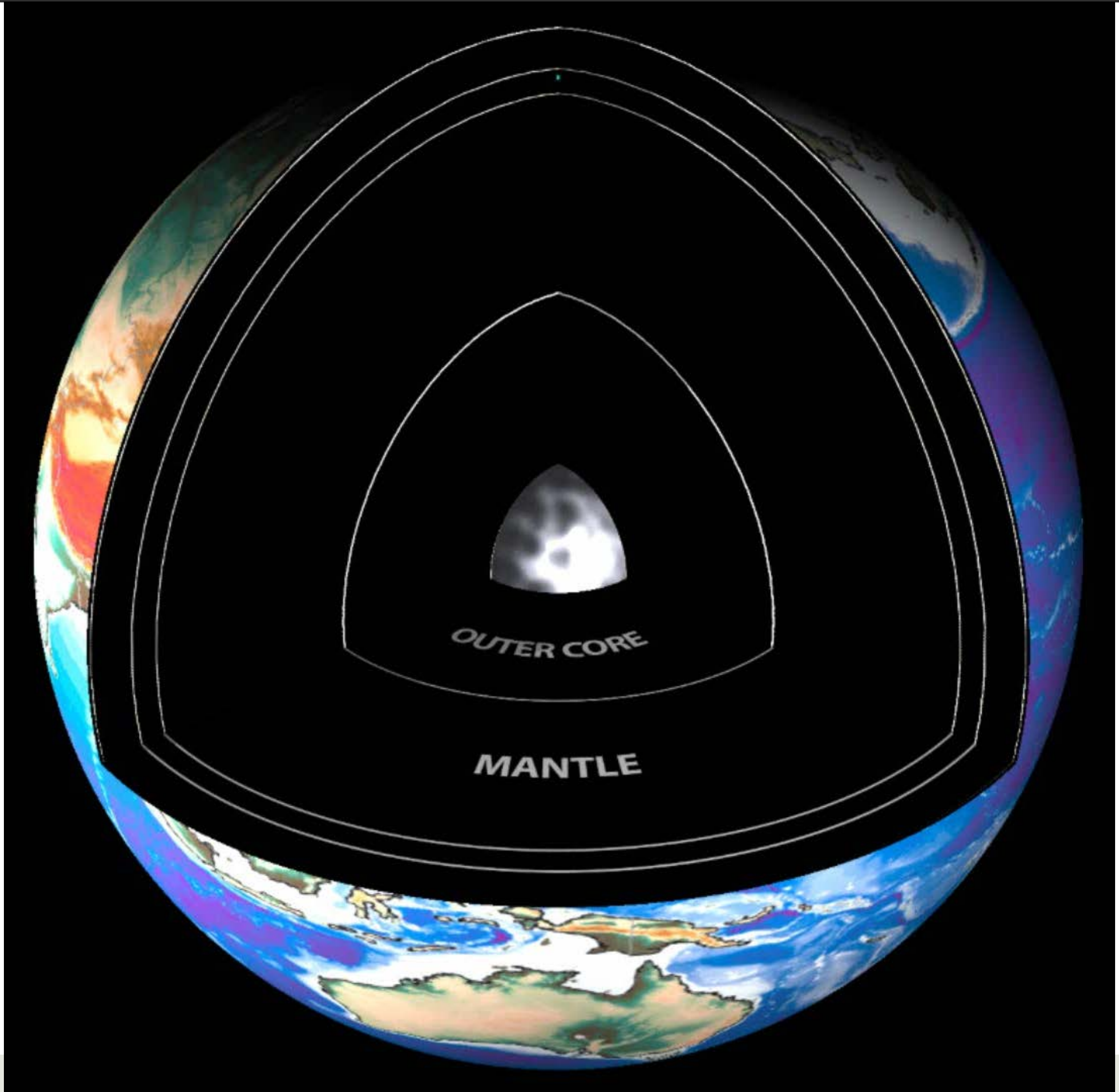


ISC GEM catalog
http://www.isc.ac.uk/iscgem/

The Seismic Wavefield

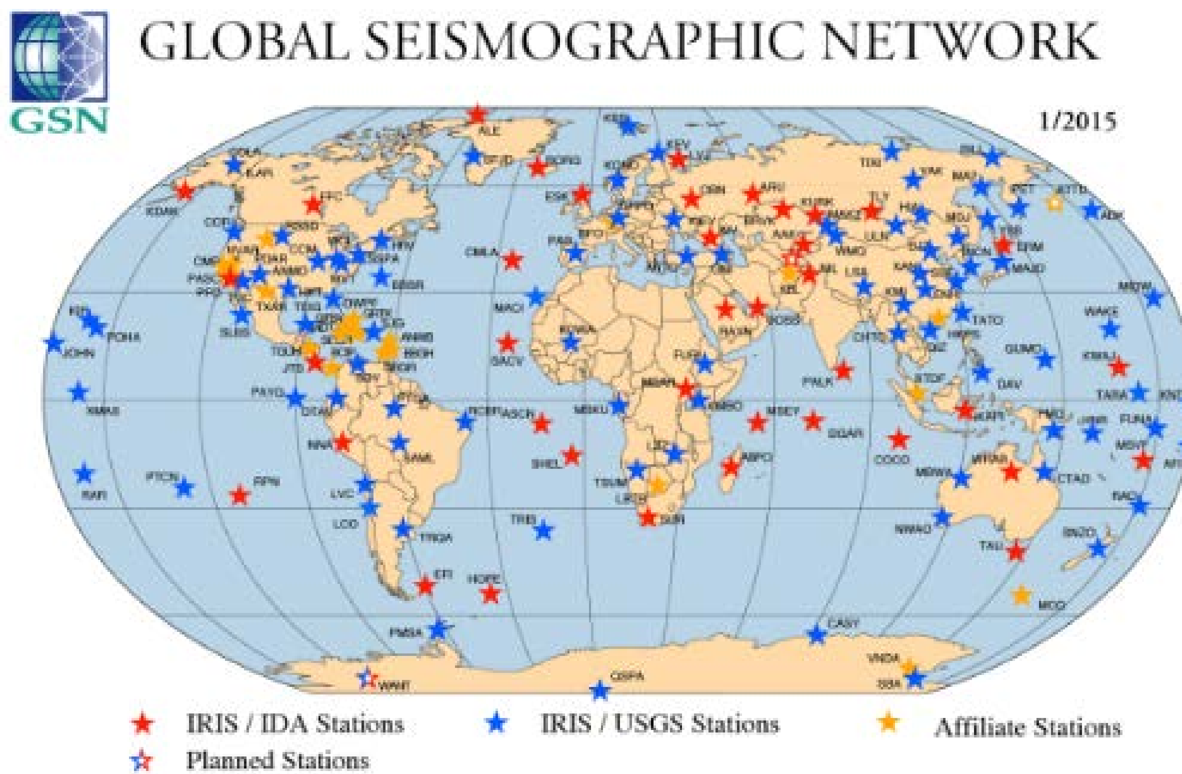
Earthquakes, explosions, ocean waves, etc. excite seismic waves that travel throughout the interior and are recorded by seismometers.

These waves carry with them information about the Earth's inaccessible interior.



The Global Seismic Network (GSN)

- High dynamic range (140 dB)
- Broadband (0.1mHz to ~10 Hz)
- Quiet instruments / sites / installations
- Real-time telemetry
- Global distribution at ~ 2000 km spacing
- Well-characterized instrument response



Basic Tomographic Problem

We can improve images by getting more or better data

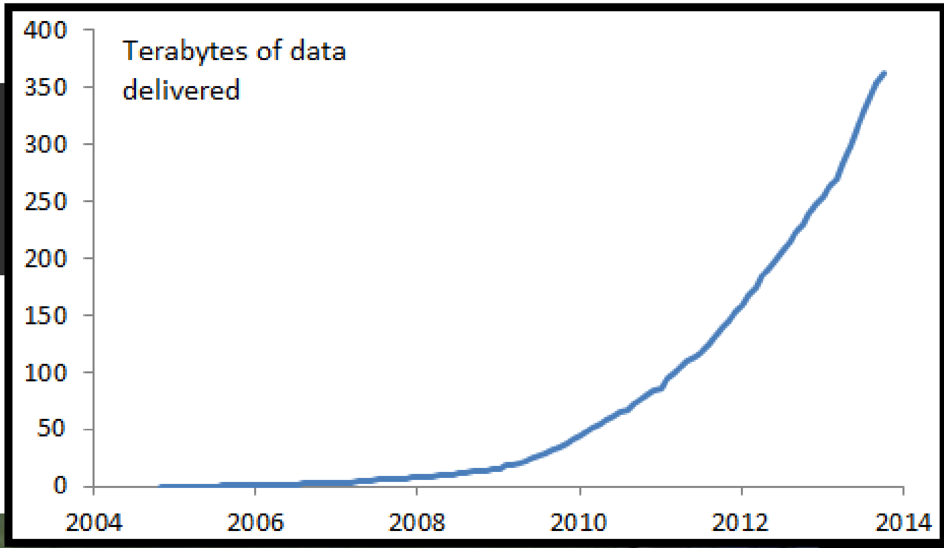
$$\mathbf{d} = \mathbf{g}(\mathbf{m})$$

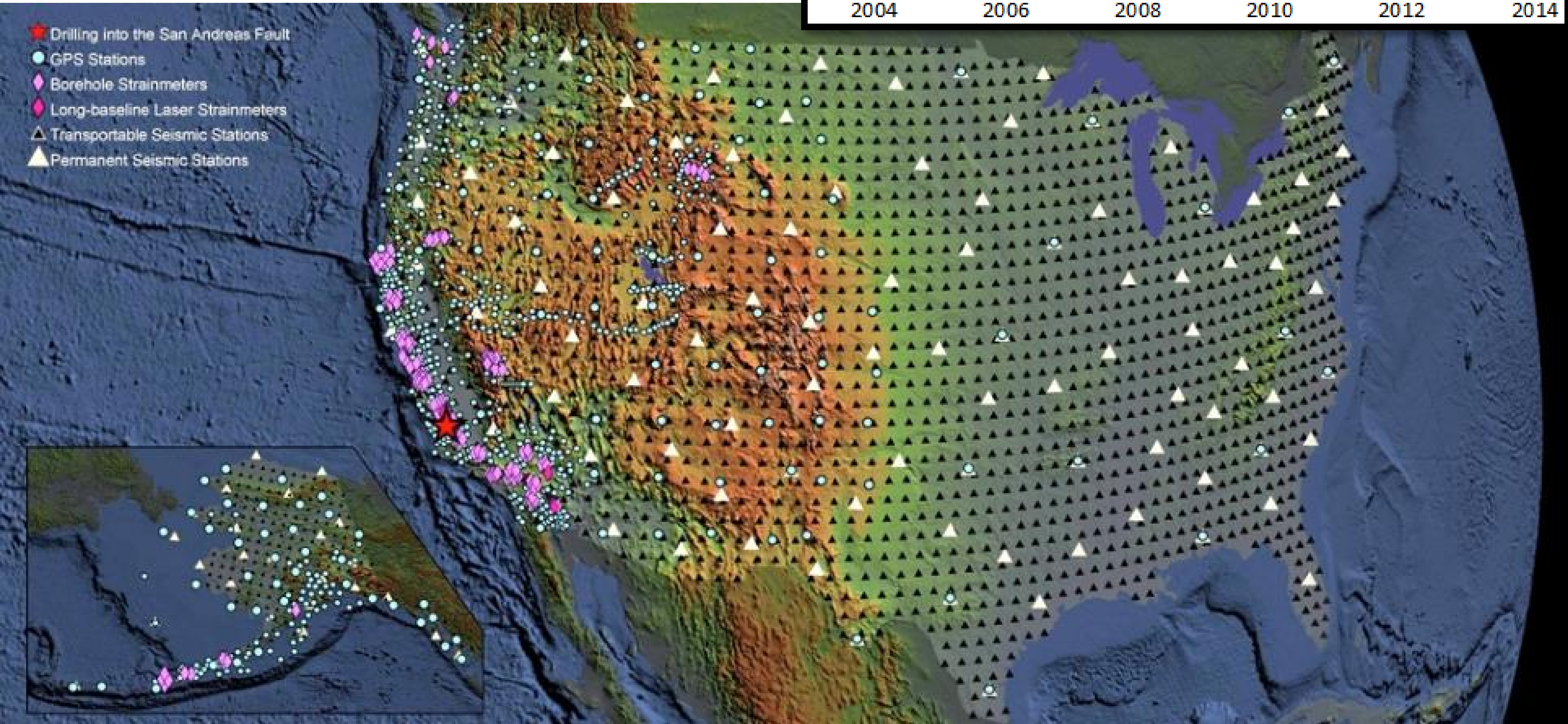
$$\mathbf{d} = \mathbf{G}\mathbf{m}$$

- **d** = seismic data (e.g. travel-times, waveforms, dispersion measurements)
- **m** = model describing spatial variations of seismic velocities or density (e.g. in blocks, splines, spherical harmonics)
- **g** = function describing how data d depend on model parameters $\mathbf{m}_k$, for linear problems $\mathbf{g}(\mathbf{m}) \rightarrow \mathbf{G}\mathbf{m}$

EarthScope

- "Apollo" project of geophysics

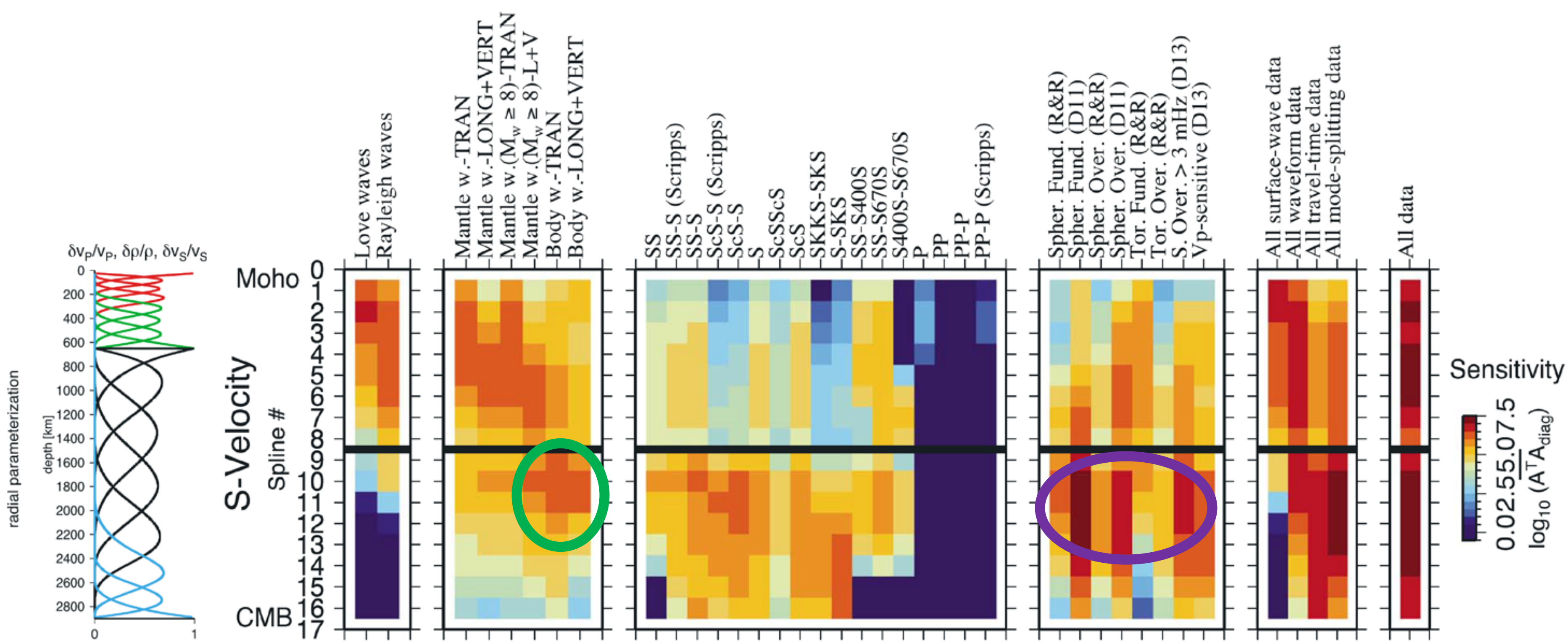




Improving P-wave Coverage

Sensitivity

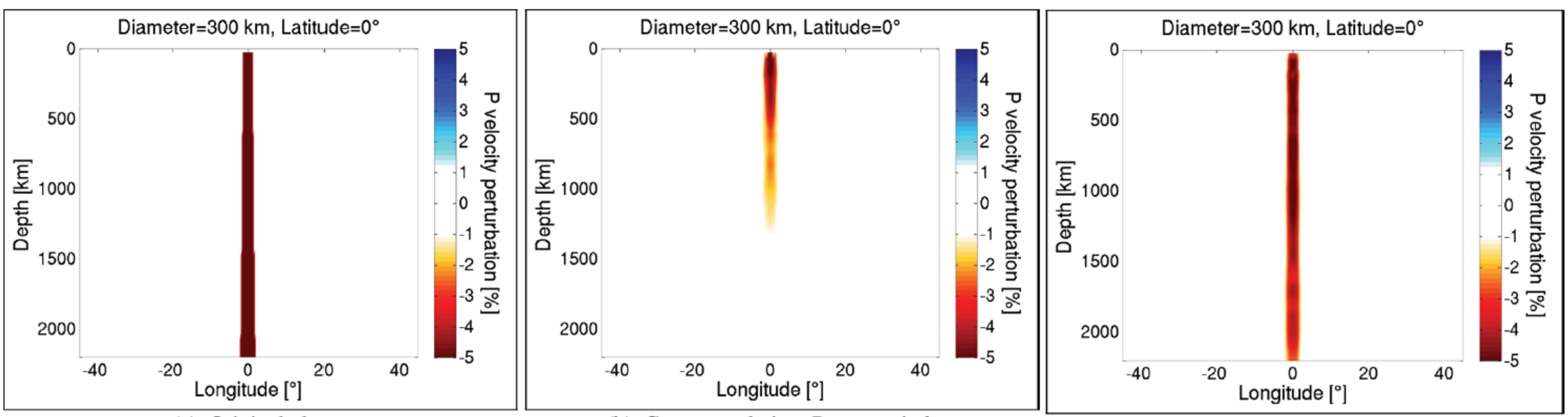
- In mid-mantle (670-1800 km), we have excellent constraints from waveforms (green), but especially recently-measured modes (purple)!



Are All Data Created Equal?

- Travel-time data is ill-suited for imaging structures that are small, far away, or have weak velocity anomalies (e.g. plume conduits)
- Using all wiggles of a seismogram (full waveform tomography) can yield better models

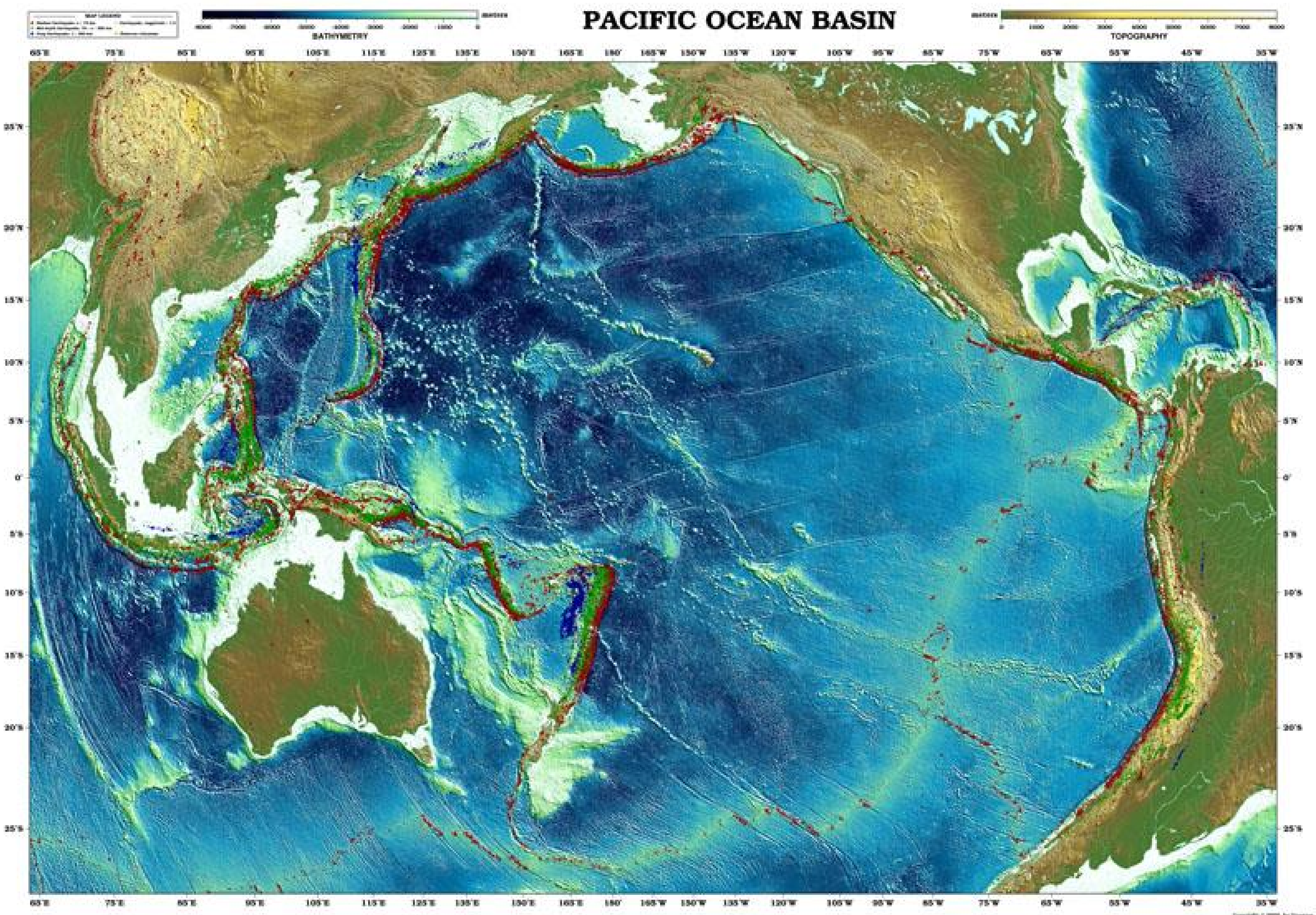
Input structure (plume) | Imaged using traveltimes | Imaged using waveforms



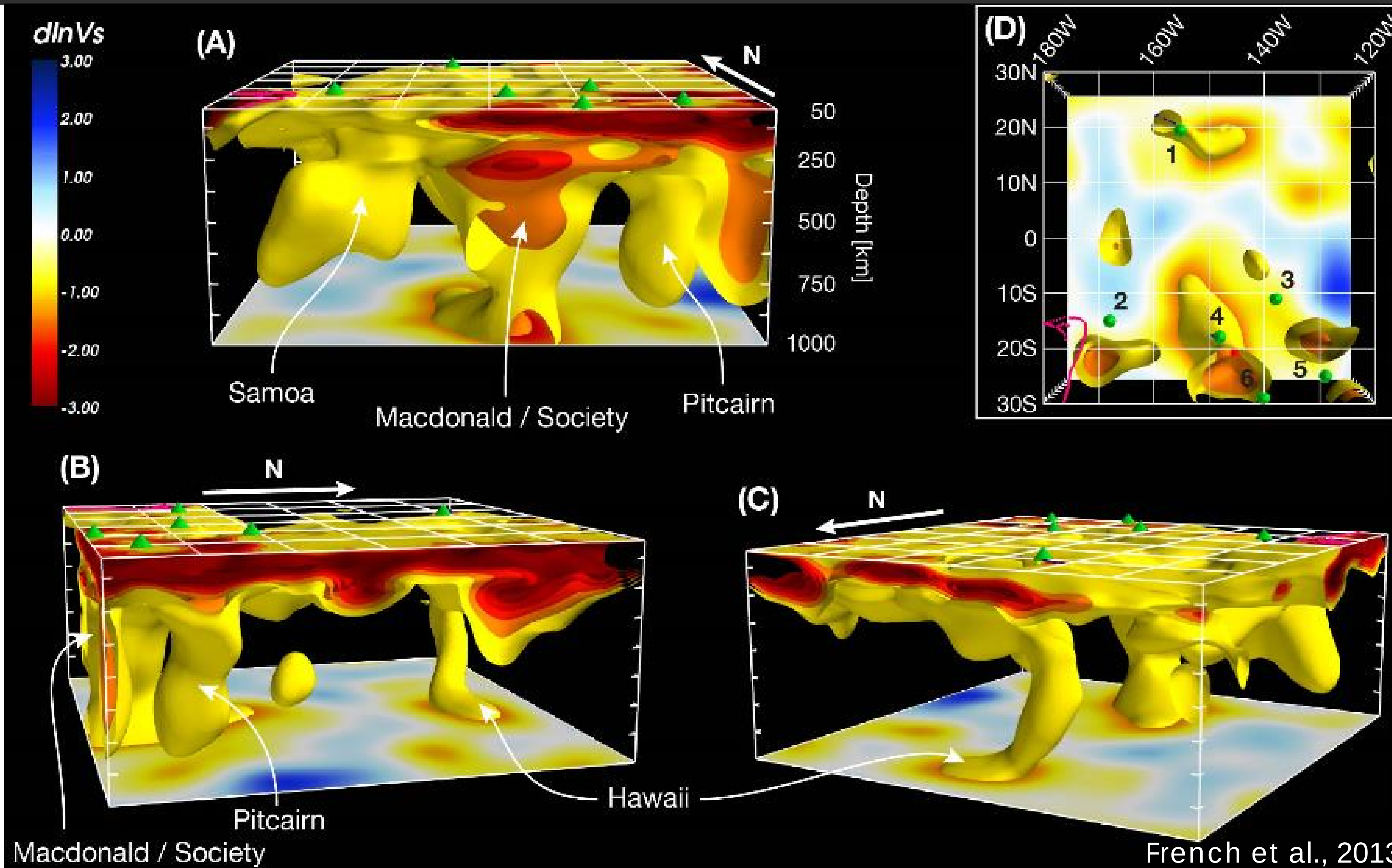
(a) Original plume (b) Cross correlation: P-wave window (d) Instantaneous phase: extended P-wave window

Rickers et al. 2012

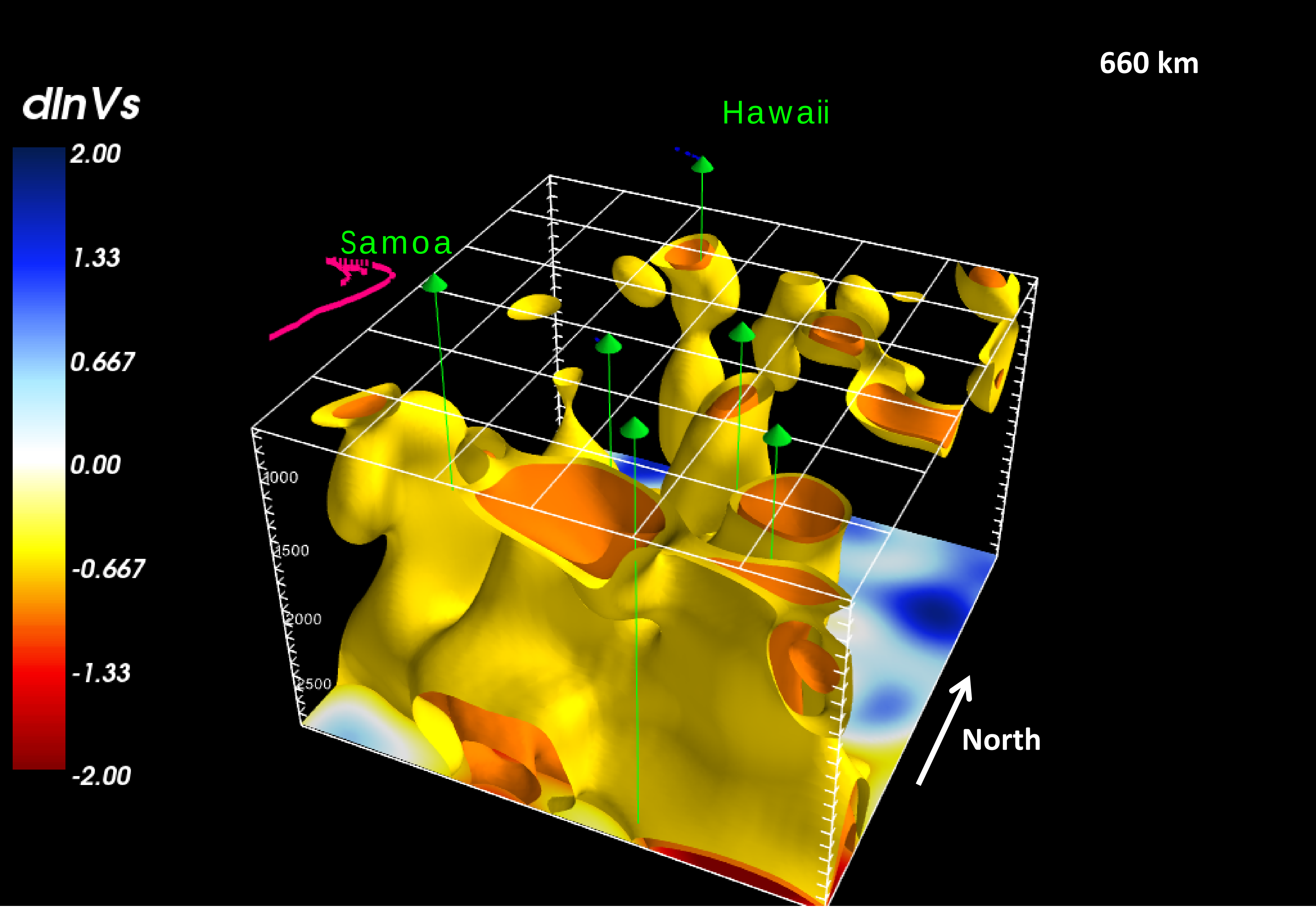
Origin of hotspot volcanism?



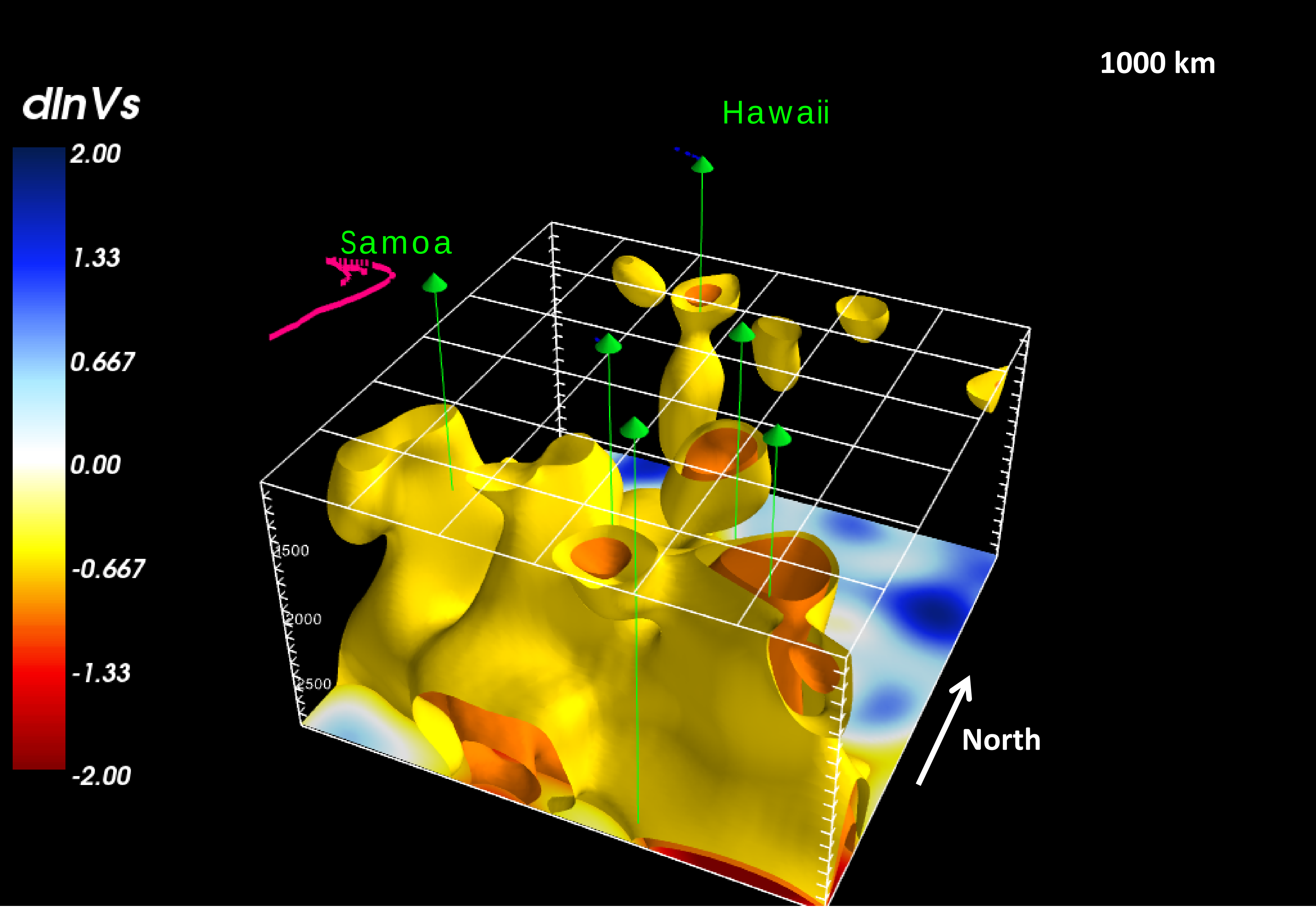
Conduits supply low velocity fingers



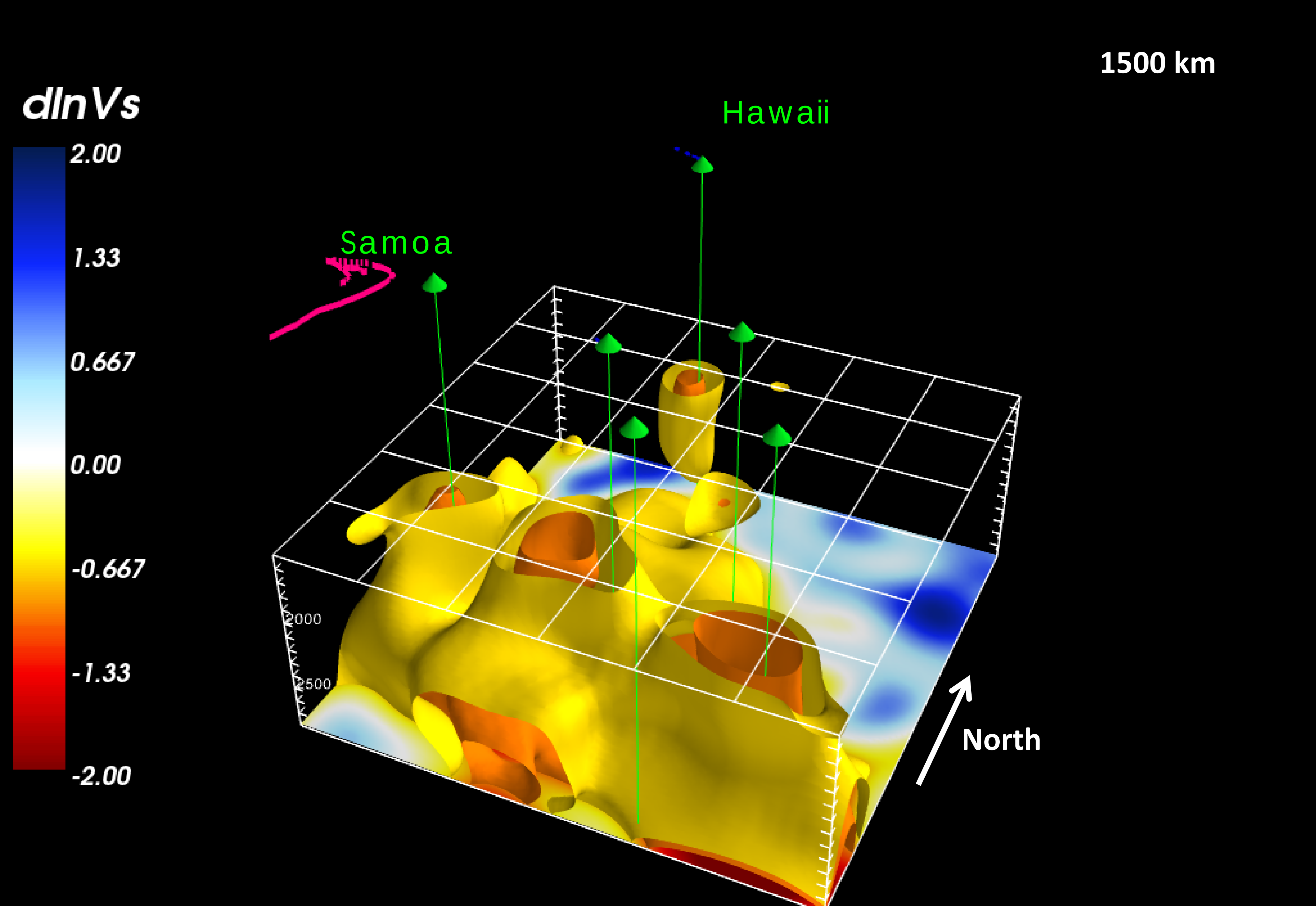
French et al., 2013



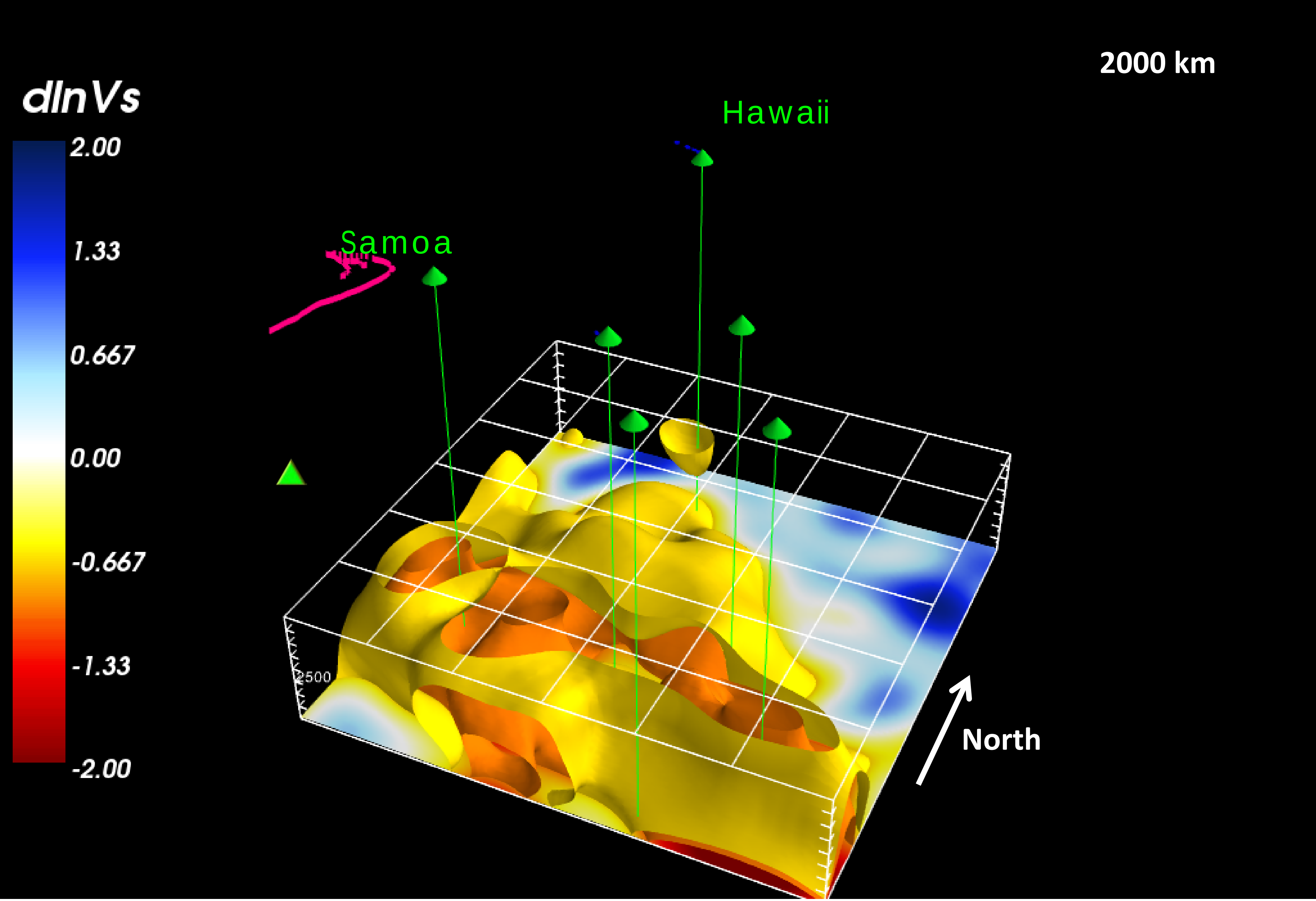
nd Romanowicz, submitted



nd Romanowicz, submitted

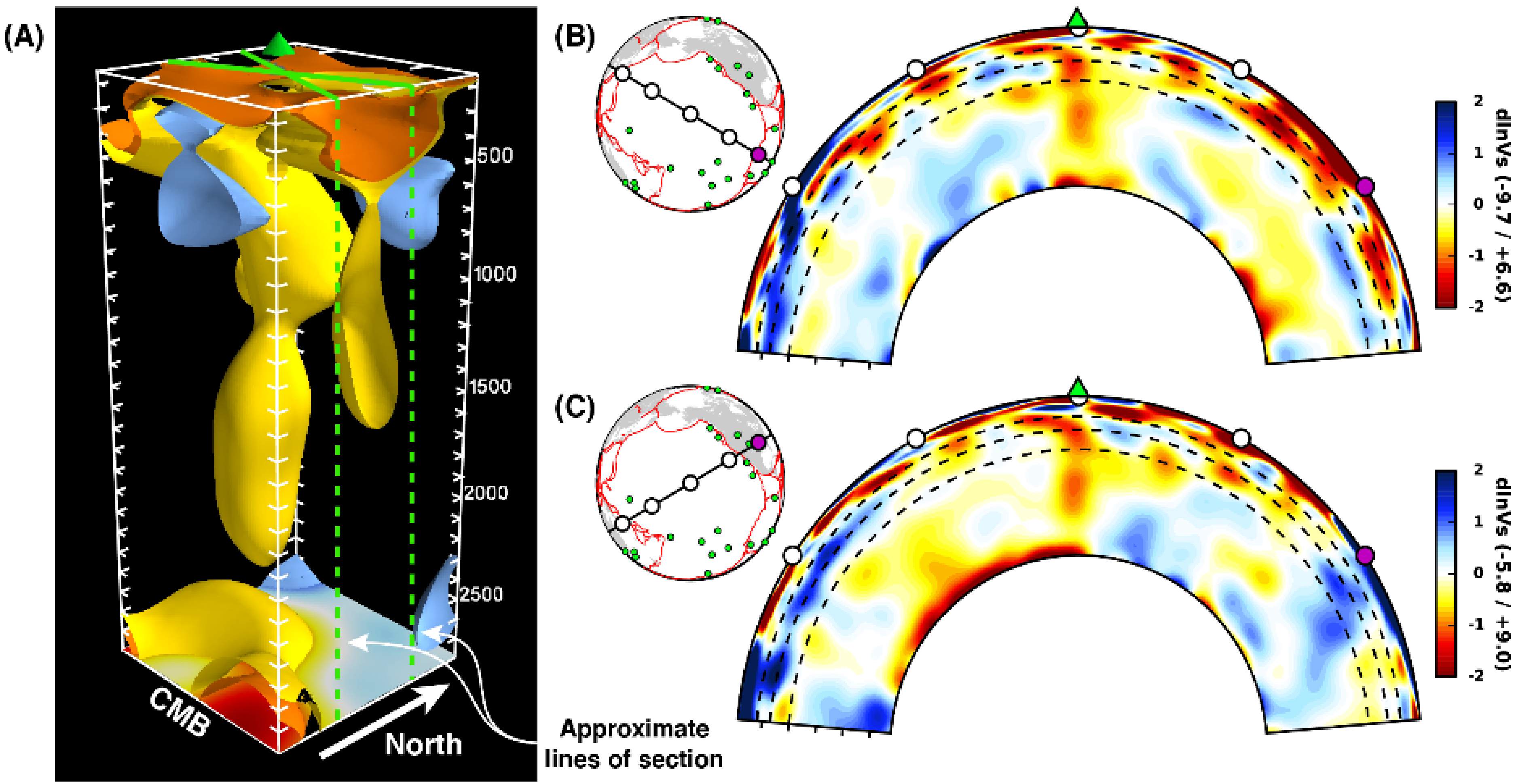


nd Romanowicz, submitted



nd Romanowicz, submitted

Hawaii



French and Romanowicz, 2015

Hybrid tomography of the mantle

- Sophisticated tomography with waveforms yields better constraints on upwellings and downwellings throughout the mantle.
- Broad (~1000 km wide) seismically slow conduits extending from base of the mantle to the LVZ (asthenosphere)
  - Too broad to be purely thermal (according to geodynamicists)
  - Must contain denser-than-average component (chemical heterogeneity)

LETTER

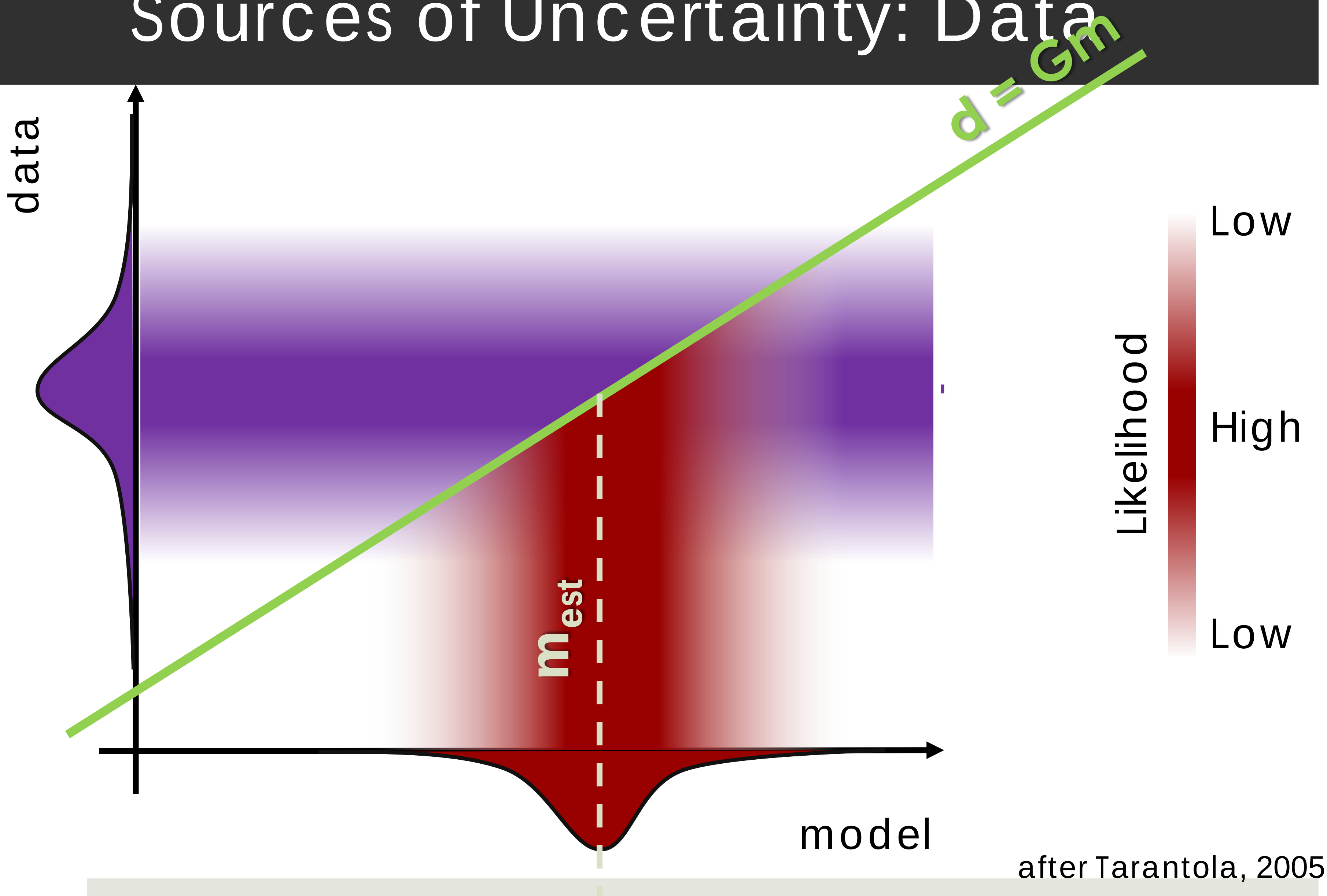
How certain can we be about this?
What is the uncertainty?
Can the hypothesis of purely thermal origin really be rejected?

doi:10.1038/nature14876

Broad plumes rooted at the base of the Earth's mantle beneath major hotspots

Scott W. French[1]† & Barbara Romanowicz[1,2,3]

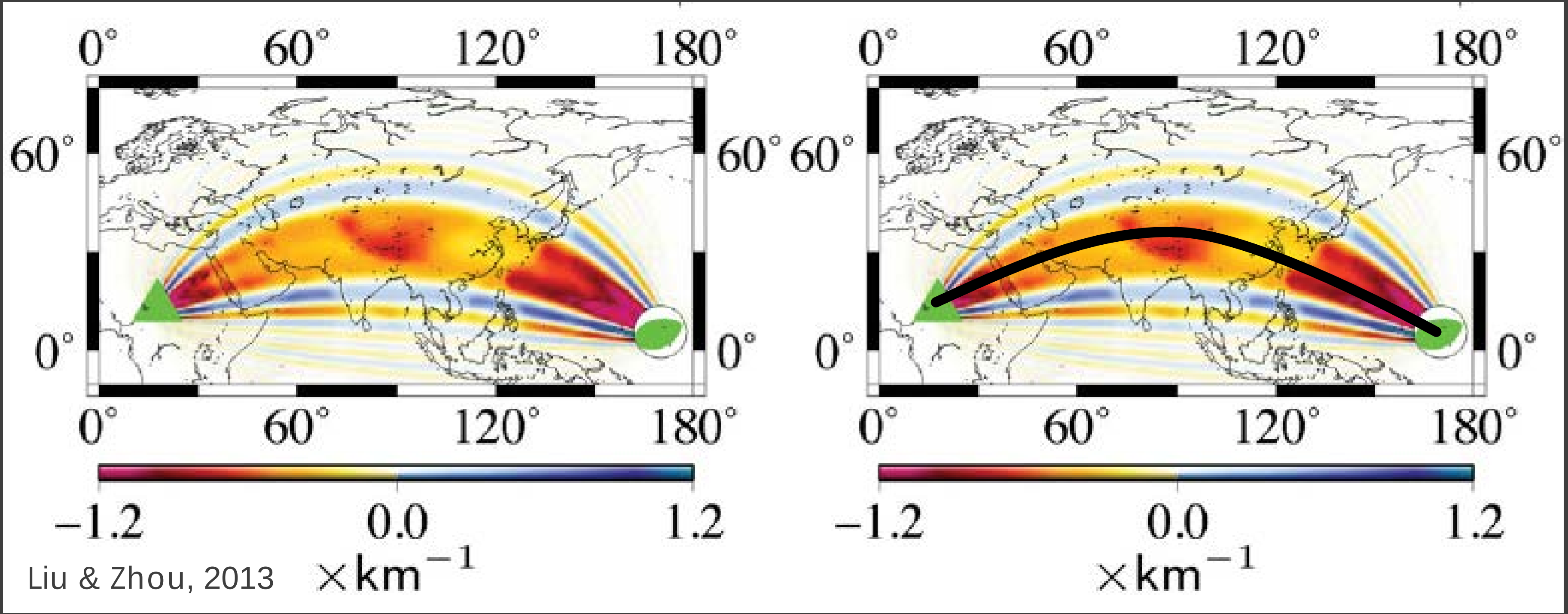
Sources of Uncertainty: Data



after Tarantola, 2005

Effect of forward modeling error

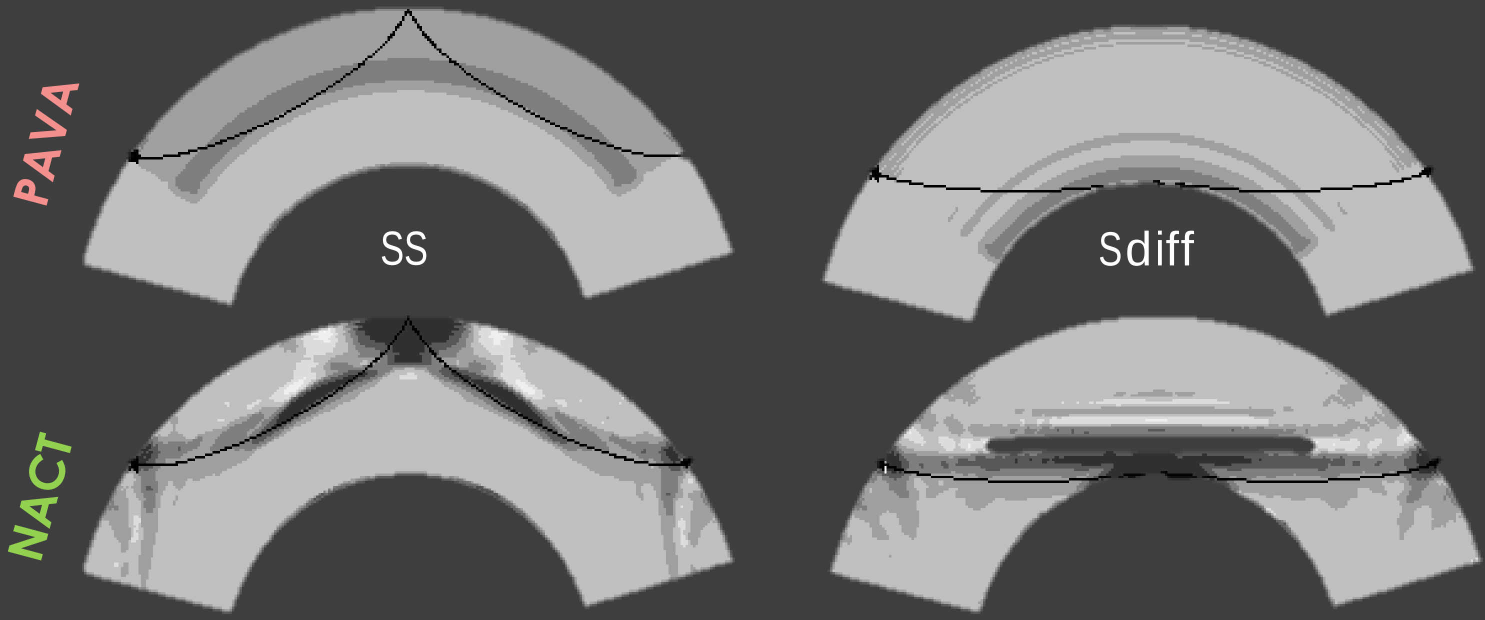
- Sometimes, we cannot (or choose not to) compute the forward problem g(m) exactly, e.g.:
  - Use rays computed for 1D Earth in 3D imaging
  - Consider only coupling along a normal mode branch (PAVA) not across branches (NACT)
  - Use perturbation theory instead of spectral element method



Liu & Zhou, 2013

Effect of forward modeling error

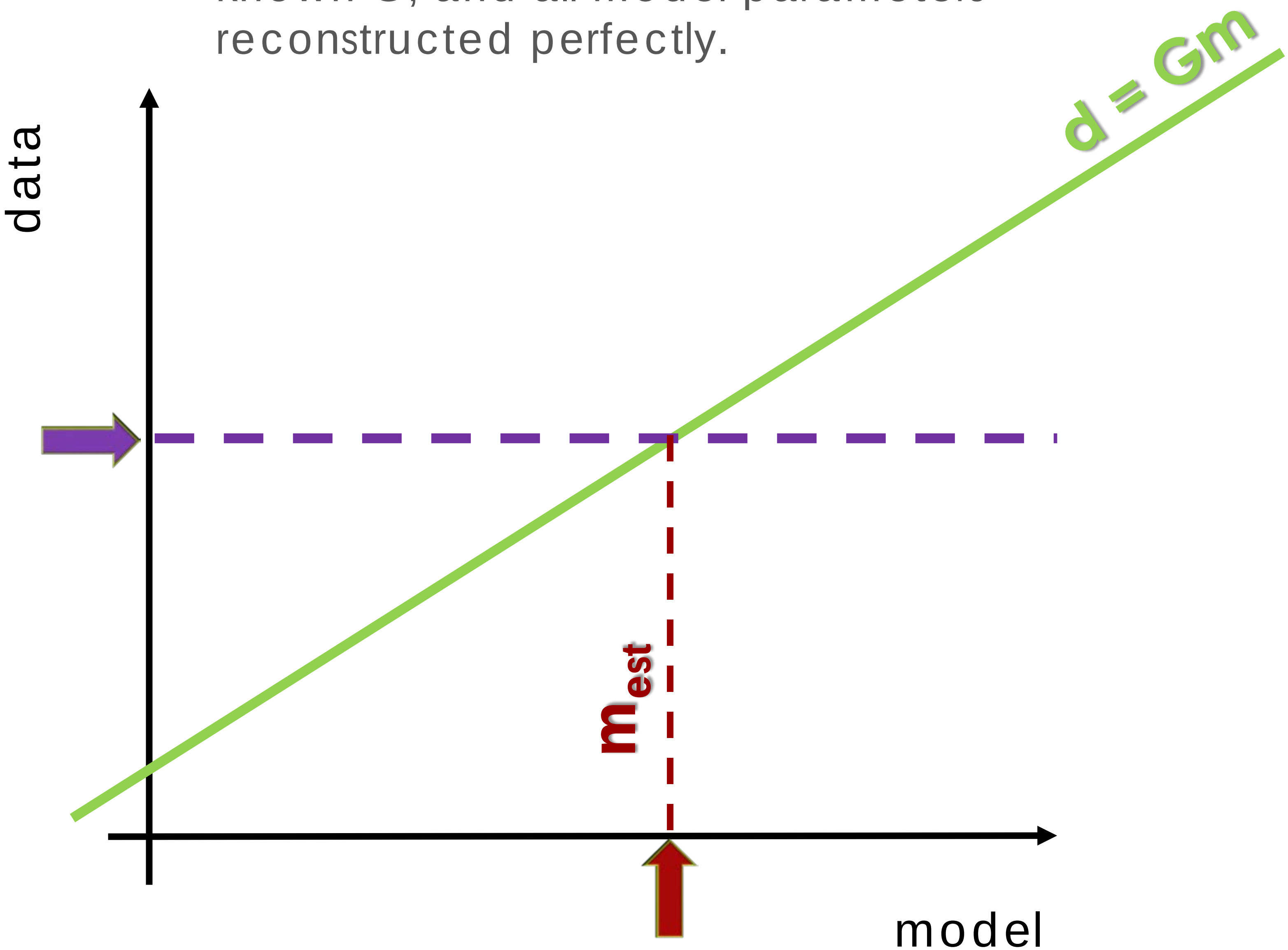
- Sometimes, we cannot (or choose not to) compute the forward problem g(m) exactly, e.g.:
  - Consider only coupling along a normal mode branch (PAVA) not across branches (NACT)
  - Use perturbation theory instead of spectral element method
  - Use rays computed for 1D Earth in 3D imaging



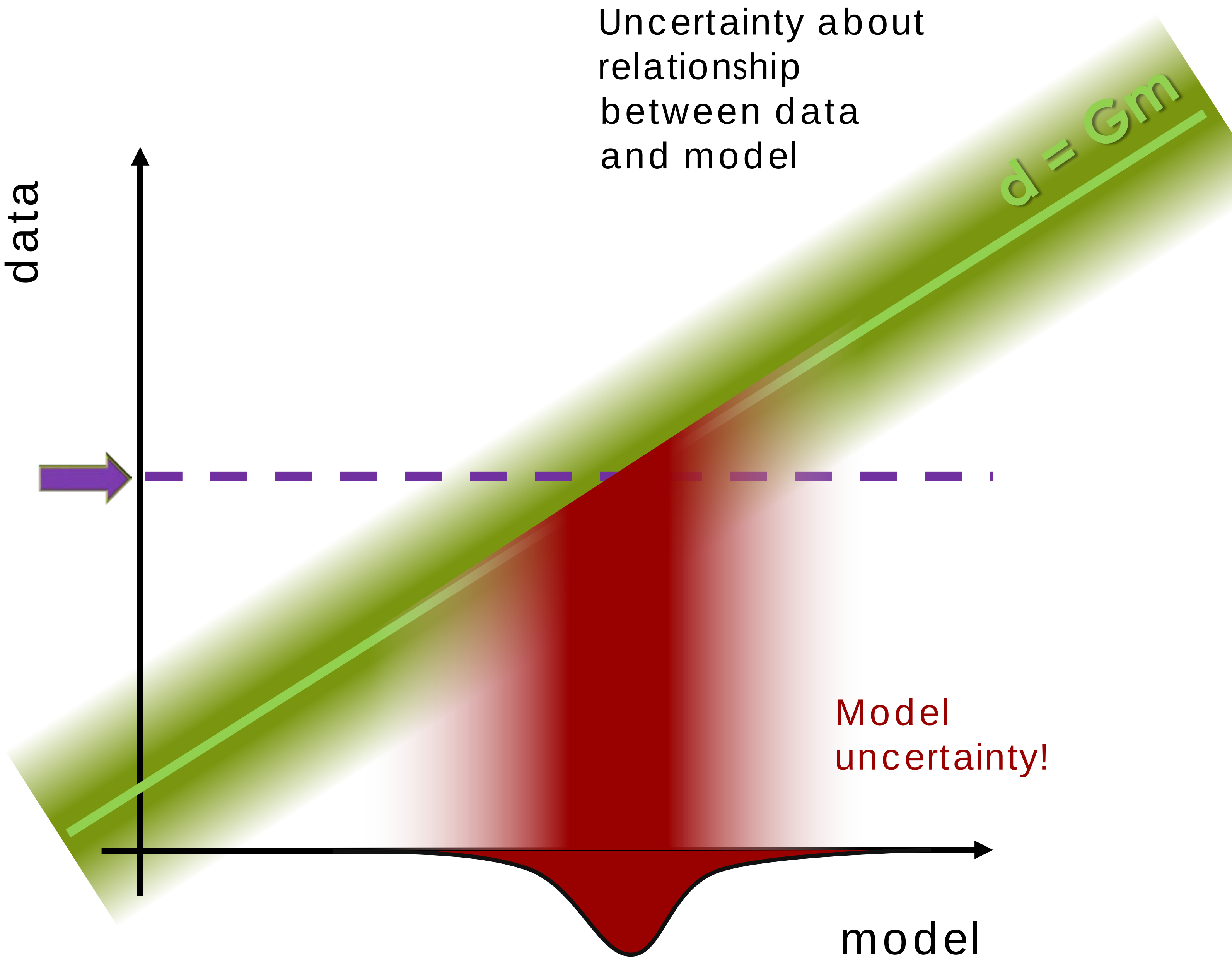
Li and Romanowicz, 1995

Perfect Inverse Problem

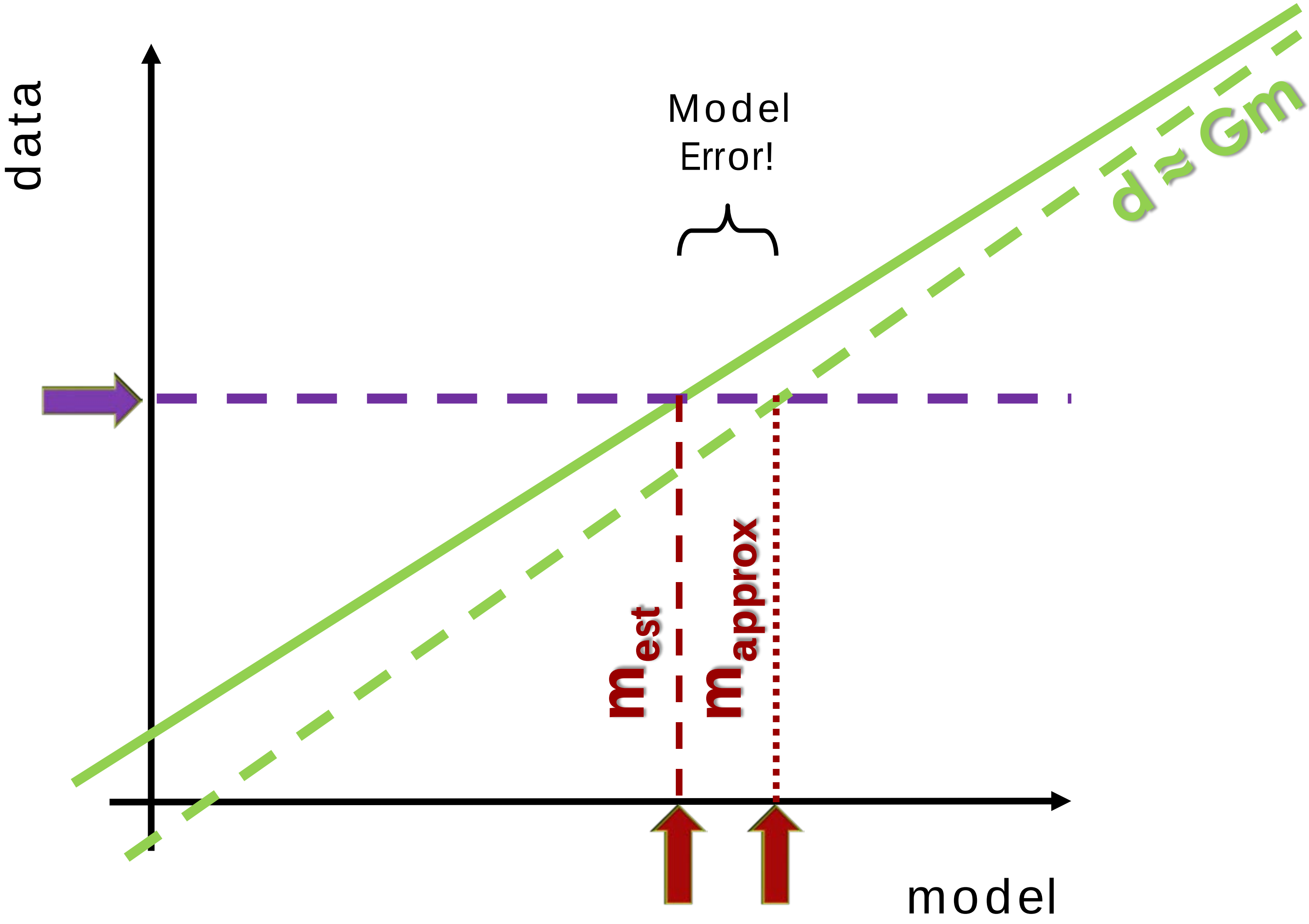
- Unrealistic illustration: Perfect data, perfectly known G, and all model parameters reconstructed perfectly.



Modeling Uncertainty

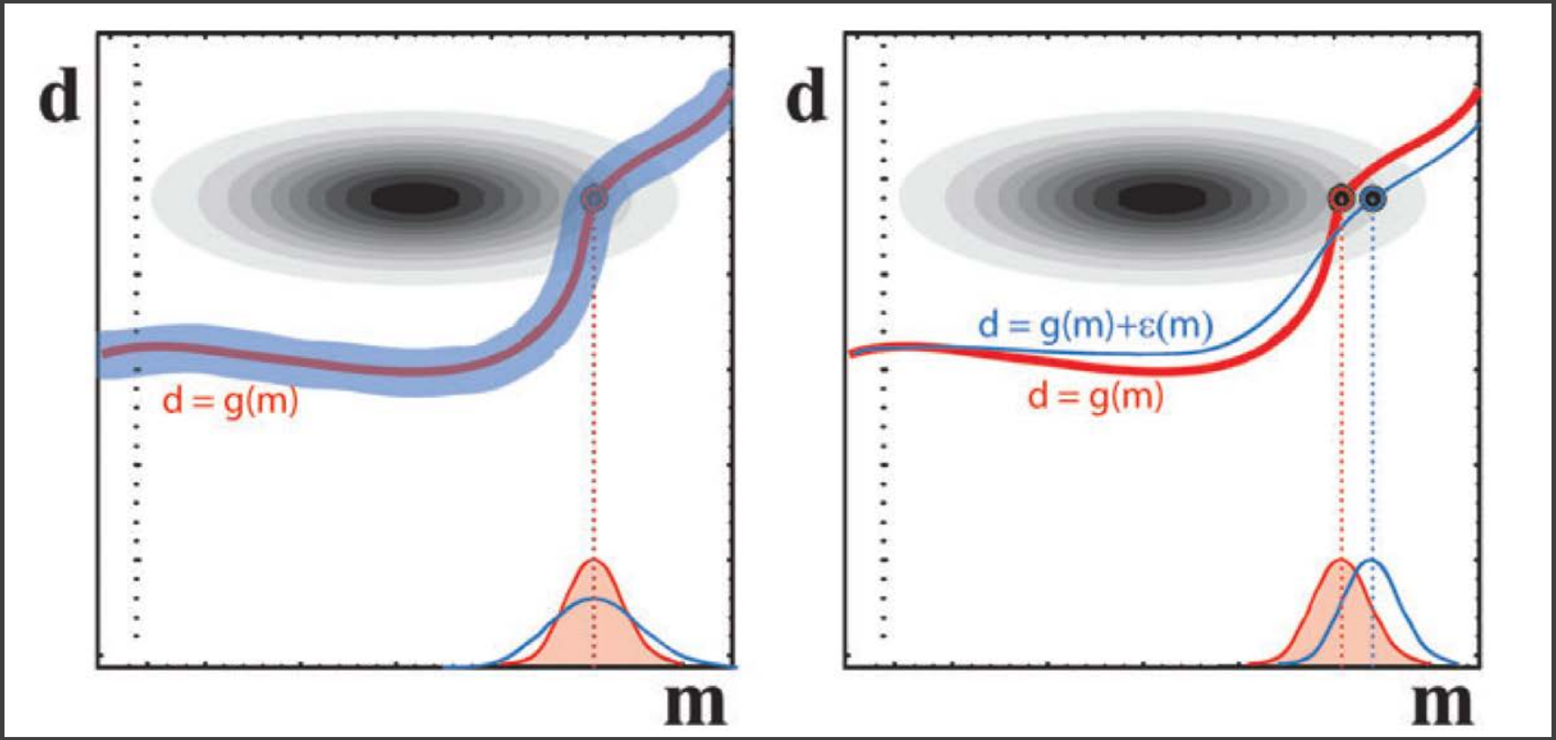


Modeling Error



Effect of forward modeling error

- Consequences of incorrectly computing g(m) can be significant: $\tilde{C}_M = (G^T(C_D + C_T)^{-1}G)^{-1}$
  - Equivalent to errors to data thereby increasing model estimate uncertainty
  - Can systematically **bias** the model estimate



Lekic and Romanowicz, 2011

Common complication...

- Often, real-world problems are mixed-determined
  - Some parameters are well constrained
  - Other parameters are poorly or completely un-constrained
- Sometimes it happens that $\det(G^T C_D^{-1} G) = 0$ and therefore $(G^T C_D^{-1} G)^{-1}$ – a.k.a. G^{-g} – does NOT exist (i.e. there are an insufficient number of linearly independent constraints to determine the unknowns)
- What do we do in that case?

…Use explicit regularization

- Regularization is the procedure by which we modify our problem to ensure that G^{-g} exists
- Our parameterization constitutes "**implicit**" regularization
  - By decreasing the number of parameters we can often ensure G^{-g} exists
  - e.g. If we only seek long wavelength variations in seismic wavespeed, then we don't need as much data to obtain a model
- **Explicit** regularization involves **introducing additional equations** based on prior knowledge

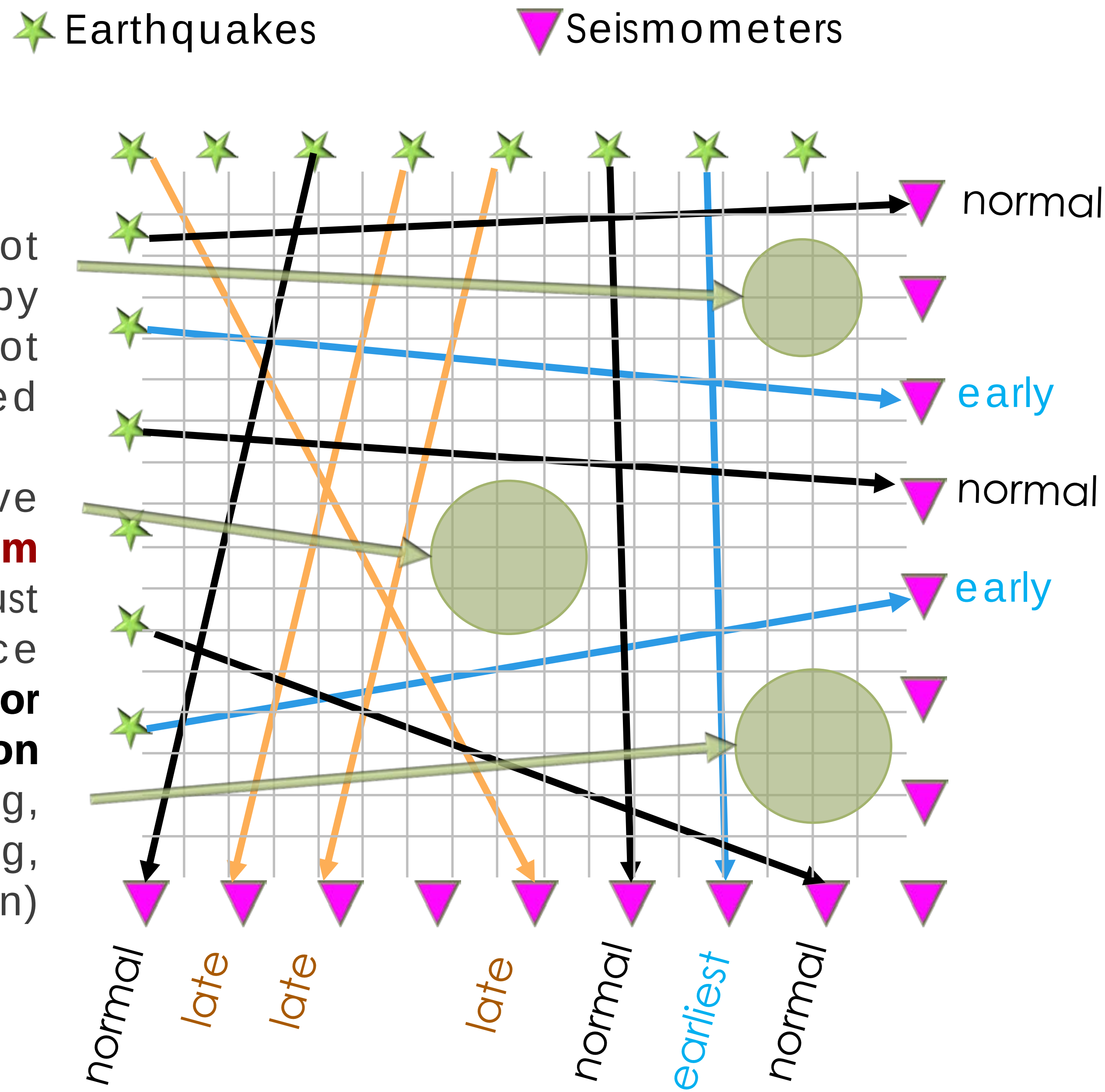
Two common forms

- We may want to find a model close to some reference model
  - minimize $(\mathbf{m}-\langle\mathbf{m}\rangle)^T(\mathbf{m}-\langle\mathbf{m}\rangle)$ in addition to minimizing $(\mathbf{d} - \mathbf{Gm})^T(\mathbf{d} - \mathbf{Gm})$
  - Norm damping
- May want to minimize roughness – size of first or second spatial derivatives of m – of the model:
  - Minimize $(\mathbf{Dm})^T(\mathbf{Dm})$ in addition to minimizing $(\mathbf{d} - \mathbf{Gm})^T(\mathbf{d} - \mathbf{Gm})$
  - Smoothing

Smoothing/Damping

Regions not sampled by data cannot be imaged

To solve **d = Gm** we must introduce **prior information** (i.e. damping, smoothing, regularization)



Prior model covariance matrix C_M

- Alternative but conceptually similar form of explicit regularization is by introducing C_M

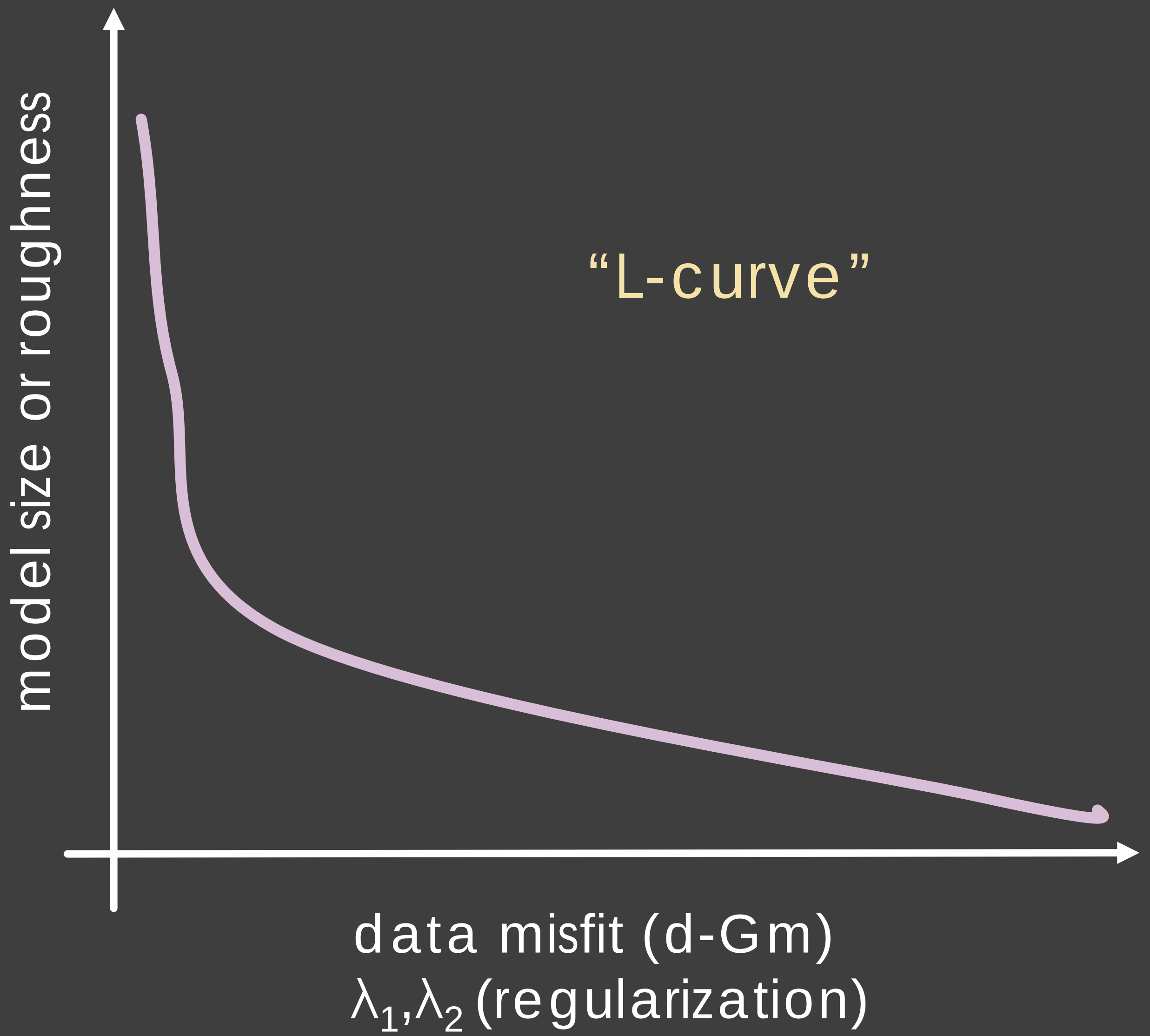
$$d = Gm \qquad \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \\ x_4 & 1 \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \end{bmatrix}$$

$$0 = Im \qquad \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \end{bmatrix}$$

$$0 = Dm \qquad [0] = [1 \quad -1] \begin{bmatrix} m_1 \\ m_2 \end{bmatrix}$$

$$m_{est} = (G^T C_D^{-1} G + \lambda_1 I + \lambda_2 \, D^T D)^{-1} \, G^T C_D^{-1} d$$

How much to regularize?



- More regularization →
  - more data misfit
  - more smoothness / smaller model
- A "sweet spot" can sometimes be determined

Prior model covariance matrix C_M

- Alternative but conceptually similar form of explicit regularization is by introducing C_M (the ignorant cousin of $\tilde{C}_M$).

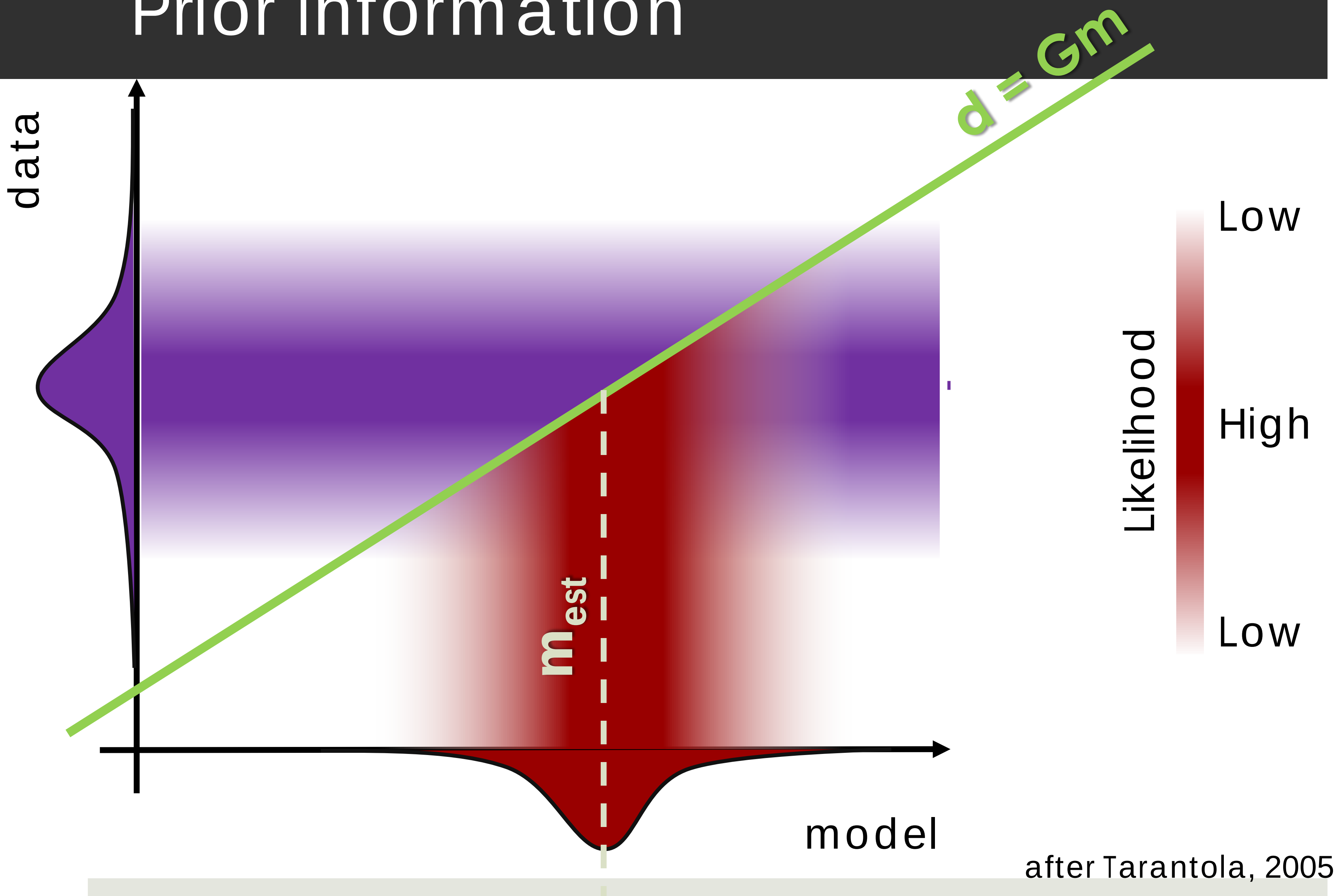
$$d = Gm \qquad \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \\ x_4 & 1 \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \end{bmatrix}$$

$$0 = C_M^{-1} m \qquad \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2 \\ \sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}^{-1} \begin{bmatrix} m_1 \\ m_2 \end{bmatrix}$$

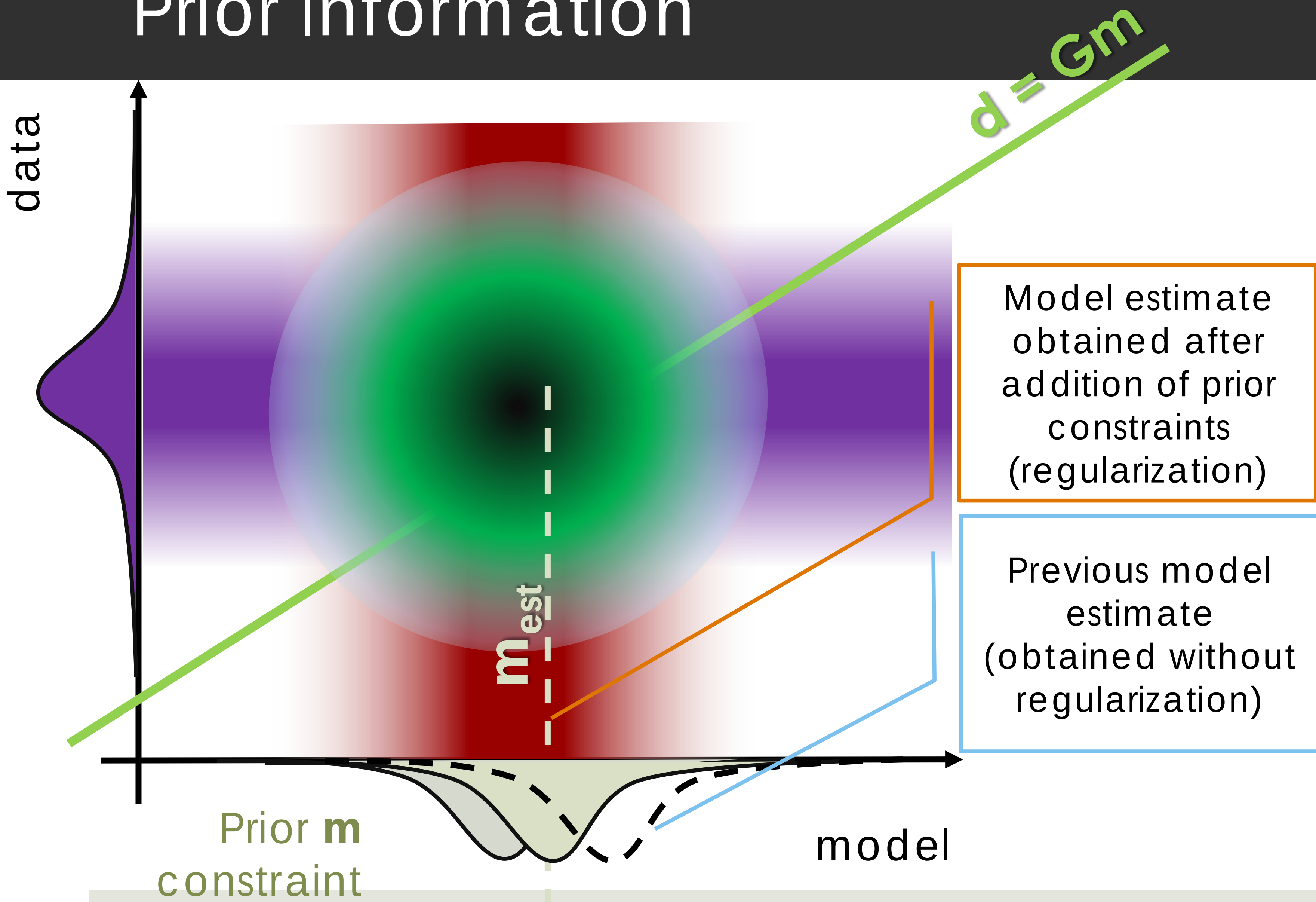
$$m_{est} = (G^T C_D^{-1} G + C_M^{-1})^{-1} \, G^T C_D^{-1} d$$

$$\tilde{C}_M = (G^T C_D^{-1} G + C_M^{-1})^{-1}$$

Prior Information



Prior Information



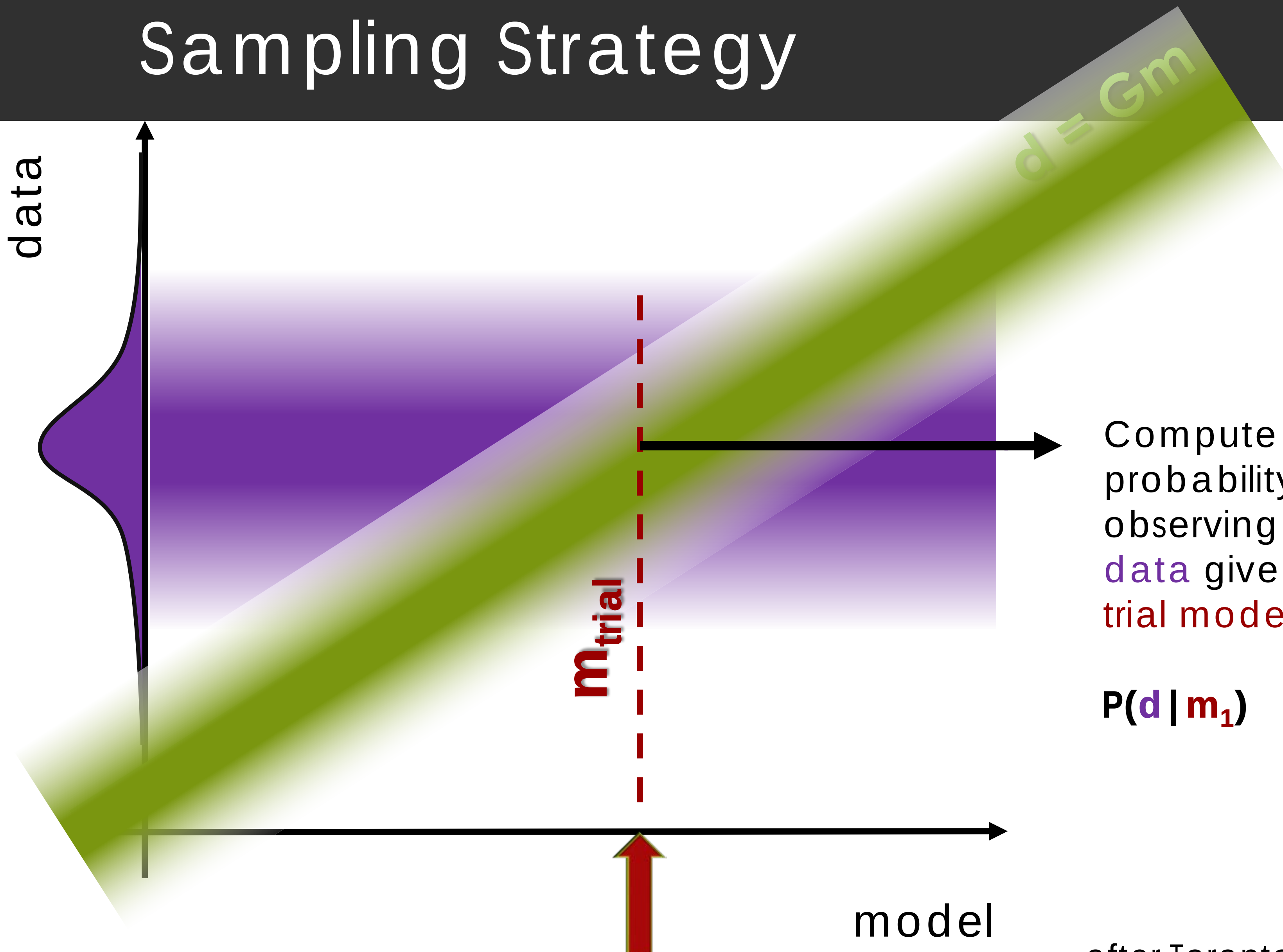
Origins of Uncertainty

- **Data** uncertainty:
  - Easy to deal with if you know it
  - Difficult to reliably estimate
- **Modeling**, i.e. computing **g(m)**:
  - Uncertainty maps to model uncertainty
  - Errors can bias estimated model
- **Prior information** about **model**:
  - Explicit damping / smoothing
  - Parameterization itself is a form of prior information

Model space search approaches

- When the relationship between data and model – i.e. g(m) – is non-linear, linear approaches can be inadequate, i.e. stuck in local minima and underestimating model error.
- Many current approaches focus on exploration of the model space
  - Eliminate need for regularization → less biased estimates of model parameters
  - Some have flexible parameterization: "transdimensional"
  - Some estimate data uncertainty: "hierarchical"
  - Yield ensemble of models that can be analyzed to map uncertainty and non-uniqueness

Sampling Strategy

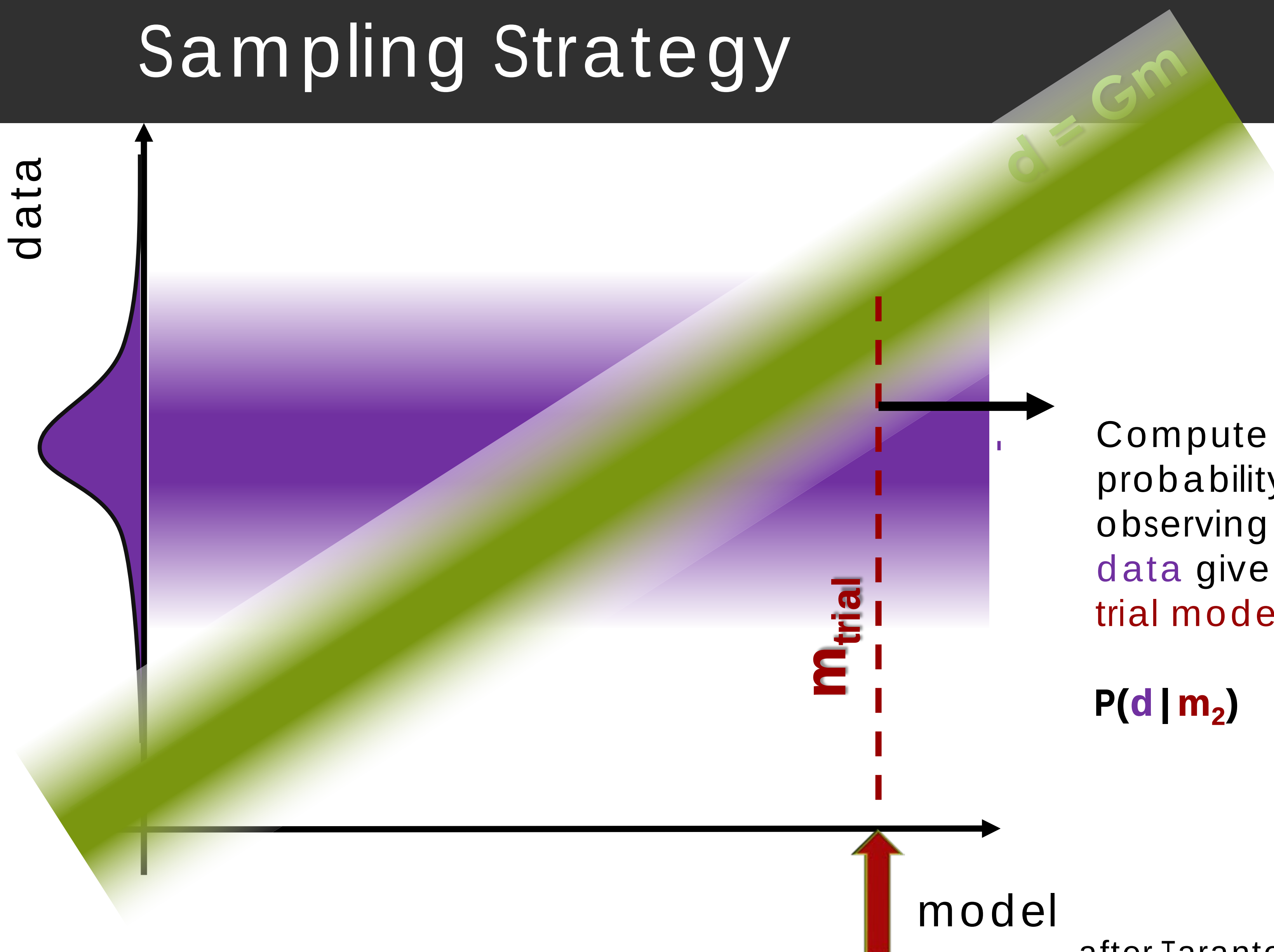


Compute probability of observing data given trial model

P(d | m_1)

after Tarantola, 2005

Sampling Strategy

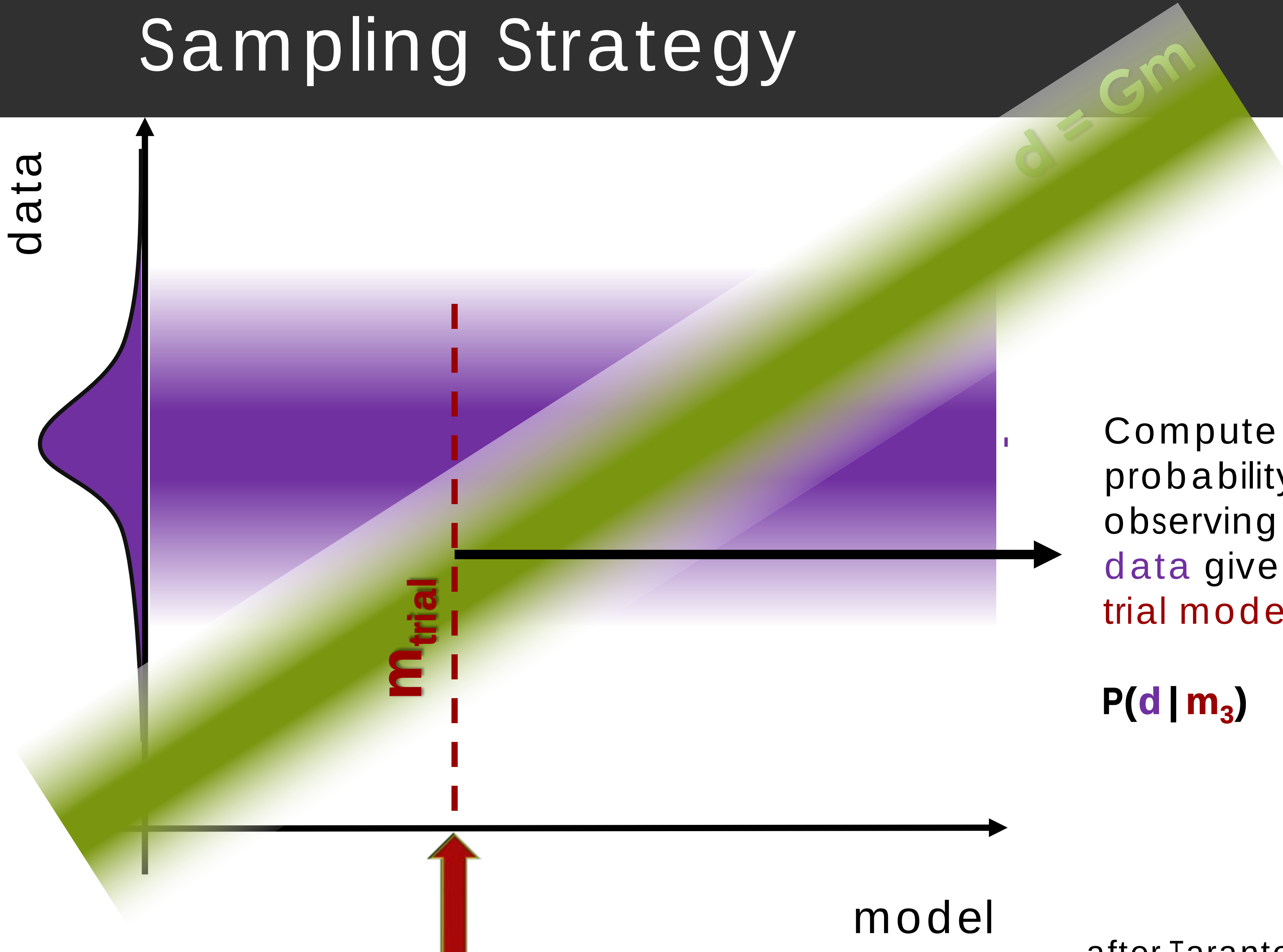


Compute probability of observing data given trial model

P(d|m₂)

after Tarantola, 2005

Sampling Strategy



Compute probability of observing data given trial model

$P(d\,|\,m_3)$

after Tarantola, 2005

Model Space Search

- Compute the probability of observing the data for many trial models → map out the **ensemble** of acceptable models
- Sampling approach is powerful because:
  - It works when the relationship between data and model – i.e. g(m) – is non-linear
  - You can sample differently parameterized models (i.e. vary number, size, shape of boxes between different models) → "transdimensional"
  - You can sample different estimates of data uncertainty → "hierarchical"
  - Eliminates need for prior information → less biased estimates of model parameters
- Analyzing the **ensemble** yields estimates of uncertainty

Model Space Search

- Compute the probability of observing the data for many trial models → map out the **ensemble** of acceptable models
- Sampling approach is **limited** because:
  - Computing the fit of sampled models to the data can be prohibitively computationally expensive
  - Approximations to the **g(m)** relationship must often be made → Increases "modeling" error
  - Convergence can be slow and difficult to assess / ensure, especially when number of unknowns is large
- We use a reversible jump Markov chain Monte Carlo (**rjMcMC)** to jump between different models parameterized in different ways.

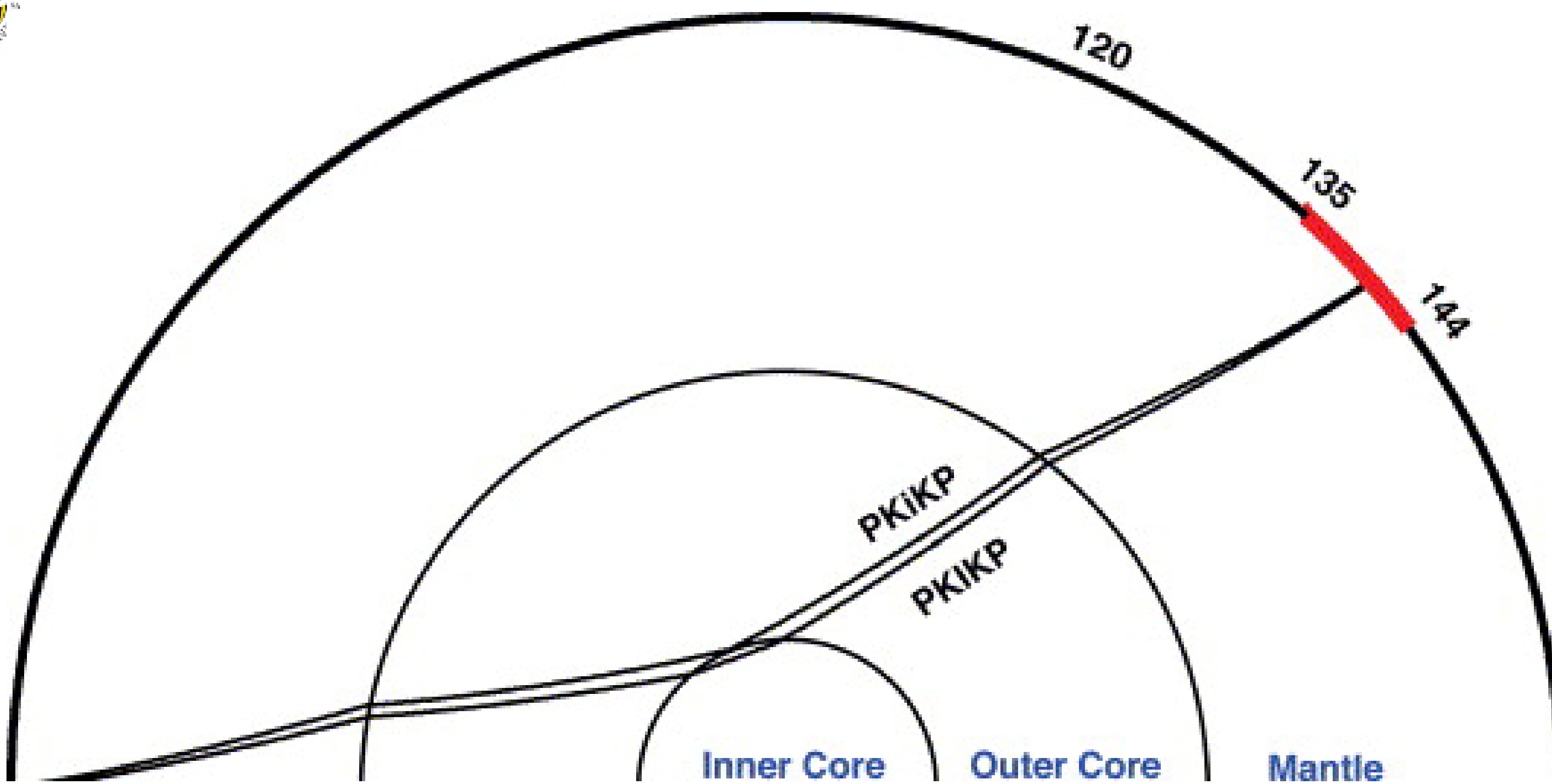
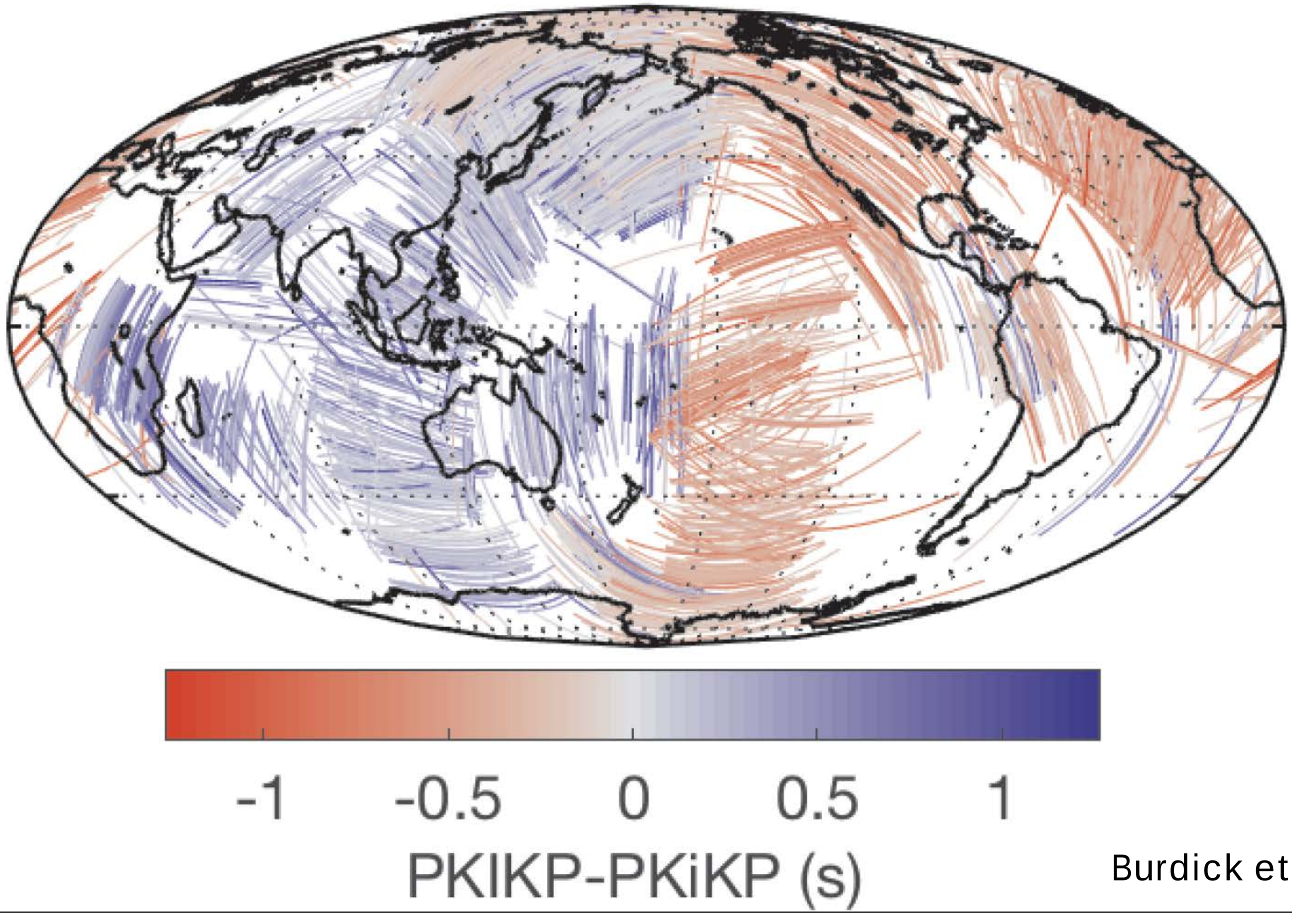


Figure from Cao and Romanowicz, 2004

Inner Core Structure

Differential travel-times of inner core sensitive phases indicate the presence of a pseudo-hemispheric east-west dichotomy in velocity, attenuation, and even anisotropy



Burdick et al. *AGU 2017*

Inner Core Structure

Differential travel-times of inner core sensitive phases indicate the presence of a pseudo-hemispheric east-west dichotomy in velocity, attenuation, and even anisotropy

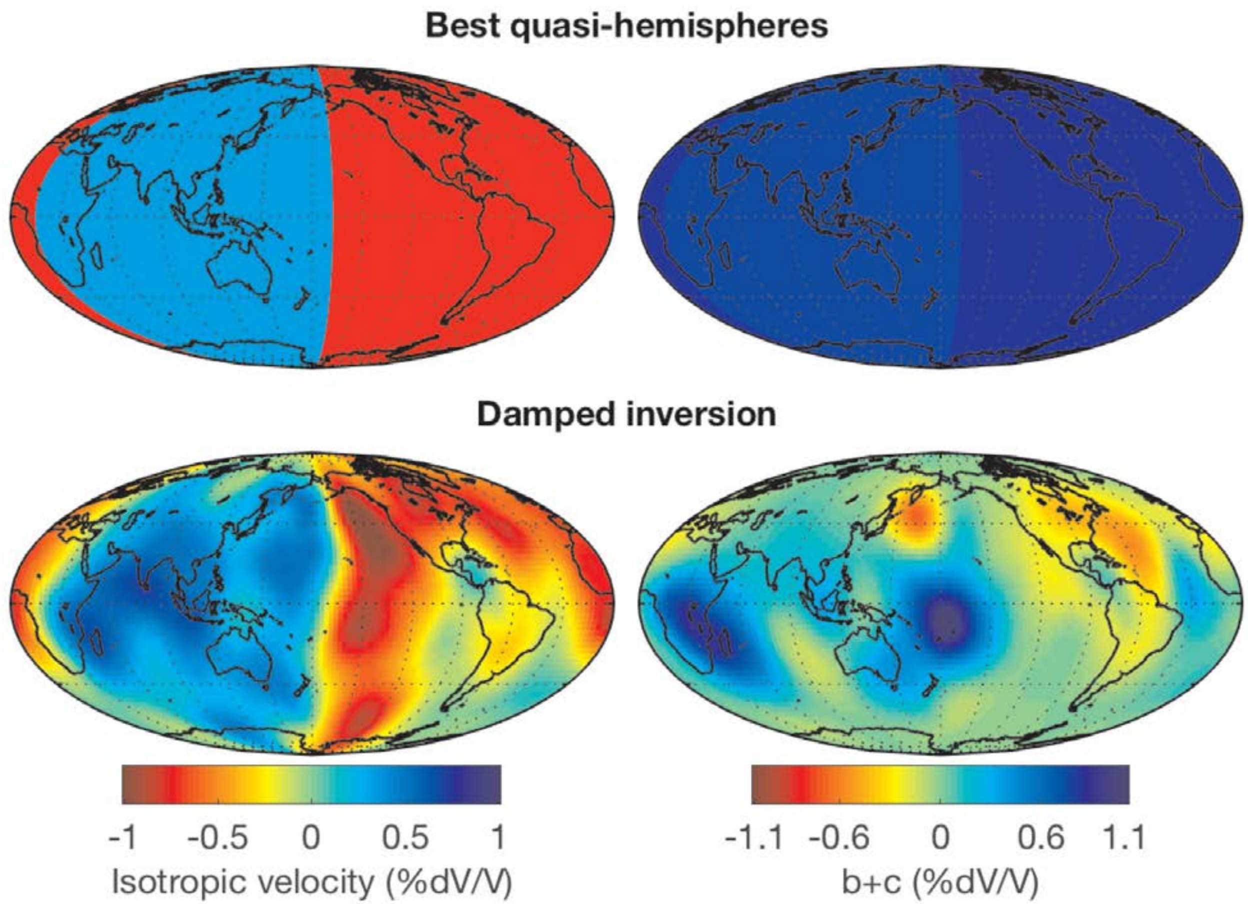
Limitations of Traditional Approach

We can infer average properties of hemispheres

But, uneven path coverage means that traditional least-square inversion requires damping / smoothing

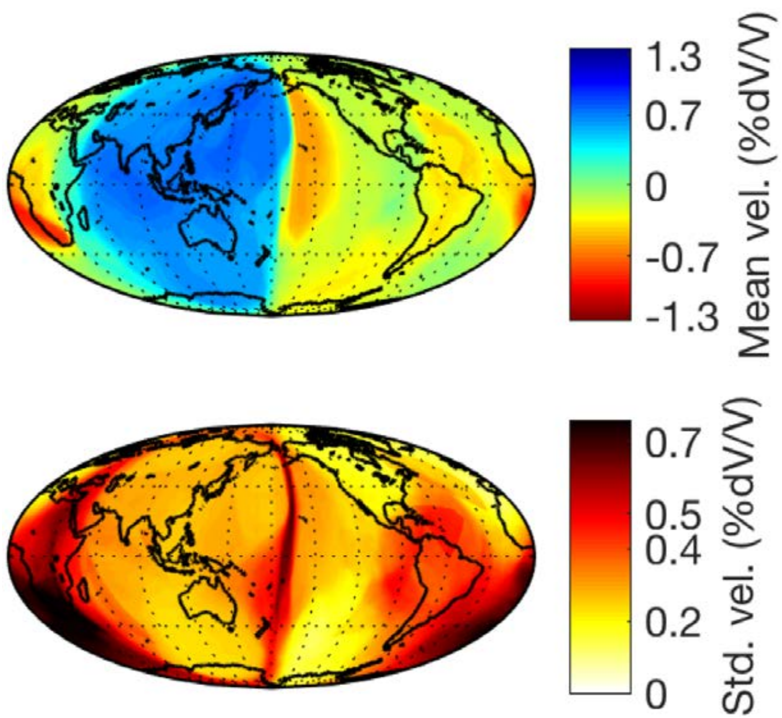
Transdimensional approach should retrieve both **long-wavelength structure and details**

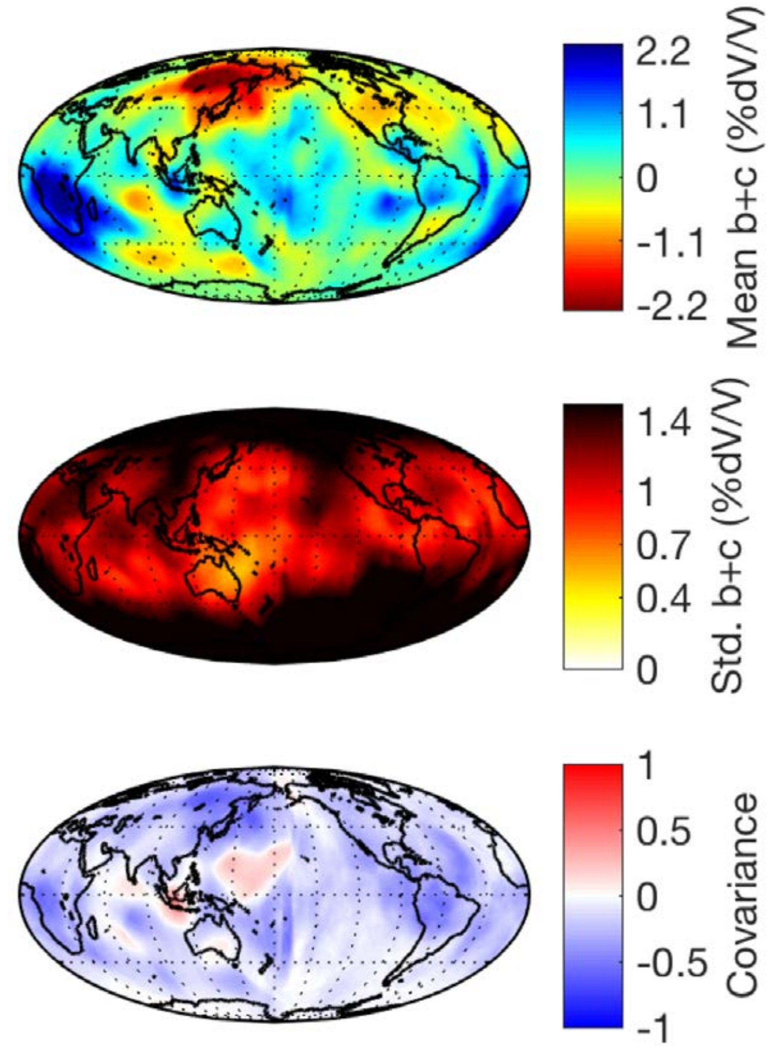
$$\frac{\Delta T}{T} = \frac{dV}{V} + b\cos^2\zeta + c\cos^4\zeta$$



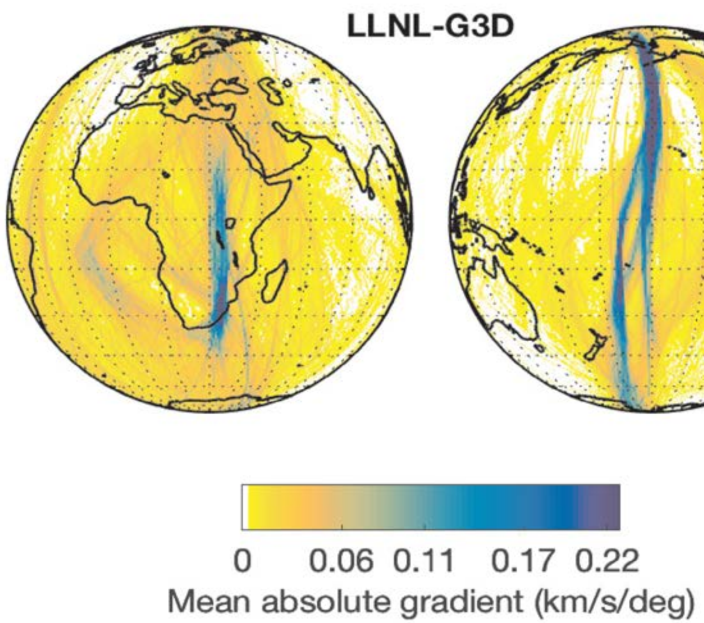
Tomography of the Inner Core

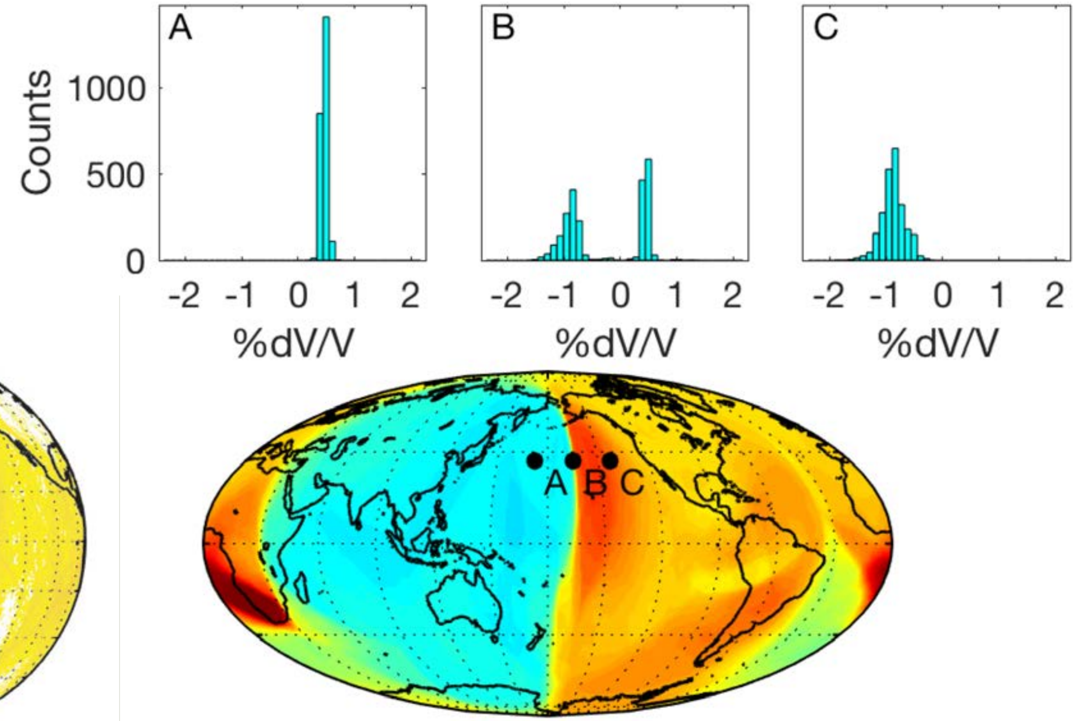
- Eastern hemisphere smaller than Western
- Bounded by abrupt velocity changes
- Anisotropic signal not hemispheric
- Strong trade-offs between anisotropy and small-scale isotropic structure





Burdick, Waszek, Lekic, in prep.





Burdick, Waszek, Lekic, in prep

Bi-modal distributions

Edges between hemispheres characterized by bi-modal distributions

Resolution analysis

- To solve the mixed-determined tomographic problem, we usually need to introduce prior information (i.e. smoothing, damping, C_M)

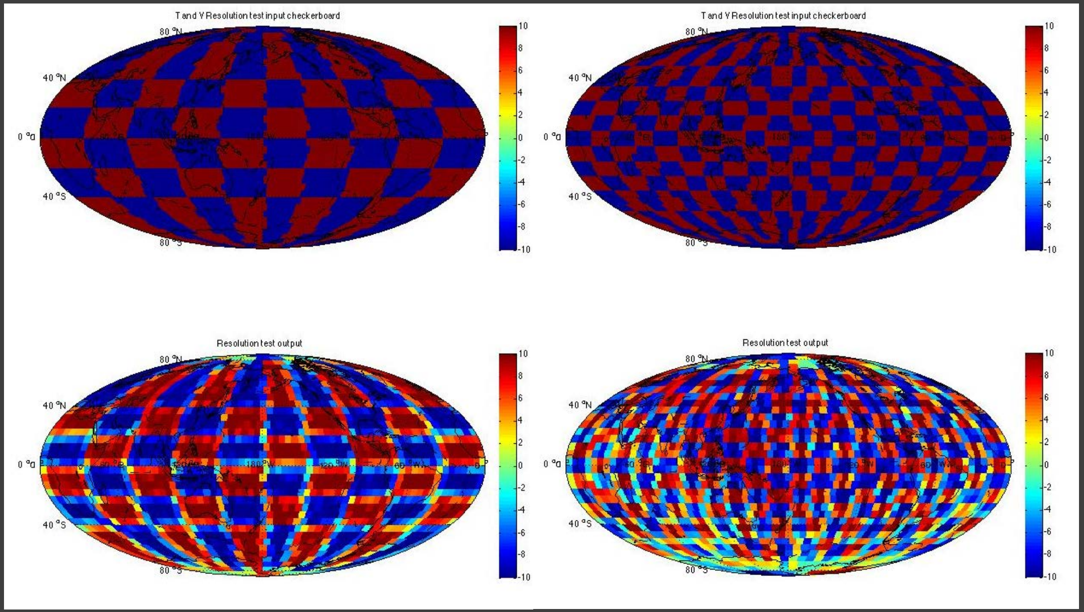
$$m = (G^T C_D^{-1} G + C_M^{-1})^{-1} G^T C_D^{-1} d$$

$$\tilde{C}_M = (G^T C_D^{-1} G + C_M^{-1})^{-1}$$

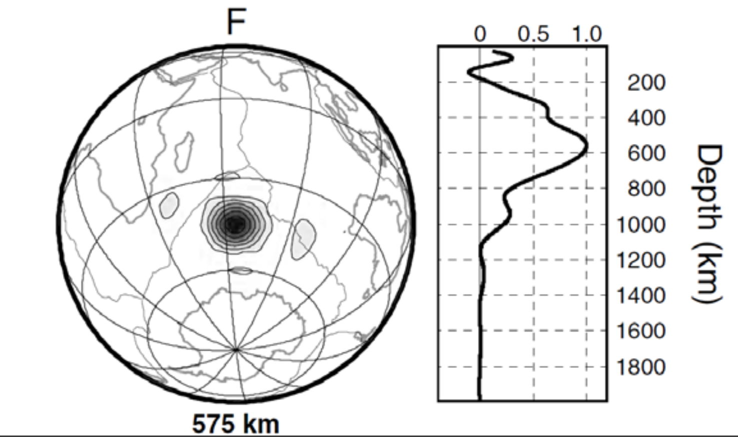
- We can see how well we would image input structures m_{test} by feeding predicted data $\mathbf{d_{test}} = \mathbf{Gm_{test}}$ into the inversion:
  - $m_{est} = (G^T C_D^{-1} G + C_M^{-1})^{-1} G^T C_D^{-1} G m_{test} = \mathbf{R} m_{test}$
  - **R** is called the "resolution" operator (or matrix)

Resolution analysis, continued.

- $m_{est} = \mathbf{R} m_{test}$
- We usually present **R** by plotting input-output pairs of “resolution tests”:
  - checkerboards of fast/slow variations or spatially localized spike of velocity perturbation
- R analysis neglects errors in computing g(m), biases in data, and fails if problem too non-linear.



Above figures from this afternoon’s tutorial!



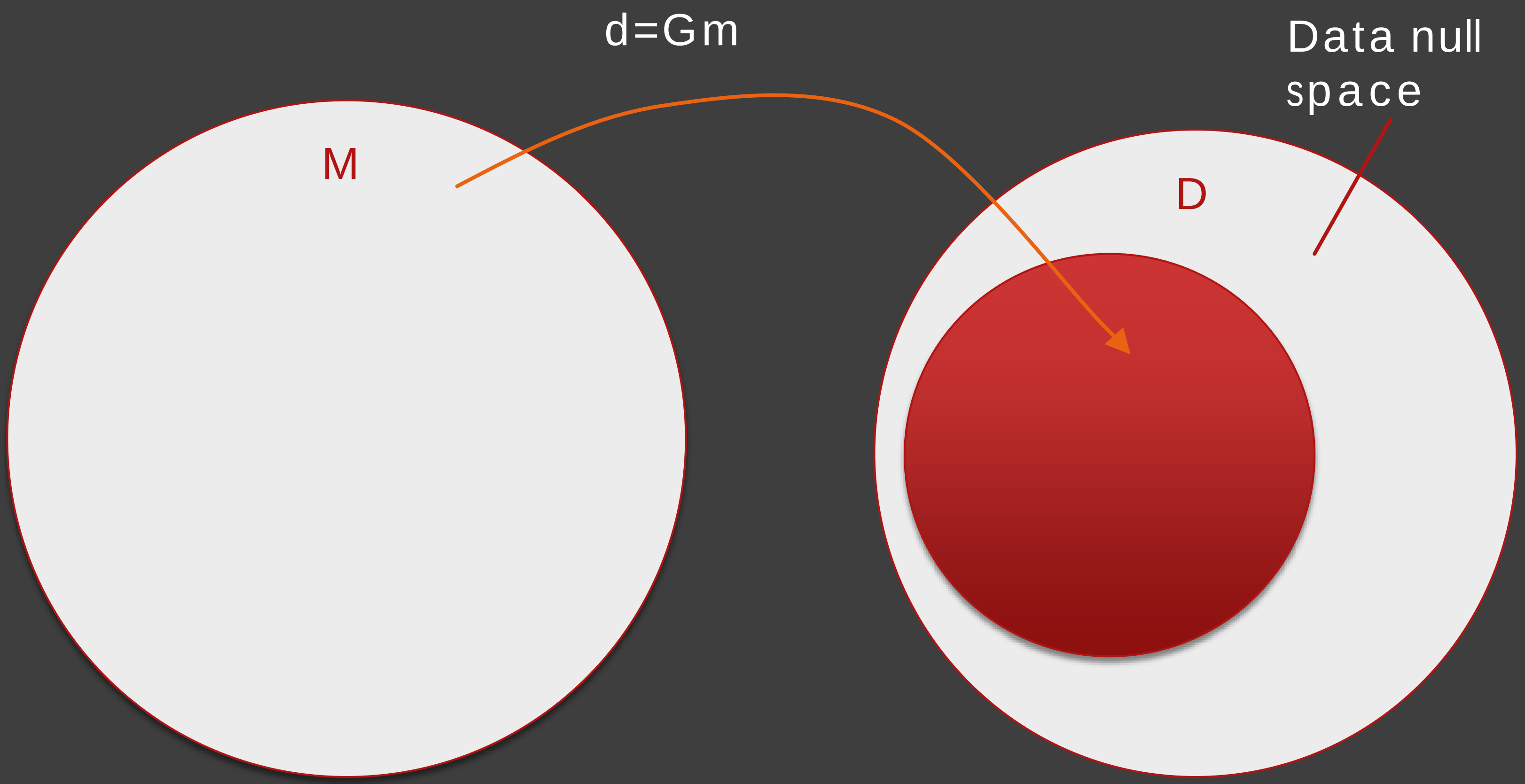
Ritsema et al. 2004

Improving Tomography Resolution

The data null space

- Linear combinations of data that cannot be predicted by any possible model vector **m**
  - For example, no simple linear theory could predict different values for a repeated measurement, but real repeated measurements will usually differ due to measurement error
- If a data null space exists, it is generally impossible to match the data exactly

Null spaces



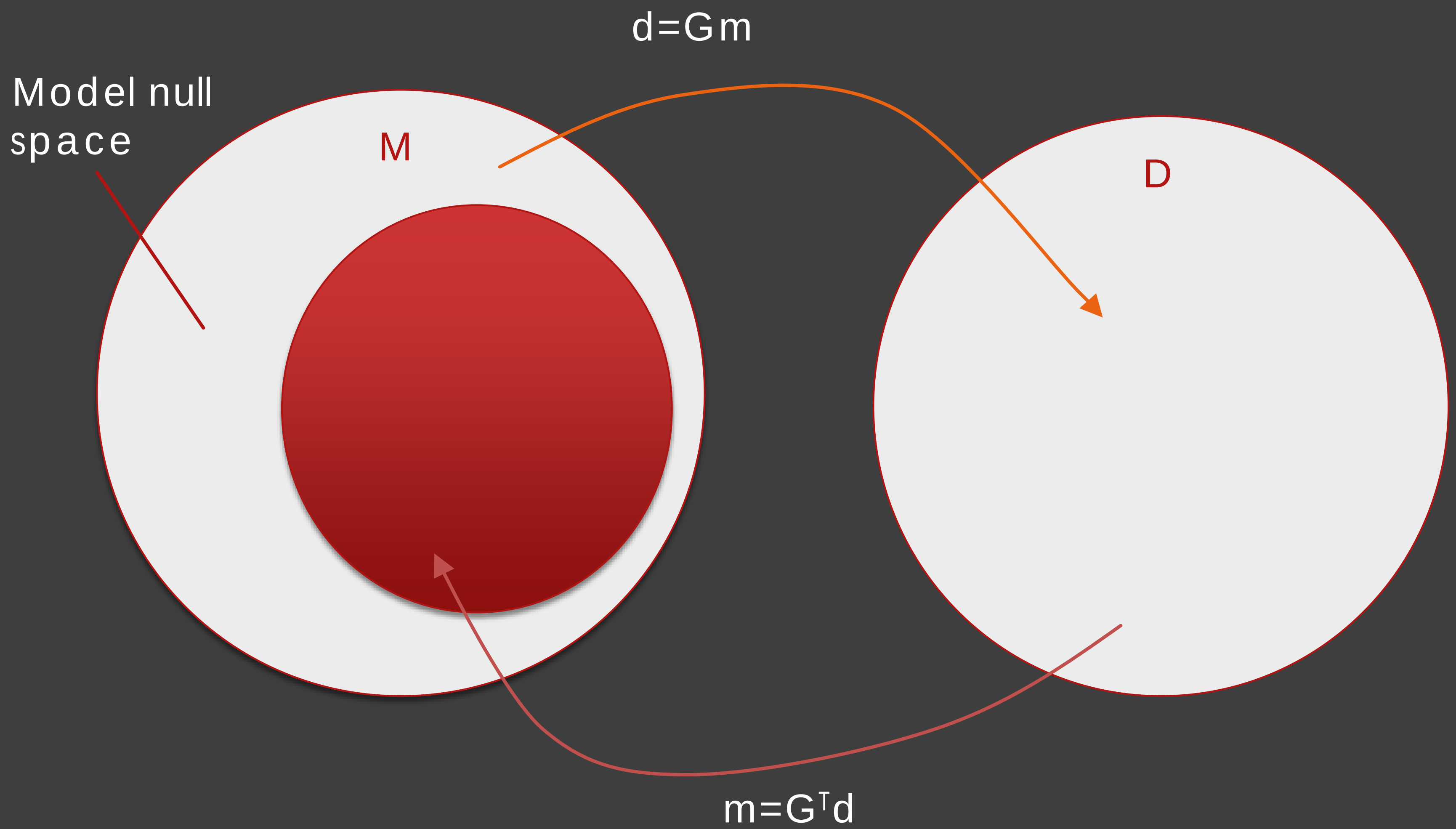
The model null space

- A model null vector is any solution to the homogenous problem

$$\mathbf{G}\mathbf{m}_{\text{null}} = 0$$

- This means we can add in an arbitrary constant times any model null vector and not affect the data misfit
- The existence of a model null space implies non-uniqueness of any inverse solution
- We can use linear algebra to map out model null space

Null spaces



Quantifying null space with Singular Value Decomposition

- SVD breaks down G matrix into a series of vectors weighted by singular values that quantify the sampling of the data and model spaces
  - Column vectors of U associated with 0 (or very near-zero) singular values are in the data null space
  - Column vectors of V associated with 0 singular values are in the the model null space

$$\mathbf{G} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^{\mathrm{T}}$$

N x N matrix with columns representing vectors that span the data space

M x M matrix with columns representing vectors that span the model space

$$\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda}_M \\ 0 \end{bmatrix}$$

If M<N, this is a M x M square diagonal matrix of the singular values of the problem

How to evaluate inferences?

- Consider data + modeling uncertainty:
  - Quality of data / processing and co-variance among data points
  - Theoretical/computational approximations to the relationship between data and model
- Consider how sensitive data is to quantity of interest:
  - e.g. waveforms contain more information than travel-times
  - e.g. density is tough compared to velocity
- Posterior model covariance matrix $\tilde{C}_M$ can tell you about trade-offs and uncertainty
  - but be mindful of limitations!

Evaluating an inverse model paper

- How well does the data sample the region being modeled? How sensitive is it to parameters of interest? Is the data any good to begin with?
- Is the problem linear or not? Can it be linearized? Should it?
- What kind of theory are they using for the forward problem?
- What inverse technique are they using? Does it make sense for the problem?
- What's the model resolution and error? Did they explain both implicit and explicit regularization choices they made and what effect they have on the model?

Q&A + Discussion

- Please ask me questions.
- Please make provocative comments.