Inference of Uncertain Parameters in Physical Models Governed by PDEs with Application to Earth's Mantle Convection

Johann Rudi[1] Georg Stadler[2] Jiashun Hu[3] Michael Gurnis[3] Omar Ghattas[4]

[1]Mathematics and Computer Science Division, Argonne National Laboratory

[2]Courant Institute of Mathematical Sciences, New York University

[3]Seismological Laboratory, California Institute of Technology

[4]Oden Institute, Jackson School of Geosciences, and Dept. of Mechanical Engineering,
The University of Texas at Austin

Outline

Statistical and Deterministic Inference

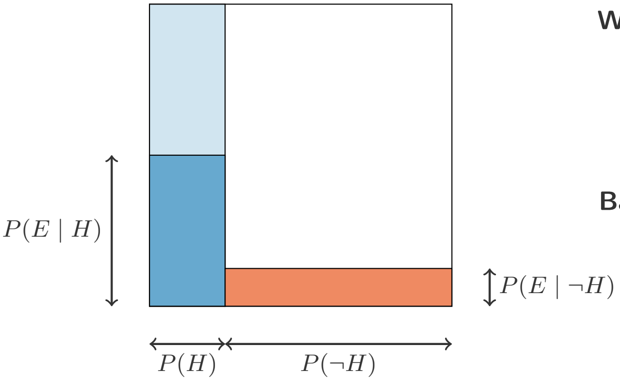
Adjoint Based Sensitivity

Earth's Mantle Convection

Simulation of Mantle Convection

Inference for Mantle Convection

Intuition for Bayes' theorem



(Credit 3Blue1Brown)

Want: Probability(Hypothesis given Evidence)

$$P(H \mid E) = \frac{P(H)P(E \mid H)}{P(H)P(E \mid H) + P(\neg H)P(E \mid \neg H)}$$

$$P(E) = P(H)P(E \mid H) + P(\neg H)P(E \mid \neg H)$$

Bayes' formula:

$$P(H \mid E) = \frac{P(H)P(E \mid H)}{P(E)}$$

Posterior $\propto$ Prior $\times$ Likelihood

Transition to a statistical inference of parameters from data

Bayes' formula for probability density (parameters m given data d):

$$\pi(m \mid d) = \frac{\pi(d \mid m)\,\pi(m)}{\pi(d)} \propto \pi(d \mid m)\,\pi(m)$$

Assuming a Gaussian noise and prior, i.e.,

$$\big(d - \mathcal{F}(m)\big) \sim \mathcal{N}(0, \mathscr{C}_{\text{noise}}) \quad \text{and} \quad m \sim \mathcal{N}(m_{\text{pr}}, \mathscr{C}_{\text{pr}})$$

where $\mathcal{F}(m)$ maps parameters to observables, Bayes' formula becomes

$$\pi(m \mid d) \propto \exp\left(-\frac{1}{2}\,\|d - \mathcal{F}(m)\|^2_{\mathscr{C}^{-1}_{\text{noise}}} - \frac{1}{2}\,\|m - m_{\text{pr}}\|^2_{\mathscr{C}^{-1}_{\text{pr}}}\right)$$

Generally, the posterior $\pi(m \mid d)$ is not Gaussian, since the dependence of the parameter-to-observable map $\mathcal{F}$ on m is not linear.

Introducing deterministic inverse problems

Forward problem: Given parameters m, find state $u(m)$ such that

$$\mathcal{A}(m,u) = f \qquad \text{(forward/model/state equation)}$$

which is well-posed with a unique solution and continuous dependence on m.

Inverse problem: Given data d, find parameters m that reduce misfit between data and state

$$d - \mathcal{B}(u(m)) \quad \text{while satisfying} \quad \mathcal{A}(m,u) = f$$

which is generally ill-posed; one remedy is adding an additional regularization term. The regularized inverse problem can be formulated as:

$$\min_m \tfrac{1}{2} \left\| d - \mathcal{B}(u(m)) \right\|^2_{\mathscr{C}_d^{-1}} + \tfrac{1}{2} \left\| m - m_0 \right\|^2_{\mathscr{C}_m^{-1}} \quad \text{subject to} \quad \mathcal{A}(m,u) = f$$

Getting to a statistical perspective on inverse problems

Recall Bayes' formula for the posterior density:

$$\pi\left(m \mid d\right) \propto \exp\left(-\frac{1}{2}\left\|d-\mathcal{F}(m)\right\|^2_{\mathscr{C}^{-1}_{\text{noise}}} - \frac{1}{2}\left\|m-m_{\text{pr}}\right\|^2_{\mathscr{C}^{-1}_{\text{pr}}}\right)$$

Finding the maximum of the posterior, the maximum a posteriori (MAP) point, amounts to

$$\begin{aligned} & \max_m \exp\left(-\frac{1}{2}\left\|d-\mathcal{F}(m)\right\|^2_{\mathscr{C}^{-1}_{\text{noise}}} - \frac{1}{2}\left\|m-m_{\text{pr}}\right\|^2_{\mathscr{C}^{-1}_{\text{pr}}}\right) \\ \Rightarrow \quad & \min_m \frac{1}{2}\left\|d-\mathcal{F}(m)\right\|^2_{\mathscr{C}^{-1}_{\text{noise}}} + \frac{1}{2}\left\|m-m_{\text{pr}}\right\|^2_{\mathscr{C}^{-1}_{\text{pr}}} \end{aligned}$$

Defining the parameter-to-observable map $\mathcal{F}$ such that it incorporates a model, we get a point estimate of the posterior density that is analogous to a deterministic inverse problem

$$\min_m \frac{1}{2}\left\|d-\mathcal{F}(u(m))\right\|^2_{\mathscr{C}^{-1}_{\text{noise}}} + \frac{1}{2}\left\|m-m_{\text{pr}}\right\|^2_{\mathscr{C}^{-1}_{\text{pr}}} \quad \text{subject to} \quad \mathcal{A}(m,u) = f$$

Inference of uncertain parameters in physical models

Goal: Ideally we want to explore the full (non-Gaussian) posterior density

$$\pi_{\text{post}}(m) \propto \exp\left(-\frac{1}{2}\left\|d - \mathcal{F}(u(m))\right\|^2_{\mathscr{C}^{-1}_{\text{noise}}} - \frac{1}{2}\left\|m - m_{\text{pr}}\right\|^2_{\mathscr{C}^{-1}_{\text{pr}}}\right) \quad \text{s.t.} \quad \mathcal{A}(m,u) = f$$

Common challenges:

- High-dimensional parameter spaces
- Quantity of data (too high or too low) and errors in data
- Data informs the parameters poorly
- Computational complexity of (physics) models makes sampling prohibitive

Inference of uncertain parameters in physical models

Goal: Ideally we want to explore the full (non-Gaussian) posterior density

$$\pi_{\text{post}}(m) \propto \exp\left(-\frac{1}{2}\left\|d - \mathcal{F}(u(m))\right\|^2_{\mathscr{C}^{-1}_{\text{noise}}} - \frac{1}{2}\left\|m - m_{\text{pr}}\right\|^2_{\mathscr{C}^{-1}_{\text{pr}}}\right) \quad \text{s.t.} \quad \mathcal{A}(m,u) = f$$

Common challenges:

- High-dimensional parameter spaces
- Quantity of data (too high or too low) and errors in data
- Data informs the parameters poorly
- Computational complexity of (physics) models makes sampling prohibitive

Computationally feasible: Find MAP point and describe local approximation of posterior around this MAP point,

$$m_{\text{MAP}} := \arg\min_{m} \frac{1}{2}\left\|d - \mathcal{F}(u(m))\right\|^2_{\mathscr{C}^{-1}_{\text{noise}}} + \frac{1}{2}\left\|m - m_{\text{pr}}\right\|^2_{\mathscr{C}^{-1}_{\text{pr}}} \quad \text{s.t.} \quad \mathcal{A}(m,u) = f$$

Outline

Statistical and Deterministic Inference

Adjoint Based Sensitivity

Earth's Mantle Convection

Simulation of Mantle Convection

Inference for Mantle Convection

Introducing the Lagrangian for the model-constrained optimization problem

Recall the optimization problem:

$$m_{\mathrm{MAP}} := \arg\min_{m} \frac{1}{2} \left\| d - \mathcal{F}(u(m)) \right\|^2_{\mathscr{C}^{-1}_{\mathrm{noise}}} + \frac{1}{2} \left\| m - m_{\mathrm{pr}} \right\|^2_{\mathscr{C}^{-1}_{\mathrm{pr}}} \quad \text{s.t.} \quad \mathcal{A}(m,u) = f$$

To derive the gradient, define a Lagrangian as the sum of the cost function and the variational form of the forward problem (Lagrange multiplier),

$$\mathcal{L}_g(m,u,v) := \frac{1}{2} \left\| d - \mathcal{F}(u(m)) \right\|^2_{\mathscr{C}^{-1}_{\mathrm{noise}}} + \frac{1}{2} \left\| m - m_{\mathrm{pr}} \right\|^2_{\mathscr{C}^{-1}_{\mathrm{pr}}} + \left[\left(\mathcal{A}(m,u)\, , v \right) - \left(f\, , v \right) \right]$$

then take variations with respect to v, u, and m and to zero:

$$\delta_v[\mathcal{L}_g]\tilde{v} \overset{!}{=} 0, \quad \delta_u[\mathcal{L}_g]\tilde{u} \overset{!}{=} 0, \quad \delta_m[\mathcal{L}_g]\tilde{m}$$

where we denote, e.g.,

$$\delta_v[\mathcal{L}_g]\tilde{v} = \delta_v[\mathcal{L}_g(m,u,v)]\tilde{v} := \lim_{\epsilon \to 0} \frac{\partial \mathcal{L}_g(m,u,v+\epsilon\tilde{v})}{\partial \epsilon}$$

Adjoint based algorithm to compute the gradient

1. Solve the (generally nonlinear) forward problem for u:

$$(\mathcal{A}(m, u)\,,\tilde{v}) = (f\,,\tilde{v}) \quad \text{for all } \tilde{v}$$

2. Solve the (linear) adjoint problem for v:

$$(\tilde{u}\,,\delta_u[\mathcal{A}]^* v) = (\delta_u[\mathcal{F}]\tilde{u}\,,d - \mathcal{F}(m, u))_{\mathscr{C}_{\text{noise}}^{-1}} \quad \text{for all } \tilde{u}$$

3. Compute the gradient

$$\mathcal{G}(\tilde{m}) = (\delta_m[\mathcal{A}]\tilde{m}\,,v) + (\tilde{m}\,,m - m_{\text{pr}})_{\mathscr{C}_{\text{pr}}^{-1}} - (\delta_m[\mathcal{F}]\tilde{m}\,,d - \mathcal{F}(m, u))_{\mathscr{C}_{\text{noise}}^{-1}}$$

Computing the gradient involves one forward and one adjoint solve, independent of the dimension of the parameters and data size, making it computationally efficient and scalable.

Outline

Statistical and Deterministic Inference

Adjoint Based Sensitivity

Earth's Mantle Convection

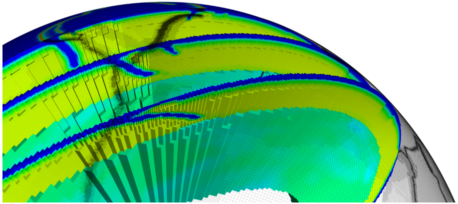
Simulation of Mantle Convection

Inference for Mantle Convection

Motivation: Understanding of Earth's fundamental properties

Simulating realistic high-resolution models:

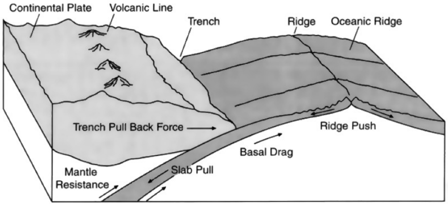
- Global mantle convection & associated plate tectonics
- Resolutions to faulted plate boundaries
- Nonlinear viscous fluid flow



Computational mesh that resolves viscosity (colors).

Learning properties of Earth's rheology:

- What we know: Observational data of plate motions at the surface, topography, surface strain rates, estimates of average viscosities, etc.
- What we want: Parameters in constitutive laws and plate coupling strength that are consistent with data and model; Uncertainties w.r.t. available data



Underlying forces (Credit: Schubert, Turcotte, Olsen).

Outline

Statistical and Deterministic Inference

Adjoint Based Sensitivity

Earth's Mantle Convection

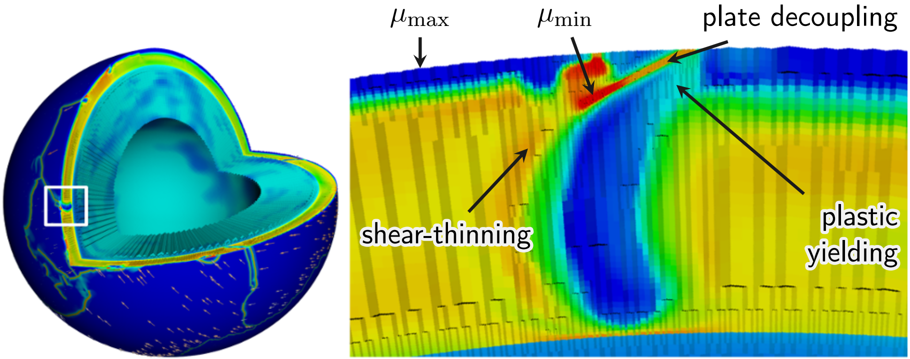
Simulation of Mantle Convection

Inference for Mantle Convection

Approach for simulation

Model PDE: present day instantaneous flow modeled by incompressible, nonlinear Stokes

Rheology / effective viscosity: shear-thinning with plastic yielding, and upper/lower bounds



Colors represent viscosity ($\mu_{\max}/\mu_{\min} = 10^6$).

Approach for simulation

Model PDE: present day instantaneous flow modeled by incompressible, nonlinear Stokes

$$-\nabla \cdot \left[\mu(\boldsymbol{x}, \dot{\varepsilon}_{\mathrm{II}}(\boldsymbol{u}))\,(\nabla \boldsymbol{u} + \nabla \boldsymbol{u}^{\mathsf{T}})\right] + \nabla p = \boldsymbol{f} \qquad \text{viscosity } \mu, \text{ RHS forcing } \boldsymbol{f}$$

$$-\nabla \cdot \boldsymbol{u} = 0 \qquad \text{velocity } \boldsymbol{u}, \text{ pressure } p$$

- Free-slip boundary conditions (tangential velocities at the surface are output of the model)
- Temperature structure from seismic models and convergence rate of slabs
- Fault weak zones based on *Slab1.0* compilation of models for subduction zones

Rheology / effective viscosity: shear-thinning with plastic yielding, and upper/lower bounds

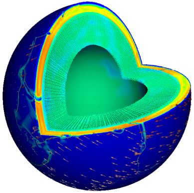
$$\mu(\boldsymbol{x}, \dot{\varepsilon}_{\mathrm{II}}(\boldsymbol{u})) := \mu_{\min} + \min\left(\frac{\tau_{\text{yield}}}{2\dot{\varepsilon}_{\mathrm{II}}(\boldsymbol{u})}\,,\, w(\boldsymbol{x}) \min\left(\mu_{\max}\,,\, a(T(\boldsymbol{x}))^{\frac{1}{n}}\,\dot{\varepsilon}_{\mathrm{II}}(\boldsymbol{u})^{\frac{1}{n}-1}\right)\right)$$

- Plates are modeled by high-viscosity: $a(T(\boldsymbol{x}))$ depends exponentially on T
- Plate boundaries are narrow zones of weak viscosity: $10^{-6} \le w(\boldsymbol{x}) \le 1$

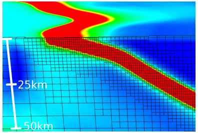
Severe challenges for parallel scalable implicit solvers

... arising from global models of Earth's mantle convection:

- Severe nonlinearity, heterogeneity, and anisotropy (upon linearization)
- Sharp viscosity gradients in narrow regions (change by $\mathcal{O}(10^6)$ in $\sim$5 km)
- Wide range of spatial scales and highly localized features, e.g., plate boundaries of size $\mathcal{O}(1\,\text{km})$ influence plate motion at continental scales of $\mathcal{O}(1000\,\text{km})$
- Adaptive refinement of computational mesh is essential
- High-order finite elements $\mathbb{Q}_k \times \mathbb{P}^{\text{disc}}_{k-1}$, order $k \geq 2$, with local mass conservation; yields a difficult to deal with discontinuous, modal pressure approximation



Colors represent viscosity, arrows show plate velocities.



Fault zone (red) between stationary and subducting plate.

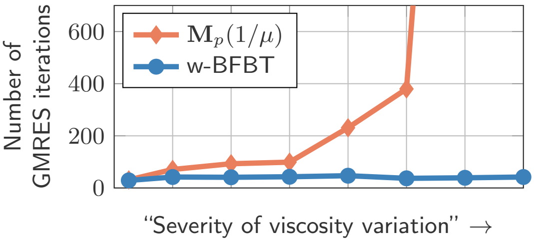
A resolved model has about 1 billion DOFs (with aggressive adaptivity), because smallest mesh elements $\sim$1 km; Runs on 10K–100K computing cores.

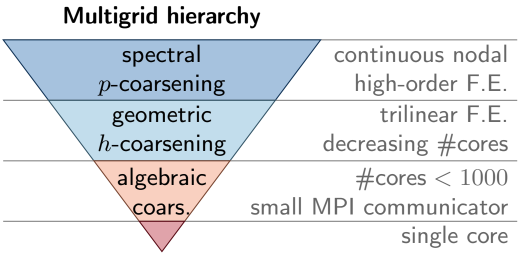
Solver highlights: Viscosity variation-robust preconditioning and scalable parallel multigrid for adaptive high-order F.E.

Stokes preconditioning: $\begin{bmatrix} \mathbf{A} & \mathbf{B}^\top \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{A}} & \mathbf{B}^\top \\ \mathbf{0} & \tilde{\mathbf{S}} \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\hat{\mathbf{u}}} \\ \tilde{\hat{\mathbf{p}}} \end{bmatrix} = \begin{bmatrix} -\mathbf{r}_1 \\ -\mathbf{r}_2 \end{bmatrix}$ $\quad \tilde{\mathbf{A}}^{-1} \approx \mathbf{A}^{-1} \rightarrow$ MG V-cycle, $\quad \tilde{\mathbf{S}}^{-1} \approx (\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^\top)^{-1}$

Schur complement preconditioning: $\tilde{\mathbf{S}}^{-1} := \underbrace{\left(\mathbf{B}\mathbf{C}_w^{-1}\mathbf{B}^\top\right)^{-1}}_{\rightarrow \text{ MG V-cycle}} \left(\mathbf{B}\mathbf{C}_w^{-1}\mathbf{A}\mathbf{D}_w^{-1}\mathbf{B}^\top\right) \underbrace{\left(\mathbf{B}\mathbf{D}_w^{-1}\mathbf{B}^\top\right)^{-1}}_{\rightarrow \text{ MG V-cycle}}$

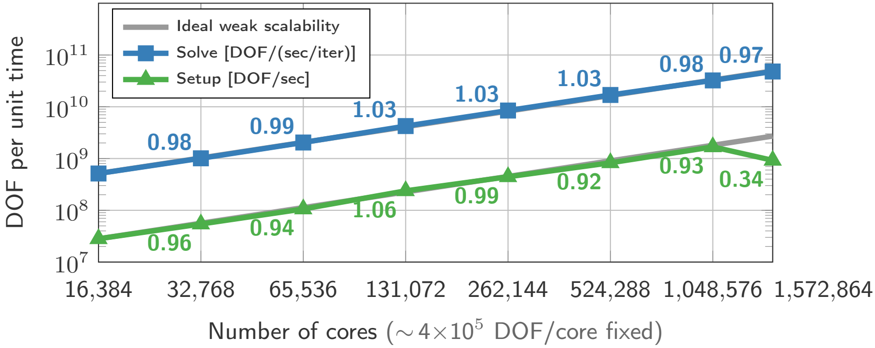
Choice of diagonal weighting matrices $\mathbf{C}_w = \mathbf{D}_w$ is critical for efficacy & robustness.





Solver highlights: Parallel weak scalability on Sequoia (BlueGene/Q)

Parallel scalability of linear iterative solvers to 1.6M cores (collaboration with IBM Research, Zurich):



- Excellent parallel scalability while maintaining optimal algorithmic performance w.r.t. mesh refinement; nearly optimal w.r.t. higher discretization order
- Full nonlinear Stokes solver: Inexact Newton–Krylov with nonlinear preconditioning and grid continuation for highly nonlinear models of Earth's mantle

Outline

Approach for inference: From data to parameters consistent with model

General procedure[1] for inference from data consistent with a model (physical, statistical, etc.):

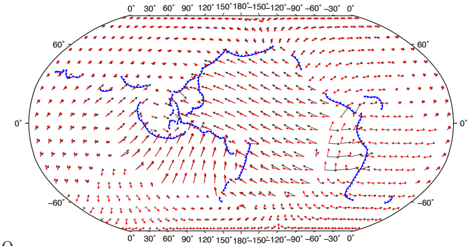
1. **Question**: Start with a scientific question, e.g., quantify forces in deep Earth.
2. **Data collection**: Gather data relevant to the question, e.g., plate motions at the surface.
3. **Model selection**: Construct a model that is *appropriate* to the problem and data, moreover, select parameters for inference, e.g., global mantle convection model with plate coupling strength as parameters.
4. **Parameter estimation**: Find the parameters in the model, which often requires minimizing the difference between model output and data (in some suitable metric).
5. **Validation**: Inspect if the model and parameters are consistent with knowledge or data that was excluded from estimation, e.g., average viscosity or topography.
6. **Prediction**: Estimate the quantities of interest and their uncertainties that are associated with the initial scientific question.

[1] Credit J. Bessac

Approach for inference: Global mantle convection & plate tectonics

Data: Challenging because of limited amount

- Current plate motion of rigid plates from GPS and magnetic anomalies recorded in bands perpendicular to seafloor spreading



Data (black) vs. model (red) (Credit: J. Hu).

Model: Challenging because of computational complexity

- Incompressible, nonlinear Stokes PDE:

$$-\nabla \cdot \left[\mu(\boldsymbol{x}, \dot{\varepsilon}_{\mathrm{II}}(\boldsymbol{u}))\,(\nabla \boldsymbol{u} + \nabla \boldsymbol{u}^{\mathsf{T}})\right] + \nabla p = \boldsymbol{f}, \quad -\nabla \cdot \boldsymbol{u} = 0$$

- Effective viscosity: $\mu(\boldsymbol{x}, \dot{\varepsilon}_{\mathrm{II}}(\boldsymbol{u})) := \mu_{\min} + \min\left(\frac{\tau_{\text{yield}}}{2\dot{\varepsilon}_{\mathrm{II}}(\boldsymbol{u})}, w(\boldsymbol{x}) \min\left(\mu_{\max}, a(T(\boldsymbol{x}))^{\frac{1}{n}} \dot{\varepsilon}_{\mathrm{II}}(\boldsymbol{u})^{\frac{1}{n}-1}\right)\right)$

Parameters: Challenging because of vastly different scales of sensitivity

- Global parameters: scaling factors, activation energy in Arrhenius law $a(T(\boldsymbol{x}))$, stress exponent n, yield strength τ_{yield}
- Local coupling strength $w(\boldsymbol{x})$, i.e., energy dissipation between adjacent plates

Systematic inversion of uncertain parameters consistent with data & model

Given:

- Model PDE (forward problem), here incompressible, nonlinear Stokes: $\mathcal{A}(m, u) = f$
- Map of model output (dependent on parameters m) to observations: $\mathcal{F}(u(m))$
- Assume data d contains normally distributed additive noise, $\mathcal{N}(0, \mathscr{C}_{\text{noise}})$
- Assume prior of the parameters m is normally distributed, $\mathcal{N}(m_{\text{pr}}, \mathscr{C}_{\text{pr}})$

Want: Description of the posterior density of the parameters (using Bayes' theorem)

$$\pi_{\text{post}}(m) \propto \exp\left(-\frac{1}{2}\,\|d - \mathcal{F}(u(m))\|^2_{\mathscr{C}^{-1}_{\text{noise}}} - \frac{1}{2}\,\|m - m_{\text{pr}}\|^2_{\mathscr{C}^{-1}_{\text{pr}}}\right)$$

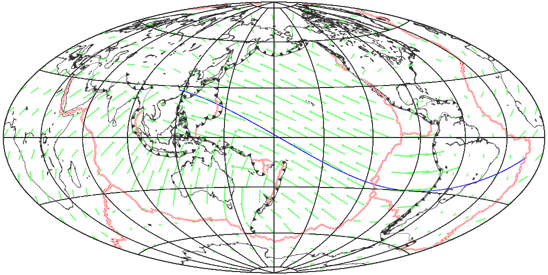
Computationally feasible: Using adjoints for gradient computation & BFGS approximation of Hessian

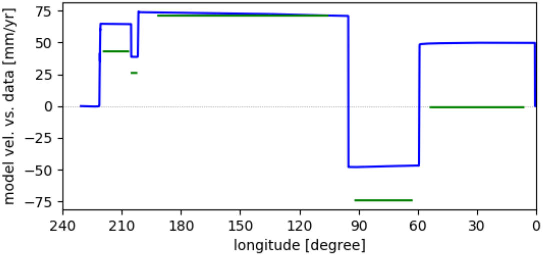
- Find the maximum of $\pi_{\text{post}}(m)$ by solving an optimization problem constraint by the model PDE:

$$\arg\min_m \mathcal{J}(m, u(m)) := \frac{1}{2}\,\|d - \mathcal{F}(u(m))\|^2_{\mathscr{C}^{-1}_{\text{noise}}} + \frac{1}{2}\,\|m - m_{\text{pr}}\|^2_{\mathscr{C}^{-1}_{\text{pr}}} \quad \text{subject to} \quad \mathcal{A}(m, u) = f$$

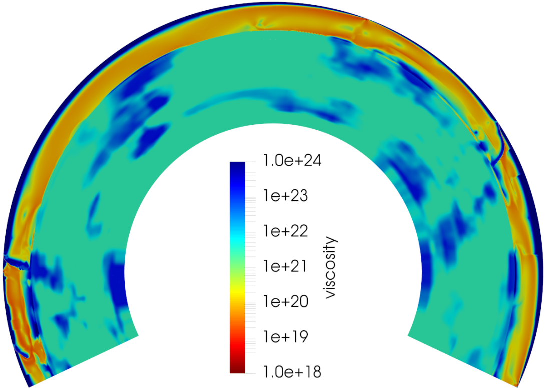
- Construct a Gaussian approximation of $\pi_{\text{post}}(m)$ around this maximum by approximating the Hessian of the optimization problem

Computational results: Inverse problem for a cross sectional model



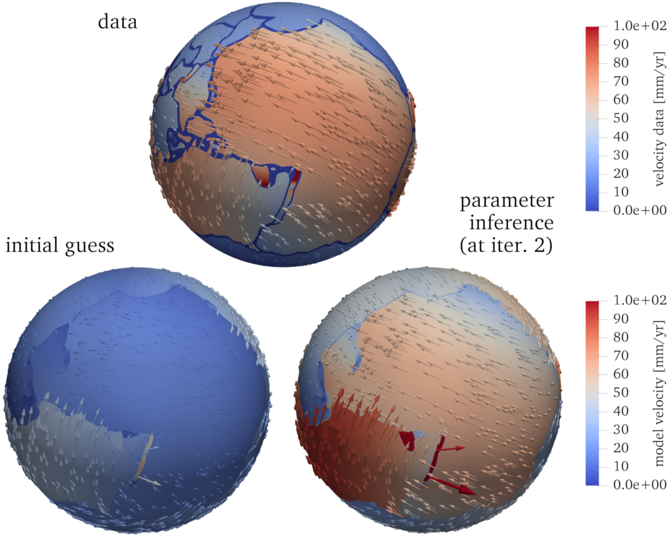


Effective viscosity after inference



- Cross sectional model is not able to adequately capture observed physics, i.e., fast motion of stationary South American plate

Computational results: Inverse problem for global Earth model



- Inversion of global scalars (scaling factors, activation energy, stress exponent, yield strength) successfully recovers main trends, i.e., velocities of large plates
- Local parameters (plate decoupling) remain challenging due to lower sensitivities

Thank you