Scalable Inference of Non-Newtonian Rheology Parameters in Earth's Mantle on HPC Platforms

Johann Rudi[1] Georg Stadler[2] Jiashun Hu[3] Micheal Gurnis[4]

[1]Department of Mathematics, Virginia Tech

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

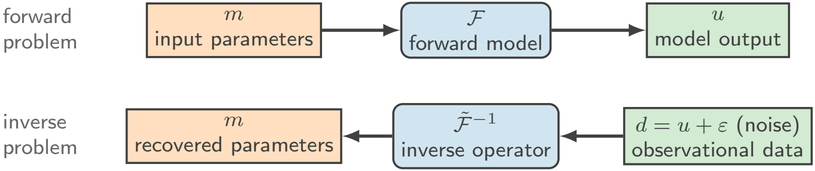
[3]Department of Earth and Space Sciences,
Southern University of Science and Technology (China)

[4]Seismological Laboratory at the Division of Geological and Planetary Sciences,
California Institute of Technology

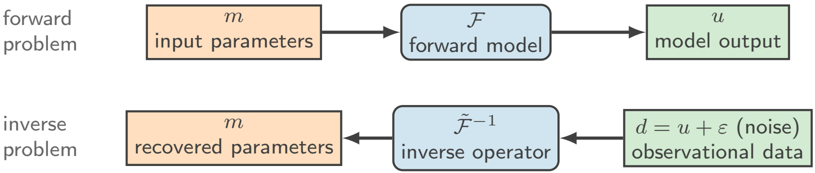
Pronouns: he/him/his/himself

Outline

Introducing forward & inverse problems



Introducing forward & inverse problems



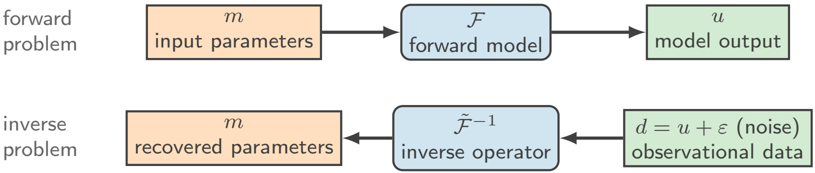
Three different ways to compute solutions of inverse problems

Sampling-based methods use randomness to explore the posterior density; typically don't need derivatives
$\rightarrow$ Markov chain Monte Carlo (MCMC)

Adjoint derivative-based methods use techniques from optimization thus require gradients/Hessians
$\rightarrow$ PDE-constrained optimization

Deep learning-based methods directly construct an inverse map, $\tilde{\mathcal{F}}^{-1}$, from data to parameters (or even posterior densities)
$\rightarrow$ artificial neural networks

Introducing forward & inverse problems



Three different ways to compute solutions of inverse problems

Sampling-based methods use randomness to explore the posterior density; typically don't need derivatives
$\rightarrow$ Markov chain Monte Carlo (MCMC)

Adjoint derivative-based methods use techniques from optimization thus require gradients/Hessians
$\rightarrow$ PDE-constrained optimization

Deep learning-based methods directly construct an inverse map, $\tilde{\mathcal{F}}^{-1}$, from data to parameters (or even posterior densities)
$\rightarrow$ artificial neural networks

Outline

Nonlinear Stokes PDE modeling Earth's mantle

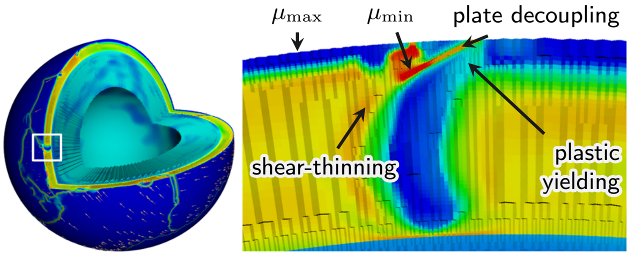
Model: Nonlinear incompressible Stokes (w/ free-slip & no-normal flow BC) models present-day instantaneous flow

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

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

Rheology / effective viscosity: Shear-thinning with plastic yielding and plate decoupling (or weakening) factor $w(\boldsymbol{x})$

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



Visualization from *rhea* code. Colors represent viscosity; Widths of plate decoupling are exaggerated.

Given: Observational data

- Current plate motion from GPS and magnetic anomalies
- Topography indicating normal traction at Earth's surface
- Plate deformation obtained from dense GPS networks
- Average viscosity in regions affected by post-glacial rebound

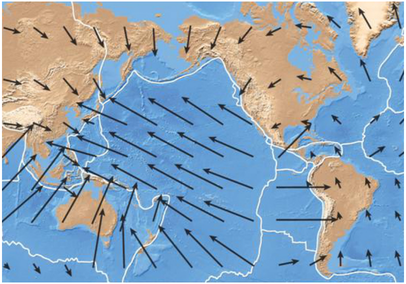


Plate motion (Credit: Pearson Prentice Hall)

Additional knowledge contributing to mantle flow models:

- Location and geometry of plates, plate boundaries, and subducting slabs (from seismicity)
- Images of present-day Earth structure (by correlating seismic wave speed with temperature)
- Rock rheology extrapolated from laboratory experiments

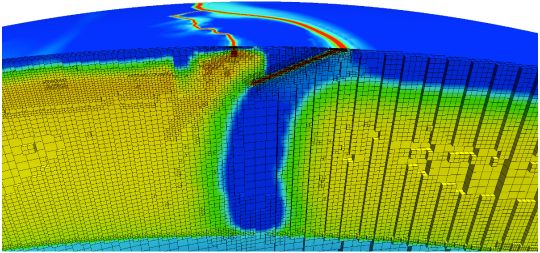
Want: Constrain parameters of mantle models

Global rheological parameters affecting viscosity and nonlinearity:

- Scaling factor of the upper mantle viscosity (down to ~ 660 km depth)
- Stress exponent controlling severity of strain rate weakening
- Yield strength governing plastic yielding phenomena

Spatially varying parameters modeling geometry of plate boundaries:

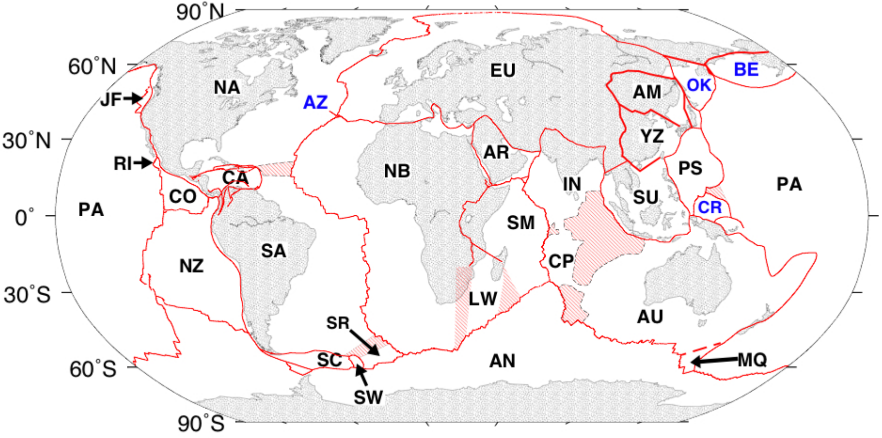
- Coupling strength / energy dissipation between plates



(Credit: L. Alisic)

Model of Earth's plate boundaries

Plate boundaries at Earth's surface *(red lines)* and plate geometries obained from MORVEL plate motion data set[1]

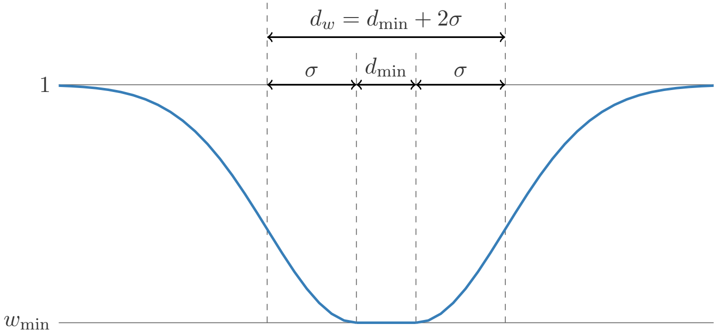


[1] DeMets, Gordon, and Argus 2010.

Parametric model for plate decoupling / weak zones

$$w(\boldsymbol{x}) \coloneqq 1 - (1 - w_{\min}) \exp\left(-\frac{\xi(\boldsymbol{x})^2}{2\sigma^2}\right) \in (0, 1]$$

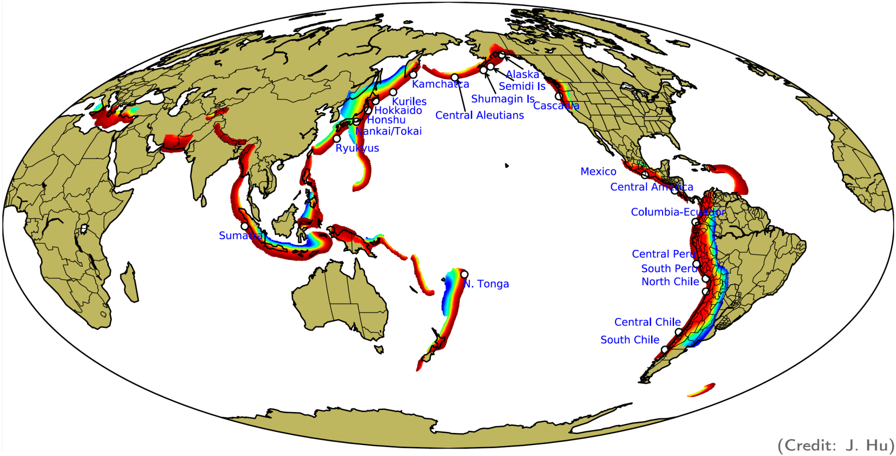
$$\xi(\boldsymbol{x}) \coloneqq \max(0, d(\boldsymbol{x}) - d_{\min}) \quad \text{and} \quad \sigma \coloneqq \frac{d_w - d_{\min}}{2}$$



Weak zone profile with width $d_w = 20\,\text{km}$, plate boundary width $d_{\min} = 5\,\text{km}$, and weak zone factor $w_{\min} = 10^{-5}$.

Geometries of decoupling surfaces of subducting plates

Surfaces $d(\boldsymbol{x})$ of the subducting plates, where *colors* indicate depth *(red is shallow and blue is deep, respectively)*



(Credit: J. Hu)

Forward solver for the nonlinear Stokes PDE

Nonlinear incompressible Stokes PDE

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

- Inexact Newton–Krylov method with grid continuation

Linearization with Newton's method, then discretization yields

$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^{\mathsf{T}} \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{u}} \\ \hat{\mathbf{p}} \end{bmatrix} = \begin{bmatrix} -\mathbf{r}_1 \\ -\mathbf{r}_2 \end{bmatrix}$$

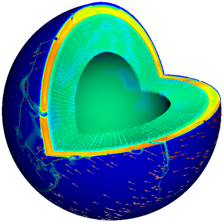
Careful design of discretization with inf-sup stable Finite Elements

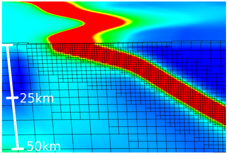
- High-order finite element shape functions ($\mathbb{Q}_k \times \mathbb{P}_{k-1}^{\text{disc}}$, $k \geq 2$)
- Locally mass conservative due to discontinuous, modal pressure
- Non-conforming hexahedral meshes with "hanging nodes"
- Adaptive mesh refinement (AMR) resolving fine-scale features of mantle

Severe challenges for parallel scalable implicit solvers

. . . arising in global mantle convection:

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





Effective viscosity *(colors)* and locally refined mesh.

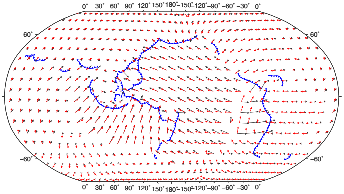
$\rightarrow$ Developing scalable non-linear & linear solvers and preconditioners took several years: Rudi et al. (2015), Rudi, Stadler, and Ghattas (2017), Rudi, Shih, and Stadler (2020).

Outline

Challenges of inferring parameters from observations

Data: Challenging because of limited amount

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

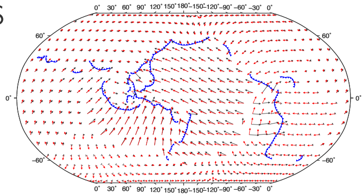


Data (black) vs. model (red) (Credit: J. Hu).

Challenges of inferring parameters from observations

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 and truncation/inexact solves

- Incompressible, nonlinear Stokes PDE:

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

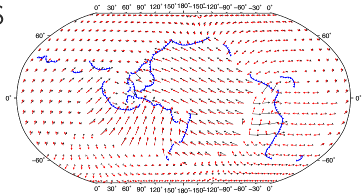
- Effective viscosity:

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

Challenges of inferring parameters from observations

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 and truncation/inexact solves

- Incompressible, nonlinear Stokes PDE:

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

- Effective viscosity:

$$\mu(\boldsymbol{x},\dot\varepsilon_{\rm II}(\boldsymbol{u})) \coloneqq \mu_{\min} + \min\left(\frac{\tau_{\rm yield}}{2\dot\varepsilon_{\rm II}(\boldsymbol{u})}\,,\, w(\boldsymbol{x})\min\left(\mu_{\max}\,,\, a(T(\boldsymbol{x}))^{\frac{1}{n}}\,\dot\varepsilon_{\rm II}(\boldsymbol{u})^{\frac{1}{n}-1}\right)\right)$$

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 $\tau_{\rm yield}$
- Local coupling strength $w_{\min}$ (i.e., energy dissipation between plates)

Formulate inverse problem in a Bayesian setting

Given: (for simplicity u now combines velocity and pressure)

- Model PDE (forward problem): $\mathcal{A}(m, u) = f$ (here nonlinear Stokes PDE)
- 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')

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

Formulate inverse problem in a Bayesian setting

Given: (for simplicity u now combines velocity and pressure)

- Model PDE (forward problem): $\mathcal{A}(m,u) = f$ (here nonlinear Stokes PDE)
- 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')

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

Computationally feasible (and thanks to Laplace approximation)

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

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

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

Newton (outer loop): Adjoints for 1st & 2nd-order derivatives

Derivation of gradient equations, using a Lagrangian as Ansatz,

1. Solve the (nonlinear) forward problem for u: (u combines velocity and pressure for simpler notation)

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

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 with respect to parameters m:

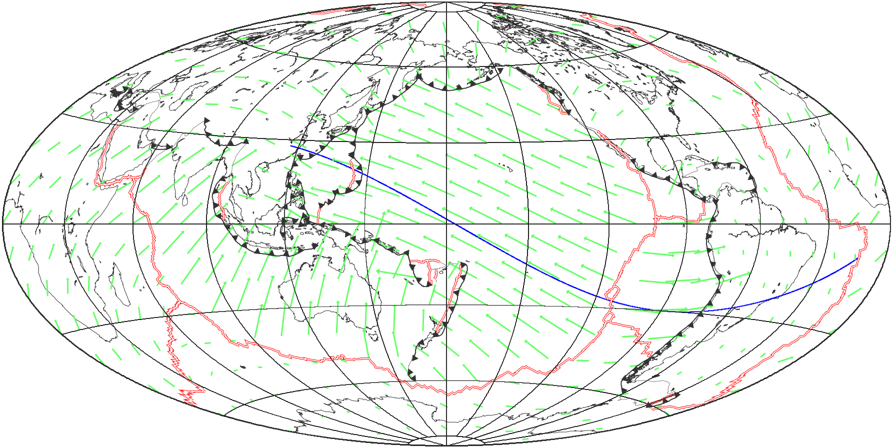
$$\mathcal{G}(\tilde{m}) = (\delta_m[\mathcal{A}](\tilde{m})\,, v) + (\tilde{m}\,, m - m_{\text{pr}})_{\mathscr{C}_{\text{pr}}^{-1}} - (\delta_m[\mathcal{F}](\tilde{m})\,, d - \mathcal{F}(m, u))_{\mathscr{C}_{\text{noise}}^{-1}}$$

Computational complexity

- One (nonlinear) forward + one (linear) adjoint solve
- Independent of the dimension of parameters & data size $\rightarrow$ scalable
- Analogous approach is used to compute Hessians

Outline

Setup: Cross section of Earth's mantle[2]



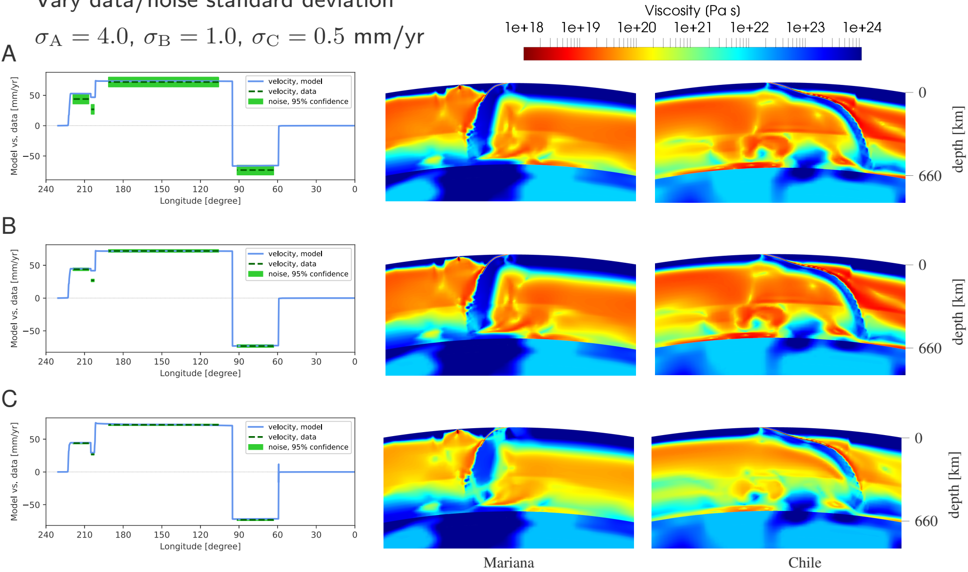
Cross section *(blue line)*; velocity vectors from MORVEL56 *(green arrows)*. (Credit: M. Gurnis)

[2] Rudi, Gurnis, and Stadler (2022), In: Geophysical Journal International.

Inference of plate decoupling for a cross section of Earth

Vary data/noise standard deviation

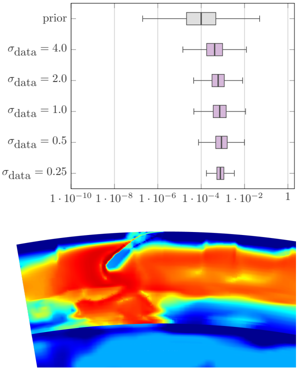
$\sigma_{\mathrm{A}} = 4.0,\ \sigma_{\mathrm{B}} = 1.0,\ \sigma_{\mathrm{C}} = 0.5$ mm/yr



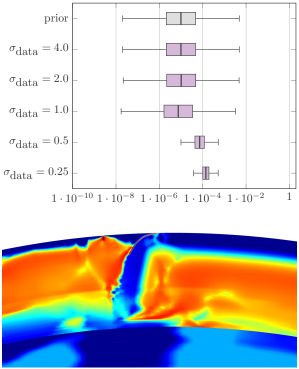
Inference of plate decoupling for a cross section of Earth

Sensitivities of the response of inferred plate coupling factors $w_{\min}$ to a (prescribed) accuracy between data and model outputs.

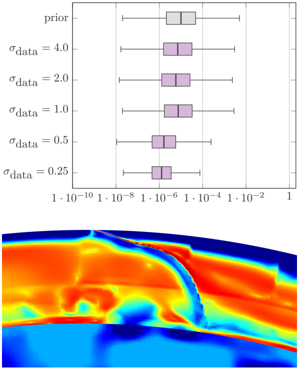
Prior and posterior distributions for $w_{\min}$



Ryuku weak zone



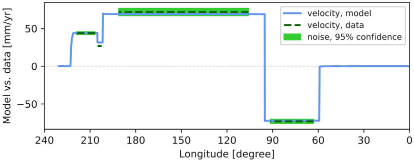
Mariana weak zone

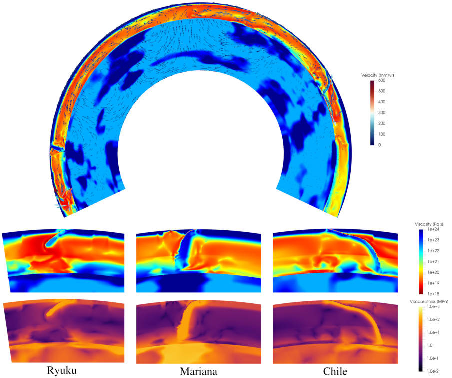


Chile weak zone

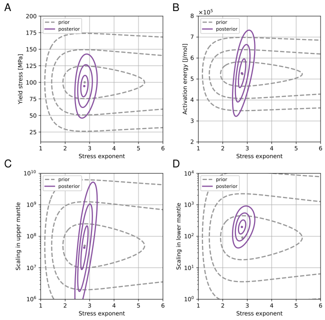
Inference for a cross section with plate-dependent weights

Plate size-dependent standard deviation

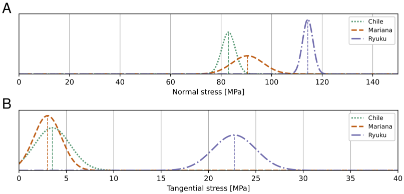




2D marginals of the (approx.) posterior



Forward sensitivities of stresses (QOIs)



Towards full-sphere inference | current & future work

Data misfit at initial guess of plate coupling parameters $w_{\min}$ at subduction faults and spatially-constant prefactors

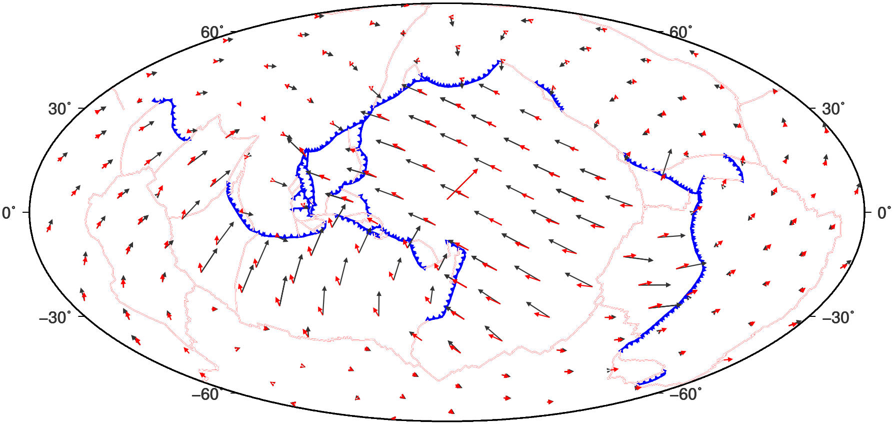


Plate motion from MORVEL65 *(black arrows)* and from model outputs *(red arrows)*. (Credit: J. Hu)

Towards full-sphere inference | current & future work

Data misfit toward optimality point (iteration 9)

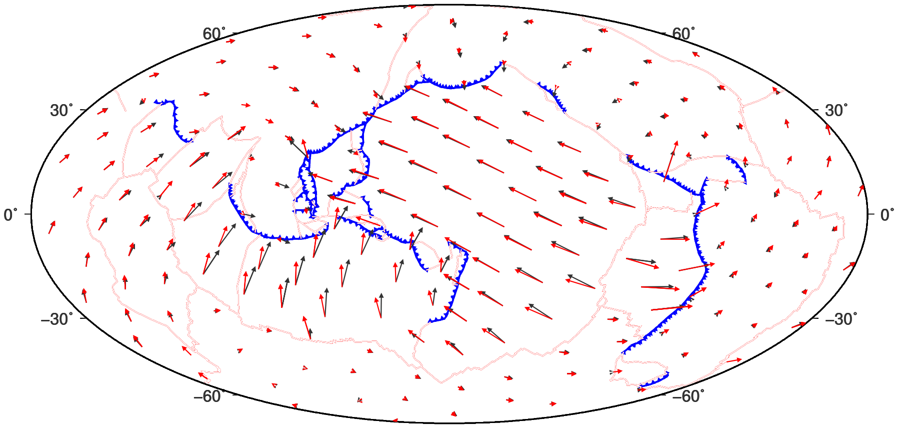
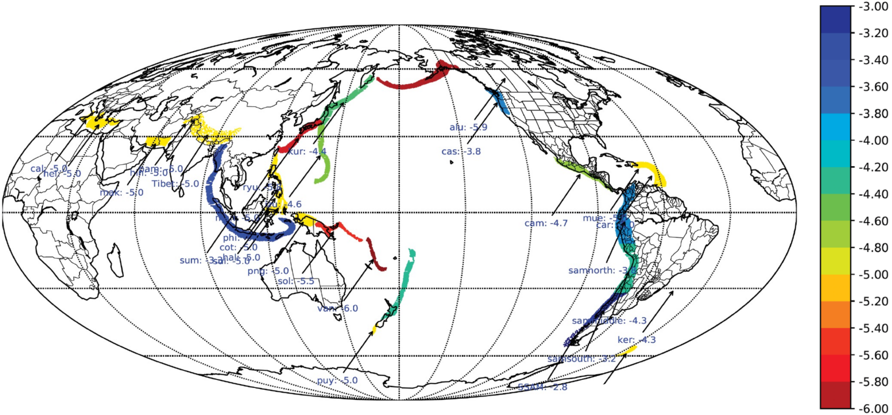


Plate motion from MORVEL65 *(black arrows)* and from model outputs *(red arrows)*. (Credit: J. Hu)

Towards full-sphere inference | current & future work

Approximately inferred plate coupling factors w_{min}



Strong variation of weak zone factors w_{min} *(colors in $\log_{10}$-scale)*. (Credit: J. Hu)

References I

DeMets, Charles, Richard G. Gordon, and Donald F. Argus (2010). "Geologically current plate motions." In: *Geophysical Journal International* 181.1, pp. 1–80.

Rudi, Johann, Michael Gurnis, and Georg Stadler (2022). "Simultaneous Inference of Plate Boundary Stresses and Mantle Rheology Using Adjoints: Large-Scale 2-D Models." In: *Geophysical Journal International* 231.1, pp. 597–614.

Rudi, Johann, A. Cristiano I. Malossi, Tobin Isaac, Georg Stadler, Michael Gurnis, Peter W. J. Staar, Yves Ineichen, Costas Bekas, Alessandro Curioni, and Omar Ghattas (2015). "An Extreme-Scale Implicit Solver for Complex PDEs: Highly Heterogeneous Flow in Earth's Mantle." In: *SC15: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*. ACM, 5:1–5:12.

References II

Rudi, Johann, Yu-Hsuan Shih, and Georg Stadler (2020). "Advanced Newton Methods for Geodynamical Models of Stokes Flow with Viscoplastic Rheologies." In: *Geochemistry, Geophysics, Geosystems* 21.9.

Rudi, Johann, Georg Stadler, and Omar Ghattas (2017). "Weighted BFBT Preconditioner for Stokes Flow Problems with Highly Heterogeneous Viscosity." In: *SIAM Journal on Scientific Computing* 39.5, S272–S297.