Derivative-free, ensemble-based optimization for inverse problems with time-averaged data and chaotic dynamics

Rebecca Gjini

Cecil H. and Ida M. Green Institute of Geophysics and Planetary Physics

Scripps Institution of Oceanography

University of California, San Diego

Collaborators:

Matthias Morzfeld (SIO/UCSD) Advisor

Oliver R.A. Dunbar (Caltech)

Tapio Schneider (Caltech)

UC San Diego | SCRIPPS INSTITUTION OF OCEANOGRAPHY

CaCAO Chaos · Computation · Analysis · Optimization

IGPP

Agenda

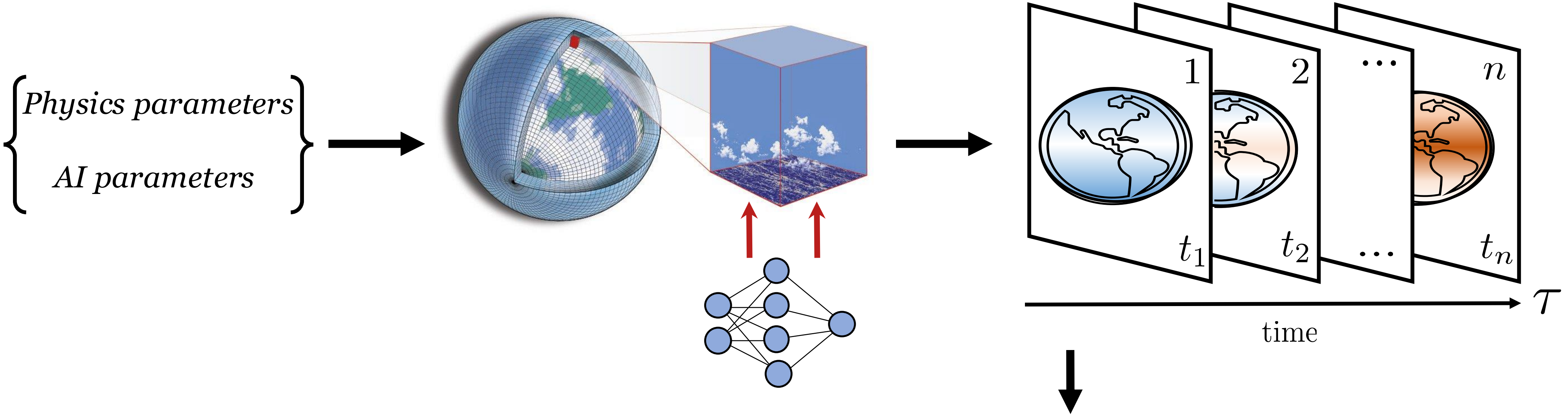
1. ***Introduction and motivation***
2. Ensemble-based optimization
3. Example problems
4. Summary and conclusions

Motivation

Improve future climate predictions

- *Calibrate global climate models (GCMs) using Earth observations*
- *A promising approach is to estimate GCM parameters from observed climate statistics**

Inputs ***Forward simulation*** ***Model output***



What model parameters minimize the misfit between the modeled and observed statistics?

$S(\cdot)$ – statistics function (e.g. temporal)

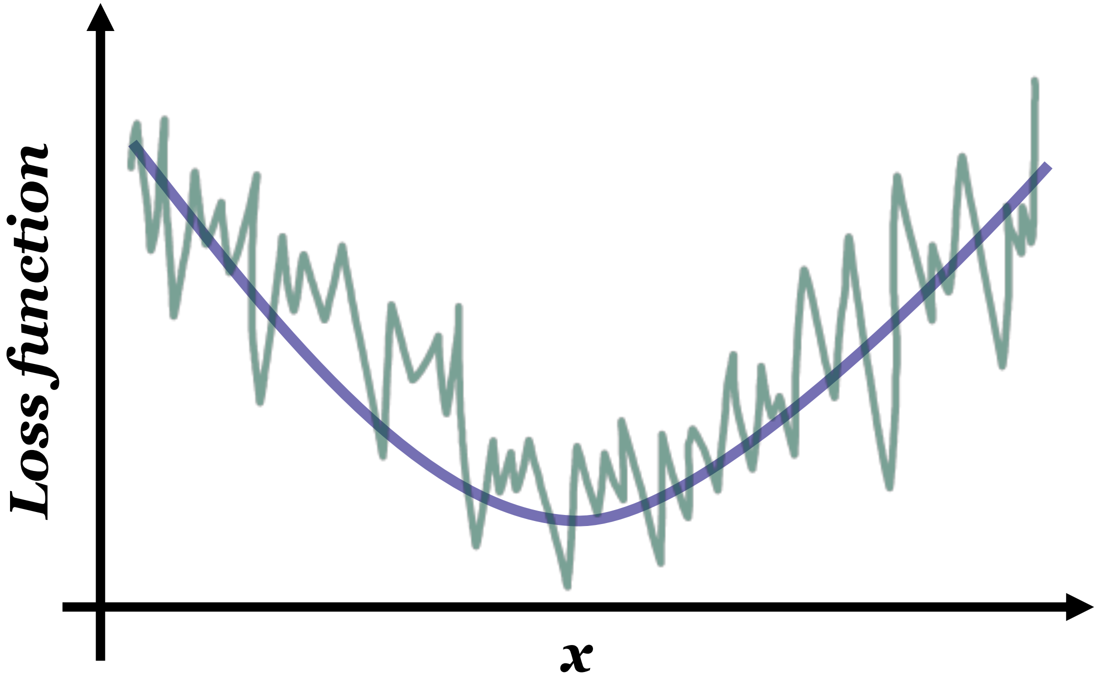
→ ***statistics (e.g. μ, C)***

*Schneider et al., 2017, 2024

GCM calibration

Difficulties:
- Computationally expensive
- Typically no derivatives
- Noisy loss functions
- High-dimensional parameter spaces
- Statistics are indirect observations

Ensemble-based optimization*
- *Main idea:* approximate derivative information using the statistics from an ensemble
- *Benefits*
    - Derivative-free
    - Smooth over noisy loss functions
- *Multiple ensemble-based optimization variants*



Which ensemble-based optimization method should I use to estimate parameters from time-averaged data?

**Bocquet and Sakov, 2014; Evensen, 1994, 2009, 2009, 2018; Chen and Oliver, 2012, 2013*

Agenda

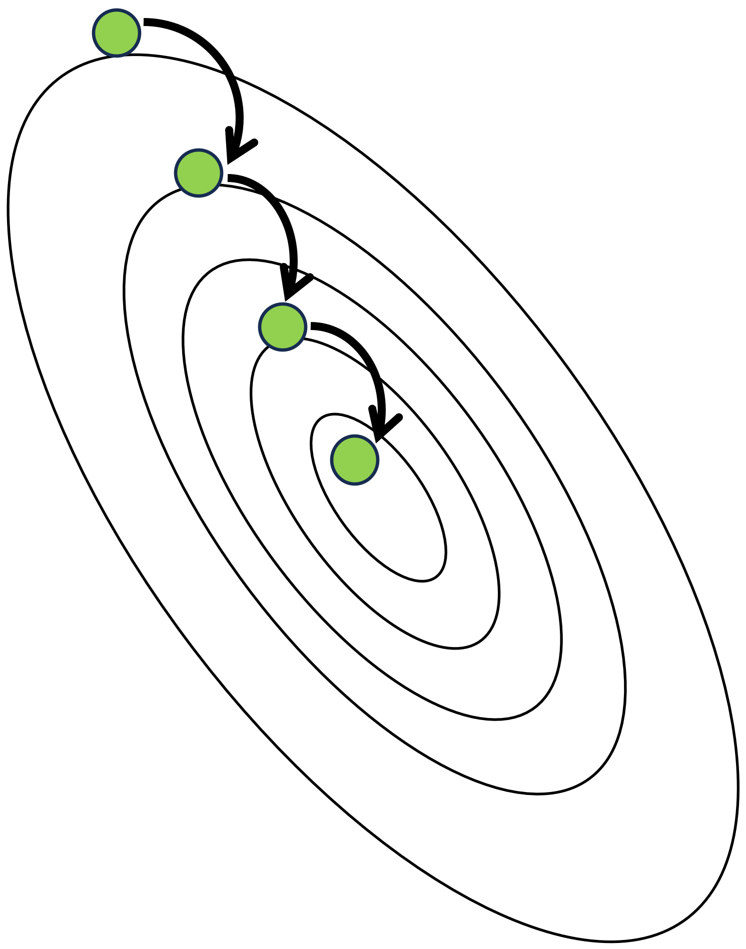
1. Introduction and motivation
2. ***Ensemble-based optimization***
3. Example problems
4. Summary and conclusions

Optimization

Goal: Find the *x* that minimizes the loss $\mathcal{L}(x)$

$$\mathcal{L}(x) = \left\| R^{-\frac{1}{2}} \left(y - \mathcal{H}(x) \right) \right\|_2^2$$

$\mathcal{H}(\cdot)$ – nonlinear model
x – parameters (AI/physics-based)
y – observations
R – observation covariance matrix



Update:

$$x^{k+1} = x^k + \alpha \delta^k$$

α – step size
δ^k – search direction

Ensemble-based optimization

Main idea

- Approximate derivative information using the statistics of an ensemble

Ensemble randomized maximum likelihood (EnRML)
(Gu and Oliver, 2007)

Iterative ensemble smoother (IES)
(Chen and Oliver, 2012, 2013)

Ensemble smoother - multiple data assimilation (ESMDA)
(Emerick and Reynolds, 2012, 2013)

Iterative ensemble Kalman smoother (IEnKS)
(Bocquet and Sakov, 2014)

.
.
.

And many other types of iterative ensemble Kalman filters*

**Evensen, 2009; Asch, Bocquet and Nodet, 2016*

Ensemble-based optimization

Main idea

- Approximate derivative information using the statistics of an ensemble

1) Ensemble $x_1^k, x_2^k, \ldots, x_{n_\mathrm{e}}^k, \qquad h_i^k = \mathcal{H}\left(x_i^k\right), \qquad h_1^k, h_2^k, \ldots, h_{n_\mathrm{e}}^k$

2) Ensemble perturbations

$$X^k = \tfrac{1}{\sqrt{n_\mathrm{e}-1}}\left[x_1^k - \bar{x}^k,\ x_2^k - \bar{x}^k,\ \cdots,\ x_{n_\mathrm{e}}^k - \bar{x}^k\right],\ \bar{x}^k = \frac{1}{n_\mathrm{e}}\sum_{i=1}^{n_\mathrm{e}} x_i^k$$

$$Y^k = \tfrac{1}{\sqrt{n_\mathrm{e}-1}}\left[h_1^k - \bar{h}^k,\ h_2^k - \bar{h}^k,\ \cdots,\ h_{n_\mathrm{e}}^k - \bar{h}^k\right],\ \bar{h}^k = \frac{1}{n_\mathrm{e}}\sum_{i=1}^{n_\mathrm{e}} h_i^k$$

3) Covariances $\quad C_{xy}^k = X^k \otimes Y^k$

$$C_{yy}^k = Y^k \otimes Y^k.$$

4) Update $\boxed{x_i^{k+1} = x_i^k + \alpha\left\{C_{xy}^k\left(C_{yy}^k + R\right)^{-1}\left(y - \left(h_i^k + \eta_i^k\right)\right)\right\}}\qquad \eta_i^k \sim \mathcal{N}\left(0, R\right)$

Problem setup

$$\mathcal{L}(x) = \left\|R^{-\frac{1}{2}}\left(y - \mathcal{H}(x)\right)\right\|_2^2$$

$\mathcal{H}(\cdot)$ – nonlinear model
x – parameters
y – observations
R – observation covariance matrix

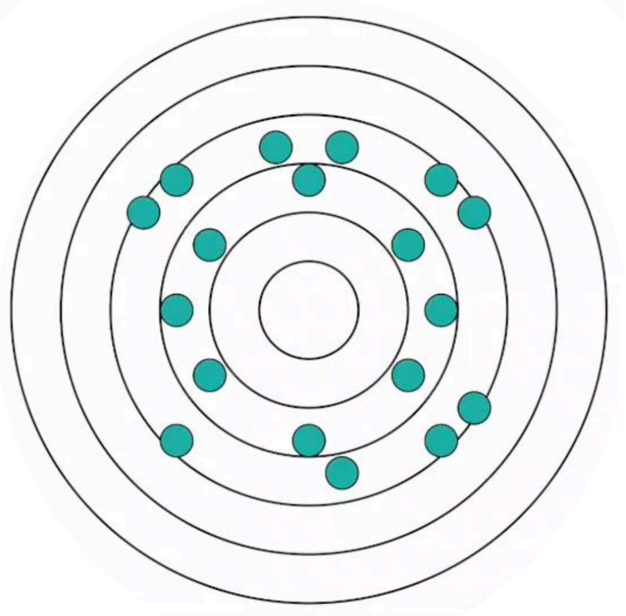
$$x^{k+1} = x^k + \alpha\delta^k$$

α – step size
δ^k – search direction

Ensemble-based optimization

Main idea

- Approximate derivative information using the statistics of an ensemble



1) Ensemble $x_1^k, x_2^k, \ldots, x_{n_\mathrm{e}}^k, \qquad h_i^k = \mathcal{H}\left(x_i^k\right), \qquad h_1^k, h_2^k, \ldots, h_{n_\mathrm{e}}^k$

2) Ensemble perturbations

$$X^k = \frac{1}{\sqrt{n_\mathrm{e}-1}}\left[x_1^k - \bar{x}^k,\ x_2^k - \bar{x}^k,\ \cdots,\ x_{n_\mathrm{e}}^k - \bar{x}^k\right],\ \bar{x}^k = \frac{1}{n_\mathrm{e}}\sum_{i=1}^{n_\mathrm{e}} x_i^k$$

$$Y^k = \frac{1}{\sqrt{n_\mathrm{e}-1}}\left[h_1^k - \bar{h}^k,\ h_2^k - \bar{h}^k,\ \cdots,\ h_{n_\mathrm{e}}^k - \bar{h}^k\right],\ \bar{h}^k = \frac{1}{n_\mathrm{e}}\sum_{i=1}^{n_\mathrm{e}} h_i^k$$

3) Covariances

$$C_{xy}^k = X^k \otimes Y^k$$
$$C_{yy}^k = Y^k \otimes Y^k.$$

Ensemble Kalman Inversion (EKI)

4) Update $\boxed{x_i^{k+1} = x_i^k + \alpha\left\{C_{xy}^k\left(C_{yy}^k + R\right)^{-1}\left(y - \left(h_i^k + \eta_i^k\right)\right)\right\}} \qquad \eta_i^k \sim \mathcal{N}(0, R)$

Iglesias et al., 2013

Ensemble methods for optimization

Iterative ensemble Kalman filter (IEKF)
(Chen and Oliver, 2013)

Tikhonov regularized EKI (TEKI)
(Chada et al., 2020)

Ensemble transform Kalman inversion (ETKI)
(Huang et al., 2022; Tippett et al., 2002; Ott et al., 2004)

Unscented Kalman inversion (UKI)
(Huang et al., 2022; Wan and Van Der Merwe, 2000)

Quasi-Newton method (QN)

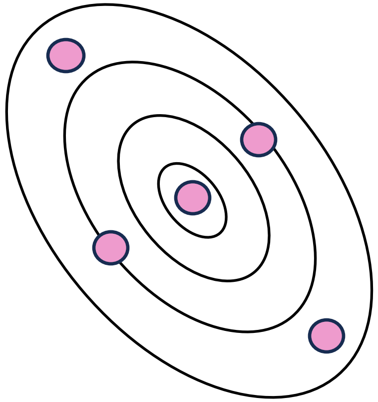
Kalman gain

$$PH^T\left(HPH^T + R\right)^{-1}$$

$$\underset{\text{parameters}}{\mathcal{L}(x)} = \left\| \begin{pmatrix} \overset{\text{observation covariance}}{R} & 0 \\ 0 & \underset{\text{Prior covariance}}{P} \end{pmatrix}^{-\frac{1}{2}} \left(\begin{pmatrix} \overset{\text{observations}}{y} \\ \underset{\text{Prior mean}}{m} \end{pmatrix} - \begin{pmatrix} \overset{\text{model}}{\mathcal{H}(x)} \\ x \end{pmatrix} \right) \right\|_2^2$$

$$\left[X^{k+1} \right] = \left[\underset{\text{ensemble perturbations}}{X^k} \right] \overset{\text{ensemble transformation}}{[T]^{n_e \times n_e}}$$

$n_e = 2n_x + 1$



How do we evaluate each algorithm?

1. Number of forward model evaluations
2. Scalability
3. Generalizability

Algorithm evaluation criteria

How can we determine how well each algorithm solves the optimization problems?

- ~~*Check the solution after convergence (within specified tolerance)*~~

- *Stop iterating when the target error is reached*
  - Each algorithm must reach the same accuracy level
  - Count the number of forward model evaluations needed to reach the target error

Root mean square error (RMSE)

- Calculate data model misfit
- Weighted by the assumed data errors

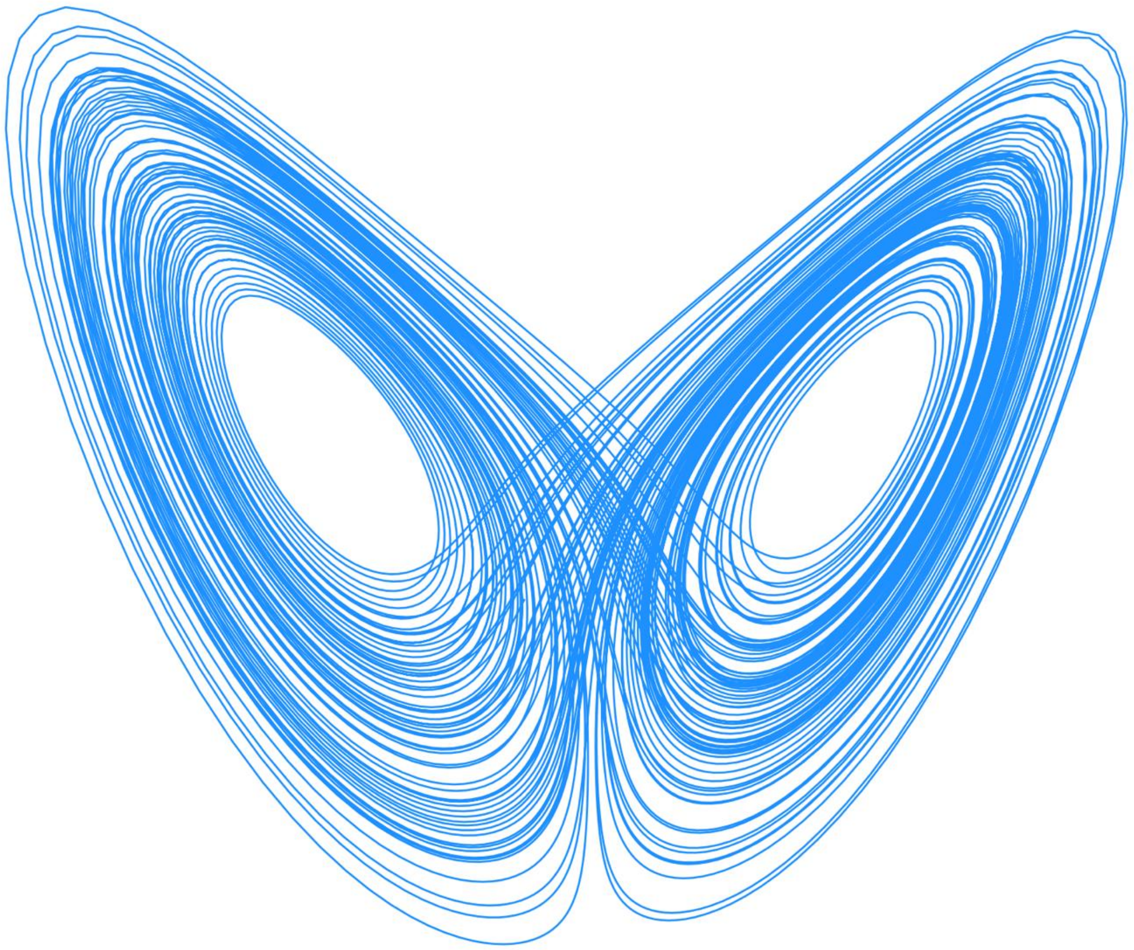
$$\mathrm{RMSE} = \sqrt{\frac{1}{n_y}\sum_{i=1}^{n_y}\frac{\left(y_i - \hat{y}_i\right)^2}{R_{ii}}}$$

Agenda

1. Introduction and motivation
2. Ensemble-based optimization
3. ***Example problems***
4. Summary and conclusions

Lorenz 63

$$\frac{\mathrm{d}z_1}{\mathrm{d}t} = \sigma(z_2 - z_1)$$
$$\frac{\mathrm{d}z_2}{\mathrm{d}t} = \rho z_1 - z_2 - z_1 z_3$$
$$\frac{\mathrm{d}z_3}{\mathrm{d}t} = z_1 z_2 - \beta z_3$$

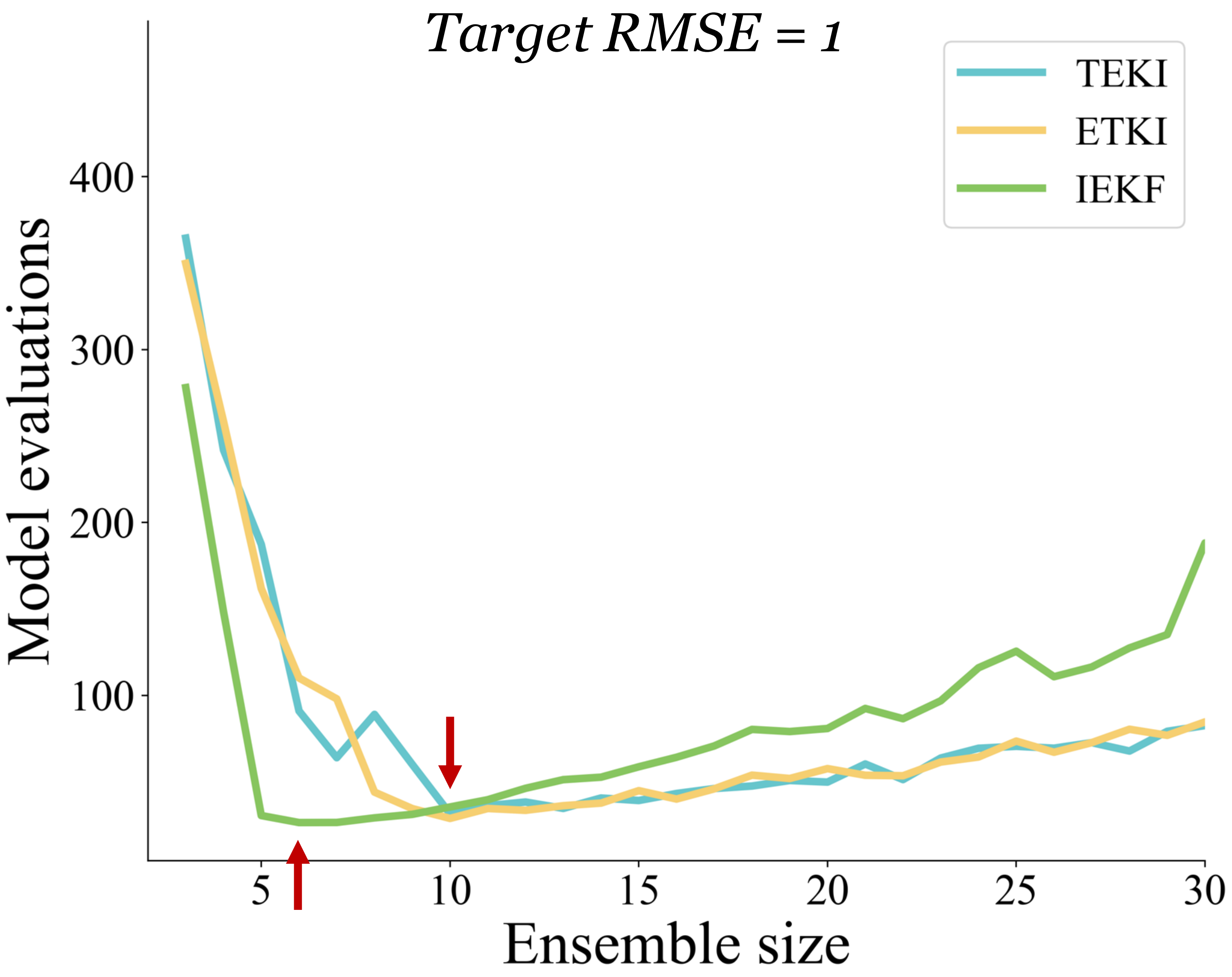


Problem setup

$(\sigma, \rho, \beta) = (10, 28, 8/3)$
$n_x = 2$
$\tau = 10$
$y = (\bar{z_1}, \bar{z_2}, \bar{z_3}, s_1^2, s_2^2, s_3^2, s_{12}, s_{13}, s_{23})$

Cleary et al., 2021

What ensemble size should I use?



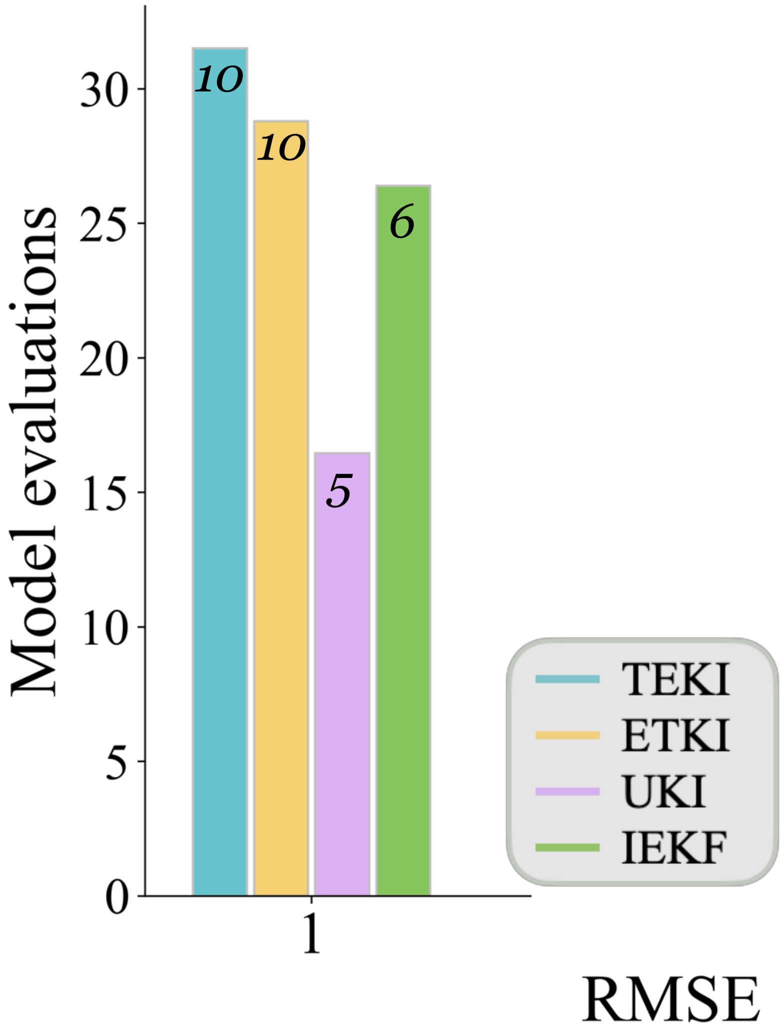
Small n_e

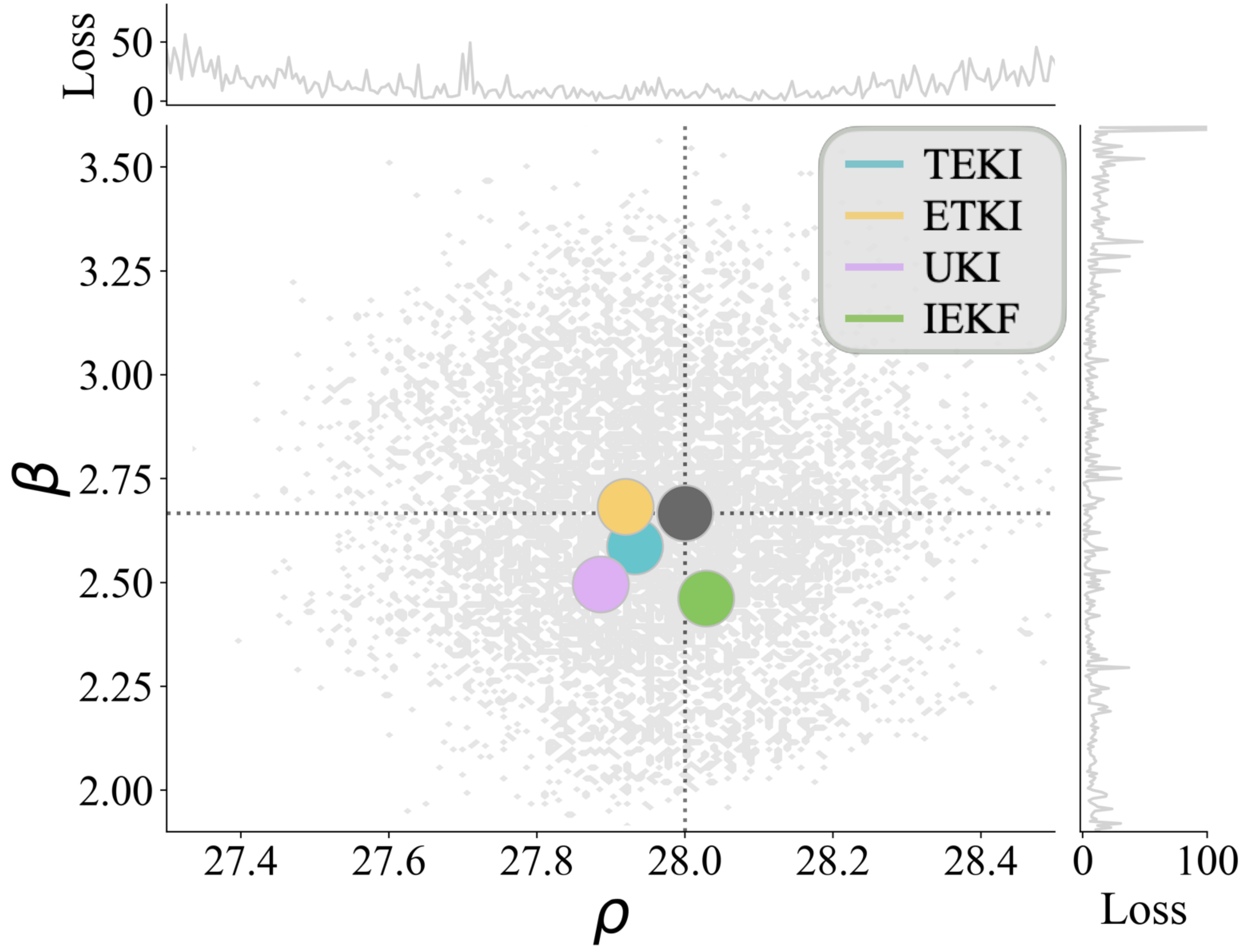
worse gradients
more iterations

Large n_e

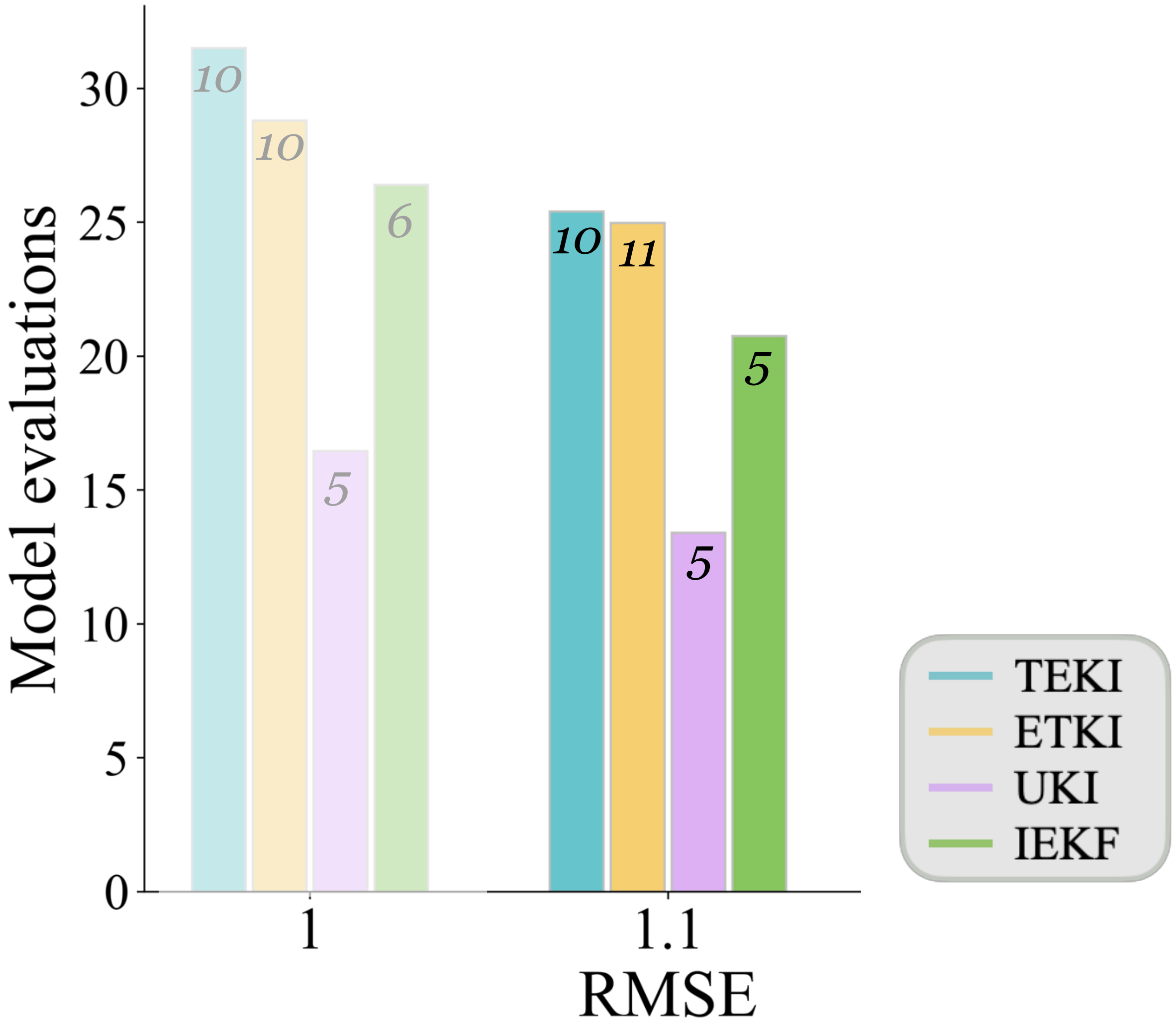
better gradients
fewer iterations

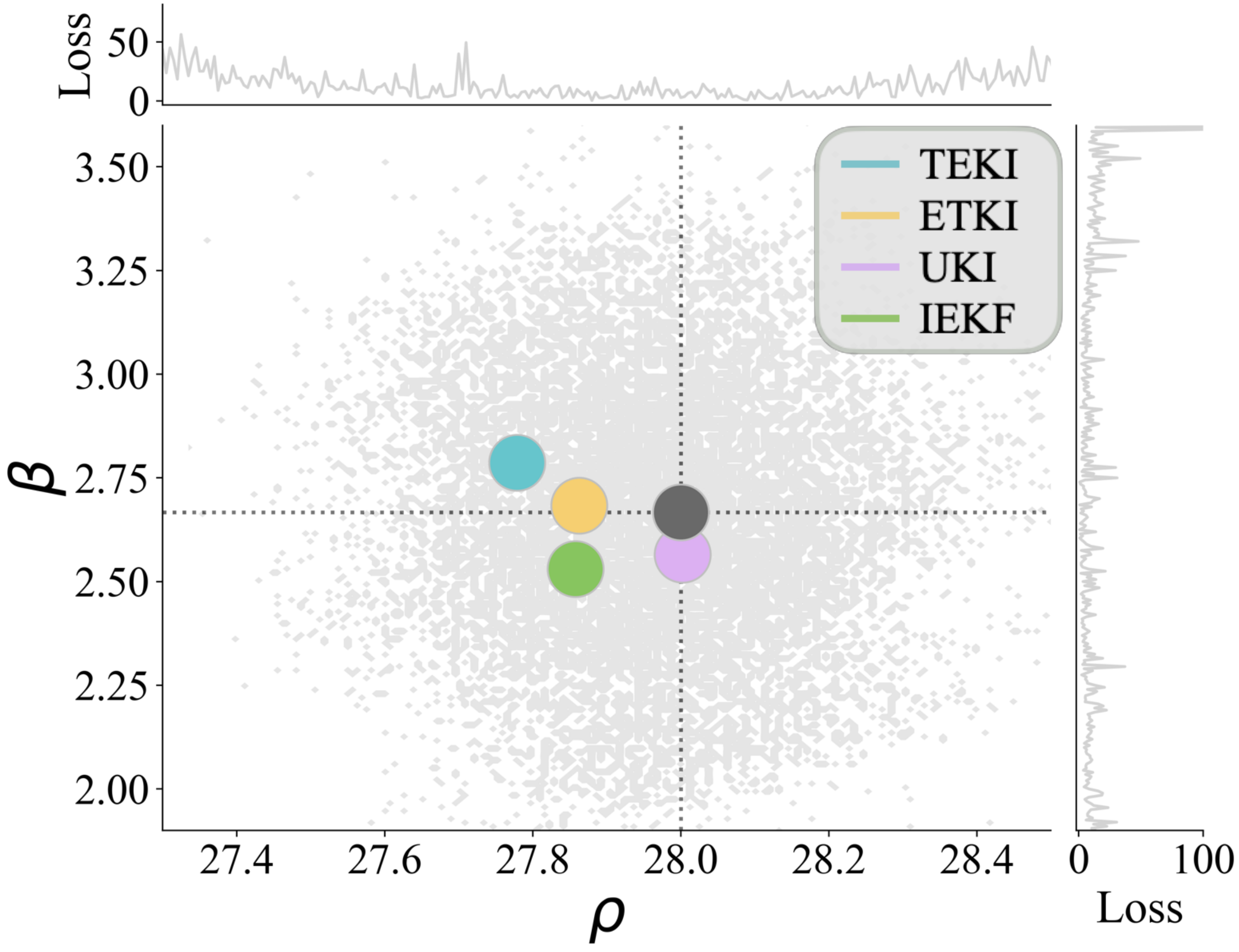
Lorenz 63 results



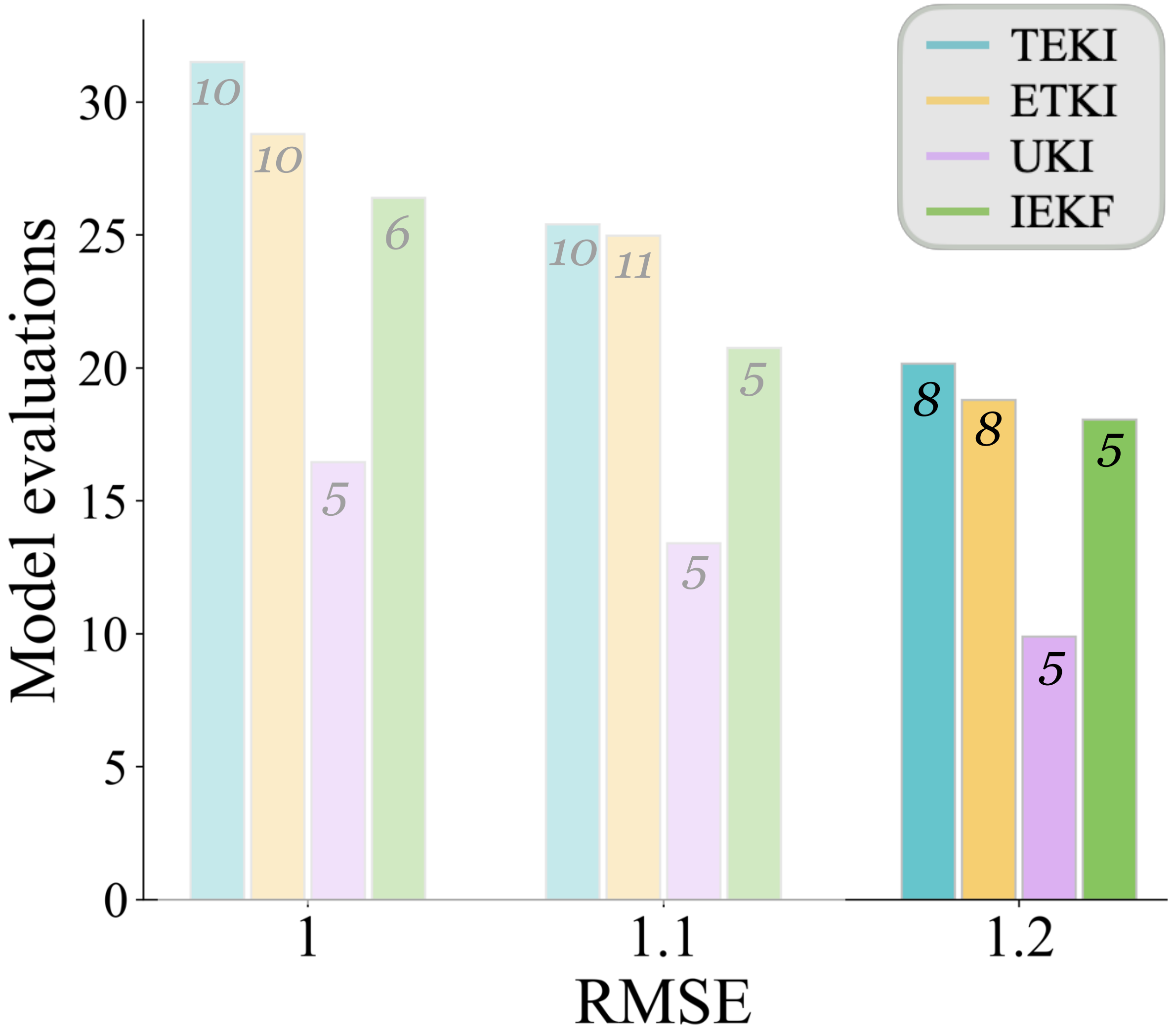


Lorenz 63 results

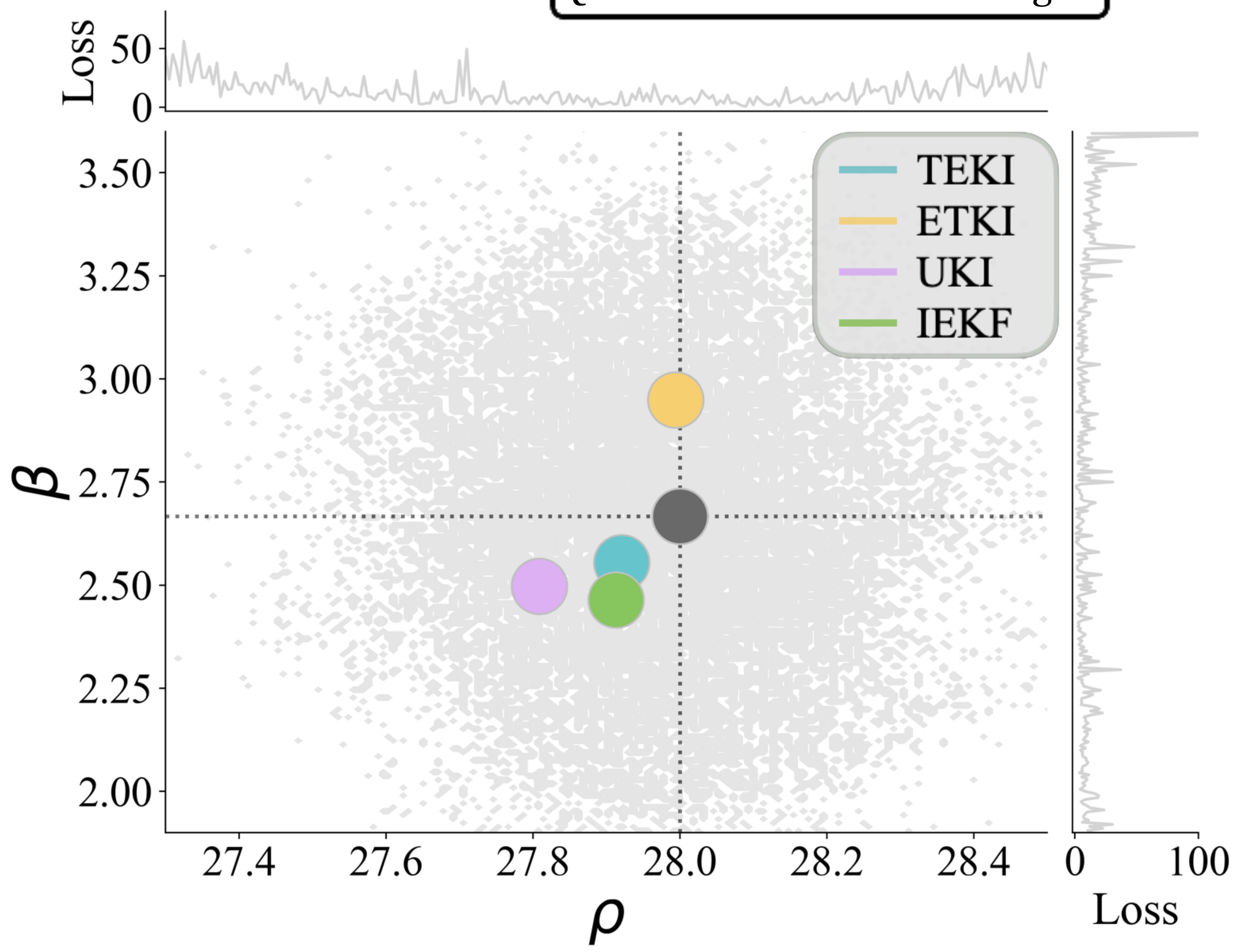




Lorenz 63 results



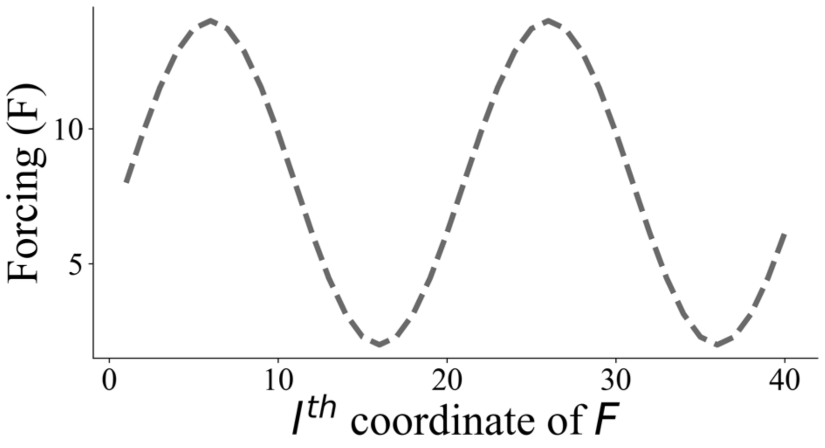
QN did not reach RMSE targets



Modified Lorenz 96

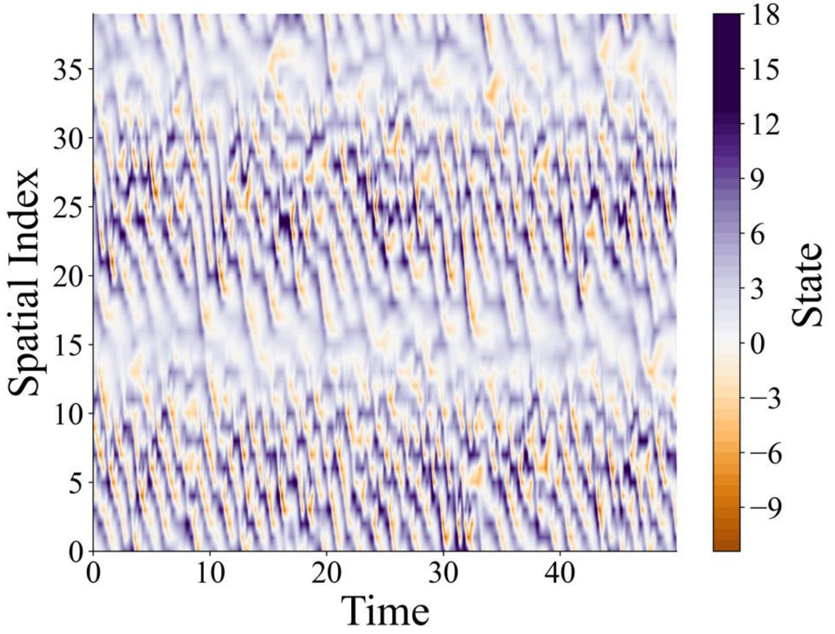
$$\frac{\mathrm{d}z_l}{\mathrm{d}t} = z_{l-1}(z_{l+1} - z_{l-2}) - z_l + \boxed{F_l,} \quad l = 1, \ldots, d$$

$$z_0 = z_d, \qquad z_{d+1} = z_1, \qquad z_{-1} = z_{d-1}$$



Problem setup

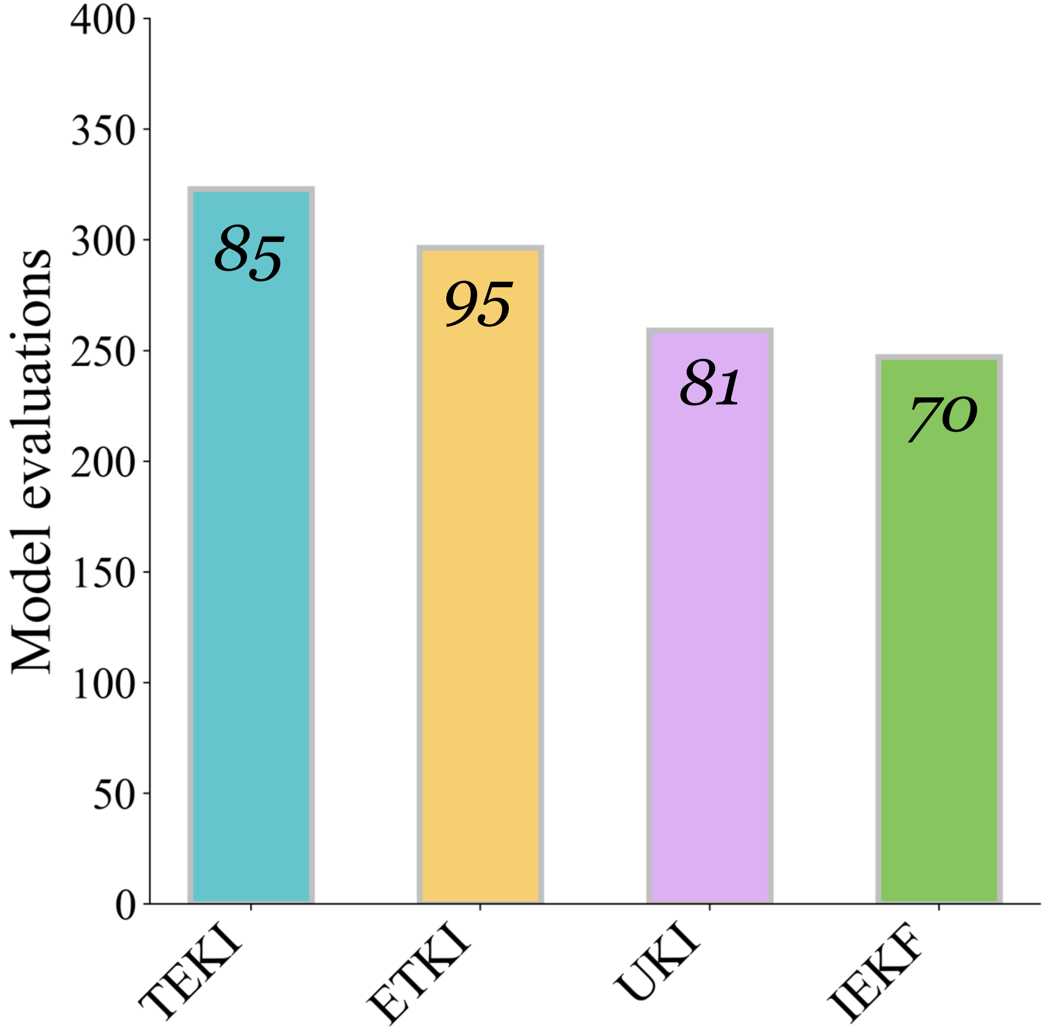
z_l – l^{th}coordinate of the current state
F_l – l^{th}coordinate of a sinusoid
$n_x = 40$
$\tau = 50$
$d = 40$
$y = (\bar{z}_1, \ldots, \bar{z}_{40}, s_1, \ldots, s_{40})$



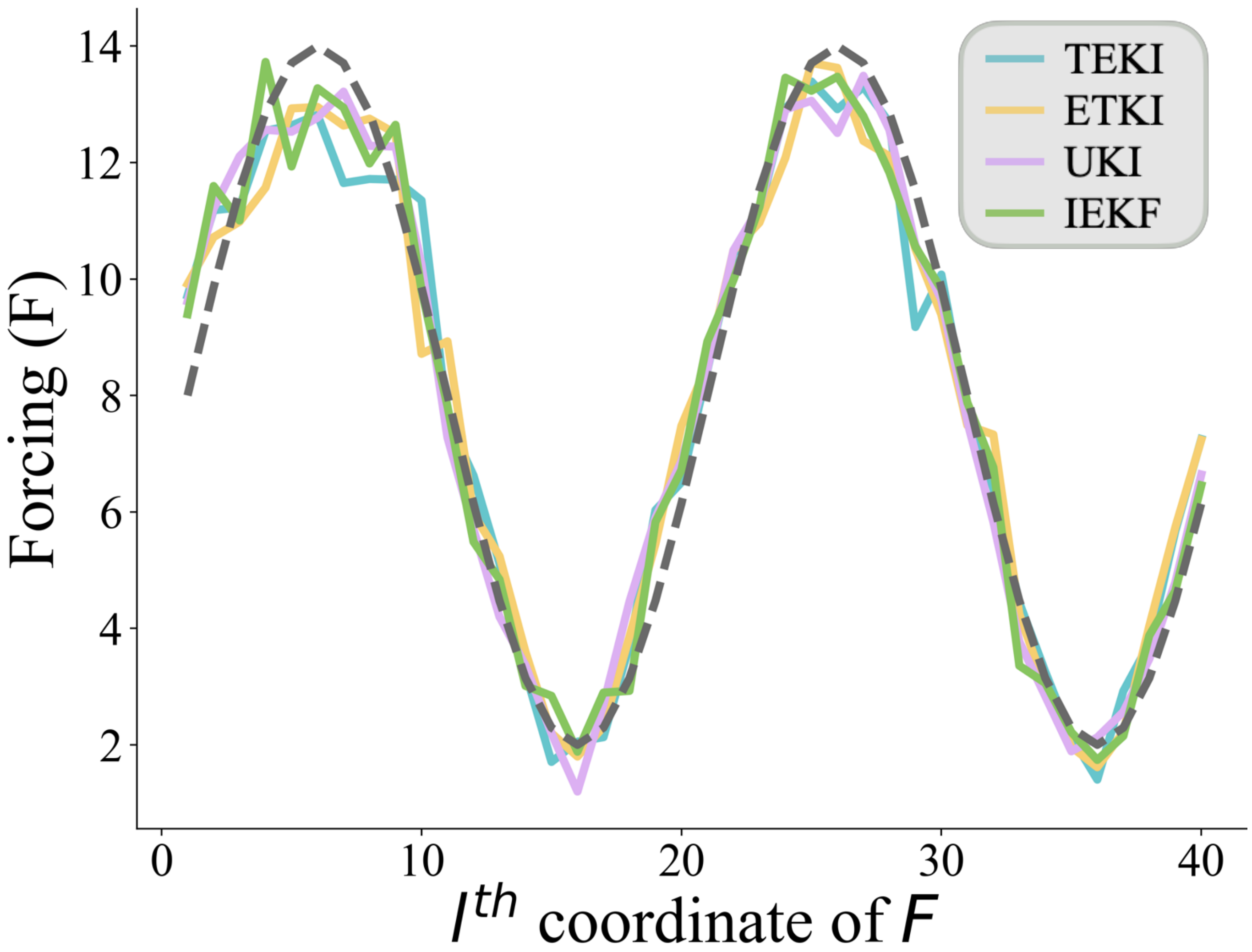
Vishny et al., 2024

Modified Lorenz 96 results

Target RMSE = 1



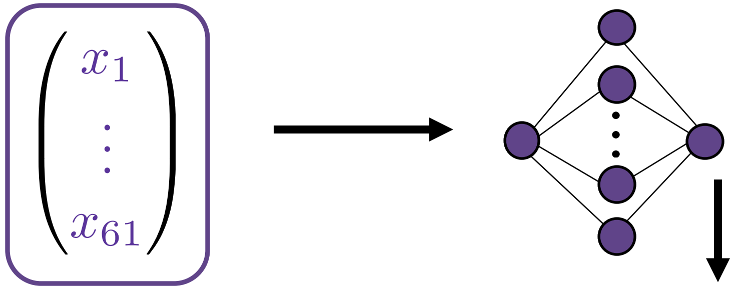
QN did not reach RMSE targets

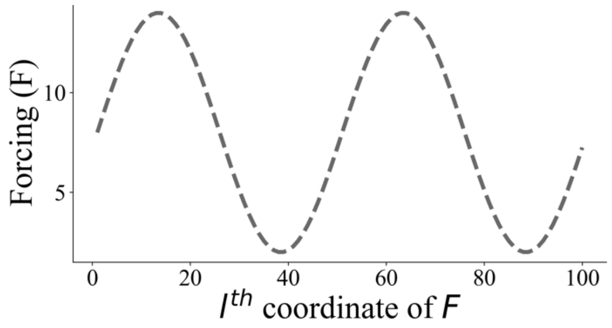


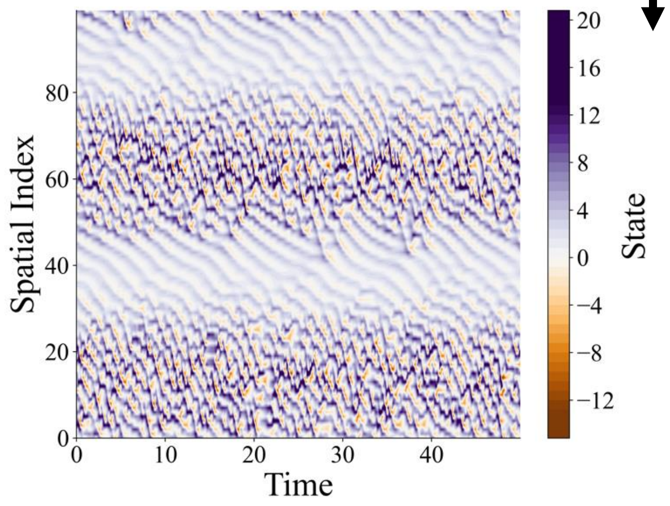
Modified Lorenz 96 with neural network

$$\frac{\mathrm{d}z_l}{\mathrm{d}t} = z_{l-1}(z_{l+1} - z_{l-2}) - z_l + F_l, \quad l = 1, \ldots, d$$

$$z_0 = z_d, \qquad z_{d+1} = z_1, \qquad z_{-1} = z_{d-1}$$







Problem setup

z_l – l^{th}coordinate of the current state

F_l – l^{th}coordinate of a sinusoid

$n_x = 61$

$\tau = 50$

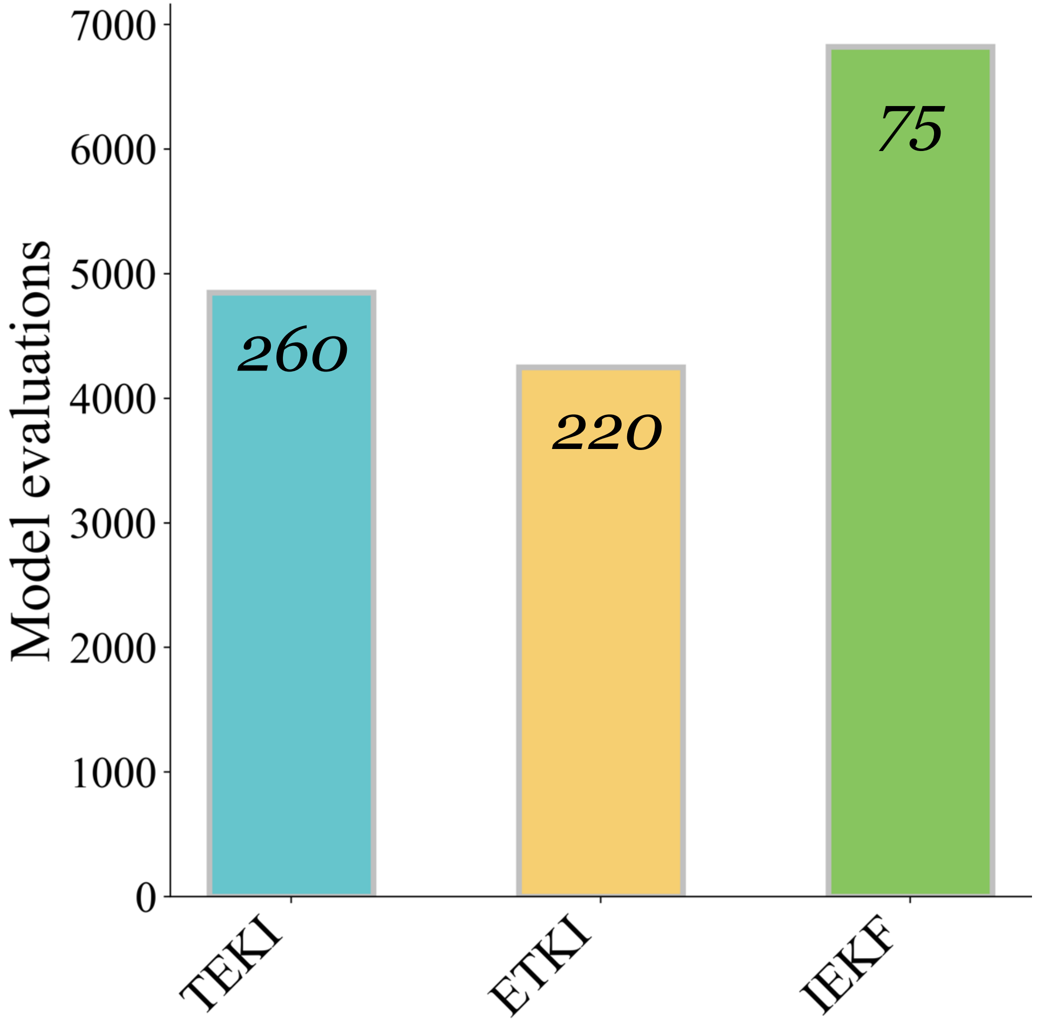
$d = 100$

$y = (\bar{z}_1, \ldots, \bar{z}_{100}, s_1, \ldots, s_{100})$

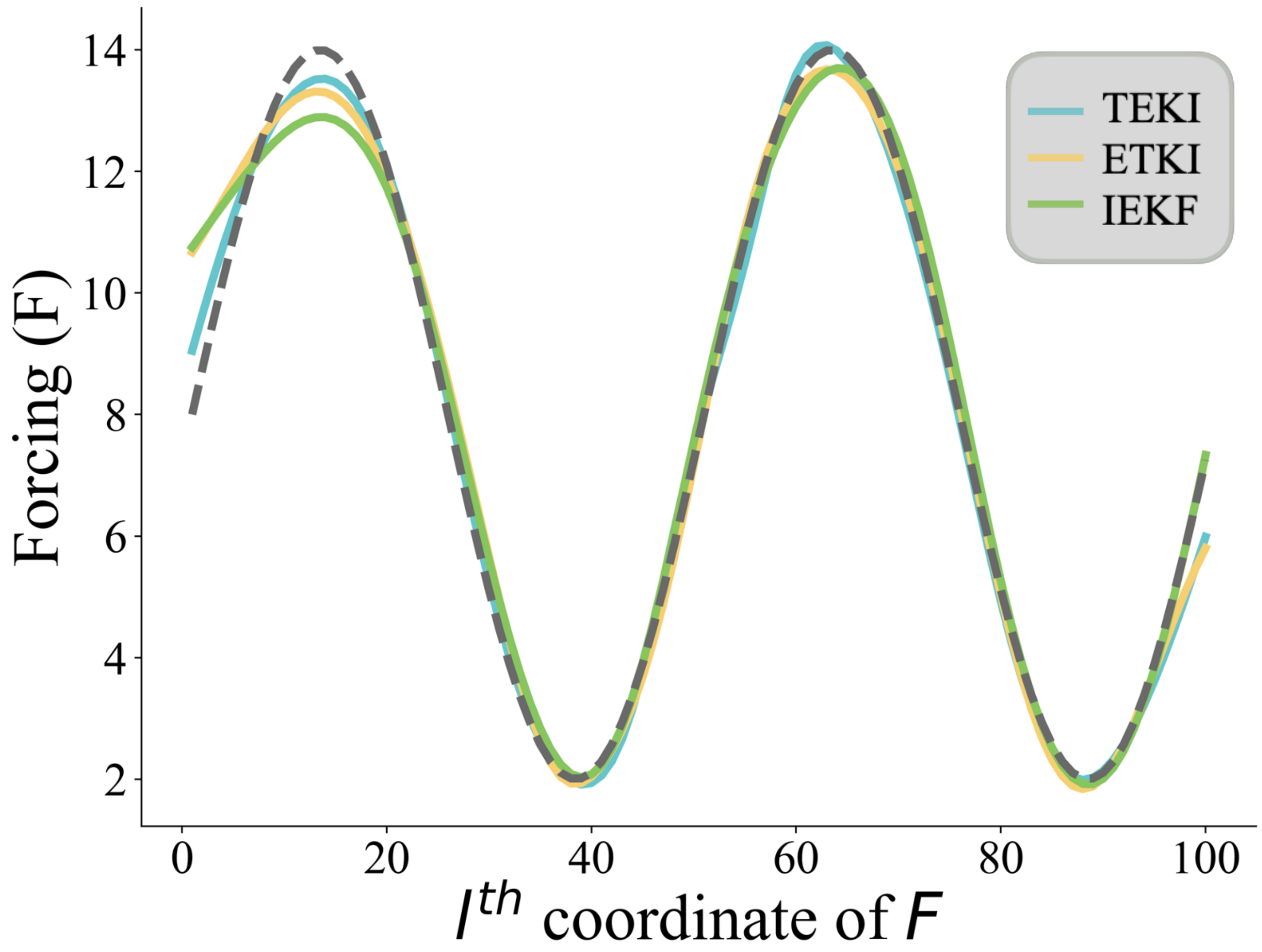
Vishny et al., 2024

Modified Lorenz 96 with neural network results

Target RMSE = 1



UKI and QN did not reach RMSE targets



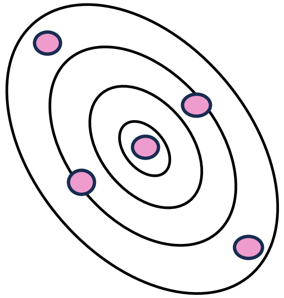
Agenda

1. Introduction and motivation
2. Ensemble-based optimization
3. Example problems
4. ***Summary and conclusions***

Conclusions

Q: Which ensemble-based method is best for estimating parameters from statistics?

- ***UKI***: *Best for problems with small parameter-space dimensions*
    - Does not scale well with the size of the parameter space



- ***IEKF***: *Best for applicable problems*
    - Consistently outperforms ETKI and TEKI for both larger and smaller parameter-space dimensions
    - Does not do well with weak priors

Kalman gain

$$PH^T\left(HPH^T + R\right)^{-1}$$

$$\left[X^{k+1}\right] = \left[X^{k}\right][T]^{n_e \times n_e}$$

- ***ETKI***: *Best for uninformative priors*
    - Consistently outperforms *TEKI*
    - Estimate parameters even with a weak prior
    - Most robust for climate science problems

$$\mathcal{L}(x) = \left\| \begin{pmatrix} R & 0 \\ 0 & P \end{pmatrix}^{-\frac{1}{2}} \left(\begin{pmatrix} y \\ m \end{pmatrix} - \begin{pmatrix} \mathcal{H}(x) \\ x \end{pmatrix} \right) \right\|_2^2$$

- ***Ensemble methods are the right choice for estimating parameters from statistics***
    - Not considered: Localization, inflation, and other algorithm enhancements

UC San Diego | SCRIPPS INSTITUTION OF OCEANOGRAPHY

Thank you

Email: *rgjini@ucsd.edu*

Thank you to the
Office of Naval Research
for supporting this work

Photo credit: Eric Jepsen

Iterative ensemble Kalman filter (IEFK)

1) Ensemble $x_1^k, x_2^k, \ldots, x_{n_e}^k, \qquad h_i^k = \mathcal{H}\left(x_i^k\right), \qquad h_1^k, h_2^k, \ldots, h_{n_e}^k$

2) Ensemble perturbations

$$X^k = \tfrac{1}{\sqrt{n_e - 1}}\left[x_1^k - \bar{x}^k,\ x_2^k - \bar{x}^k,\ \cdots,\ x_{n_e}^k - \bar{x}^k\right],\ \bar{x}^k = \frac{1}{n_e}\sum_{i=1}^{n_e} x_i^k$$

$$Y^k = \tfrac{1}{\sqrt{n_e - 1}}\left[h_1^k - \bar{h}^k,\ h_2^k - \bar{h}^k,\ \cdots,\ h_{n_e}^k - \bar{h}^k\right],\ \bar{h}^k = \frac{1}{n_e}\sum_{i=1}^{n_e} h_i^k$$

3) Covariances

$$C_{xy}^k = X^k \otimes Y^k$$
$$C_{yy}^k = Y^k \otimes Y^k.$$

4) Update

$$\boxed{x_j^{k+1} = x_j^k + \alpha\left\{K^k\left(y - \left(h_j^k + \eta_j^k\right)\right) + \left(I - K^k H^k\right)\left(m - \left(x_j^k + \zeta_j^k\right)\right)\right\}}$$

$$K^k = \left(P \otimes H^k\right)\left(H^k\left(P \otimes H^k\right) + R\right)^{-1}$$
$$H^k = Y^k\left(X^k\right)^{-\dagger}$$

$$\eta_i^k \sim \mathcal{N}(0, R)$$
$$\zeta_i^k \sim \mathcal{N}(0, P)$$

Problem setup

$$\mathcal{L}(x) = \left\|R^{-\frac{1}{2}}\left(y - \mathcal{H}(x)\right)\right\|_2^2$$

$\mathcal{H}(\cdot)$ – nonlinear model
x – parameters
y – observations
R – observation covariance matrix

$$x^{k+1} = x^k + \alpha\delta^k$$

α – step size
δ^k – search direction

Ensemble transform Kalman inversion (ETKI)

1) Ensemble $x_1^k, x_2^k, \ldots, x_{n_e}^k, \qquad h_i^k = \mathcal{H}\left(x_i^k\right), \qquad h_1^k, h_2^k, \ldots, h_{n_e}^k$

2) Ensemble perturbations

$$X^k = \frac{1}{\sqrt{n_e - 1}} \left[x_1^k - \bar{x}^k,\ x_2^k - \bar{x}^k,\ \cdots,\ x_{n_e}^k - \bar{x}^k\right],\ \bar{x}^k = \frac{1}{n_e}\sum_{i=1}^{n_e} x_i^k$$

$$Y^k = \frac{1}{\sqrt{n_e - 1}} \left[h_1^k - \bar{h}^k,\ h_2^k - \bar{h}^k,\ \cdots,\ h_{n_e}^k - \bar{h}^k\right],\ \bar{h}^k = \frac{1}{n_e}\sum_{i=1}^{n_e} h_i^k$$

4) Update

$$\bar{x}^{k+1} = \bar{x}^k + X^k \left(I + \left(Y^k\right)^T R^{-1} Y^k\right)^{-1} \left(Y^k\right)^T R^{-1} \left(y - \bar{h}^k\right)$$

$$X^{k+1} = X^k \left(I + \left(Y^k\right)^T R^{-1} Y^k\right)^{-\frac{1}{2}}$$

$$x_j^{k+1} = \bar{x}^{k+1} + X^{k+1}\sqrt{n_e - 1}.$$

Problem setup

$$\mathcal{L}(x) = \left\| R^{-\frac{1}{2}} \left(y - \mathcal{H}(x)\right) \right\|_2^2$$

$\mathcal{H}(\cdot)$ – nonlinear model
x – parameters
y – observations
R – observation covariance matrix

$$x^{k+1} = x^k + \alpha \delta^k$$

α – step size
δ^k – search direction

Unscented Kalman inversion (UKI)

1) Ensemble

$$x_0^k = \bar{x}^k$$

$$h_i^k = \mathcal{H}\left(x_i^k\right)$$

$$x_j^k = \bar{x}^k + a\sqrt{n_x}\left[\hat{C}_{xx}^k\right]_j \quad (1 \leq j \leq n_x)$$

$$x_{j+n_x}^k = \bar{x}^k - a\sqrt{n_x}\left[\hat{C}_{xx}^k\right]_j \quad (1 \leq j \leq n_x)$$

2) Ensemble perturbations

$$X^k = \tfrac{1}{\sqrt{n_e - 1}}\left[x_1^k - \bar{x}^k,\ x_2^k - \bar{x}^k,\ \cdots,\ x_{n_e}^k - \bar{x}^k\right], \quad \bar{x}^k = \frac{1}{n_e}\sum_{i=1}^{n_e} x_i^k$$

$$Y^k = \tfrac{1}{\sqrt{n_e - 1}}\left[h_1^k - \bar{h}^k,\ h_2^k - \bar{h}^k,\ \cdots,\ h_{n_e}^k - \bar{h}^k\right], \quad \bar{h}^k = h_0^k$$

3) Covariances

$$C_{xy}^k = X^k \otimes Y^k$$

$$C_{yy}^k = Y^k \otimes Y^k$$

$$C_{xx}^k = X^k \otimes X^k$$

4) Update

$$\boxed{\begin{aligned} \bar{x}^{k+1} &= \bar{x}^k + \hat{C}_{xy}^k\left(\hat{C}_{yy}^k + R\right)^{-1}\left(y - \bar{h}^k\right) \\ \hat{C}_{xx}^{k+1} &= \hat{C}_{xx}^k - \hat{C}_{xy}^k\left(\hat{C}_{yy}^k + R\right)^{-1} \otimes \hat{C}_{xy}^k. \end{aligned}}$$

Problem setup

$$\mathcal{L}(x) = \left\|R^{-\frac{1}{2}}\left(y - \mathcal{H}(x)\right)\right\|_2^2$$

$\mathcal{H}(\cdot)$ – nonlinear model
x – parameters
y – observations
R – observation covariance matrix

$$x^{k+1} = x^k + \alpha\delta^k$$

α – step size
δ^k – search direction