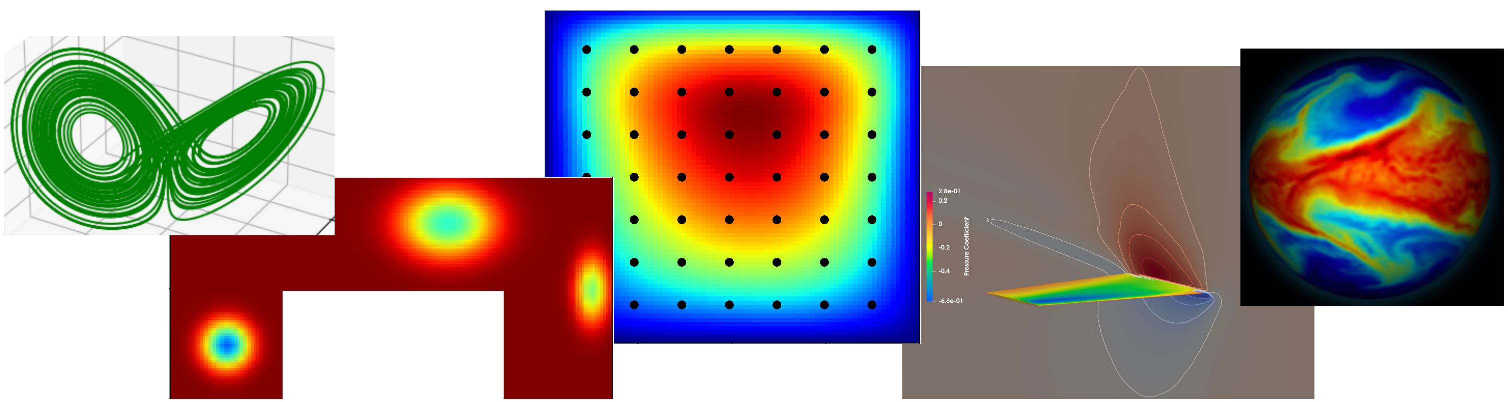
Unscented Kalman Inversion

Daniel Zhengyu Huang, Tapio Schneider, Andrew M. Stuart

California Institute of Technology

EnKF Workshop 2021



Code: https://github.com/Zhengyu-Huang/InverseProblems.jl

INVERSE PROBLEM

Caltech

- Inverse Problem

$$y = \mathcal{G}(\theta) + \eta \qquad \eta \sim \mathcal{N}(0, \Sigma_\eta)$$

- Optimization approach

$$\Phi\,(\theta; y) = \frac{1}{2} \parallel \Sigma_\eta^{-\frac{1}{2}} (y - \mathcal{G}(\theta) \parallel^2$$

$$\Phi_R(\theta; y) = \Phi\,(\theta; y) + \frac{1}{2} \parallel \Sigma_0^{-\frac{1}{2}} (\theta - r_0) \parallel^2$$

- Bayesian (probabilistic) approach

$$\mu(d\theta) \propto \exp\big(-\Phi\,(\theta; y)\big)\ \mu_0(d\theta)$$

KALMAN INVERSION

Caltech

- Kalman filtering (real time n)

evolution: $x_{n+1} = \mathcal{F}(x_n) + \omega_{n+1}$ $\omega_{n+1} \sim \mathcal{N}(0, \Sigma_\omega)$

observation: $y_{n+1} = \mathcal{G}(x_{n+1}) + \nu_{n+1}$ $\nu_{n+1} \sim \mathcal{N}(0, \Sigma_\nu)$

- Kalman inversion (artificial time n)

evolution: $\theta_{n+1} = r + \alpha(\theta_n - r) + \omega_{n+1}$ $\omega_{n+1} \sim \mathcal{N}(0, \Sigma_\omega)$

observation: $y_{n+1} = \mathcal{G}(\theta_{n+1}) + \nu_{n+1}$ $\nu_{n+1} \sim \mathcal{N}(0, \Sigma_\nu)$

linear “identity” dynamics

repeated observation $y_{n+1} = y$

free parameters $\alpha \in (0,1],\ r,\ \Sigma_\omega, \Sigma_\nu$

GAUSSIAN APRROXIMATION ALGORITHM

Caltech

- Kalman inversion

Let denote $Y_n = \{y_1, y_2, \cdots y_n\}$ and approximate $\theta_n | Y_n \sim \mathcal{N}(m_n, C_n)$

Hope that conditional distribution of $\theta_n | Y_n \rightarrow \mu(d\theta)$

- Prediction analysis procedure

Prediction: $\theta_n | Y_n \rightarrow \theta_{n+1} | Y_n \sim \mathcal{N}(\widehat{m}_{n+1}, \hat{C}_{n+1})$

$$\widehat{m}_{n+1} = r + \alpha(m_n - r) \qquad \hat{C}_{n+1} = \alpha^2 C_n + \Sigma_\omega$$

GAUSSIAN APRROXIMATION ALGORITHM

Caltech

- Prediction analysis procedure

Analysis: $\theta_{n+1}|Y_n \rightarrow \{\theta_{n+1}, y_{n+1}\}|Y_n \rightarrow \theta_{n+1}|Y_{n+1} \sim \mathcal{N}(m_{n+1}, C_{n+1})$

$$\{\theta_{n+1}, y_{n+1}\} \mid Y_n \sim \left(\begin{bmatrix} \widehat{m}_{n+1} \\ \hat{y}_{n+1} \end{bmatrix}, \begin{bmatrix} \hat{C}_{n+1} & \hat{C}_{n+1}^{\theta p} \\ \hat{C}_{n+1}^{\theta p^T} & \hat{C}_{n+1}^{pp} \end{bmatrix} \right)$$

$$\hat{y}_{n+1} = \mathbb{E}[\mathcal{G}(\theta_{n+1})|Y_n]$$

$$\hat{C}_{n+1}^{\theta p} = \mathrm{Cov}[\theta_{n+1}, \mathcal{G}(\theta_{n+1})|Y_n]$$

$$\hat{C}_{n+1}^{pp} = \mathrm{Cov}[\mathcal{G}(\theta_{n+1})|Y_n] + \Sigma_\nu$$

$$m_{n+1} = \widehat{m}_{n+1} + \hat{C}_{n+1}^{\theta p}(\hat{C}_{n+1}^{pp})^{-1}(y_{n+1} - \hat{y}_{n+1})$$

$$C_{n+1} = \hat{C}_{n+1} - \hat{C}_{n+1}^{\theta p}(\hat{C}_{n+1}^{pp})^{-1}\hat{C}_{n+1}^{\theta p^T}$$

GAUSSIAN APRROXIMATION ALGORITHM

Caltech

- Kalman inversion

Proposition (Affine invariance)

We may write the algorithm as

$$(m_{n+1}, C_{n+1}) = F(m_n, C_n;\ \mathcal{G}, r, \Sigma_\omega)$$

Consider any invertible affine mapping $x^* = Ax + b$, and define

$$m_n^* = Am_n + b \qquad r^* = Ar + b$$

$$\mathcal{G}^*(\theta) = \mathcal{G}\big(A^{-1}(\theta - b)\big) \qquad \Sigma_\omega^* = A\Sigma_\omega A^{\mathrm{T}} \qquad C_n^* = AC_nA^{\mathrm{T}}$$

The algorithm is invariant and satisfies

$$(m_{n+1}^*, C_{n+1}^*) = F(m_n^*, C_n^*; \mathcal{G}^*, r^*, \Sigma_\omega^*)$$

KALMAN FILTERS

Caltech

- Extended Kalman filter (linearization)

$$\mathcal{G}(\theta_{n+1}) \approx \mathcal{G}(\widehat{m}_{n+1}) + d\mathcal{G}(\widehat{m}_{n+1})(\theta_{n+1} - \widehat{m}_{n+1})$$

$$\hat{y}_{n+1} = \mathcal{G}(\widehat{m}_{n+1})$$

$$\hat{C}_{n+1}^{\theta p} = \hat{C}_{n+1} d\mathcal{G}(\widehat{m}_{n+1})^T$$

$$\hat{C}_{n+1}^{pp} = d\mathcal{G}(\widehat{m}_{n+1})\hat{C}_{n+1} d\mathcal{G}(\widehat{m}_{n+1})^T + \Sigma_\nu$$

- Ensemble Kalman filter (Monte Carlo sampling)

$$\hat{y}_{n+1} = \frac{1}{J}\sum_{j=1}^{J} y_{n+1}^j \qquad y_{n+1}^j = \mathcal{G}(\hat{\theta}_{n+1}^j)$$

$$\hat{C}_{n+1}^{\theta p} = \frac{1}{J-1}\sum_{j=1}^{J}(\hat{\theta}_{n+1}^j - \widehat{m}_{n+1})(y_{n+1}^j - \hat{y}_{n+1})^T$$

$$\hat{C}_{n+1}^{pp} = \frac{1}{J-1}\sum_{j=1}^{J}(y_{n+1}^j - \hat{y}_{n+1})(y_{n+1}^j - \hat{y}_{n+1})^T + \Sigma_\nu$$

KALMAN FILTERS

Caltech

- Unscented Kalman filter (quadrature rule)

Definition (Modified unscented transform)

Let denote Gaussian random variable $\theta \sim \mathcal{N}(m, C) \in R^{N_\theta}$, $2N_\theta + 1$ sigma points are chosen deterministically

$$\theta^0 = m \quad \theta^j = m + c_j[\sqrt{C}]_j \quad \theta^j = m - c_j[\sqrt{C}]_j \quad (1 \leq j \leq N_\theta)$$

where $[\sqrt{C}]_j$ is the jth column of the Cholesky factor of C. The approximations are

$$\mathbb{E}[\mathcal{G}_i(\theta)] \approx \mathcal{G}_i(\theta^0)$$

$$\mathrm{Cov}[\mathcal{G}_1(\theta), \mathcal{G}_2(\theta)] \approx \sum_{j=1}^{2N_\theta} W_j^c \left(\mathcal{G}_1(\theta^j) - \mathbb{E}[\mathcal{G}_1(\theta)]\right)\left(\mathcal{G}_2(\theta^j) - \mathbb{E}[\mathcal{G}_2(\theta)]\right)^T$$

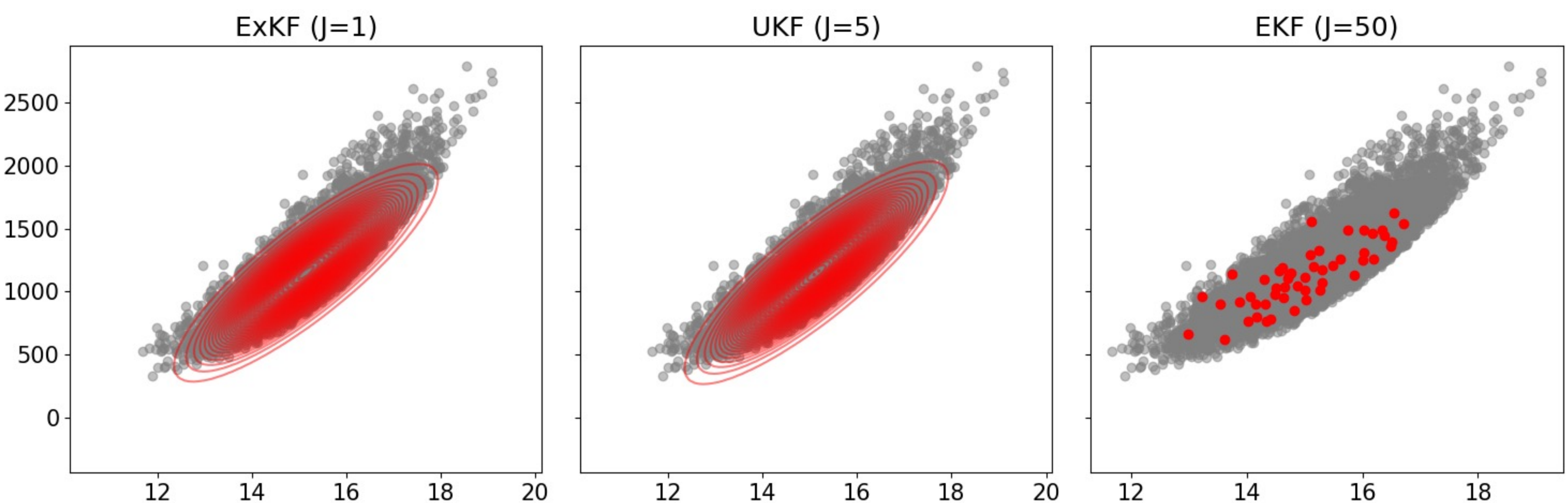
$$c_j = \sqrt{N_\theta + \lambda} \quad W_j^c = \frac{1}{\sqrt{N_\theta + \lambda}} \quad \lambda = a^2 N_\theta - N_\theta \quad a = \min\left\{\sqrt{\frac{4}{N_\theta + \kappa}}, 1\right\}$$

KALMAN FILTERS

Caltech

Numerical demonstration

$$\theta \sim \left(\begin{bmatrix} 10 \\ 10 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \qquad \mathcal{G}(\theta) = \begin{bmatrix} 1 + \sqrt{\theta_{(1)}^2 + \theta_{(2)}^2} \\ \exp \frac{\theta_{(1)}}{2} + \theta_{(2)}^3 \end{bmatrix}$$



UNSCENTED KALMAN INVERSION

- Linear setting ($\mathcal{G}(\theta) = G \cdot \theta$)

Theorem (Exponential convergence)

Assume that $\Sigma_\omega \succ 0$. Assume further that $\alpha \in (0, 1)$ or $\alpha = 1$ and G has empty null-space, the steady state equation

$$C_\infty^{-1} = G^T \Sigma_\nu^{-1} G + (\alpha^2 C_\infty + \Sigma_\omega)^{-1}$$

has a unique solution $C_\infty \succ 0$.

The pair (m_n, C_n) converges exponentially fast to (m_∞, C_∞).

Furthermore m_∞ is the minimizer of the Tikhonov regularized least squares function given by

$$\Phi_R(\theta; y) = \frac{1}{2} \| \Sigma_\nu^{-\frac{1}{2}} (y - G\theta) \|^2 + \frac{1-\alpha}{2} \| \hat{C}_\infty^{-\frac{1}{2}} (\theta - r) \|^2$$

where

$$\hat{C}_\infty = \alpha^2 C_\infty + \Sigma_\omega$$

UNSCENTED KALMAN INVERSION

- Linear setting ($\mathcal{G}(\theta) = G \cdot \theta$)

Let assume the posterior covariance $C_* = (G^T \Sigma_\eta^{-1} G)^{-1}$ exists. If we choose

$$r = r_0 \qquad \Sigma_\nu = 2\Sigma_\eta \qquad \Sigma_\omega = (2 - \alpha^2) C_*$$

Then we have

$$C_\infty = C_*$$

And the Tikhonov regularized least squares function becomes

$$\Phi_R(\theta; y) = \frac{1}{4} \| \Sigma_\eta^{-\frac{1}{2}} G(\theta - G^{-1}(y)) \|^2 + \frac{1-\alpha}{4} \| \Sigma_\eta^{-\frac{1}{2}} G(\theta - r_0) \|^2$$

UNSCENTED KALMAN INVERSION

- Linear setting ($\mathcal{G}(\theta) = G \cdot \theta$)

When the problem is over/well determined ($\mathcal{G}$ has empty null-space)

$$r = r_0 \qquad \Sigma_\nu = 2\Sigma_\eta \qquad \Sigma_\omega = (2-\alpha^2)C_n$$

When the problem is ill-posed

$$r = r_0 \qquad \Sigma_\nu = 2\Sigma_\eta \qquad \Sigma_\omega = (2-\alpha^2)C_0$$

α is the regularization parameter

UNSCENTED KALMAN INVERSION

Caltech

- Nonlinear Setting

Proposition (ExKI Levenberg-Marquardt Connection)

For the nonlinear least squares problem

$$\Phi\,(\theta; y) = \frac{1}{2} \parallel \Sigma_\nu^{-\frac{1}{2}}\,(y \;-\; \mathcal{G}(\theta) \parallel^2$$

The Levenberg-Marquardt algorithm solves it as

$$(d\mathcal{G}(\theta_n)^T \Sigma_\nu^{-1} d\mathcal{G}(\theta_n) + \lambda_n I)\delta\theta_n = d\mathcal{G}(\theta_n)^T \Sigma_\nu^{-1}\,(y - \mathcal{G}(\theta_n))$$

The extended Kalman inversion solves it as

$$(d\mathcal{G}(\theta_n)^T \Sigma_\nu^{-1} d\mathcal{G}(\theta_n) + (C_n + \Sigma_\omega\,)^{-1})\delta\theta_n = d\mathcal{G}(\theta_n)^T \Sigma_\nu^{-1}\left(y - \mathcal{G}(\theta_n)\right)$$

Nonlinear Setting

Proposition (Averaging property)

Let denote Gaussian random variable $\theta \sim \mathcal{N}(m, C) \in R^{N_\theta}$. For any nonlinear function $\mathcal{G}$, we define the averaged function $\mathcal{F}\mathcal{G}$ and averaged gradient $\mathcal{F}d\mathcal{G}$ at m as follows

$$\mathcal{F}\mathcal{G}(m, C) = \mathbb{E}[\mathcal{G}(\theta)] \qquad \mathcal{F}d\mathcal{G}(m, C) = \frac{\partial \mathcal{F}_u\mathcal{G}(m, C)}{\partial m}$$

Then we have

$$\mathcal{F}d\mathcal{G}(m, C) = \mathrm{Cov}[\mathcal{G}(\theta), \theta] \cdot C^{-1}$$

UNSCENTED KALMAN INVERSION

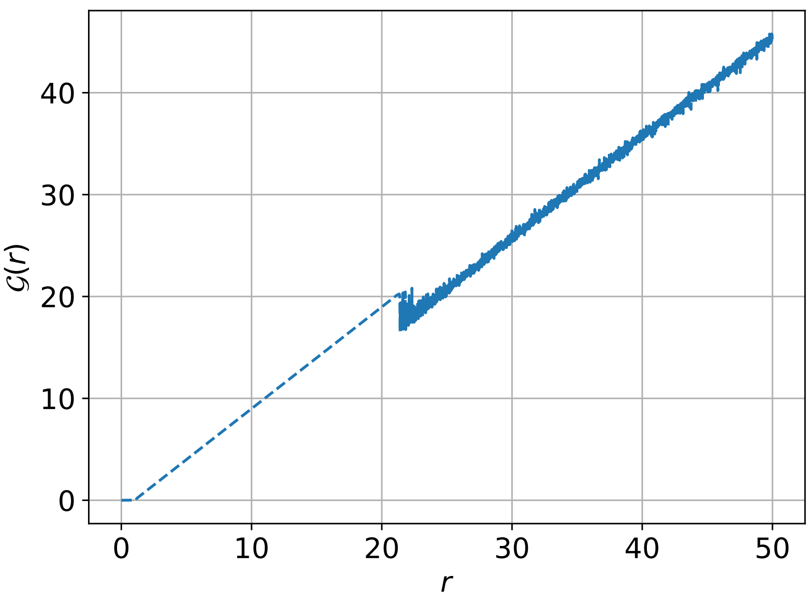
Caltech

Numerical demonstration

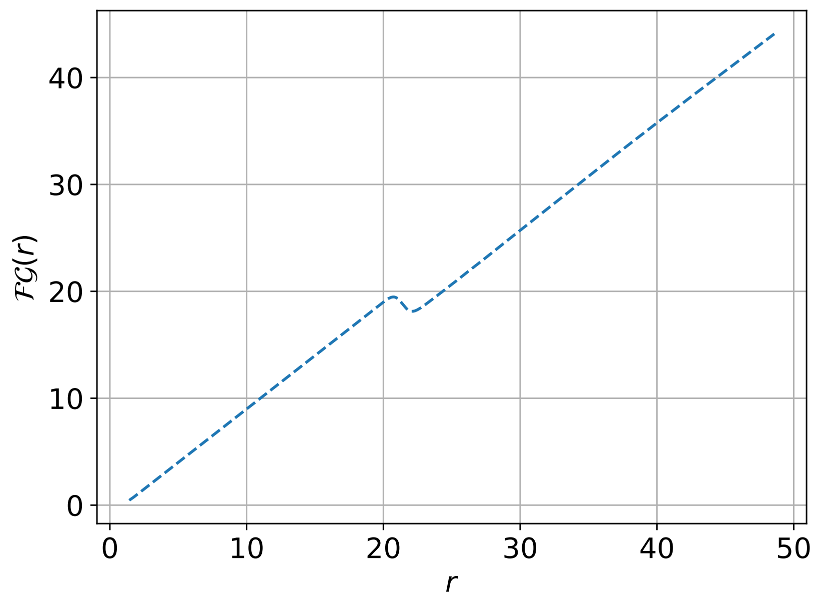
$$\frac{dx_1}{dt} = \sigma(x_2 - x_1) \qquad \frac{dx_2}{dt} = x_1(r - x_3) - x_2 \qquad \frac{dx_3}{dt} = x_1 x_2 - \beta x_3$$

$$y = \mathcal{G}(r) + \eta \qquad \mathcal{G}(r) = \frac{1}{20}\int_{30}^{50} x_3(t)dt \qquad (\sigma = 10 \ \ \beta = \frac{8}{3})$$

Gradient-based adjoint method



UKF



UNSCENTED KALMAN INVERSION

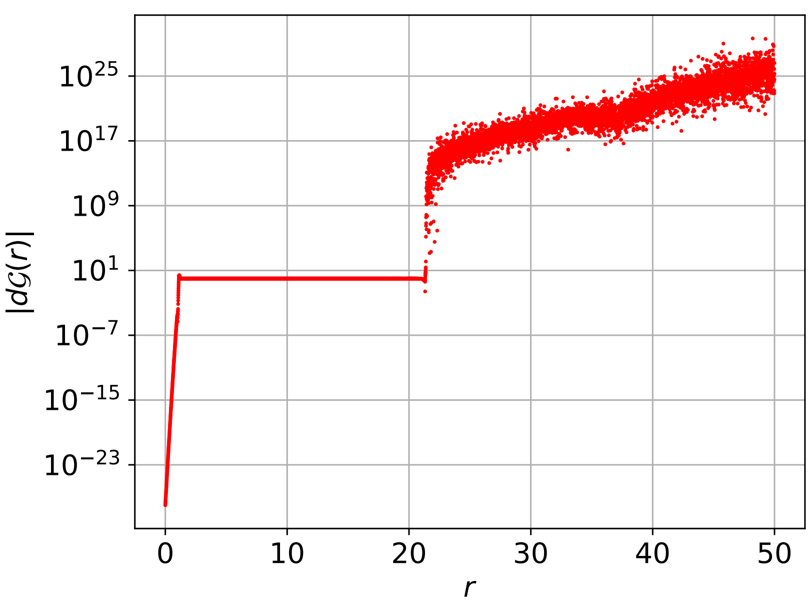
Caltech

- Numerical demonstration

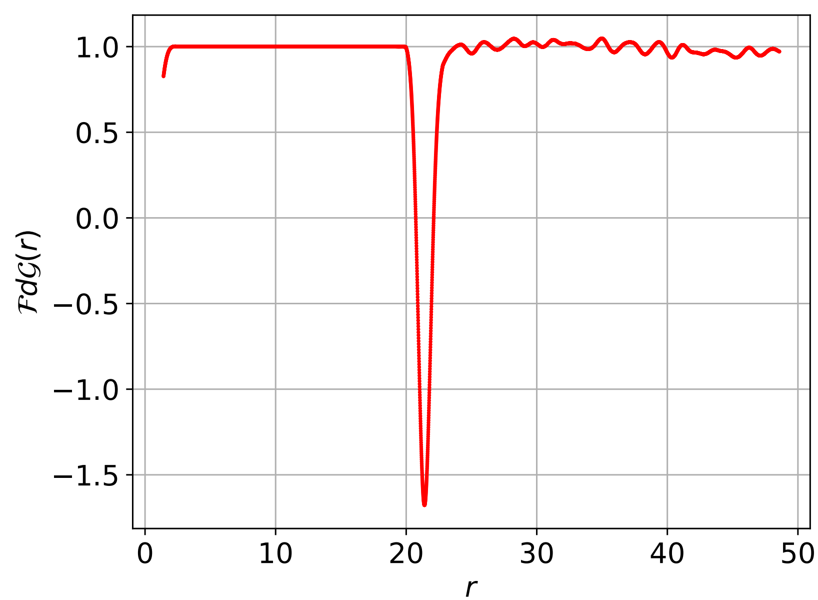
$$\frac{dx_1}{dt} = \sigma(x_2 - x_1) \qquad \frac{dx_2}{dt} = x_1(r - x_3) - x_2 \qquad \frac{dx_3}{dt} = x_1 x_2 - \beta x_3$$

$$y = \mathcal{G}(r) + \eta \qquad \mathcal{G}(r) = \frac{1}{20}\int_{30}^{50} x_3(t)dt \qquad (\sigma = 10 \;\; \beta = \frac{8}{3})$$

Gradient-based adjoint method



UKF



UNSCENTED KALMAN INVERSION

Caltech

- Nonlinear setting

Theorem (Posterior approximation)

Assume $\mathcal{G}$ is bijection and satisfies the Lipschitz property

$$\left|\left|\det\frac{d\mathcal{G}^{-1}(\theta_1)}{d\theta}\right| - \left|\det\frac{d\mathcal{G}^{-1}(\theta_2)}{d\theta}\right|\right| \leq c_0 \parallel \theta_1 - \theta_2 \parallel^{c_1}$$

and other regularization assumptions. We have error bounds

$$\parallel m - m_\infty \parallel_\infty = \mathcal{O}(\rho(\Sigma_\eta)^{c_1}\sqrt{\det\Sigma_\eta})$$

$$\parallel C - C_\infty \parallel_\infty = \mathcal{O}(\rho(\Sigma_\eta)^{c_1}\sqrt{\det\Sigma_\eta})$$

Here m and C are posterior mean and covariance with an improper uniform prior, and m_∞ and C_∞ are converged mean and covariance.

UNSCENTED KALMAN INVERSION

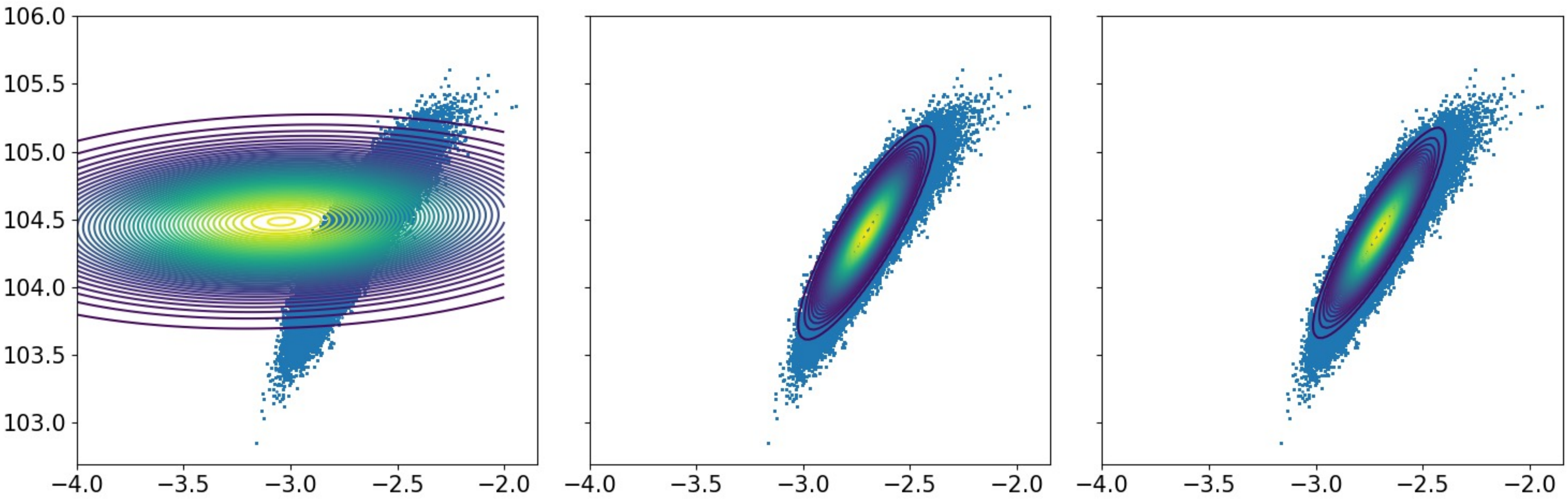
Caltech

Numerical demonstration

$$\frac{d}{dx}\left(e^{\theta_{(1)}}\frac{d}{dx}p(x)\right) = 1 \qquad x \in [0,1] \qquad p(0) = 1 \text{ and } p(1) = \theta_{(2)}$$

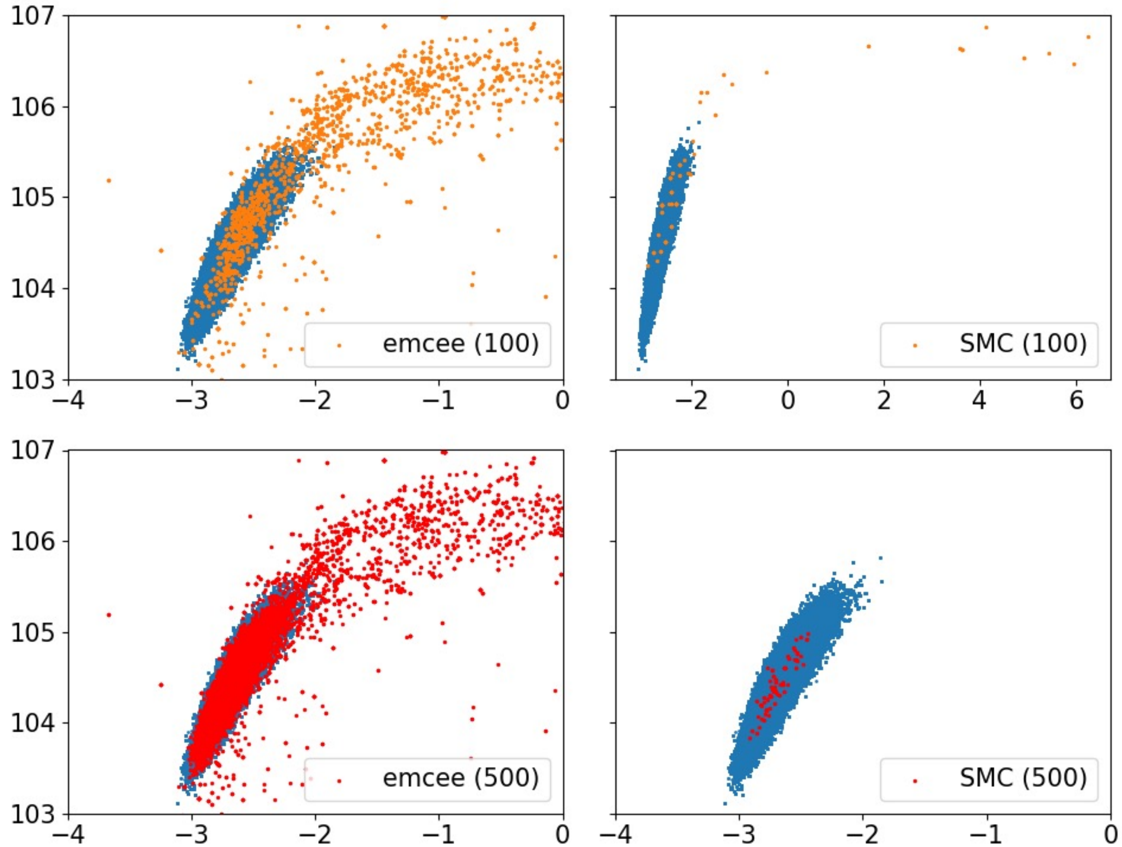
$$y = \mathcal{G}(\theta) + \eta \qquad \mathcal{G}(\theta) = \begin{bmatrix} p(0.25;\theta) \\ p(0.75;\theta) \end{bmatrix}$$

UKI approximation with $J = 5$ at iteration 5, 10, and 15



- Numerical demonstration

Affined invariant Markov chain Monte Carlo and sequential Monte-Carlo with $J = 100$



INITIAL CONDITION RECOVERY

Caltech

- Barotropic vorticity equation on the surface of the Earth

$$\frac{\partial \omega}{\partial t} = -v \cdot \nabla(\omega + f)$$

$$\nabla^2 \psi = \omega \qquad v = k \times \nabla \psi$$

where ω and ψ are (absolute) vorticity and streamfunction,

v is the no-divergent flow velocity, k is the unit vector in the

radial direction and f is the Coriolis force.

INITIAL CONDITION RECOVERY

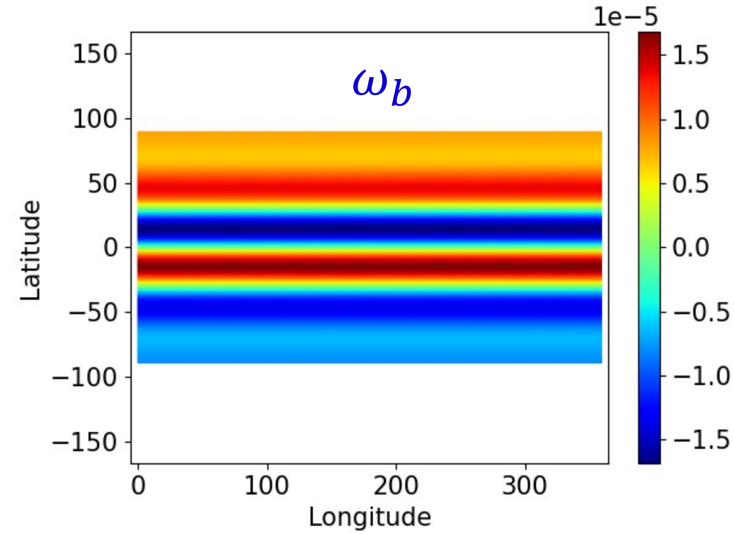
Caltech

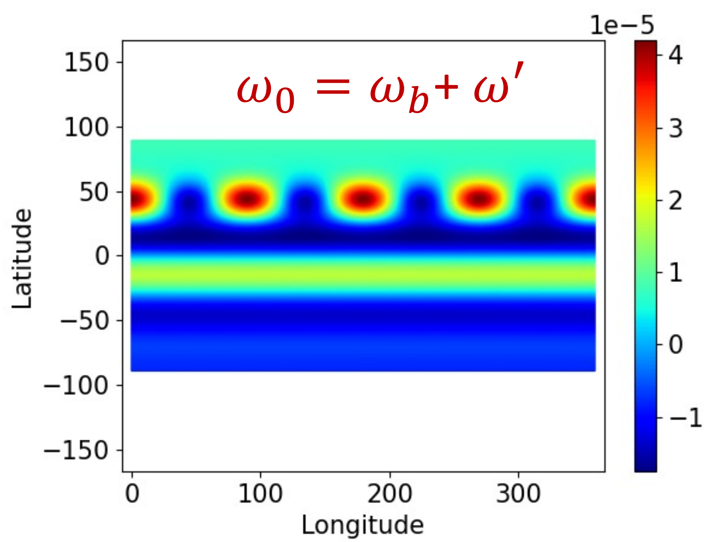
- Initial condition (superposition of u_b and ω')

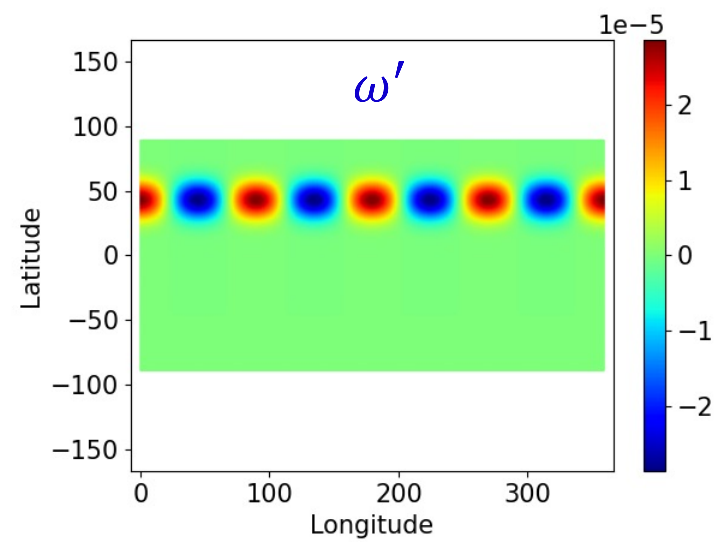
$$u_b = 25\cos(\phi) - 30\cos^3(\phi) + 300\sin^2(\phi)\cos^6(\phi)$$

$$\omega' = \frac{8\times10^{-5}}{2}\cos(\phi)\, e^{-\left(\frac{\phi-15^\circ}{45^\circ}\right)^2}\cos(4\lambda)$$

where ϕ and λ are latitude and longitude.





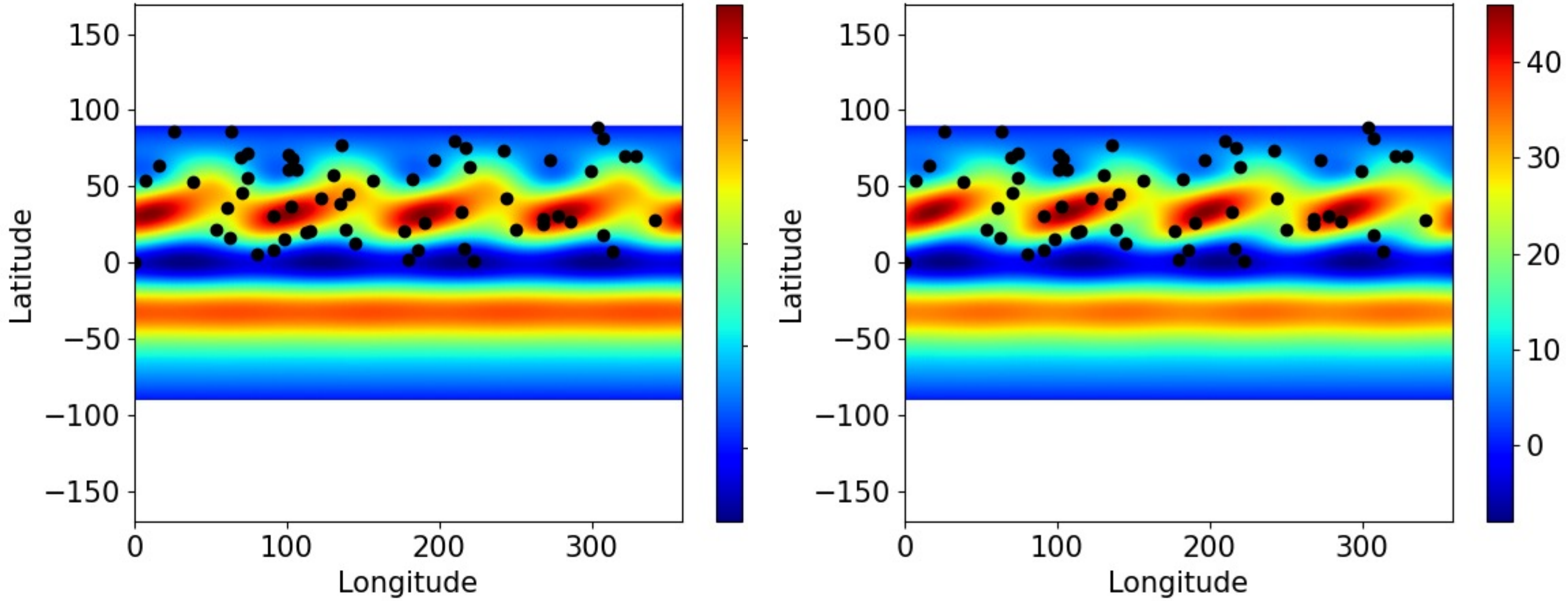


INITIAL CONDITION RECOVERY

Caltech

- Observation

50 random pointwise measurements in the north hemisphere at t $= 12h$ and t $= 24h$



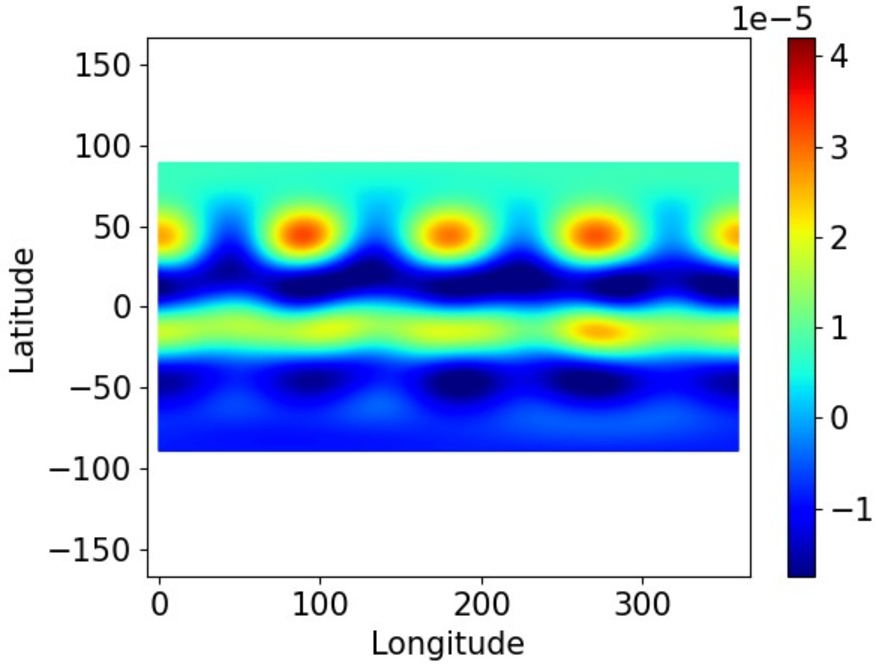
$$y_{obs} = y_{ref} + 5\% y_{ref} \odot \mathcal{N}(0, I)$$

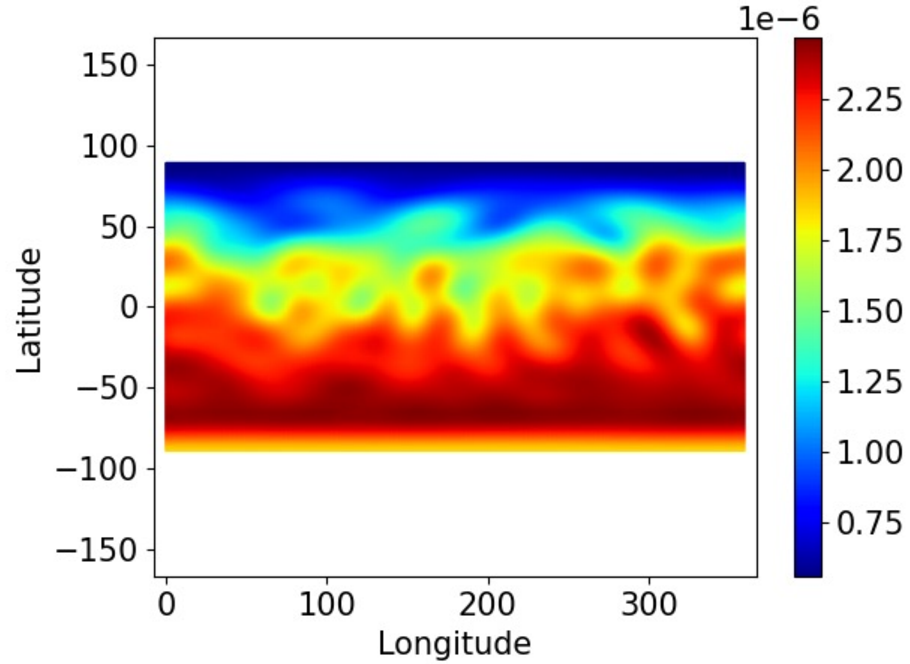
INITIAL CONDITION RECOVERY

Caltech

➢ Unscented Kalman inversion

- Ill-posedness (no observation at the southern hemisphere)
  Regularization: $r = \omega_b$ and $\alpha = 0.5$
- $N_\theta = \mathcal{O}(10^5)$
  Low-rank approximation: constrain in the subspace spanned by the first $N_r = 63$ spectral modes





LEARNING SUBGRID-SCALE PARAMETERS

- General circulation model (GCM)

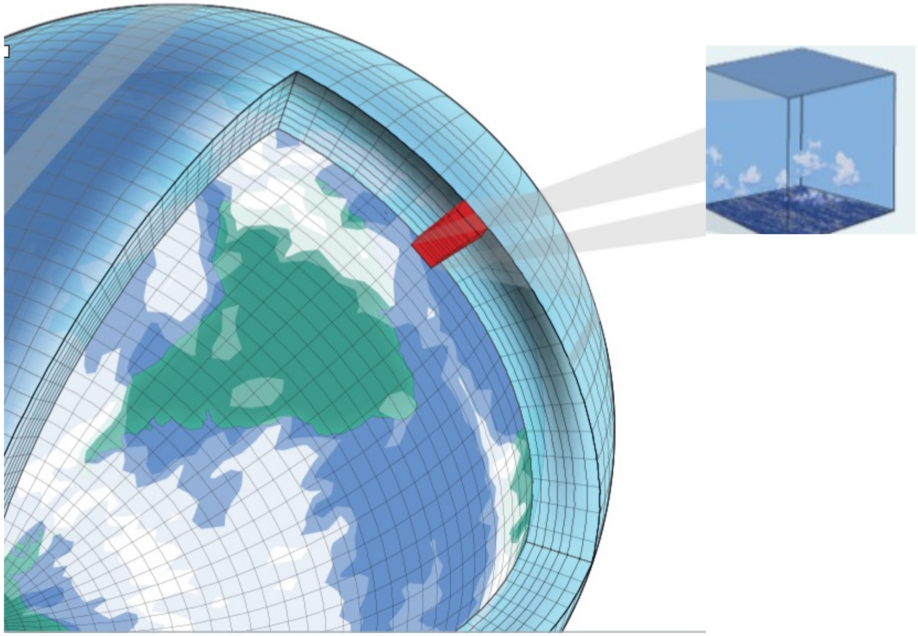
$$\frac{\partial \rho}{\partial t} + \nabla(\rho v) + \frac{\partial \rho w}{\partial z} = 0$$

$$\frac{Dv}{Dt} + \Omega k \times v + \frac{\nabla p}{\rho} + \nabla \Phi = F$$

$$\frac{DT}{Dt} - \frac{RTw}{C_p p} = Q$$

$$\frac{\partial p}{\partial z} = -\rho g$$

$$p = \rho R T$$



where ρ is the density, v and w are horizontal and vertical velocities, T is the temperature, p is the pressure, and Φ is the geopotential, R is the gas constant C_p is the heat capacity at constant pressure, k is the unit vertical vector, Q is the radiation model source term for temperature.

LEARNING SUBGRID-SCALE PARAMETERS

Caltech

- Closure Model (Radiation Modal)

$$Q = -k_T(\phi,\sigma)\left(T - T_{eq}(\phi,p)\right)$$

$$k_T = k_a + (k_s - k_a)\max(0, \frac{\sigma-\sigma_b}{1-\sigma_b}) \cos^4(\phi)$$

$$T_{eq} = \max\{200\mathrm{K}, [315\mathrm{K} - \Delta T_y \sin^2(\phi) - \Delta\theta_z \log(\frac{p}{p_0})\cos^2(\phi)](\frac{p}{p_0})^\kappa\}$$

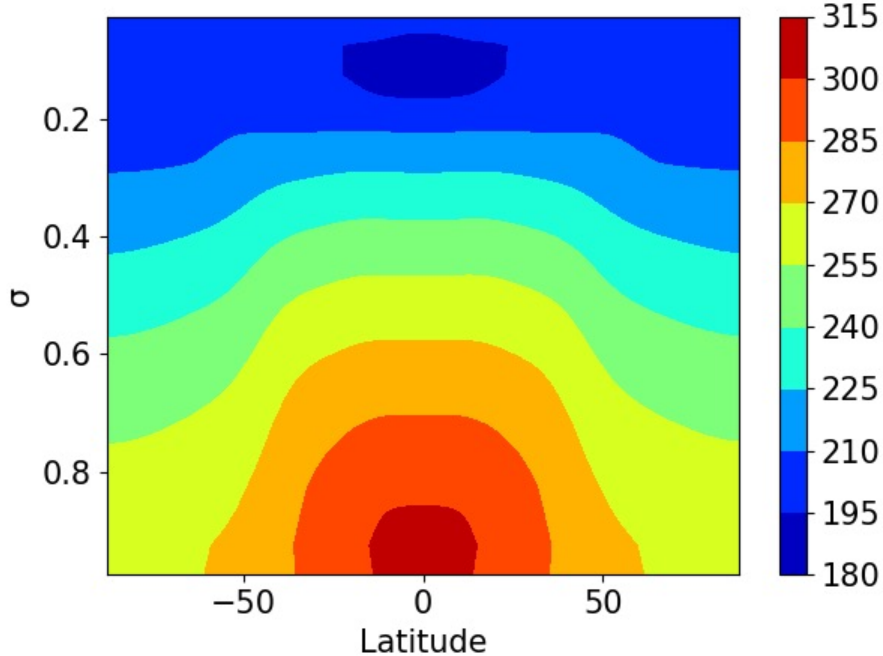
$$k_a = \frac{1}{40\mathrm{day}} \qquad k_s = \frac{1}{4\mathrm{day}} \qquad \Delta T_y = 60\mathrm{K} \quad \Delta\theta_z = 10\mathrm{K}$$

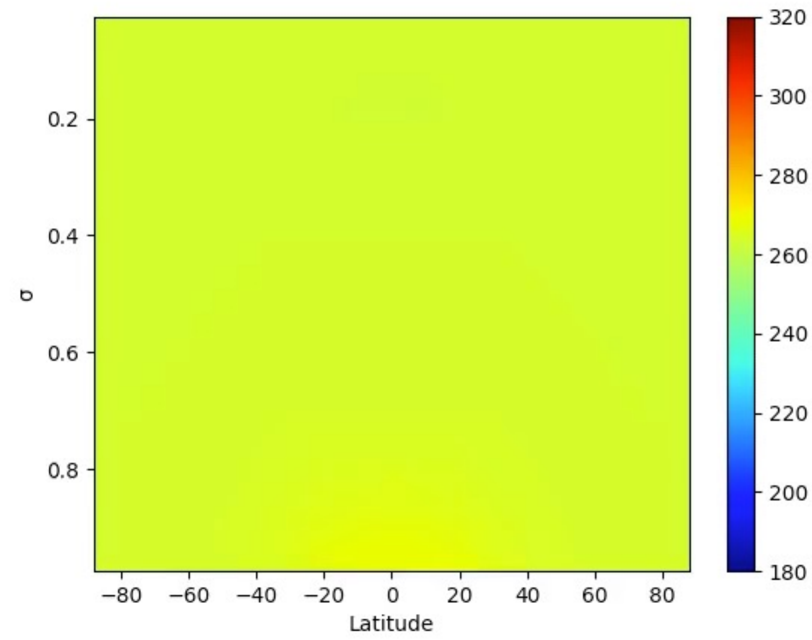
where ϕ is the latitude, and σ is the height coordinate.

LEARNING SUBGRID-SCALE PARAMETERS

- Observation

Zonally/temporally averaged temperature of 1000 days



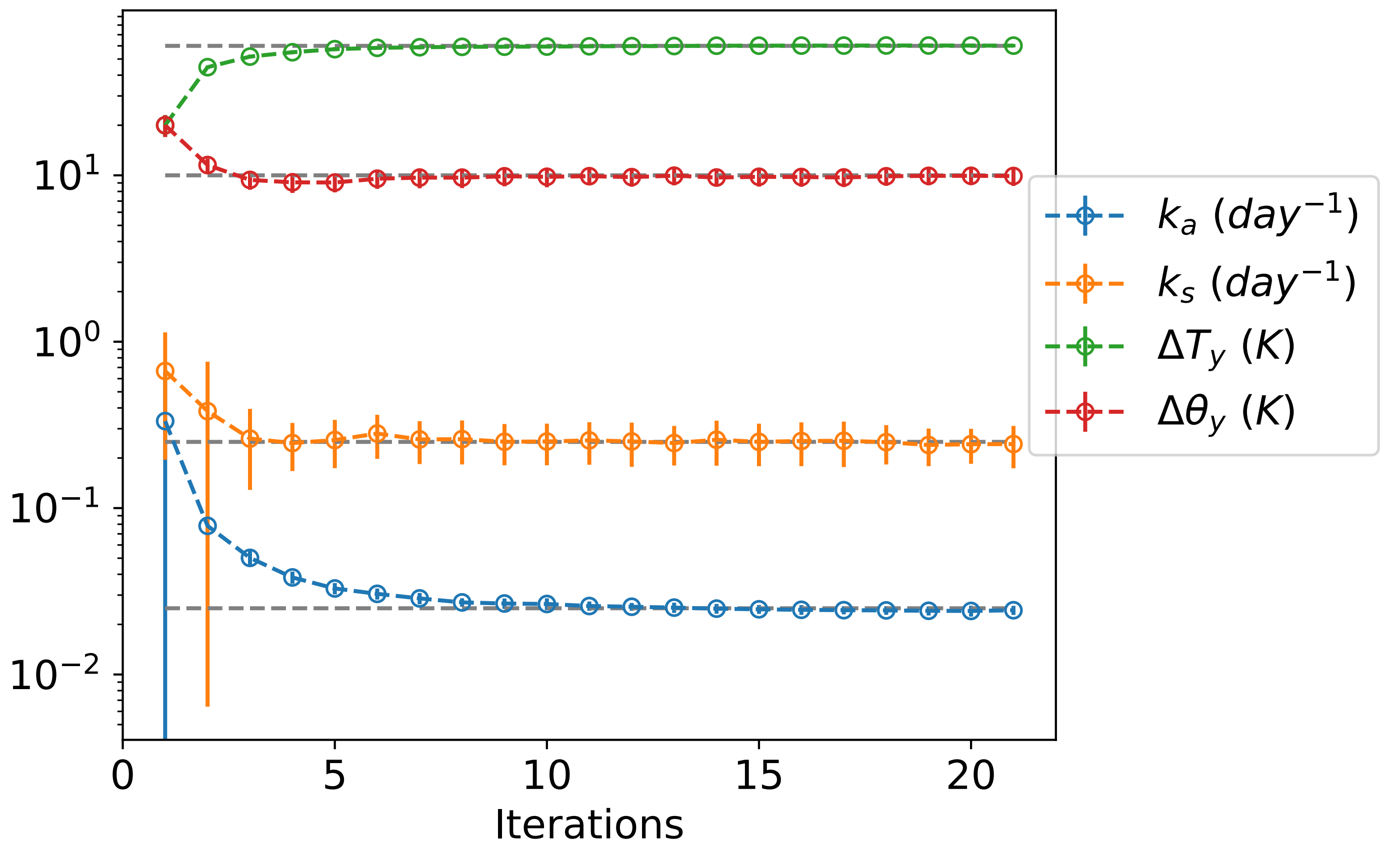


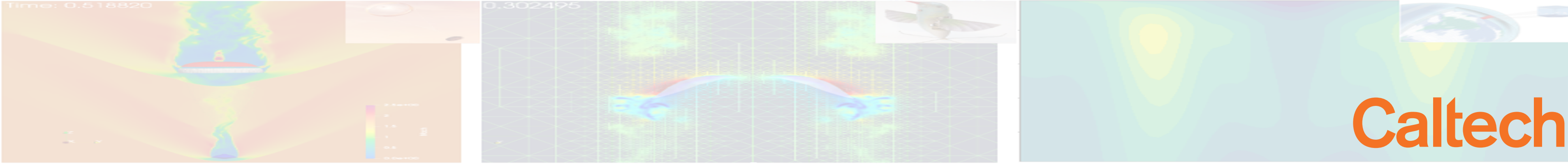
LEARNING SUBGRID-SCALE PARAMETERS

Caltech

- Unscented Kalman inversion

$J = 9$





- Unscented Kalman inversion is an effective tool for derivative-free inversion

Thank you!