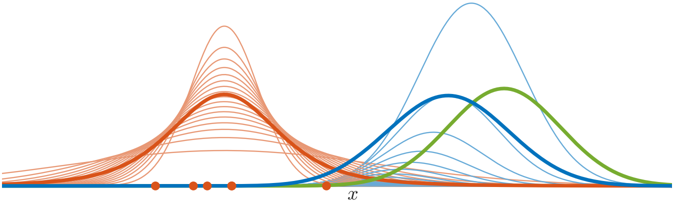
Adaptive covariance inflation in the EnKF by Gaussian scale mixtures



Patrick N. Raanes, Marc Bocquet, Alberto Carrassi

patrick.n.raanes@gmail.com

EnKF Workshop, Bergen, May 2018

Questions ∼answered in paper

- **Nonlinear** models cause **sampling error**.
  - Why and how?
  - Can it be dissociated from **non-Gaussianity**?
- Does the **inherent bias** $\left(\mathbb{E}[\mathrm{tr}(\bar{\mathbf{P}}^a)] < \mathrm{tr}(\mathbf{P}^a)\right)$ cause
  - **collapse** ?
  - **divergence** ?
- Other reasons for **inflating** in nonlinear contexts.
- **Linear** models **attenuate** sampling error. How?
- Is the covariance factor $\frac{1}{N-1}$ optimal?
- How does **localization** affect inflation ?
- How should inflation be defined as a parameter, rather than just a target statistic?
- How does the **feedback** of the EnKF-N compare to "**unbiased**" updates.
- What is the bias of the **estimator** $\hat{\beta}_{\mathbf{R}}$? Why is it better than $\hat{\beta}_{\mathbf{I}}$ or $\hat{\beta}_{\mathrm{ML}}$?

Questions ∼answered in paper

- **Nonlinear** models cause **sampling error**.
  - Why and how?
  - Can it be dissociated from **non-Gaussianity**?
- Does the **inherent bias** $\Big(\mathbb{E}[\mathrm{tr}(\bar{\mathbf{P}}^a)] < \mathrm{tr}(\mathbf{P}^a)\Big)$ cause
  - **collapse** ?
  - **divergence** ?
- Other reasons for **inflating** in nonlinear contexts.
- **Linear** models **attenuate** sampling error. How?
- Is the covariance factor $\frac{1}{N-1}$ optimal?
- How does **localization** affect inflation ?
- How should inflation be defined as a parameter,
  rather than just a target statistic?
- How does the **feedback** of the EnKF-N compare to "**unbiased**" updates.
- What is the bias of the **estimator** $\hat{\beta}_{\mathrm{R}}$? Why is it better than $\hat{\beta}_{\mathbf{I}}$ or $\hat{\beta}_{\mathrm{ML}}$?

Questions ∼answered in paper

- **Nonlinear** models cause **sampling error**.
  - Why and how?
  - Can it be dissociated from **non-Gaussianity**?
- Does the **inherent bias** $\Big(\mathbb{E}[\mathrm{tr}(\bar{\mathbf{P}}^a)] < \mathrm{tr}(\mathbf{P}^a)\Big)$ cause
  - **collapse** ?
  - **divergence** ?
- Other reasons for **inflating** in nonlinear contexts.
- **Linear** models **attenuate** sampling error. How?
- Is the covariance factor $\frac{1}{N-1}$ optimal?
- How does **localization** affect inflation ?
- How should inflation be defined as a parameter,
  rather than just a target statistic?
- How does the **feedback** of the EnKF-N compare to "**unbiased**" updates.
- What is the bias of the **estimator** $\hat{\beta}_{\mathrm{R}}$? Why is it better than $\hat{\beta}_{\mathbf{I}}$ or $\hat{\beta}_{\mathrm{ML}}$?

Questions ∼answered in paper

- **Nonlinear** models cause **sampling error**.
  - Why and how?
  - Can it be dissociated from **non-Gaussianity**?
- Does the **inherent bias** $\left(\mathbb{E}[\mathrm{tr}(\bar{\mathbf{P}}^a)] < \mathrm{tr}(\mathbf{P}^a)\right)$ cause
  - **collapse** ?
  - **divergence** ?
- Other reasons for **inflating** in nonlinear contexts.
- **Linear** models **attenuate** sampling error. How?
- Is the covariance factor $\frac{1}{N-1}$ optimal?
- How does **localization** affect inflation ?
- How should inflation be defined as a parameter,
  rather than just a target statistic?
- How does the **feedback** of the EnKF-N compare to "**unbiased**" updates.
- What is the bias of the **estimator** $\hat{\beta}_{\mathrm{R}}$? Why is it better than $\hat{\beta}_{\mathrm{I}}$ or $\hat{\beta}_{\mathrm{ML}}$?

Questions ∼answered in paper

- **Nonlinear** models cause **sampling error**.
  - Why and how?
  - Can it be dissociated from **non-Gaussianity**?
- Does the **inherent bias** $\left(\mathbb{E}[\mathrm{tr}(\bar{\mathbf{P}}^a)] < \mathrm{tr}(\mathbf{P}^\mathsf{a}) \right)$ cause
  - **collapse** ?
  - **divergence** ?
- Other reasons for **inflating** in nonlinear contexts.
- **Linear** models **attenuate** sampling error. How?
- Is the covariance factor $\frac{1}{N-1}$ optimal?
- How does **localization** affect inflation ?
- How should inflation be defined as a parameter, rather than just a target statistic?
- How does the **feedback** of the EnKF-N compare to "**unbiased**" updates.
- What is the bias of the **estimator** $\hat{\beta}_\mathrm{R}$? Why is it better than $\hat{\beta}_\mathrm{I}$ or $\hat{\beta}_\mathrm{ML}$?

Questions $\sim$answered in paper

- **Nonlinear** models cause **sampling error**.
  - Why and how?
  - Can it be dissociated from **non-Gaussianity**?
- Does the **inherent bias** $\Big(\mathbb{E}[\mathrm{tr}(\bar{\mathbf{P}}^a)] < \mathrm{tr}(\mathbf{P}^{\mathsf{a}})\Big)$ cause
  - **collapse** ?
  - **divergence** ?
- Other reasons for **inflating** in nonlinear contexts.
- **Linear** models **attenuate** sampling error. How?
- Is the covariance factor $\frac{1}{N-1}$ optimal?
- How does **localization** affect inflation ?
- How should inflation be defined as a parameter, rather than just a target statistic?
- How does the **feedback** of the EnKF-N compare to "**unbiased**" updates.
- What is the bias of the **estimator** $\hat{\beta}_{\mathrm{R}}$? Why is it better than $\hat{\beta}_{\mathrm{I}}$ or $\hat{\beta}_{\mathrm{ML}}$?

Questions ~answered in paper

- **Nonlinear** models cause **sampling error**.
  - Why and how?
  - Can it be dissociated from **non-Gaussianity**?
- Does the **inherent bias** $\left(\mathbb{E}[\mathrm{tr}(\bar{\mathbf{P}}^a)] < \mathrm{tr}(\mathbf{P}^{\mathsf{a}}) \right)$ cause
  - **collapse** ?
  - **divergence** ?
- Other reasons for **inflating** in nonlinear contexts.
- **Linear** models **attenuate** sampling error. How?
- Is the covariance factor $\frac{1}{N-1}$ optimal?
- How does **localization** affect inflation ?
- How should inflation be defined as a parameter,
  rather than just a target statistic?
- How does the **feedback** of the EnKF-N compare to "**unbiased**" updates.
- What is the bias of the **estimator** $\hat{\beta}_{\mathrm{R}}$? Why is it better than $\hat{\beta}_{\mathrm{I}}$ or $\hat{\beta}_{\mathrm{ML}}$?

Questions ∼answered in paper

- **Nonlinear** models cause **sampling error**.
    - Why and how?
    - Can it be dissociated from **non-Gaussianity**?
- Does the **inherent bias** $\Big(\mathbb{E}[\mathrm{tr}(\bar{\mathbf{P}}^a)] < \mathrm{tr}(\mathbf{P}^\mathsf{a}) \Big)$ cause
    - **collapse** ?
    - **divergence** ?
- Other reasons for **inflating** in nonlinear contexts.
- **Linear** models **attenuate** sampling error. How?
- Is the covariance factor $\frac{1}{N-1}$ optimal?
- How does **localization** affect inflation ?
- How should inflation be defined as a parameter,
  rather than just a target statistic?
- How does the **feedback** of the EnKF-N compare to "**unbiased**" updates.
- What is the bias of the **estimator** $\hat{\beta}_\mathrm{R}$? Why is it better than $\hat{\beta}_\mathbf{I}$ or $\hat{\beta}_\mathrm{ML}$?

Questions ∼answered in paper

- **Nonlinear** models cause **sampling error**.
  - Why and how?
  - Can it be dissociated from **non-Gaussianity**?
- Does the **inherent bias** $\left(\mathbb{E}[\mathrm{tr}(\bar{\mathbf{P}}^a)] < \mathrm{tr}(\mathbf{P}^{\mathsf{a}}) \right)$ cause
  - **collapse** ?
  - **divergence** ?
- Other reasons for **inflating** in nonlinear contexts.
- **Linear** models **attenuate** sampling error. How?
- Is the covariance factor $\frac{1}{N-1}$ optimal?
- How does **localization** affect inflation ?
- How should inflation be defined as a parameter,
  rather than just a target statistic?
- How does the **feedback** of the EnKF-N compare to "**unbiased**" updates.
- What is the bias of the **estimator** $\hat{\beta}_{\mathrm{R}}$? Why is it better than $\hat{\beta}_{\mathrm{I}}$ or $\hat{\beta}_{\mathrm{ML}}$?

Questions ∼answered in paper

- **Nonlinear** models cause **sampling error**.
  - Why and how?
  - Can it be dissociated from **non-Gaussianity**?
- Does the **inherent bias** $\left(\mathbb{E}[\mathrm{tr}(\bar{\mathbf{P}}^a)] < \mathrm{tr}(\mathbf{P}^{\mathsf{a}}) \right)$ cause
  - **collapse** ?
  - **divergence** ?
- Other reasons for **inflating** in nonlinear contexts.
- **Linear** models **attenuate** sampling error. How?
- Is the covariance factor $\frac{1}{N-1}$ optimal?
- How does **localization** affect inflation ?
- How should inflation be defined as a parameter,
  rather than just a target statistic?
- How does the **feedback** of the EnKF-N compare to "**unbiased**" updates.
- What is the bias of the **estimator** $\hat{\beta}_{\mathrm{R}}$? Why is it better than $\hat{\beta}_{\mathrm{I}}$ or $\hat{\beta}_{\mathrm{ML}}$?

Questions ∼answered in paper

- **Nonlinear** models cause **sampling error**.
  - Why and how?
  - Can it be dissociated from **non-Gaussianity**?
- Does the **inherent bias** $\left(\mathbb{E}[\mathrm{tr}(\bar{\mathbf{P}}^a)] < \mathrm{tr}(\mathbf{P}^{\mathsf{a}})\right)$ cause
  - **collapse** ?
  - **divergence** ?
- Other reasons for **inflating** in nonlinear contexts.
- **Linear** models **attenuate** sampling error. How?
- Is the covariance factor $\frac{1}{N-1}$ optimal?
- How does **localization** affect inflation ?
- How should inflation be defined as a parameter, rather than just a target statistic?
- How does the **feedback** of the EnKF-N compare to "**unbiased**" updates.
- What is the bias of the **estimator** $\hat{\beta}_{\mathrm{R}}$? Why is it better than $\hat{\beta}_{\mathrm{I}}$ or $\hat{\beta}_{\mathrm{ML}}$?

Questions ∼answered in paper

- **Nonlinear** models cause **sampling error**.
  - Why and how?
  - Can it be dissociated from **non-Gaussianity**?
- Does the **inherent bias** $\Big(\mathbb{E}[\mathrm{tr}(\bar{\mathbf{P}}^a)] < \mathrm{tr}(\mathbf{P}^\mathsf{a})\Big)$ cause
  - **collapse** ?
  - **divergence** ?
- Other reasons for **inflating** in nonlinear contexts.
- **Linear** models **attenuate** sampling error. How?
- Is the covariance factor $\frac{1}{N-1}$ optimal?
- How does **localization** affect inflation ?
- How should inflation be defined as a parameter,
  rather than just a target statistic?
- How does the **feedback** of the EnKF-N compare to "**unbiased**" updates.
- What is the bias of the **estimator** $\hat{\beta}_\mathrm{R}$? Why is it better than $\hat{\beta}_\mathrm{I}$ or $\hat{\beta}_\mathrm{ML}$?

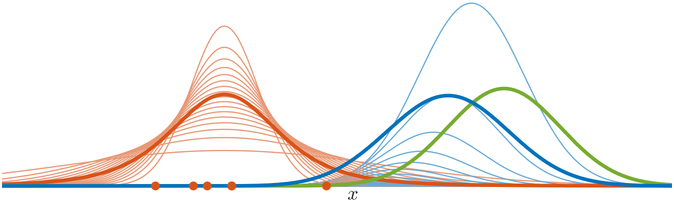
Questions ~answered in paper

- **Nonlinear** models cause **sampling error**.
  - Why and how?
  - Can it be dissociated from **non-Gaussianity**?
- Does the **inherent bias** $\left(\mathbb{E}[\mathrm{tr}(\bar{\mathbf{P}}^a)] < \mathrm{tr}(\mathbf{P}^\mathsf{a}) \right)$ cause
  - **collapse** ?
  - **divergence** ?
- Other reasons for **inflating** in nonlinear contexts.
- **Linear** models **attenuate** sampling error. How?
- Is the covariance factor $\frac{1}{N-1}$ optimal?
- How does **localization** affect inflation ?
- How should inflation be defined as a parameter,
  rather than just a target statistic?
- How does the **feedback** of the EnKF-N compare to "**unbiased**" updates.
- What is the bias of the **estimator** $\hat{\beta}_\mathrm{R}$? Why is it better than $\hat{\beta}_\mathrm{I}$ or $\hat{\beta}_\mathrm{ML}$?

Questions ∼answered in paper

- **Nonlinear** models cause **sampling error**.
  - Why and how?
  - Can it be dissociated from **non-Gaussianity**?
- Does the **inherent bias** $\left(\mathbb{E}[\mathrm{tr}(\bar{\mathbf{P}}^a)] < \mathrm{tr}(\mathbf{P}^\mathsf{a}) \right)$ cause
  - **collapse** ?
  - **divergence** ?
- Other reasons for **inflating** in nonlinear contexts.
- **Linear** models **attenuate** sampling error. How?
- Is the covariance factor $\frac{1}{N-1}$ optimal?
- How does **localization** affect inflation ?
- How should inflation be defined as a parameter,
  rather than just a target statistic?
- How does the **feedback** of the EnKF-N compare to "**unbiased**" updates.
- What is the bias of the **estimator** $\hat{\beta}_\mathbf{R}$? Why is it better than $\hat{\beta}_\mathbf{I}$ or $\hat{\beta}_\mathsf{ML}$?

Adaptive covariance inflation in the EnKF by Gaussian scale mixtures



Patrick N. Raanes, Marc Bocquet, Alberto Carrassi

patrick.n.raanes@gmail.com

NERSC EmblA NordForsk IRIS École des Ponts ParisTech eDF

EnKF Workshop, Bergen, May 2018

Outline

- Idealistic contexts (with sampling error)
  - Revisiting the EnKF assumptions
    $\implies$ Gaussian scale mixture (EnKF-$\mathcal{N}$)
- With model error
  - Survey inflation estimation
    - ETKF-adaptive
    - EAKF-adaptive
  - EnKF-$\mathcal{N}$ hybrid
  - Benchmarks

Outline

- Idealistic contexts (with sampling error)
  - Revisiting the EnKF assumptions
    $\Longrightarrow$ Gaussian scale mixture (EnKF-N)
- With model error
  - Survey inflation estimation
    - ETKF-adaptive
    - EAKF-adaptive
  - EnKF-N hybrid
  - Benchmarks

Outline

- Idealistic contexts (with sampling error)
  - Revisiting the EnKF assumptions

    $\Longrightarrow$ Gaussian scale mixture (EnKF-N)
- With model error
  - Survey inflation estimation
    - ETKF-adaptive
    - EAKF-adaptive
  - EnKF-N hybrid
  - Benchmarks

Outline

- Idealistic contexts (with sampling error)
  - Revisiting the EnKF assumptions

    $\Longrightarrow$ Gaussian scale mixture (EnKF-N)
- With model error
  - Survey inflation estimation
    - ETKF-adaptive
    - EAKF-adaptive
  - EnKF-N hybrid
  - Benchmarks

Outline

- Idealistic contexts (with sampling error)
  - Revisiting the EnKF assumptions

    $\Longrightarrow$ Gaussian scale mixture (EnKF-N)
- With model error
  - Survey inflation estimation
    - ETKF-adaptive
    - EAKF-adaptive
  - EnKF-N hybrid
  - Benchmarks

Outline

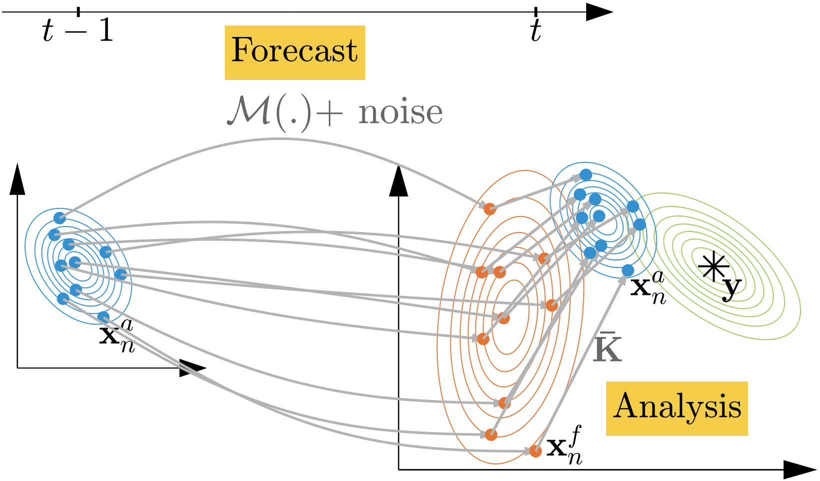
- Idealistic contexts (with sampling error)
  - Revisiting the EnKF assumptions

    $\Longrightarrow$ Gaussian scale mixture (EnKF-N)
- With model error
  - Survey inflation estimation
    - ETKF-adaptive
    - EAKF-adaptive
  - EnKF-N hybrid
  - Benchmarks

Idealistic contexts (EnKF-N)

Assume $\mathcal{M}, \mathcal{H}, \mathbf{Q}, \mathbf{R}$ are perfectly known,
and $p(\boldsymbol{x})$ and $p(\boldsymbol{y}|\boldsymbol{x})$ are always Gaussian.

EnKF



$t-1$
t
Forecast
$\mathcal{M}(.)+$ noise
$\mathbf{x}_n^a$
$\mathbf{x}_n^a$
$\mathbf{y}$
$\bar{\mathbf{K}}$
Analysis
$\mathbf{x}_n^f$

Revisiting EnKF assumptions

Denote $\boldsymbol{y}_{\text{prior}}$ all prior information on the "true" state, $\boldsymbol{x} \in \mathbb{R}^M$, and suppose that, with known mean ($\boldsymbol{b}$) and cov ($\mathbf{B}$),

$$p(\boldsymbol{x}|\boldsymbol{y}_{\text{prior}}) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{b},\mathbf{B}) \,. \tag{1}$$

Computational costs induce:

$$\approx p(\boldsymbol{x}|\mathbf{E}) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{b},\mathbf{B})$$

$\Longrightarrow$ "true" moments, $\boldsymbol{b}$ and $\mathbf{B}$, are unknowns,
to be estimated from $\mathbf{E}$.

Ensemble $\mathbf{E} = \begin{bmatrix} \boldsymbol{x}_1, & \ldots & \boldsymbol{x}_n, & \ldots & \boldsymbol{x}_N \end{bmatrix}$ *also* from (1) and iid.

Revisiting EnKF assumptions

Denote $\boldsymbol{y}_{\text{prior}}$ all prior information on the "true" state, $\boldsymbol{x} \in \mathbb{R}^M$, and suppose that, with known mean (b) and cov ($\mathbf{B}$),

$$p(\boldsymbol{x}|\boldsymbol{y}_{\text{prior}}) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{b}, \mathbf{B})\,. \tag{1}$$

Computational costs induce:

$$\approx p(x|\mathbf{E}) = \mathcal{N}(x|b, \mathbf{B})$$

$\Longrightarrow$ "true" moments, b and $\mathbf{B}$, are unknowns, to be estimated from $\mathbf{E}$.

Ensemble $\mathbf{E} = \begin{bmatrix} x_1, & \dots & x_n, & \dots & x_N \end{bmatrix}$ *also* from (1) and iid.

Revisiting EnKF assumptions

Denote $\boldsymbol{y}_{\text{prior}}$ all prior information on the "true" state, $\boldsymbol{x} \in \mathbb{R}^M$, and suppose that, with known mean $(\boldsymbol{b})$ and cov $(\mathbf{B})$,

$$p(\boldsymbol{x}|\boldsymbol{y}_{\text{prior}}) \quad = \quad \mathcal{N}(\boldsymbol{x}|\boldsymbol{b}, \mathbf{B})\,. \tag{1}$$

Computational costs induce:

$$\approx p(\boldsymbol{x}|\mathbf{E}) \quad = \quad \mathcal{N}(\boldsymbol{x}|\boldsymbol{b}, \mathbf{B})$$

$\Longrightarrow$ "true" moments, $\boldsymbol{b}$ and $\mathbf{B}$, are unknowns, to be estimated from $\mathbf{E}$.

Ensemble $\mathbf{E} = \begin{bmatrix} \boldsymbol{x}_1, & \ldots & \boldsymbol{x}_n, & \ldots & \boldsymbol{x}_N \end{bmatrix}$ *also* from (1) and iid.

Revisiting EnKF assumptions

Denote $\boldsymbol{y}_{\text{prior}}$ all prior information on the "true" state, $\boldsymbol{x} \in \mathbb{R}^M$, and suppose that, with known mean $(\boldsymbol{b})$ and cov $(\mathbf{B})$,

$$p(\boldsymbol{x}|\boldsymbol{y}_{\text{prior}}) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{b}, \mathbf{B})\,. \tag{1}$$

Computational costs induce:

$$\approx p(\boldsymbol{x}|\mathbf{E}) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{b}, \mathbf{B})$$

$\Longrightarrow$ "true" moments, b and $\mathbf{B}$, are unknowns, to be estimated from $\mathbf{E}$.

Ensemble $\mathbf{E} = \left[\boldsymbol{x}_1, \quad \dots \quad \boldsymbol{x}_n, \quad \dots \quad \boldsymbol{x}_N\right]$ *also* from (1) and iid.

Revisiting EnKF assumptions

Denote $\boldsymbol{y}_{\text{prior}}$ all prior information on the "true" state, $\boldsymbol{x} \in \mathbb{R}^M$, and suppose that, with known mean $(\boldsymbol{b})$ and cov $(\mathbf{B})$,

$$p(\boldsymbol{x}|\boldsymbol{y}_{\text{prior}}) \quad = \quad \mathcal{N}(\boldsymbol{x}|\boldsymbol{b},\mathbf{B})\,. \tag{1}$$

Computational costs induce:

$$\approx p(\boldsymbol{x}|\mathbf{E}) \quad = \quad \mathcal{N}(\boldsymbol{x}|\boldsymbol{b},\mathbf{B})$$

$\Longrightarrow$ "true" moments, $\boldsymbol{b}$ and $\mathbf{B}$, are unknowns,
to be estimated from $\mathbf{E}$.

Ensemble $\mathbf{E} = \left[\boldsymbol{x}_1, \quad \ldots \quad \boldsymbol{x}_n, \quad \ldots \quad \boldsymbol{x}_N\right]$ *also* from (1) and iid.

Revisiting EnKF assumptions

Denote $\boldsymbol{y}_{\text{prior}}$ all prior information on the "true" state, $\boldsymbol{x} \in \mathbb{R}^M$, and suppose that, with known mean $(\boldsymbol{b})$ and cov $(\mathbf{B})$,

$$p(\boldsymbol{x}|\boldsymbol{y}_{\text{prior}}) \quad = \quad \mathcal{N}(\boldsymbol{x}|\boldsymbol{b}, \mathbf{B})\,. \tag{1}$$

Computational costs induce:

$$\approx p(\boldsymbol{x}|\mathbf{E}) \quad = \quad \mathcal{N}(\boldsymbol{x}|\boldsymbol{b}, \mathbf{B})$$

$\Longrightarrow$ "true" moments, $\boldsymbol{b}$ and $\mathbf{B}$, are unknowns, to be estimated from $\mathbf{E}$.

Ensemble $\mathbf{E} = \Big[\boldsymbol{x}_1, \quad \ldots \quad \boldsymbol{x}_n, \quad \ldots \quad \boldsymbol{x}_N\Big]$ *also* from (1) and iid.

Revisiting EnKF assumptions

Denote $\boldsymbol{y}_{\text{prior}}$ all prior information on the "true" state, $\boldsymbol{x} \in \mathbb{R}^M$, and suppose that, with known mean $(\boldsymbol{b})$ and cov $(\mathbf{B})$,

$$p(\boldsymbol{x}|\boldsymbol{y}_{\text{prior}}) \quad = \quad \mathcal{N}(\boldsymbol{x}|\boldsymbol{b}, \mathbf{B})\,. \tag{1}$$

Computational costs induce:

$$\approx p(\boldsymbol{x}|\mathbf{E}) \quad = \quad \mathcal{N}(\boldsymbol{x}|\boldsymbol{b}, \mathbf{B})$$

$\implies$ "true" moments, $\boldsymbol{b}$ and $\mathbf{B}$, are unknowns,
to be estimated from $\mathbf{E}$.

Ensemble $\mathbf{E} = \left[\boldsymbol{x}_1, \quad \ldots \quad \boldsymbol{x}_n, \quad \ldots \quad \boldsymbol{x}_N\right]$ *also* from (1) and iid.

Revisiting EnKF assumptions

Denote $\boldsymbol{y}_{\text{prior}}$ all prior information on the "true" state, $\boldsymbol{x} \in \mathbb{R}^M$, and suppose that, with known mean $(\boldsymbol{b})$ and cov $(\mathbf{B})$,

$$p(\boldsymbol{x}|\boldsymbol{y}_{\text{prior}}) \quad = \quad \mathcal{N}(\boldsymbol{x}|\boldsymbol{b}, \mathbf{B})\,. \tag{1}$$

Computational costs induce:

$$\approx p(\boldsymbol{x}|\mathbf{E}) \quad = \quad \iint \mathcal{N}(\boldsymbol{x}|\boldsymbol{b}, \mathbf{B}) \, p(\boldsymbol{b}, \mathbf{B}|\mathbf{E}) \, \mathrm{d}\boldsymbol{b} \, \mathrm{d}\mathbf{B}$$

$\implies$ "true" moments, $\boldsymbol{b}$ and $\mathbf{B}$, are unknowns,
to be estimated from $\mathbf{E}$.

Ensemble $\mathbf{E} = \begin{bmatrix} \boldsymbol{x}_1, & \ldots & \boldsymbol{x}_n, & \ldots & \boldsymbol{x}_N \end{bmatrix}$ *also* from (1) and iid.

Revisiting EnKF assumptions

Denote $\boldsymbol{y}_{\text{prior}}$ all prior information on the "true" state, $\boldsymbol{x} \in \mathbb{R}^M$, and suppose that, with known mean $(\boldsymbol{b})$ and cov $(\mathbf{B})$,

$$p(\boldsymbol{x}|\boldsymbol{y}_{\text{prior}}) \quad = \quad \mathcal{N}(\boldsymbol{x}|\boldsymbol{b}, \mathbf{B})\,. \tag{1}$$

Computational costs induce:

$$\approx p(\boldsymbol{x}|\mathbf{E}) \quad = \iint \mathcal{N}(\boldsymbol{x}|\boldsymbol{b}, \mathbf{B})\, p(\boldsymbol{b}, \mathbf{B}|\mathbf{E})\, \mathrm{d}\boldsymbol{b}\, \mathrm{d}\mathbf{B}$$

$\implies$ "true" moments, $\boldsymbol{b}$ and $\mathbf{B}$, are unknowns,
to be estimated from $\mathbf{E}$.

Ensemble $\mathbf{E} = \begin{bmatrix} \boldsymbol{x}_1, & \ldots & \boldsymbol{x}_n, & \ldots & \boldsymbol{x}_N \end{bmatrix}$ *also* from (1) and iid.

EnKF prior

But

$$p(\boldsymbol{x}|\mathbf{E}) = \int_{\mathcal{B}} \int_{\mathbb{R}^M} \mathcal{N}(\boldsymbol{x}|\boldsymbol{b}, \mathbf{B})\, p(\boldsymbol{b}, \mathbf{B}|\mathbf{E})\, \mathrm{d}\boldsymbol{b}\, \mathrm{d}\mathbf{B} \tag{2}$$

Recover standard EnKF by assuming $N=\infty$ so that

$$p(\boldsymbol{b}, \mathbf{B}|\mathbf{E}) = \delta(\boldsymbol{b} - \bar{\boldsymbol{x}})\delta(\mathbf{B} - \bar{\mathbf{B}}) \,,$$

where

$$\bar{\boldsymbol{x}} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}_n \,, \qquad \bar{\mathbf{B}} = \frac{1}{N-1}\sum_{n=1}^{N} (\boldsymbol{x}_n - \bar{\boldsymbol{x}})(\boldsymbol{x}_n - \bar{\boldsymbol{x}})^{\mathsf{T}} \,. \tag{3}$$

The EnKF-N does not make this approximation.

EnKF prior

But

$$p(\boldsymbol{x}|\mathbf{E}) = \int\limits_{\mathcal{B}} \int\limits_{\mathbb{R}^M} \mathcal{N}(\boldsymbol{x}|\boldsymbol{b}, \mathbf{B})\, p(\boldsymbol{b}, \mathbf{B}|\mathbf{E})\, \mathrm{d}\boldsymbol{b}\, \mathrm{d}\mathbf{B} \tag{2}$$

Recover standard EnKF by assuming $N{=}\infty$ so that

$$p(\boldsymbol{b}, \mathbf{B}|\mathbf{E}) = \delta(\boldsymbol{b} - \bar{\boldsymbol{x}})\delta(\mathbf{B} - \bar{\mathbf{B}})\,,$$

where

$$\bar{\boldsymbol{x}} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}_n\,, \qquad \bar{\mathbf{B}} = \frac{1}{N-1}\sum_{n=1}^{N} (\boldsymbol{x}_n - \bar{\boldsymbol{x}})\,(\boldsymbol{x}_n - \bar{\boldsymbol{x}})^{\mathsf{T}}\ . \tag{3}$$

The EnKF-N does not make this approximation.

EnKF prior

But

$$p(\boldsymbol{x}|\mathbf{E}) = \int_{\mathcal{B}} \int_{\mathbb{R}^M} \mathcal{N}(\boldsymbol{x}|\boldsymbol{b}, \mathbf{B})\, p(\boldsymbol{b}, \mathbf{B}|\mathbf{E})\, \mathrm{d}\boldsymbol{b}\, \mathrm{d}\mathbf{B} \tag{2}$$

Recover standard EnKF by assuming $N{=}\infty$ so that

$$p(\boldsymbol{b}, \mathbf{B}|\mathbf{E}) = \delta(\boldsymbol{b} - \bar{\boldsymbol{x}})\delta(\mathbf{B} - \bar{\mathbf{B}})\,,$$

where

$$\bar{\boldsymbol{x}} = \frac{1}{N} \sum_{n=1}^{N} \boldsymbol{x}_n\,, \qquad \bar{\mathbf{B}} = \frac{1}{N-1} \sum_{n=1}^{N} (\boldsymbol{x}_n - \bar{\boldsymbol{x}})(\boldsymbol{x}_n - \bar{\boldsymbol{x}})^{\mathsf{T}}\ . \tag{3}$$

The EnKF-N does not make this approximation.

EnKF-N via scale mixture

$$\text{Prior:} \quad p(\boldsymbol{x}|\mathbf{E}) = \iint \mathcal{N}(\boldsymbol{x}|\boldsymbol{b}, \mathbf{B})\, p(\boldsymbol{b}, \mathbf{B}|\mathbf{E})\, \mathrm{d}\boldsymbol{b}\, \mathrm{d}\mathbf{B} \tag{4}$$

(5)

(6)

(8)

EnKF-N via scale mixture

$$\text{Prior:} \quad p(\boldsymbol{x}|\mathbf{E}) = \iint \mathcal{N}(\boldsymbol{x}|\boldsymbol{b}, \mathbf{B})\, p(\boldsymbol{b}, \mathbf{B}|\mathbf{E})\, \mathrm{d}\boldsymbol{b}\, \mathrm{d}\mathbf{B} \tag{4}$$

$$\vdots$$

$$\vdots$$

$$\tag{5}$$

$$\vdots$$

$$\vdots \tag{6}$$

$$\vdots$$

$$\propto \left(1 + \tfrac{1}{N-1} \|\boldsymbol{x} - \bar{\boldsymbol{x}}\|^2_{\varepsilon_N \bar{\mathbf{B}}}\right)^{-N/2} \tag{7}$$

$$\tag{8}$$

EnKF-N via scale mixture

Prior: $$p(\boldsymbol{x}|\mathbf{E}) = \iint \mathcal{N}(\boldsymbol{x}|\boldsymbol{b},\mathbf{B})\, p(\boldsymbol{b},\mathbf{B}|\mathbf{E})\,\mathrm{d}\boldsymbol{b}\,\mathrm{d}\mathbf{B} \tag{4}$$

$$\vdots$$

$$\propto \int\limits_{\alpha>0} \mathcal{N}\big(\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|_{\varepsilon_N \bar{\mathbf{B}}}|0,\alpha\big)\; p(\alpha|\mathbf{E})\,\mathrm{d}\alpha \tag{5}$$

$$\vdots$$

$$\vdots \tag{6}$$

$$\vdots$$

$$\propto \left(1 + \tfrac{1}{N-1}\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|^2_{\varepsilon_N \bar{\mathbf{B}}}\right)^{-N/2} \tag{7}$$

(8)

EnKF-N via scale mixture

Prior: $$p(\boldsymbol{x}|\mathbf{E}) = \iint \mathcal{N}(\boldsymbol{x}|\boldsymbol{b},\mathbf{B})\, p(\boldsymbol{b},\mathbf{B}|\mathbf{E})\,\mathrm{d}\boldsymbol{b}\,\mathrm{d}\mathbf{B} \tag{4}$$

$$\vdots$$

$$\propto \int_{\alpha>0} \mathcal{N}\big(\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|_{\varepsilon_N \bar{\mathbf{B}}}|0,\alpha\big)\; p(\alpha|\mathbf{E})\,\mathrm{d}\alpha \tag{5}$$

$$\vdots$$

$$\vdots \tag{6}$$

$$\vdots$$

$$\propto \left(1+\tfrac{1}{N-1}\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|^2_{\varepsilon_N \bar{\mathbf{B}}}\right)^{-N/2} \tag{7}$$

Posterior: $$p(\boldsymbol{x}|\mathbf{E},\boldsymbol{y}) \propto p(\boldsymbol{x}|\mathbf{E})\; \mathcal{N}\big(\boldsymbol{y}|\mathbf{H}\boldsymbol{x},\mathbf{R}\big) \tag{8}$$

EnKF-N via scale mixture

$$\text{Prior:} \quad p(\boldsymbol{x}|\mathbf{E}) = \iint \mathcal{N}(\boldsymbol{x}|\boldsymbol{b}, \mathbf{B})\, p(\boldsymbol{b}, \mathbf{B}|\mathbf{E})\, \mathrm{d}\boldsymbol{b}\, \mathrm{d}\mathbf{B} \tag{4}$$

$$\vdots$$

$$\propto \int\limits_{\alpha>0} \mathcal{N}\big(\|\boldsymbol{x} - \bar{\boldsymbol{x}}\|_{\varepsilon_N \bar{\mathbf{B}}} | 0, \alpha\big)\ p(\alpha|\mathbf{E})\, \mathrm{d}\alpha \tag{5}$$

$$\vdots$$

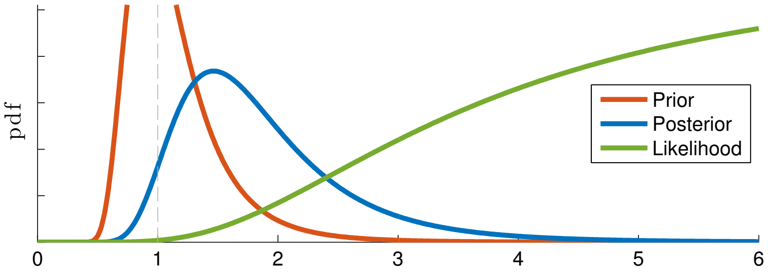
$$\propto \mathcal{N}\big(\boldsymbol{x}|\bar{\boldsymbol{x}}, \alpha(\boldsymbol{x})\bar{\mathbf{B}}\big)\ \tilde{p}\big(\alpha(\boldsymbol{x})|\mathbf{E}\big) \tag{6}$$

$$\vdots$$

$$\propto \left(1 + \tfrac{1}{N-1}\|\boldsymbol{x} - \bar{\boldsymbol{x}}\|^2_{\varepsilon_N \bar{\mathbf{B}}}\right)^{-N/2} \tag{7}$$

$$\text{Posterior:} \quad p(\boldsymbol{x}|\mathbf{E}, \boldsymbol{y}) \propto p(\boldsymbol{x}|\mathbf{E})\ \mathcal{N}\big(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}, \mathbf{R}\big) \tag{8}$$

Mixing distributions – $p(\alpha|\ldots)$



Prior: $$p(\alpha|\mathbf{E}) = \chi^{-2}(\alpha|1, N-1)$$

Likelihood: $$p(\boldsymbol{x}_\star, \boldsymbol{y}|\alpha, \mathbf{E}) \propto \exp\left(-\frac{1}{2}\|\boldsymbol{y}-\mathbf{H}\bar{\boldsymbol{x}}\|^2_{\alpha\varepsilon_N \mathbf{H}\bar{\mathbf{B}}\mathbf{H}^\mathsf{T}+\mathbf{R}}\right)$$

$\Longrightarrow$ Posterior: $$p(\boldsymbol{x}_\star, \alpha|\boldsymbol{y}, \mathbf{E}) \propto \exp\left(-\frac{1}{2}\, D(\alpha)\right)$$

Summary – Perfect model scenario

- Even with a perfect model, Gaussian forecasts, and a deterministic EnKF, "sampling error" arises for $N < \infty$ due to nonlinearity, and inflation is necessary.
- Not assuming $\bar{\mathbf{B}} = \mathbf{B}$ as in the EnKF leads to a Gaussian scale mixture.
- This leads to an adaptive inflation scheme, nullifying the need to tune the inflation factor, and yielding very strong benchmarks in idealistic settings.
- Excellent training for EnKF theory.
  Especially general-purpose inflation estimation.

Summary – Perfect model scenario

- Even with a perfect model, Gaussian forecasts, and a deterministic EnKF, "sampling error" arises for $N < \infty$ due to nonlinearity, and inflation is necessary.
- Not assuming $\bar{\mathrm{B}} = \mathrm{B}$ as in the EnKF leads to a Gaussian scale mixture.
- This leads to an adaptive inflation scheme, nullifying the need to tune the inflation factor, and yielding very strong benchmarks in idealistic settings.
- Excellent training for EnKF theory.
  Especially general-purpose inflation estimation.

Summary – Perfect model scenario

- Even with a perfect model, Gaussian forecasts, and a deterministic EnKF, "sampling error" arises for $N < \infty$ due to nonlinearity, and inflation is necessary.
- Not assuming $\bar{\mathbf{B}} = \mathbf{B}$ as in the EnKF leads to a Gaussian scale mixture.
- This leads to an adaptive inflation scheme, nullifying the need to tune the inflation factor, and yielding very strong benchmarks in idealistic settings.
- Excellent training for EnKF theory. Especially general-purpose inflation estimation.

Summary – Perfect model scenario

- Even with a perfect model, Gaussian forecasts, and a deterministic EnKF, "sampling error" arises for $N < \infty$ due to nonlinearity, and inflation is necessary.
- Not assuming $\bar{\mathbf{B}} = \mathbf{B}$ as in the EnKF leads to a Gaussian scale mixture.
- This leads to an adaptive inflation scheme, nullifying the need to tune the inflation factor, and yielding very strong benchmarks in idealistic settings.
- Excellent training for EnKF theory. Especially general-purpose inflation estimation.

Summary – Perfect model scenario

- Even with a perfect model, Gaussian forecasts, and a deterministic EnKF, "sampling error" arises for $N < \infty$ due to nonlinearity, and inflation is necessary.
- Not assuming $\bar{\mathbf{B}} = \mathbf{B}$ as in the EnKF leads to a Gaussian scale mixture.
- This leads to an adaptive inflation scheme, nullifying the need to tune the inflation factor, and yielding very strong benchmarks in idealistic settings.
- Excellent training for EnKF theory.
  Especially general-purpose inflation estimation.

Summary – Perfect model scenario

- Even with a perfect model, Gaussian forecasts, and a deterministic EnKF, "sampling error" arises for $N < \infty$ due to nonlinearity, and inflation is necessary.
- Not assuming $\bar{\mathbf{B}} = \mathbf{B}$ as in the EnKF leads to a Gaussian scale mixture.
- This leads to an adaptive inflation scheme, nullifying the need to tune the inflation factor, and yielding very strong benchmarks in idealistic settings.
- Excellent training for EnKF theory.
  Especially general-purpose inflation estimation.

With model error

Because all models are wrong.

Fundamentals

- Suppose $\boldsymbol{x}_n \sim \mathcal{N}(\boldsymbol{b}, \mathbf{B}/\beta)$, and $N = \infty$.
  Then there's no mixture, but simply

$$p(\boldsymbol{x}|\beta,) = \mathcal{N}(\boldsymbol{x}|\bar{\boldsymbol{x}}, \beta\bar{\mathbf{B}}) . \tag{9}$$

Fundamentals

- Suppose $\boldsymbol{x}_n \sim \mathcal{N}(\boldsymbol{b}, \mathbf{B}/\beta)$, and $N = \infty$.
  Then there's no mixture, but simply

$$p(\boldsymbol{x}|\beta,) = \mathcal{N}\left(\boldsymbol{x}|\bar{\boldsymbol{x}}, \beta\bar{\mathbf{B}}\right). \tag{9}$$

Fundamentals

- Suppose $\boldsymbol{x}_n \sim \mathcal{N}(\boldsymbol{b}, \mathbf{B}/\beta)$, and $N = \infty$.

  Then there's no mixture, but simply

$$p(\boldsymbol{x}|\beta,) = \mathcal{N}\left(\boldsymbol{x}|\bar{\boldsymbol{x}}, \beta\bar{\mathbf{B}}\right). \tag{9}$$

Fundamentals

- Suppose $\boldsymbol{x}_n \sim \mathcal{N}(\boldsymbol{b}, \mathbf{B}/\beta)$, and $N = \infty$.
  Then there's no mixture, but simply

$$p(\boldsymbol{x}|\beta, \mathbf{E}) = \mathcal{N}\left(\boldsymbol{x}|\bar{\boldsymbol{x}}, \beta\bar{\mathbf{B}}\right) . \tag{9}$$

- Recall

$$p(\boldsymbol{y}|\boldsymbol{x}) = \mathcal{N}\left(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}, \mathbf{R}\right) .$$

- Then

$$\begin{aligned} p(\boldsymbol{y}|\beta) &= \mathcal{N}\left(\boldsymbol{y} \,|\, \mathbf{H}\bar{\boldsymbol{x}}, \bar{\mathbf{C}}(\beta)\right) , \\ \text{where} \quad \bar{\mathbf{C}}(\beta) &= \beta\mathbf{H}\bar{\mathbf{B}}\mathbf{H}^{\mathsf{T}} + \mathbf{R} , \end{aligned} \tag{10}$$

Fundamentals

- Suppose $\boldsymbol{x}_n \sim \mathcal{N}(\boldsymbol{b}, \mathbf{B}/\beta)$, and $N = \infty$.
  Then there's no mixture, but simply

$$p(\boldsymbol{x}|\beta, \cancel{\mathbf{E}}) = \mathcal{N}\left(\boldsymbol{x}|\bar{\boldsymbol{x}}, \beta\bar{\mathbf{B}}\right). \tag{9}$$

- Recall

$$p(\boldsymbol{y}|\boldsymbol{x}) = \mathcal{N}\left(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}, \mathbf{R}\right).$$

- Then

$$p(\boldsymbol{y}|\beta) = \mathcal{N}\left(\boldsymbol{y}\,\middle|\,\mathbf{H}\bar{\boldsymbol{x}}, \bar{\mathbf{C}}(\beta)\right),$$

where $\bar{\mathbf{C}}(\beta) = \beta\mathbf{H}\bar{\mathbf{B}}\mathbf{H}^{\mathsf{T}} + \mathbf{R}$, (10)

Fundamentals

- Suppose $\boldsymbol{x}_n \sim \mathcal{N}(\boldsymbol{b}, \mathbf{B}/\beta)$, and $N = \infty$.
  Then there's no mixture, but simply

$$p(\boldsymbol{x}|\beta, \cancel{\mathbf{E}}) = \mathcal{N}\left(\boldsymbol{x}|\bar{\boldsymbol{x}}, \beta\bar{\mathbf{B}}\right). \tag{9}$$

- Recall

$$p(\boldsymbol{y}|\boldsymbol{x}) = \mathcal{N}\left(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}, \mathbf{R}\right).$$

- Then

$$p(\boldsymbol{y}|\beta) = \mathcal{N}\left(\boldsymbol{y} \,\middle|\, \mathbf{H}\bar{\boldsymbol{x}}, \bar{\mathbf{C}}(\beta)\right),$$

$$\text{where} \quad \bar{\mathbf{C}}(\beta) = \beta\mathbf{H}\bar{\mathbf{B}}\mathbf{H}^{\mathsf{T}} + \mathbf{R}, \tag{10}$$

Fundamentals

- Suppose $\boldsymbol{x}_n \sim \mathcal{N}(\boldsymbol{b}, \mathbf{B}/\beta)$, and $N = \infty$.
  Then there's no mixture, but simply

$$p(\boldsymbol{x}|\beta, \cancel{\mathbf{E}}) = \mathcal{N}\left(\boldsymbol{x}|\bar{\boldsymbol{x}}, \beta\bar{\mathbf{B}}\right). \tag{9}$$

- Recall

$$p(\boldsymbol{y}|\boldsymbol{x}) = \mathcal{N}\left(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}, \mathbf{R}\right).$$

- Then

$$p(\boldsymbol{y}|\beta) = \mathcal{N}\left(\boldsymbol{y} \,\middle|\, \mathbf{H}\bar{\boldsymbol{x}}, \bar{\mathbf{C}}(\beta)\right) \quad ,$$

$$\text{where} \qquad \bar{\mathbf{C}}(\beta) = \beta\mathbf{H}\bar{\mathbf{B}}\mathbf{H}^{\mathsf{T}} + \mathbf{R}\,, \tag{10}$$

Fundamentals

- Suppose $\boldsymbol{x}_n \sim \mathcal{N}(\boldsymbol{b}, \mathbf{B}/\beta)$, and $N = \infty$.
  Then there's no mixture, but simply

$$p(\boldsymbol{x}|\beta, \cancel{\mathbf{E}}) = \mathcal{N}\left(\boldsymbol{x}|\bar{\boldsymbol{x}}, \beta\bar{\mathbf{B}}\right). \tag{9}$$

- Recall

$$p(\boldsymbol{y}|\boldsymbol{x}) = \mathcal{N}\left(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}, \mathbf{R}\right).$$

- Then

$$p(\boldsymbol{y}|\beta) = \mathcal{N}\left(\boldsymbol{y} \,\middle|\, \mathbf{H}\bar{\boldsymbol{x}}, \bar{\mathbf{C}}(\beta)\right) = \mathcal{N}\left(\bar{\boldsymbol{\delta}} \,\middle|\, \mathbf{0}, \bar{\mathbf{C}}(\beta)\right),$$

$$\text{where} \qquad \begin{aligned} \bar{\mathbf{C}}(\beta) &= \beta\mathbf{H}\bar{\mathbf{B}}\mathbf{H}^{\mathsf{T}} + \mathbf{R}, \\ \bar{\boldsymbol{\delta}} &= \boldsymbol{y} - \mathbf{H}\bar{\boldsymbol{x}}. \end{aligned} \tag{10}$$

ETKF adaptive inflation

- Again,

$$p(y|\beta) = \mathcal{N}\left(\bar{\delta} \,\middle|\, 0, \bar{\mathbf{C}}(\beta)\right), \tag{11}$$

$$\text{where} \qquad \bar{\mathbf{C}}(\beta) = \beta \mathbf{H}\bar{\mathbf{B}}\mathbf{H}^\mathsf{T} + \mathbf{R} \approx \bar{\delta}\bar{\delta}^\mathsf{T}. \tag{12}$$

- "yielding" (Wang and Bishop, 2003)

$$\hat{\beta}_{\mathbf{R}} = \frac{\|\bar{\delta}\|^2_{\mathbf{R}}/P - 1}{\bar{\sigma}^2},$$

  where $P = \text{length}(y)$ and $\bar{\sigma}^2 = \text{tr}(\mathbf{H}\bar{\mathbf{B}}\mathbf{H}^\mathsf{T}\mathbf{R}^{-1})/P$.

- Also considered: $\hat{\beta}_{\mathbf{I}}$, $\hat{\beta}_{\mathbf{H}\bar{\mathbf{B}}\mathbf{H}^\mathsf{T}}$, $\hat{\beta}_{\bar{\mathbf{C}}(1)}$, ML, VB (EM).

ETKF adaptive inflation

- Again,

$$p(\boldsymbol{y}|\beta) = \mathcal{N}\big(\bar{\boldsymbol{\delta}} \,\big|\, \mathbf{0}, \bar{\mathbf{C}}(\beta)\big) \,, \tag{11}$$

where $$\bar{\mathbf{C}}(\beta) = \beta \mathbf{H}\bar{\mathbf{B}}\mathbf{H}^\mathsf{T} + \mathbf{R} \approx \overline{\delta}\,\overline{\delta}^\mathsf{T} \,. \tag{12}$$

- "yielding" (Wang and Bishop, 2003)

$$\hat{\beta}_{\mathbf{R}} = \frac{\|\bar{\delta}\|^2_{\mathbf{R}}/P - 1}{\bar{\sigma}^2} \,,$$

where $P = \text{length}(y)$ and $\bar{\sigma}^2 = \text{tr}(\mathbf{H}\bar{\mathbf{B}}\mathbf{H}^\mathsf{T}\mathbf{R}^{-1})/P$.

- Also considered: $\hat{\beta}_{\mathbf{I}}$, $\hat{\beta}_{\mathbf{H}\bar{\mathbf{B}}\mathbf{H}^\mathsf{T}}$, $\hat{\beta}_{\bar{\mathbf{C}}(1)}$, ML, VB (EM).

ETKF adaptive inflation

- Again,

$$p(\boldsymbol{y}|\beta) = \mathcal{N}\big(\bar{\boldsymbol{\delta}} \,\big|\, \mathbf{0}, \bar{\mathbf{C}}(\beta)\big)\,, \tag{11}$$

where

$$\bar{\mathbf{C}}(\beta) = \beta \mathbf{H}\bar{\mathbf{B}}\mathbf{H}^\mathsf{T} + \mathbf{R} \approx \bar{\boldsymbol{\delta}}\bar{\boldsymbol{\delta}}^\mathsf{T}\,. \tag{12}$$

- "yielding" (Wang and Bishop, 2003)

$$\hat{\beta}_{\mathbf{R}} = \frac{\|\bar{\boldsymbol{\delta}}\|^2_{\mathbf{R}}/P - 1}{\bar{\sigma}^2}\,,$$

where $P = \text{length}(\boldsymbol{y})$ and $\bar{\sigma}^2 = \text{tr}(\mathbf{H}\bar{\mathbf{B}}\mathbf{H}^\mathsf{T}\mathbf{R}^{-1})/P$.

- Also considered: $\hat{\beta}_{\mathbf{I}}$, $\hat{\beta}_{\mathbf{H}\bar{\mathbf{B}}\mathbf{H}^\mathsf{T}}$, $\hat{\beta}_{\bar{\mathbf{C}}(1)}$, ML, VB (EM).

ETKF adaptive inflation

- Again,

$$p(\boldsymbol{y}|\beta) = \mathcal{N}(\bar{\boldsymbol{\delta}} \,|\, \mathbf{0}, \bar{\mathbf{C}}(\beta)) \,, \tag{11}$$

where $$\bar{\mathbf{C}}(\beta) = \beta \mathbf{H}\bar{\mathbf{B}}\mathbf{H}^{\mathsf{T}} + \mathbf{R} \approx \bar{\boldsymbol{\delta}}\bar{\boldsymbol{\delta}}^{\mathsf{T}} \,. \tag{12}$$

- "yielding" (Wang and Bishop, 2003)

$$\hat{\beta}_{\mathbf{R}} = \frac{\|\bar{\boldsymbol{\delta}}\|^2_{\mathbf{R}}/P - 1}{\bar{\sigma}^2} \,,$$

where $P = \text{length}(\boldsymbol{y})$ and $\bar{\sigma}^2 = \mathrm{tr}(\mathbf{H}\bar{\mathbf{B}}\mathbf{H}^{\mathsf{T}}\mathbf{R}^{-1})/P$.

- Also considered: $\hat{\beta}_{\mathbf{I}}$, $\hat{\beta}_{\mathbf{H}\bar{\mathbf{B}}\mathbf{H}^{\mathsf{T}}}$, $\hat{\beta}_{\bar{\mathbf{C}}(1)}$, ML, VB (EM).

ETKF adaptive inflation

- Again,
$$p(\boldsymbol{y}|\beta) = \mathcal{N}\left(\bar{\boldsymbol{\delta}} \,\middle|\, \mathbf{0}, \bar{\mathbf{C}}(\beta)\right), \tag{11}$$
where $$\bar{\mathbf{C}}(\beta) = \beta \mathbf{H}\bar{\mathbf{B}}\mathbf{H}^\mathsf{T} + \mathbf{R} \approx \bar{\boldsymbol{\delta}}\bar{\boldsymbol{\delta}}^\mathsf{T}. \tag{12}$$

- "yielding" (Wang and Bishop, 2003)
$$\hat{\beta}_{\mathbf{R}} = \frac{\|\bar{\boldsymbol{\delta}}\|^2_{\mathbf{R}}/P - 1}{\bar{\sigma}^2},$$
where $P = \text{length}(\boldsymbol{y})$ and $\bar{\sigma}^2 = \mathrm{tr}(\mathbf{H}\bar{\mathbf{B}}\mathbf{H}^\mathsf{T}\mathbf{R}^{-1})/P$.

- Also considered: $\hat{\beta}_{\mathbf{I}}$, $\hat{\beta}_{\mathbf{H}\bar{\mathbf{B}}\mathbf{H}^\mathsf{T}}$, $\hat{\beta}_{\bar{\mathbf{C}}(1)}$, ML, VB (EM).

ETKF adaptive inflation

- Again,

$$p(\boldsymbol{y}|\beta) = \mathcal{N}\left(\bar{\boldsymbol{\delta}} \,\middle|\, \mathbf{0}, \bar{\mathbf{C}}(\beta)\right), \tag{11}$$

where $$\bar{\mathbf{C}}(\beta) = \beta \mathbf{H}\bar{\mathbf{B}}\mathbf{H}^{\mathsf{T}} + \mathbf{R} \approx \bar{\boldsymbol{\delta}}\bar{\boldsymbol{\delta}}^{\mathsf{T}}. \tag{12}$$

- "yielding" (Wang and Bishop, 2003)

$$\hat{\beta}_{\mathbf{R}} = \frac{\|\bar{\boldsymbol{\delta}}\|^2_{\mathbf{R}}/P - 1}{\bar{\sigma}^2},$$

  where $P = \text{length}(\boldsymbol{y})$ and $\bar{\sigma}^2 = \mathrm{tr}(\mathbf{H}\bar{\mathbf{B}}\mathbf{H}^{\mathsf{T}}\mathbf{R}^{-1})/P$.

- Also considered: $\hat{\beta}_{\mathbf{I}}$, $\hat{\beta}_{\mathbf{H}\bar{\mathbf{B}}\mathbf{H}^{\mathsf{T}}}$, $\hat{\beta}_{\bar{\mathbf{C}}(1)}$, ML, VB (EM).

ETKF adaptive inflation

- Again,

$$p(\boldsymbol{y}|\beta) = \mathcal{N}\left(\bar{\boldsymbol{\delta}} \,\middle|\, \mathbf{0}, \bar{\mathbf{C}}(\beta)\right), \tag{11}$$

where
$$\bar{\mathbf{C}}(\beta) = \beta \mathbf{H}\bar{\mathbf{B}}\mathbf{H}^{\mathsf{T}} + \mathbf{R} \approx \bar{\boldsymbol{\delta}}\bar{\boldsymbol{\delta}}^{\mathsf{T}}. \tag{12}$$

- "yielding" (Wang and Bishop, 2003)

$$\hat{\beta}_{\mathbf{R}} = \frac{\|\bar{\boldsymbol{\delta}}\|^2_{\mathbf{R}}/P - 1}{\bar{\sigma}^2},$$

where $P = \text{length}(\boldsymbol{y})$ and $\bar{\sigma}^2 = \mathrm{tr}(\mathbf{H}\bar{\mathbf{B}}\mathbf{H}^{\mathsf{T}}\mathbf{R}^{-1})/P$.

- Also considered: $\hat{\beta}_{\mathbf{I}}$, $\hat{\beta}_{\mathbf{H}\bar{\mathbf{B}}\mathbf{H}^{\mathsf{T}}}$, $\hat{\beta}_{\bar{\mathbf{C}}(1)}$, ML, VB (EM).

Renouncing Gaussianity

- Assume $\mathbf{H}\bar{\mathbf{B}}\mathbf{H}^\mathsf{T} \propto \mathbf{R}$.
- The likelihood $p(\boldsymbol{y}|\beta) = \mathcal{N}\left(\bar{\boldsymbol{\delta}} \,\middle|\, \mathbf{0}, \bar{\mathbf{C}}(\beta)\right)$ becomes

$$p(\boldsymbol{y}|\beta) \propto \chi^{+2}\left(\|\bar{\boldsymbol{\delta}}\|^2_{\mathbf{R}}/P \,\middle|\, (1+\bar{\sigma}^2\beta), P\right). \qquad (13)$$

- Surprise !!!: $\operatorname{argmax} p(\boldsymbol{y}|\beta) = \hat{\beta}_{\mathbf{R}}$.
- A further approximation is fitted:

$$p(\boldsymbol{y}|\beta) \approx \chi^{+2}(\hat{\beta}_{\mathbf{R}}|\beta, \hat{\nu}) . \qquad (14)$$

- Likelihood (14) fits mode of (13). Fitting curvature $\Longrightarrow \hat{\nu}$ $\Longrightarrow$ same variance as in Miyoshi (2011) !!!
- Likelihood (14) conjugate to $p(\beta) = \chi^{-2}(\beta|\beta^\mathsf{f}, \nu^\mathsf{f})$, yielding

$$\nu^\mathsf{a} = \nu^\mathsf{f} + \hat{\nu} , \qquad (15)$$

$$\beta^\mathsf{a} = (\nu^\mathsf{f}\beta^\mathsf{f} + \hat{\nu}\hat{\beta}_{\mathbf{R}})/\nu^\mathsf{a} , \qquad (16)$$

again, as in Miyoshi (2011).

Renouncing Gaussianity

- Assume $\mathbf{H}\bar{\mathbf{B}}\mathbf{H}^{\mathsf{T}} \propto \mathbf{R}$.
- The likelihood $p(y|\beta) = \mathcal{N}(\bar{\delta} \,|\, \mathbf{0}, \bar{\mathbf{C}}(\beta))$ becomes

$$p(y|\beta) \propto \chi^{+2}\left(\|\bar{\delta}\|^2_{\mathbf{R}}/P \,\middle|\, (1+\bar{\sigma}^2\beta), P\right). \tag{13}$$

- Surprise !!!: $\operatorname{argmax} p(y|\beta) = \hat{\beta}_{\mathbf{R}}$.
- A further approximation is fitted:

$$p(y|\beta) \approx \chi^{+2}(\hat{\beta}_{\mathbf{R}}|\beta, \hat{\nu}). \tag{14}$$

- Likelihood (14) fits mode of (13). Fitting curvature $\implies \hat{\nu}$ $\implies$ same variance as in Miyoshi (2011) !!!
- Likelihood (14) conjugate to $p(\beta) = \chi^{-2}(\beta|\beta^{\mathrm{f}}, \nu^{\mathrm{f}})$, yielding

$$\nu^{\mathrm{a}} = \nu^{\mathrm{f}} + \hat{\nu}, \tag{15}$$

$$\beta^{\mathrm{a}} = (\nu^{\mathrm{f}}\beta^{\mathrm{f}} + \hat{\nu}\hat{\beta}_{\mathbf{R}})/\nu^{\mathrm{a}}, \tag{16}$$

- again, as in Miyoshi (2011).

Renouncing Gaussianity

- Assume $\mathbf{H}\bar{\mathbf{B}}\mathbf{H}^{\mathsf{T}} \propto \mathbf{R}$.
- The likelihood $p(\boldsymbol{y}|\beta) = \mathcal{N}\big(\bar{\boldsymbol{\delta}} \,\big|\, \mathbf{0}, \bar{\mathbf{C}}(\beta)\big)$ becomes

$$p(\boldsymbol{y}|\beta) \propto \chi^{+2}\Big(\|\bar{\boldsymbol{\delta}}\|^2_{\mathbf{R}}/P \,\Big|\, (1+\bar{\sigma}^2\beta), P\Big) \,. \tag{13}$$

- Surprise !!!: $\operatorname{argmax} p(\boldsymbol{y}|\beta) = \hat{\beta}_{\mathbf{R}}$.
- A further approximation is fitted:

$$p(\boldsymbol{y}|\beta) \approx \chi^{+2}(\hat{\beta}_{\mathbf{R}}|\beta, \hat{\nu}) \,. \tag{14}$$

- Likelihood (14) fits mode of (13). Fitting curvature $\implies \hat{\nu}$ $\implies$ same variance as in Miyoshi (2011) !!!
- Likelihood (14) conjugate to $p(\beta) = \chi^{-2}(\beta|\beta^{\mathsf{f}}, \nu^{\mathsf{f}})$, yielding

$$\nu^{\mathsf{a}} = \nu^{\mathsf{f}} + \hat{\nu} \,, \tag{15}$$

$$\beta^{\mathsf{a}} = (\nu^{\mathsf{f}}\beta^{\mathsf{f}} + \hat{\nu}\hat{\beta}_{\mathbf{R}})/\nu^{\mathsf{a}} \,, \tag{16}$$

again, as in Miyoshi (2011).

Renouncing Gaussianity

- Assume $\mathbf{H}\bar{\mathbf{B}}\mathbf{H}^{\mathsf{T}} \propto \mathbf{R}$.
- The likelihood $p(\boldsymbol{y}|\beta) = \mathcal{N}\big(\bar{\boldsymbol{\delta}} \,\big|\, \mathbf{0}, \bar{\mathbf{C}}(\beta)\big)$ becomes

$$p(\boldsymbol{y}|\beta) \propto \chi^{+2}\Big(\|\bar{\boldsymbol{\delta}}\|^2_{\mathbf{R}}/P \,\Big|\, (1+\bar{\sigma}^2\beta), P\Big)\,. \tag{13}$$

- Surprise !!!: $\operatorname{argmax} p(\boldsymbol{y}|\beta) = \hat{\beta}_{\mathbf{R}}$.
- A further approximation is fitted:

$$p(\boldsymbol{y}|\beta) \approx \chi^{+2}(\hat{\beta}_{\mathbf{R}}|\beta, \hat{\nu})\,. \tag{14}$$

- Likelihood (14) fits mode of (13). Fitting curvature $\implies \hat{\nu}$ $\implies$ same variance as in Miyoshi (2011) !!!
- Likelihood (14) conjugate to $p(\beta) = \chi^{-2}(\beta|\beta^{\mathsf{f}}, \nu^{\mathsf{f}})$, yielding

$$\nu^{\mathsf{a}} = \nu^{\mathsf{f}} + \hat{\nu}\,, \tag{15}$$

$$\beta^{\mathsf{a}} = (\nu^{\mathsf{f}}\beta^{\mathsf{f}} + \hat{\nu}\hat{\beta}_{\mathbf{R}})/\nu^{\mathsf{a}}\,, \tag{16}$$

again, as in Miyoshi (2011).

Renouncing Gaussianity

- Assume $\mathbf{H}\bar{\mathbf{B}}\mathbf{H}^{\mathsf{T}} \propto \mathbf{R}$.
- The likelihood $p(\boldsymbol{y}|\beta) = \mathcal{N}\big(\bar{\boldsymbol{\delta}} \,\big|\, \mathbf{0}, \bar{\mathbf{C}}(\beta)\big)$ becomes

$$p(\boldsymbol{y}|\beta) \propto \chi^{+2}\Big(\|\bar{\boldsymbol{\delta}}\|^2_{\mathbf{R}}/P \,\Big|\, (1+\bar{\sigma}^2\beta), P\Big)\,. \tag{13}$$

- Surprise !!!: $\operatorname{argmax} p(\boldsymbol{y}|\beta) = \hat{\beta}_{\mathbf{R}}$,
- A further approximation is fitted:

$$p(\boldsymbol{y}|\beta) \approx \chi^{+2}(\hat{\beta}_{\mathbf{R}}|\beta, \hat{\nu})\,. \tag{14}$$

- Likelihood (14) fits mode of (13). Fitting curvature $\implies \hat{\nu}$ $\implies$ same variance as in Miyoshi (2011) !!!
- Likelihood (14) conjugate to $p(\beta) = \chi^{-2}(\beta|\beta^{\mathsf{f}}, \nu^{\mathsf{f}})$, yielding

$$\nu^{\mathsf{a}} = \nu^{\mathsf{f}} + \hat{\nu}\,, \tag{15}$$

$$\beta^{\mathsf{a}} = (\nu^{\mathsf{f}}\beta^{\mathsf{f}} + \hat{\nu}\hat{\beta}_{\mathbf{R}})/\nu^{\mathsf{a}}\,, \tag{16}$$

again, as in Miyoshi (2011).

Renouncing Gaussianity

- Assume $\mathbf{H}\bar{\mathbf{B}}\mathbf{H}^\mathsf{T} \propto \mathbf{R}$.
- The likelihood $p(\boldsymbol{y}|\beta) = \mathcal{N}\big(\bar{\boldsymbol{\delta}} \,\big|\, \mathbf{0}, \bar{\mathbf{C}}(\beta)\big)$ becomes

$$p(\boldsymbol{y}|\beta) \propto \chi^{+2}\Big(\|\bar{\boldsymbol{\delta}}\|^2_{\mathbf{R}}/P \,\Big|\, (1+\bar{\sigma}^2\beta), P\Big)\,. \tag{13}$$

- Surprise !!!: $\operatorname{argmax} p(\boldsymbol{y}|\beta) = \hat{\beta}_{\mathbf{R}}$,
- A further approximation is fitted:

$$p(\boldsymbol{y}|\beta) \approx \chi^{+2}(\hat{\beta}_{\mathbf{R}}|\beta, \hat{\nu})\,. \tag{14}$$

- Likelihood (14) fits mode of (13). Fitting curvature $\Longrightarrow \hat{\nu}$ $\Longrightarrow$ same variance as in Miyoshi (2011) !!!
- Likelihood (14) conjugate to $p(\beta) = \chi^{-2}(\beta|\beta^f, \nu^f)$, yielding

$$\nu^a = \nu^f + \hat{\nu}\,, \tag{15}$$

$$\beta^a = (\nu^f\beta^f + \hat{\nu}\hat{\beta}_{\mathbf{R}})/\nu^a\,, \tag{16}$$

again, as in Miyoshi (2011).

Renouncing Gaussianity

- Assume $\mathbf{H}\bar{\mathbf{B}}\mathbf{H}^{\mathsf{T}} \propto \mathbf{R}$.
- The likelihood $p(\boldsymbol{y}|\beta) = \mathcal{N}\big(\bar{\boldsymbol{\delta}} \,\big|\, \mathbf{0}, \bar{\mathbf{C}}(\beta)\big)$ becomes

$$p(\boldsymbol{y}|\beta) \propto \chi^{+2}\Big(\|\bar{\boldsymbol{\delta}}\|^2_{\mathbf{R}}/P \,\Big|\, (1+\bar{\sigma}^2\beta), P\Big)\,. \tag{13}$$

- Surprise !!!: $\operatorname{argmax} p(\boldsymbol{y}|\beta) = \hat{\beta}_{\mathbf{R}}$,
- A further approximation is fitted:

$$p(\boldsymbol{y}|\beta) \approx \chi^{+2}(\hat{\beta}_{\mathbf{R}}|\beta, \hat{\nu})\,. \tag{14}$$

- Likelihood (14) fits mode of (13). Fitting curvature $\Longrightarrow$ $\hat{\nu}$ $\Longrightarrow$ same variance as in Miyoshi (2011) !!!
- Likelihood (14) conjugate to $p(\beta) = \chi^{-2}(\beta|\beta^{\mathsf{f}}, \nu^{\mathsf{f}})$, yielding

$$\nu^{\mathsf{a}} = \nu^{\mathsf{f}} + \hat{\nu}\,, \tag{15}$$

$$\beta^{\mathsf{a}} = (\nu^{\mathsf{f}}\beta^{\mathsf{f}} + \hat{\nu}\hat{\beta}_{\mathbf{R}})/\nu^{\mathsf{a}}\,, \tag{16}$$

again, as in Miyoshi (2011).

Renouncing Gaussianity

- Assume $\mathbf{H}\bar{\mathbf{B}}\mathbf{H}^{\mathsf{T}} \propto \mathbf{R}$.
- The likelihood $p(\boldsymbol{y}|\beta) = \mathcal{N}\left(\bar{\boldsymbol{\delta}} \,\middle|\, \mathbf{0}, \bar{\mathbf{C}}(\beta)\right)$ becomes

$$p(\boldsymbol{y}|\beta) \propto \chi^{+2}\left(\|\bar{\boldsymbol{\delta}}\|^2_{\mathbf{R}}/P \,\middle|\, (1+\bar{\sigma}^2\beta), P\right). \tag{13}$$

- Surprise !!!: $\operatorname{argmax} p(\boldsymbol{y}|\beta) = \hat{\beta}_{\mathbf{R}}$,
- A further approximation is fitted:

$$p(\boldsymbol{y}|\beta) \approx \chi^{+2}(\hat{\beta}_{\mathbf{R}}|\beta, \hat{\nu}). \tag{14}$$

- Likelihood (14) fits mode of (13). Fitting curvature $\Longrightarrow$ $\hat{\nu}$ $\Longrightarrow$ same variance as in Miyoshi (2011) !!!
- Likelihood (14) conjugate to $p(\beta) = \chi^{-2}(\beta|\beta^{\mathsf{f}}, \nu^{\mathsf{f}})$, yielding

$$\nu^{\mathsf{a}} = \nu^{\mathsf{f}} + \hat{\nu}, \tag{15}$$

$$\beta^{\mathsf{a}} = (\nu^{\mathsf{f}}\beta^{\mathsf{f}} + \hat{\nu}\hat{\beta}_{\mathbf{R}})/\nu^{\mathsf{a}}, \tag{16}$$

again, as in Miyoshi (2011).

Renouncing Gaussianity

- Assume $\mathbf{H}\bar{\mathbf{B}}\mathbf{H}^{\mathsf{T}} \propto \mathbf{R}$.
- The likelihood $p(\boldsymbol{y}|\beta) = \mathcal{N}\big(\bar{\boldsymbol{\delta}} \,\big|\, \mathbf{0}, \bar{\mathbf{C}}(\beta)\big)$ becomes

$$p(\boldsymbol{y}|\beta) \propto \chi^{+2}\Big(\|\bar{\boldsymbol{\delta}}\|^2_{\mathbf{R}}/P \,\Big|\, (1+\bar{\sigma}^2\beta), P\Big)\,. \tag{13}$$

- Surprise !!!: $\operatorname{argmax} p(\boldsymbol{y}|\beta) = \hat{\beta}_{\mathbf{R}}$,
- A further approximation is fitted:

$$p(\boldsymbol{y}|\beta) \approx \chi^{+2}(\hat{\beta}_{\mathbf{R}}|\beta, \hat{\nu})\,. \tag{14}$$

- Likelihood (14) fits mode of (13). Fitting curvature $\Longrightarrow$ $\hat{\nu}$ $\Longrightarrow$ same variance as in Miyoshi (2011) !!!
- Likelihood (14) conjugate to $p(\beta) = \chi^{-2}(\beta|\beta^{\mathsf{f}}, \nu^{\mathsf{f}})$, yielding

$$\nu^{\mathsf{a}} = \nu^{\mathsf{f}} + \hat{\nu}\,, \tag{15}$$

$$\beta^{\mathsf{a}} = (\nu^{\mathsf{f}}\beta^{\mathsf{f}} + \hat{\nu}\hat{\beta}_{\mathbf{R}})/\nu^{\mathsf{a}}\,, \tag{16}$$

again, as in Miyoshi (2011).

Renouncing Gaussianity

- Assume $\mathbf{H}\bar{\mathbf{B}}\mathbf{H}^{\mathsf{T}} \propto \mathbf{R}$.
- The likelihood $p(\boldsymbol{y}|\beta) = \mathcal{N}(\bar{\boldsymbol{\delta}} \,|\, \mathbf{0}, \bar{\mathbf{C}}(\beta))$ becomes

$$p(\boldsymbol{y}|\beta) \propto \chi^{+2}\Big(\|\bar{\boldsymbol{\delta}}\|^2_{\mathbf{R}}/P \,\Big|\, (1+\bar{\sigma}^2\beta), P\Big)\,. \tag{13}$$

- Surprise !!!: $\operatorname{argmax} p(\boldsymbol{y}|\beta) = \hat{\beta}_{\mathbf{R}}$,
- A further approximation is fitted:

$$p(\boldsymbol{y}|\beta) \approx \chi^{+2}(\hat{\beta}_{\mathbf{R}}|\beta, \hat{\nu})\,. \tag{14}$$

- Likelihood (14) fits mode of (13). Fitting curvature $\Longrightarrow \hat{\nu}$ $\Longrightarrow$ same variance as in Miyoshi (2011) !!!
- Likelihood (14) conjugate to $p(\beta) = \chi^{-2}(\beta|\beta^{\mathsf{f}}, \nu^{\mathsf{f}})$, yielding

$$\nu^{\mathsf{a}} = \nu^{\mathsf{f}} + \hat{\nu}\,, \tag{15}$$

$$\beta^{\mathsf{a}} = (\nu^{\mathsf{f}}\beta^{\mathsf{f}} + \hat{\nu}\hat{\beta}_{\mathbf{R}})/\nu^{\mathsf{a}}\,, \tag{16}$$

again, as in Miyoshi (2011).

Renouncing Gaussianity

- Assume $\mathbf{H}\bar{\mathbf{B}}\mathbf{H}^{\mathsf{T}} \propto \mathbf{R}$.
- The likelihood $p(\boldsymbol{y}|\beta) = \mathcal{N}\big(\bar{\boldsymbol{\delta}} \,\big|\, \mathbf{0}, \bar{\mathbf{C}}(\beta)\big)$ becomes

$$p(\boldsymbol{y}|\beta) \propto \chi^{+2}\Big(\|\bar{\boldsymbol{\delta}}\|^2_{\mathbf{R}}/P \,\Big|\, (1+\bar{\sigma}^2\beta), P\Big)\,. \tag{13}$$

- Surprise !!!: $\operatorname{argmax} p(\boldsymbol{y}|\beta) = \hat{\beta}_{\mathbf{R}}$,
- A further approximation is fitted:

$$p(\boldsymbol{y}|\beta) \approx \chi^{+2}(\hat{\beta}_{\mathbf{R}}|\beta, \hat{\nu})\,. \tag{14}$$

- Likelihood (14) fits mode of (13). Fitting curvature $\Longrightarrow \hat{\nu}$ $\Longrightarrow$ same variance as in Miyoshi (2011) !!!
- Likelihood (14) conjugate to $p(\beta) = \chi^{-2}(\beta|\beta^{\mathsf{f}}, \nu^{\mathsf{f}})$, yielding

$$\nu^{\mathsf{a}} = \nu^{\mathsf{f}} + \hat{\nu}\,, \tag{15}$$

$$\beta^{\mathsf{a}} = (\nu^{\mathsf{f}}\beta^{\mathsf{f}} + \hat{\nu}\hat{\beta}_{\mathbf{R}})/\nu^{\mathsf{a}}\,, \tag{16}$$

again, as in Miyoshi (2011).

EAKF adaptive inflation

- Anderson (2007) assigns Gaussian prior:
$$p(\beta) = \mathcal{N}(\beta|\beta^{\mathrm{f}}, V^{\mathrm{f}})\,, \tag{17}$$

- and fits the posterior by a "Gaussian":
$$p(\beta|y_i) \approx \mathcal{N}(\beta|\hat{\beta}_{\mathrm{MAP}}, V^{\mathrm{a}})\,, \tag{18}$$
where $\hat{\beta}_{\mathrm{MAP}}$ and V^{a} are fitted using the *exact* posterior ("easy" by virtue of serial update).

- Gharamti (2017) improves via χ^{-2} and χ^{+2} (Gamma).

EAKF adaptive inflation

- Anderson (2007) assigns Gaussian prior:
$$p(\beta) = \mathcal{N}(\beta|\beta^{\mathsf{f}}, V^{\mathsf{f}})\,, \tag{17}$$

- and fits the posterior by a "Gaussian":
$$p(\beta|y_i) \approx \mathcal{N}(\beta|\hat{\beta}_{\mathsf{MAP}}, V^{\mathsf{a}})\,, \tag{18}$$
where $\hat{\beta}_{\mathsf{MAP}}$ and V^{a} are fitted using the *exact* posterior ("easy" by virtue of serial update).

- Gharamti (2017) improves via χ^{-2} and χ^{+2} (Gamma).

EAKF adaptive inflation

- Anderson (2007) assigns Gaussian prior:
$$p(\beta) = \mathcal{N}(\beta|\beta^{\mathsf{f}}, V^{\mathsf{f}})\,, \tag{17}$$

- and fits the posterior by a "Gaussian":
$$p(\beta|y_i) \approx \mathcal{N}(\beta|\hat{\beta}_{\mathsf{MAP}}, V^{\mathsf{a}})\,, \tag{18}$$

  where $\hat{\beta}_{\mathsf{MAP}}$ and V^{a} are fitted using the *exact* posterior ("easy" by virtue of serial update).

- Gharamti (2017) improves via χ^{-2} and χ^{+2} (Gamma).

EAKF adaptive inflation

- Anderson (2007) assigns Gaussian prior:

$$p(\beta) = \mathcal{N}(\beta|\beta^\mathsf{f}, V^\mathsf{f})\,, \tag{17}$$

- and fits the posterior by a "Gaussian":

$$p(\beta|y_i) \approx \mathcal{N}(\beta|\hat{\beta}_\mathsf{MAP}, V^\mathsf{a})\,, \tag{18}$$

  where $\hat{\beta}_\mathsf{MAP}$ and V^a are fitted using the *exact* posterior ("easy" by virtue of serial update).

- Gharamti (2017) improves via χ^{-2} and χ^{+2} (Gamma).

EAKF adaptive inflation

- Anderson (2007) assigns Gaussian prior:
$$p(\beta) = \mathcal{N}(\beta|\beta^{\mathsf{f}}, V^{\mathsf{f}})\,, \tag{17}$$

- and fits the posterior by a "Gaussian":
$$p(\beta|y_i) \approx \mathcal{N}(\beta|\hat{\beta}_{\mathsf{MAP}}, V^{\mathsf{a}})\,, \tag{18}$$
where $\hat{\beta}_{\mathsf{MAP}}$ and V^{a} are fitted using the *exact* posterior ("easy" by virtue of serial update).

- Gharamti (2017) improves via χ^{-2} and χ^{+2} (Gamma).

EnKF-N hybrid

- Use two inflation factors: α and β,
  dedicated to sampling and model error, respectively.
- For β, pick simplest (and ~best) scheme: $\hat{\beta}_{\mathbf{R}}$.
- Algorithm:
  - Find β (via $\hat{\beta}_{\mathbf{R}}$)
  - Find α given β (via EnKF-N)
- Potential improvements:
  - Determining (α, β) jointly (simultaneously).
  - Rather than fitting the likelihood parameters, fit posterior parameters (similarly to EAKF).
  - Matching moments via quadrature
  - Non-parametric (grid- or MC- based)
  - De-biasing $\hat{\beta}_{\mathbf{R}}$

Testing "improvements" did not yield significant gains.

EnKF-N hybrid

- Use two inflation factors: α and β,

  dedicated to sampling and model error, respectively.

- For β, pick simplest (and ~best) scheme: $\hat{\beta}_{\mathbf{R}}$.

- Algorithm:

  Find β (via $\hat{\beta}_{\mathbf{R}}$)

  Find α given β (via EnKF-N)

- Potential improvements:
  - Determining (α, β) jointly (simultaneously).
  - Rather than fitting the likelihood parameters, fit posterior parameters (similarly to EAKF).
  - Matching moments via quadrature
  - Non-parametric (grid- or MC- based)
  - De-biasing $\hat{\beta}_{\mathbf{R}}$.

Testing "improvements" did not yield significant gains.

EnKF-N hybrid

- Use two inflation factors: α and β,
  dedicated to sampling and model error, respectively.
- For β, pick simplest (and ~best) scheme: $\hat{\beta}_{\mathbf{R}}$.
- Algorithm:

  Find β (via $\hat{\beta}_{\mathbf{R}}$)

  Find α given β (via EnKF-N)
- Potential improvements:
  - Determining (α, β) jointly (simultaneously).
  - Rather than fitting the likelihood parameters, fit posterior parameters (similarly to EAKF).
  - Matching moments via quadrature
  - Non-parametric (grid- or MC- based)
  - De-biasing $\hat{\beta}_{\mathbf{R}}$

Testing "improvements" did not yield significant gains.

EnKF-N hybrid

- Use two inflation factors: α and β,
  dedicated to sampling and model error, respectively.
- For β, pick simplest (and $\sim$best) scheme: $\hat{\beta}_{\mathbf{R}}$.
- Algorithm:

  Find β (via $\hat{\beta}_{\mathbf{R}}$)

  Find α given β (via EnKF-N)
- Potential improvements:
  - Determining (α, β) jointly (simultaneously).
  - Rather than fitting the likelihood parameters, fit posterior parameters (similarly to EAKF).
  - Matching moments via quadrature
  - Non-parametric (grid- or MC- based)
  - De-biasing $\hat{\beta}_{\mathbf{R}}$

  Testing "improvements" did not yield significant gains.

EnKF-N hybrid

- Use two inflation factors: α and β,
  dedicated to sampling and model error, respectively.
- For β, pick simplest (and $\sim$best) scheme: $\hat{\beta}_{\mathbf{R}}$.
- Algorithm:

  Find β (via $\hat{\beta}_{\mathbf{R}}$)

  Find α given β (via EnKF-N)
- Potential improvements:
  - Determining (α, β) jointly (simultaneously).
  - Rather than fitting the likelihood parameters, fit posterior parameters (similarly to EAKF).
  - Matching moments via quadrature
  - Non-parametric (grid- or MC- based)
  - De-biasing $\hat{\beta}_{\mathbf{R}}$.

Testing "improvements" did not yield significant gains.

EnKF-N hybrid

- Use two inflation factors: α and β,
  dedicated to sampling and model error, respectively.
- For β, pick simplest (and $\sim$best) scheme: $\hat{\beta}_{\mathbf{R}}$.
- Algorithm:

  Find β (via $\hat{\beta}_{\mathbf{R}}$)

  Find α given β (via EnKF-N)

- Potential improvements:
  - Determining (α, β) jointly (simultaneously).
  - Rather than fitting the likelihood parameters, fit posterior parameters (similarly to EAKF).
  - Matching moments via quadrature
  - Non-parametric (grid- or MC- based)
  - De-biasing $\hat{\beta}_{\mathbf{R}}$.

Testing "improvements" did not yield significant gains.

EnKF-N hybrid

- Use two inflation factors: α and β,
  dedicated to sampling and model error, respectively.
- For β, pick simplest (and $\sim$best) scheme: $\hat{\beta}_{\mathbf{R}}$.
- Algorithm:

  Find β (via $\hat{\beta}_{\mathbf{R}}$)

  Find α given β (via EnKF-N)
- Potential improvements:
  - Determining (α, β) jointly (simultaneously).
  - Rather than fitting the likelihood parameters, fit posterior parameters (similarly to EAKF).
  - Matching moments via quadrature
  - Non-parametric (grid- or MC- based)
  - De-biasing $\hat{\beta}_{\mathbf{R}}$.

Testing "improvements" did not yield significant gains.

EnKF-N hybrid

- Use two inflation factors: α and β,
  dedicated to sampling and model error, respectively.
- For β, pick simplest (and $\sim$best) scheme: $\hat{\beta}_{\mathbf{R}}$.
- Algorithm:

  Find β (via $\hat{\beta}_{\mathbf{R}}$)

  Find α given β (via EnKF-N)
- Potential improvements:
  - Determining (α, β) jointly (simultaneously).
  - Rather than fitting the likelihood parameters, fit posterior parameters (similarly to EAKF).
  - Matching moments via quadrature
  - Non-parametric (grid- or MC- based)
  - De-biasing $\hat{\beta}_{\mathbf{R}}$.

Testing "improvements" did not yield significant gains.

EnKF-N hybrid

- Use two inflation factors: α and β,
  dedicated to sampling and model error, respectively.
- For β, pick simplest (and $\sim$best) scheme: $\hat{\beta}_{\mathbf{R}}$.
- Algorithm:

  Find β (via $\hat{\beta}_{\mathbf{R}}$)

  Find α given β (via EnKF-N)
- Potential improvements:
  - Determining (α, β) jointly (simultaneously).
  - Rather than fitting the likelihood parameters, fit posterior parameters (similarly to EAKF).
  - Matching moments via quadrature
  - Non-parametric (grid- or MC- based)
  - De-biasing $\hat{\beta}_{\mathbf{R}}$

Testing "improvements" did not yield significant gains.

EnKF-N hybrid

- Use two inflation factors: α and β, dedicated to sampling and model error, respectively.
- For β, pick simplest (and ∼best) scheme: $\hat{\beta}_{\mathbf{R}}$.
- Algorithm:

  Find β (via $\hat{\beta}_{\mathbf{R}}$)

  Find α given β (via EnKF-N)
- Potential improvements:
  - Determining (α, β) jointly (simultaneously).
  - Rather than fitting the likelihood parameters, fit posterior parameters (similarly to EAKF).
  - Matching moments via quadrature
  - Non-parametric (grid- or MC- based)
  - De-biasing $\hat{\beta}_{\mathbf{R}}$.

Testing "improvements" did not yield significant gains.

EnKF-N hybrid

- Use two inflation factors: α and β,
  dedicated to sampling and model error, respectively.
- For β, pick simplest (and $\sim$best) scheme: $\hat{\beta}_{\mathbf{R}}$.
- Algorithm:

  Find β (via $\hat{\beta}_{\mathbf{R}}$)

  Find α given β (via EnKF-N)
- Potential improvements:
  - Determining (α, β) jointly (simultaneously).
  - Rather than fitting the likelihood parameters, fit posterior parameters (similarly to EAKF).
  - Matching moments via quadrature
  - Non-parametric (grid- or MC- based)
  - De-biasing $\hat{\beta}_{\mathbf{R}}$.

Testing "improvements" did not yield significant gains.

EnKF-N hybrid

- Use two inflation factors: α and β,
  dedicated to sampling and model error, respectively.
- For β, pick simplest (and $\sim$best) scheme: $\hat{\beta}_{\mathbf{R}}$.
- Algorithm:

  Find β (via $\hat{\beta}_{\mathbf{R}}$)

  Find α given β (via EnKF-N)
- Potential improvements:
  - Determining (α, β) jointly (simultaneously).
  - Rather than fitting the likelihood parameters, fit posterior parameters (similarly to EAKF).
  - Matching moments via quadrature
  - Non-parametric (grid- or MC- based)
  - De-biasing $\hat{\beta}_{\mathbf{R}}$

Testing "improvements" did not yield significant gains.

EnKF-N hybrid

- Use two inflation factors: α and β,
  dedicated to sampling and model error, respectively.
- For β, pick simplest (and $\sim$best) scheme: $\hat{\beta}_{\mathbf{R}}$.
- Algorithm:

  Find β (via $\hat{\beta}_{\mathbf{R}}$)

  Find α given β (via EnKF-N)
- Potential improvements:
  - Determining (α, β) jointly (simultaneously).
  - Rather than fitting the likelihood parameters, fit posterior parameters (similarly to EAKF).
  - Matching moments via quadrature
  - Non-parametric (grid- or MC- based)
  - De-biasing $\hat{\beta}_{\mathbf{R}}$

Testing "improvements" did not yield significant gains.

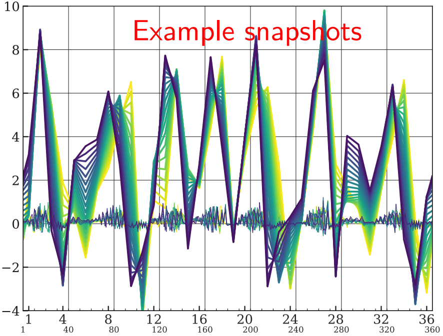
Two-layer Lorenz-96

Evolution

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = \psi_i^+(\boldsymbol{x}) + F - h\frac{c}{b}\sum_{j=1}^{10} z_{j+10(i-1)}\,, \qquad i = 1, \ldots, 36,$$

$$\frac{\mathrm{d}z_j}{\mathrm{d}t} = \frac{c}{b}\psi_j^-(b\boldsymbol{z}) + 0 + h\frac{c}{b}x_{1+(j-1)//10}\,, \qquad j = 1, \ldots, 360,$$

where ψ_i is the single-layer Lorenz-96 dynamics.



$$\mathrm{RMSE} = \frac{1}{T}\sum_{t=1}^{T}\sqrt{\frac{1}{M}\|\bar{x}_t - x_t\|_2^2}\,.$$

$N = 20$, no localization.

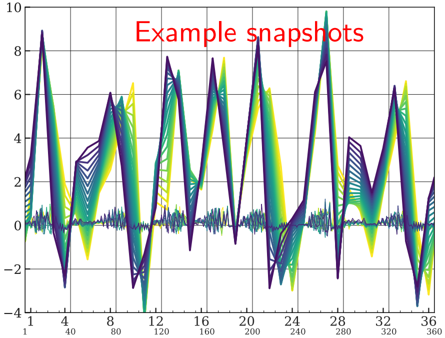
Two-layer Lorenz-96

Evolution

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = \psi_i^+(\boldsymbol{x}) + F - h\frac{c}{b}\sum_{j=1}^{10} z_{j+10(i-1)}\,, \qquad i = 1, \ldots, 36,$$

$$\frac{\mathrm{d}z_j}{\mathrm{d}t} = \frac{c}{b}\psi_j^-(b\boldsymbol{z}) + 0 + h\frac{c}{b}x_{1+(j-1)//10}\,, \qquad j = 1, \ldots, 360,$$

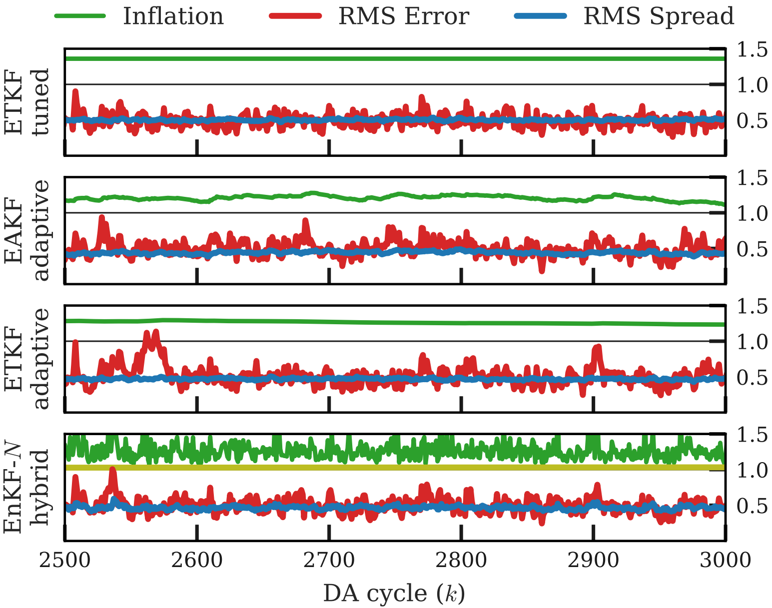
where ψ_i is the single-layer Lorenz-96 dynamics.



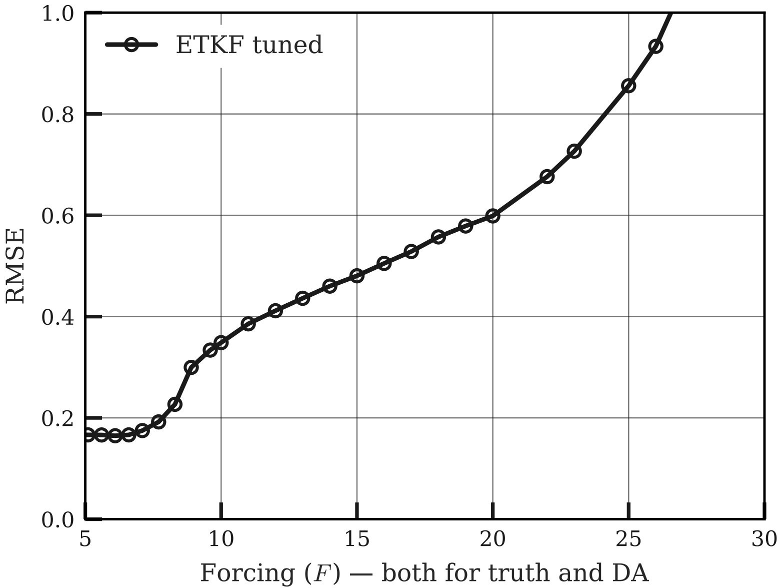
$$\text{RMSE} = \frac{1}{T}\sum_{t=1}^{T}\sqrt{\frac{1}{M}\|\bar{\boldsymbol{x}}_t - \boldsymbol{x}_t\|_2^2}\,.$$

$N = 20$, no localization.

Illustration of time series

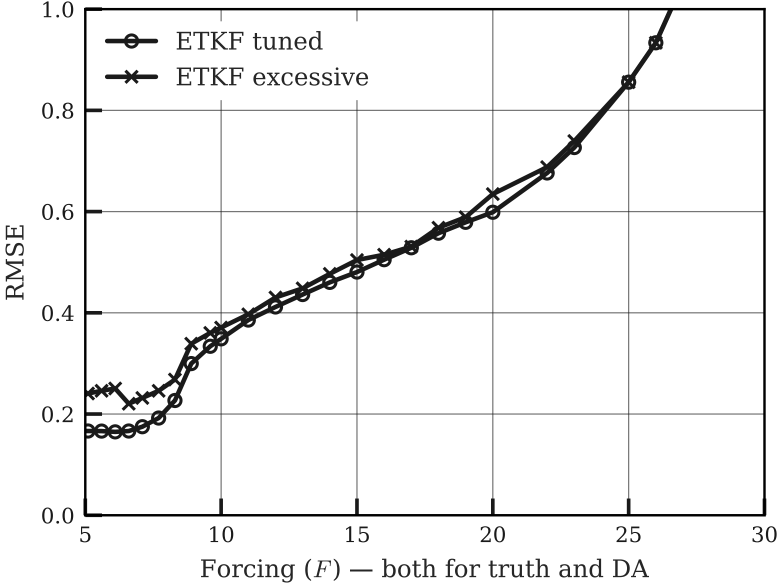


Benchmarks



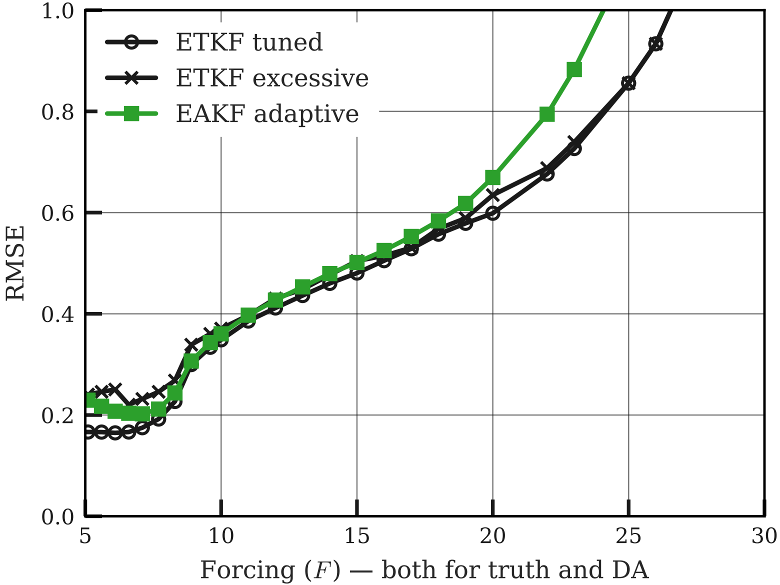
ETKF tuned
RMSE
Forcing (F) — both for truth and DA
1.0
0.8
0.6
0.4
0.2
0.0
5
10
15
20
25
30

Benchmarks



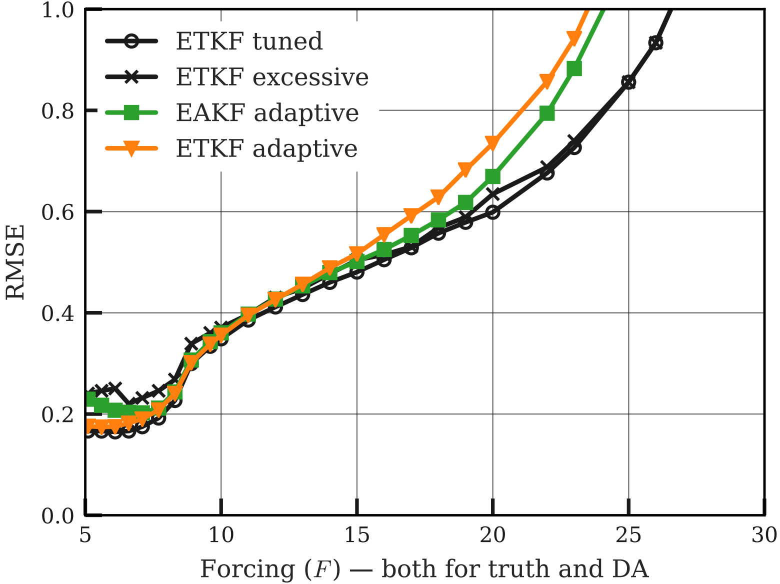
ETKF tuned
ETKF excessive
RMSE
1.0
0.8
0.6
0.4
0.2
0.0
5
10
15
20
25
30
Forcing (F) — both for truth and DA

Benchmarks

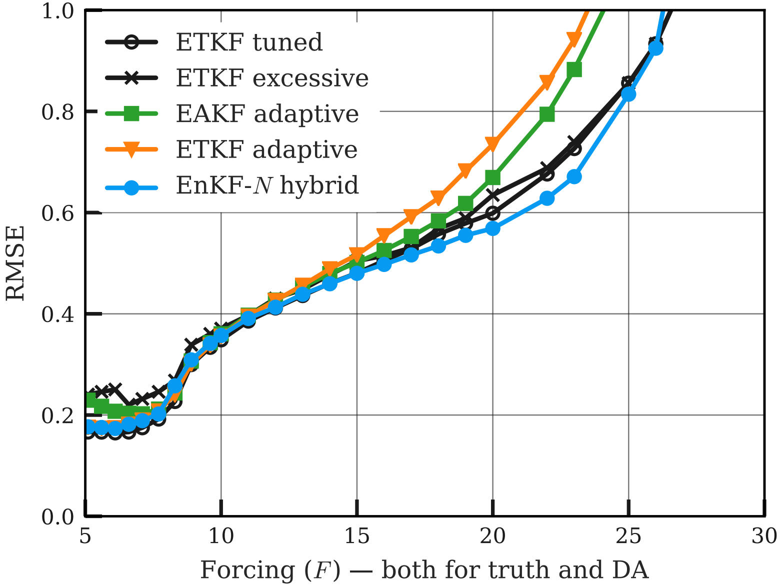


ETKF tuned
ETKF excessive
EAKF adaptive
RMSE
1.0
0.8
0.6
0.4
0.2
0.0
5
10
15
20
25
30
Forcing (F) — both for truth and DA

Benchmarks

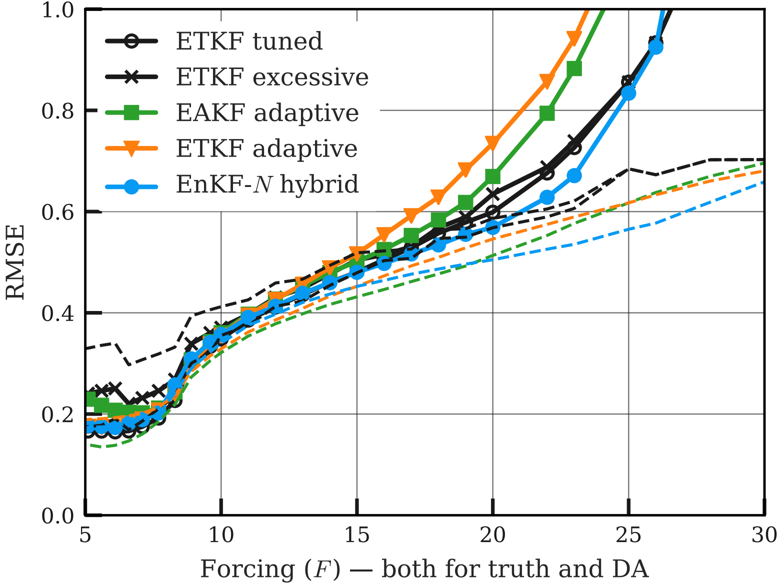

ETKF tuned
ETKF excessive
EAKF adaptive
ETKF adaptive
RMSE
0.0
0.2
0.4
0.6
0.8
1.0
5
10
15
20
25
30
Forcing (F) — both for truth and DA

Benchmarks



ETKF tuned
ETKF excessive
EAKF adaptive
ETKF adaptive
EnKF-N hybrid
RMSE
0.0
0.2
0.4
0.6
0.8
1.0
5
10
15
20
25
30
Forcing (F) — both for truth and DA

Benchmarks



ETKF tuned
ETKF excessive
EAKF adaptive
ETKF adaptive
EnKF-N hybrid
RMSE
Forcing (F) — both for truth and DA
0.0
0.2
0.4
0.6
0.8
1.0
5
10
15
20
25
30

Summary

Paper highlights:

- Cataloguing of **reasons to inflate**.
- **Inflation-centric** re-derivation of the dual **EnKF-N**.
- Formal **survey of adaptive inflation** methods.
- A simple **hybrid of EnKF-N and $\hat{\beta}_{\mathrm{R}}$**, which is shown to systematically (but moderately) improve filter accuracy (no re-tuning!).

Summary

Paper highlights:

- Cataloguing of **reasons to inflate**.
- **Inflation-centric** re-derivation of the dual **EnKF-N**.
- Formal **survey of adaptive inflation** methods.
- A simple **hybrid of EnKF-N and $\hat{\beta}_{\mathrm{R}}$**, which is shown to systematically (but moderately) improve filter accuracy (no re-tuning!).

Summary

Paper highlights:

- Cataloguing of **reasons to inflate**.
- **Inflation-centric** re-derivation of the dual **EnKF-N**.
- Formal **survey of adaptive inflation** methods.
- A simple **hybrid of EnKF-N and** $\hat{\beta}_{\mathbf{R}}$, which is shown to systematically (but moderately) improve filter accuracy (no re-tuning!).

Summary

Paper highlights:

- Cataloguing of **reasons to inflate**.
- **Inflation-centric** re-derivation of the dual **EnKF-**N.
- Formal **survey of adaptive inflation** methods.
- A simple **hybrid of EnKF-**N **and** $\hat{\beta}_{\mathrm{R}}$, which is shown to systematically (but moderately) improve filter accuracy (no re-tuning!).

Summary

Paper highlights:

- Cataloguing of **reasons to inflate**.
- **Inflation-centric** re-derivation of the dual **EnKF-N**.
- Formal **survey of adaptive inflation** methods.
- A simple **hybrid of EnKF-N and $\hat{\beta}_{\mathbf{R}}$**, which is shown to systematically (but moderately) improve filter accuracy (no re-tuning!).

Summary

Paper highlights:

- Cataloguing of **reasons to inflate**.
- **Inflation-centric** re-derivation of the dual **EnKF-**N.
- Formal **survey of adaptive inflation** methods.
- A simple **hybrid of EnKF-**N **and** $\hat{\beta}_{\mathbf{R}}$, which is shown to systematically (but moderately) improve filter accuracy (no re-tuning!).

References 1

- `https://github.com/nansencenter/DAPPER`
- 2011: **Marc Bocquet.** *Ensemble Kalman filtering without the intrinsic need for inflation*. Nonlinear Processes in Geophysics.
- 2012: **Marc Bocquet and Pavel Sakov.** *Combining inflation-free and iterative ensemble Kalman filters for strongly nonlinear systems*. Nonlinear Processes in Geophysics
- 2015: **Marc Bocquet, Patrick N. Raanes, and Alexis Hannart**. *Expanding the validity of the ensemble Kalman filter without the intrinsic need for inflation*. Nonlinear Processes in Geophysics
- 2018: **Patrick N. Raanes, Marc Bocquet, and Alberto Carrassi.** *Adaptive covariance inflation in the ensemble Kalman filter by Gaussian scale mixtures*. QJRMS (minor rev), `arxiv.org/abs/1801.08474`

References 2

Jeffrey L. Anderson. An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus A*, 59(2):210–224, 2007.

M. E. Gharamti. Enhanced adaptive inflation algorithm for ensemble filters. *Monthly Weather Review*, In review(0):0–0, 2017.

Takemasa Miyoshi. The Gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter. *Monthly Weather Review*, 139(5):1519–1535, 2011.

Xuguang Wang and Craig H. Bishop. A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *Journal of the Atmospheric Sciences*, 60(9):1140–1158, 2003.

UKF reinventing localization

2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)

Multiple Sigma-point Kalman Smoothers for High-dimensional State-Space Models

Jordi Vilà-Valls*, Pau Closas†, Ángel F. García-Fernández‡ and Carles Fernández-Prades*
*Centre Tecnològic de Telecomunicacions de Catalunya (CTTC/CERCA), 08860, Barcelona, Spain
†Dept. of Electrical and Computer Eng., Northeastern University, Boston, MA 02115, USA
‡ Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool, UK.
E-mail: {jvila,cfernandez}@cttc.cat, closas@northeastern.edu, angel.garcia-fernandez@liverpool.ac.uk

***Abstract*—This article presents a new multiple state-partitioning solution to the Bayesian smoothing problem in nonlinear high-dimensional Gaussian systems. The key idea is to partition the original state into several low-dimensional subspaces, and apply an individual smoother to each of them. The main goal is to reduce the state dimension each filter has to explore, to reduce the curse of dimensionality and eventual loss of accuracy. We provide the theoretical multiple smoothing formulation and a new nested sigma-point approximation to the resulting smoothing solution. The performance of the new approach is shown for the 40-dimensional Lorenz model.**

I. Introduction

In general, we are interested in nonlinear Gaussian state-space models (SSM), which are expressed as

$$\mathbf{x}_k = \mathbf{f}_{k-1}(\mathbf{x}_{k-1}) + \mathbf{v}_{k-1} \,, \qquad \mathbf{v}_{k-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{k-1}), \tag{1}$$

$$\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{n}_k \,, \qquad \mathbf{n}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k), \tag{2}$$

where $\mathbf{x}_k \in \mathbb{R}^{n_x}$ are the hidden states of the system, $\mathbf{y}_k \in \mathbb{R}^{n_y}$ the measurements at time k, $\mathbf{f}_{k-1}(\cdot)$ and $\mathbf{h}_k(\cdot)$ are the nonlinear process and measurement functions, and both Gaussian noises are assumed to be independent. The Bayesian smoothing solution is given by the marginal distri-

with $\mathbf{x}_k = [\mathbf{x}_k^{(1)}, \ldots, \mathbf{x}_k^{(S)}]$. The subspace process functions $\mathbf{f}_{k-1}^{(s)}(\cdot)$ can be different and the independent s-th subspace Gaussian process noise is $\mathbf{v}_{k-1}^{(s)} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{k-1}^{(s)})$. The main idea is to partition the original state in several subspaces, and apply a low dimensional individual filter to each subspace, directly reducing the dimension each filter must explore. In this approach, we are interested in the subspace marginal smoothed posterior, $p(\mathbf{x}_k^{(s)}|\mathbf{y}_{1:N})$. In the multiple state-partioning framework, we make the approximation that the different subspaces are independent, which is typically accurate in applications such as multiple target tracking. Mathematically, this implies that the joint smoothing posterior is $p(\mathbf{x}_k^{(s)}, \mathbf{x}_k^{(-s)}|\mathbf{y}_{1:N}) \approx p(\mathbf{x}_k^{(s)}|\mathbf{y}_{1:N})p(\mathbf{x}_k^{(-s)}|\mathbf{y}_{1:N})$. In this contribution we extend previous results on MQKF [11] to the smoothing problem, and propose a new nested sigma-point approximation to the smoothing marginal posterior integrals.

II. Multiple Gaussian Smoothing

A. Background on Multiple Gaussian Filtering

As done in standard Bayesian filtering, the s-th subspace posterior can be recursively computed in two steps: *prediction*

Appendix

Parametric distributions – Table

Table 2: Parametric probability distributions. As elsewhere in the paper, $\boldsymbol{b}, \boldsymbol{x} \in \mathbb{R}^M$, $\mathbf{B}, \mathbf{S} \in \mathcal{B}$, $s, \beta > 0$, and it is assumed that $\nu > M$. The constants are $c_\mathcal{N} = (2\pi)^{-M/2}$, $c_t = \frac{\Gamma(\frac{\nu+M}{2})}{(\pi\nu)^{M/2}\Gamma(\nu/2)}$, $c_\mathcal{W} = \frac{\nu^{\nu/2}}{2^{\nu M/2}\Gamma_M(\nu/2)}$, and $c_\chi = c_\mathcal{W}$ with $M = 1$. The (unlisted) variance of element (i,j) of $\mathbf{B}$ with the Wishart distribution is $(s_{ij}^2 + s_{ii}s_{jj})/\nu$, where s_{ij} is element (i,j) of $\mathbf{S}$. The variances of the inverse-Wishart distribution are asymptotically, for $\nu \to \infty$, the same.

Name	**Symbol**	**Probability density function**	**Mean**	**Mode**	**(Co)Var**
Gauss./Normal	$\mathcal{N}(\boldsymbol{x}\|\boldsymbol{b}, \mathbf{B})$	$= c_\mathcal{N} \|\mathbf{B}\|^{-1/2} \exp\left(-\frac{1}{2}\|\boldsymbol{x} - \boldsymbol{b}\|_\mathbf{B}^2\right)$	$\boldsymbol{b}$	$\boldsymbol{b}$	$\mathbf{B}$
t distribution	$\boldsymbol{t}(\boldsymbol{x}\|\nu; \boldsymbol{b}, \mathbf{B})$	$= c_t \|\mathbf{B}\|^{-1/2}\left(1 + \frac{1}{\nu}\|\boldsymbol{x} - \boldsymbol{b}\|_\mathbf{B}^2\right)^{-(\nu+M)/2}$	$\boldsymbol{b}$	$\boldsymbol{b}$	$\frac{\nu}{\nu-2}\mathbf{B}$
Wishart	$\mathcal{W}^{+1}(\mathbf{B}\|\mathbf{S}, \nu)$	$= c_\mathcal{W} \|\mathbf{S}\|^{-\nu/2} \|\mathbf{B}\|^{(\nu-M-1)/2} e^{-\operatorname{tr}(\nu\mathbf{B}\mathbf{S}^{-1})/2}$	$\mathbf{S}$	$\frac{\nu-M-1}{\nu}\mathbf{S}$	
Inv-Wishart	$\mathcal{W}^{-1}(\mathbf{B}\|\mathbf{S}, \nu)$	$= c_\mathcal{W} \|\mathbf{S}\|^{\nu/2} \|\mathbf{B}\|^{-(\nu+M+1)/2} e^{-\operatorname{tr}(\nu\mathbf{S}\mathbf{B}^{-1})/2}$	$\frac{\nu}{\nu-M-1}\mathbf{S}$	$\frac{\nu}{\nu+M+1}\mathbf{S}$	
Chi-square	$\chi^{+2}(\beta\|s, \nu)$	$= c_\chi\, s^{-\nu/2} \beta^{\nu/2-1} e^{-\nu\beta/2s}$	s	$\frac{\nu-2}{\nu}s$	$2s^2/\nu$
Inv-chi-sq.	$\chi^{-2}(\beta\|s, \nu)$	$= c_\chi\, s^{\nu/2} \beta^{-\nu/2-1} e^{-\nu s/2\beta}$	$\frac{\nu}{\nu-2}s$	$\frac{\nu}{\nu+2}s$	$\frac{2(\nu s)^2}{(\nu-2)^2(\nu-4)}$

Parametric distributions – Properties

Property 1 The ("scaled") chi-square distribution is equivalent to the Gamma distribution:

$$\chi^{\pm 2}(\beta|s,\nu) = \mathrm{Gamma}^{\pm 1}(\beta|\nu/2, \nu s^{\mp 1}/2)\,, \tag{70}$$

where the switch sign $\pm$ has been used to represent both the regular and inverse distributions. The χ parameterization has been preferred for the notational simplicity of the relations of Properties 2 to 4.

Property 2 Asymptotic normality. If $\beta \sim \chi^{\pm 2}(s,\nu)$, then the distribution of $\sqrt{\nu}(\beta - s)$ converges to $\mathcal{N}(0, 2s^2)$ as $\nu \to \infty$. This shows that s is a location parameter, while $2s^2/\nu$ plays the role of variance, which is why this paper prefers referring to ν as "certainty" instead of "degree of freedom". The asymptotic result for χ^{+2} is a well known consequence of the central limit theorem, since β may then be written as an average of random variables. The result for χ^{-2} is less known, but can be shown by through the pointwise convergence of the pdf of $\sqrt{\nu}(\beta - s)$, normalized by its value at 0.

Property 3 In the univariate case ($M = 1$),

$$\mathcal{W}^{\pm 1}(\beta|s,\nu) = \chi^{\pm 2}(\beta|s,\nu)\,. \tag{71}$$

Property 4 Reciprocity. If $t = 1/\beta$:

$$\begin{aligned} p(\beta) &= \chi^{-2}(\beta|s,\nu) \\ \iff p(t) &= \chi^{+2}(t|1/s,\nu)\,. \end{aligned} \tag{72}$$

Property 5 Reciprocity. If $\mathbf{T} = \mathbf{B}^{-1}$:

$$\begin{aligned} p(\mathbf{B}) &= \mathcal{W}^{-1}(\mathbf{B}|\mathbf{S},\nu) \\ \iff p(\mathbf{T}) &= \mathcal{W}^{+1}(\mathbf{T}|\mathbf{S}^{-1},\nu)\,, \end{aligned} \tag{73}$$

as follows by the change of variables and the Jacobian $|\mathbf{T}|^{-(M+1)}$ [Muirhead, 1982, §. 2.1].

Property 6 Let $\boldsymbol{u} \neq \mathbf{0}$ be any m-dimensional vector, or an (almost never zero) random vector. If $\mathbf{T} \sim \mathcal{W}^{+1}(\mathbf{S},\nu)$ is independent of $\boldsymbol{u}$, then

$$\frac{\boldsymbol{u}^\mathsf{T}\mathbf{T}\boldsymbol{u}}{\boldsymbol{u}^\mathsf{T}\mathbf{S}\boldsymbol{u}} \sim \chi^{+2}(1,\nu)\,. \tag{74}$$

Moreover, this statistic is also independent of $\boldsymbol{u}$. Proof: Theorem 3.2.8 of Muirhead [1982].

EnKF-N mixing distribution

Instead, we assign the Jeffreys (hyper)prior:

$$p(\boldsymbol{b}, \mathbf{B}) \propto p(\mathbf{B}) \propto |\mathbf{B}|^{-(M+1)/2} \,, \tag{19}$$

and recall the likelihood:

$$p(\mathbf{E}|\boldsymbol{b}, \mathbf{B}) \propto \prod_{n=1}^{N} \mathcal{N}(\boldsymbol{x}_n|\boldsymbol{b}, \mathbf{B}) \,, \tag{20}$$

yielding

$$p(\boldsymbol{b}, \mathbf{B}|\mathbf{E}) = \underbrace{\mathcal{N}(\boldsymbol{b}|\bar{\boldsymbol{x}}, \mathbf{B}/N)}_{p(\boldsymbol{b}|\mathbf{B},\mathbf{E})} \underbrace{\mathcal{W}^{-1}(\mathbf{B}|\bar{\mathbf{B}}, N-1)}_{p(\mathbf{B}|\mathbf{E})} \,, \tag{21}$$

where $\mathcal{W}^{-1}$ is the inverse-Wishart distribution (c.f. Table 2).

EnKF-N mixing distribution

Instead, we assign the Jeffreys (hyper)prior:

$$p(\boldsymbol{b}, \mathbf{B}) \propto p(\mathbf{B}) \propto |\mathbf{B}|^{-(M+1)/2} \,, \tag{19}$$

and recall the likelihood:

$$p(\mathbf{E}|\boldsymbol{b}, \mathbf{B}) \propto \prod_{n=1}^{N} \mathcal{N}(\boldsymbol{x}_n|\boldsymbol{b}, \mathbf{B}) \,, \tag{20}$$

yielding

$$p(\boldsymbol{b}, \mathbf{B}|\mathbf{E}) = \underbrace{\mathcal{N}(\boldsymbol{b}|\bar{\boldsymbol{x}}, \mathbf{B}/N)}_{p(\boldsymbol{b}|\mathbf{B},\mathbf{E})} \underbrace{\mathcal{W}^{-1}(\mathbf{B}|\bar{\mathbf{B}}, N-1)}_{p(\mathbf{B}|\mathbf{E})} \,, \tag{21}$$

where $\mathcal{W}^{-1}$ is the inverse-Wishart distribution (c.f. Table 2).

EnKF-N mixing distribution

Instead, we assign the Jeffreys (hyper)prior:

$$p(\boldsymbol{b}, \mathbf{B}) \propto p(\mathbf{B}) \propto |\mathbf{B}|^{-(M+1)/2} \,, \tag{19}$$

and recall the likelihood:

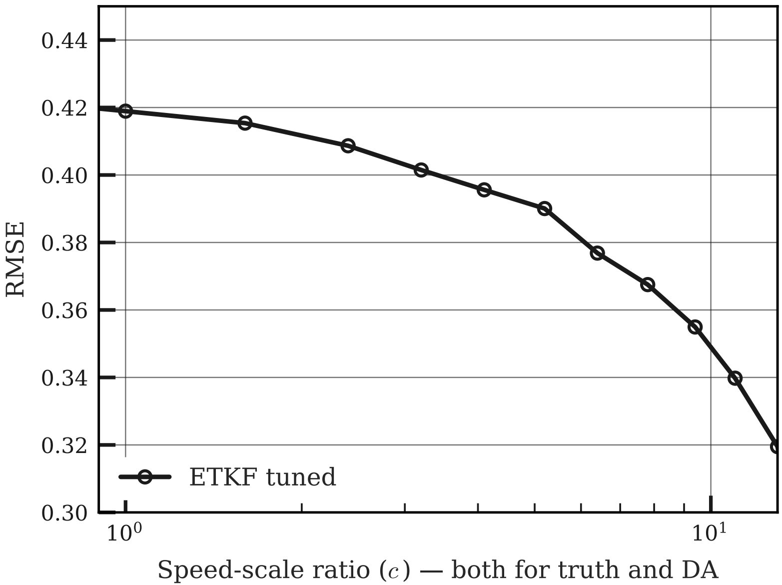
$$p(\mathbf{E}|\boldsymbol{b}, \mathbf{B}) \propto \prod_{n=1}^{N} \mathcal{N}(\boldsymbol{x}_n|\boldsymbol{b}, \mathbf{B}) \,, \tag{20}$$

yielding

$$p(\boldsymbol{b}, \mathbf{B}|\mathbf{E}) = \underbrace{\mathcal{N}(\boldsymbol{b}|\bar{\boldsymbol{x}}, \mathbf{B}/N)}_{p(\boldsymbol{b}|\mathbf{B},\mathbf{E})} \underbrace{\mathcal{W}^{-1}(\mathbf{B}|\bar{\mathbf{B}}, N-1)}_{p(\mathbf{B}|\mathbf{E})} \,, \tag{21}$$

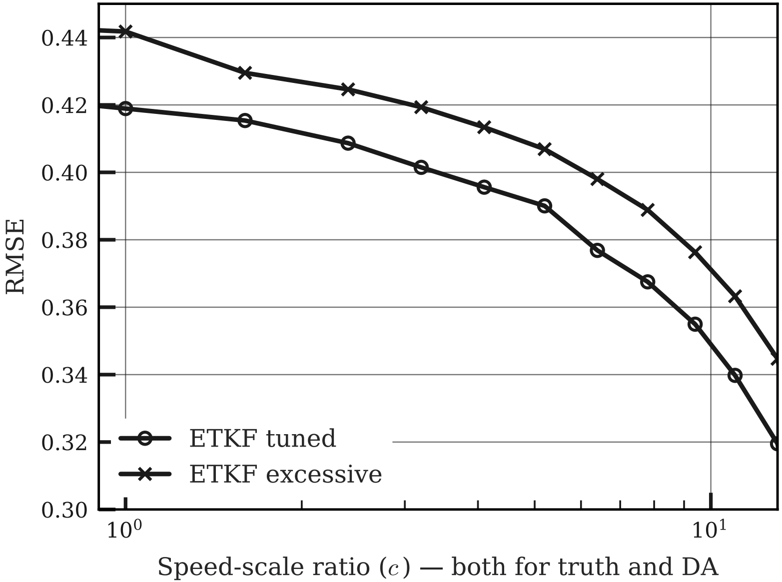
where $\mathcal{W}^{-1}$ is the inverse-Wishart distribution (c.f. Table 2).

Benchmarks

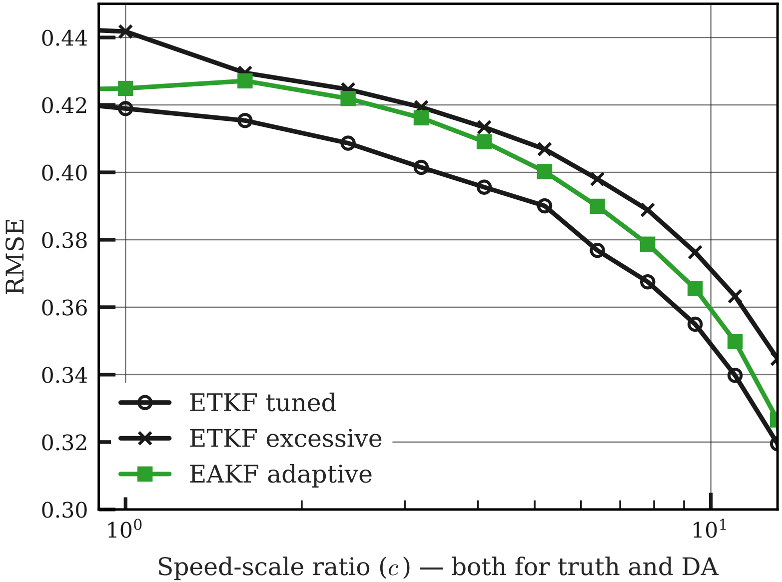


RMSE
0.44
0.42
0.40
0.38
0.36
0.34
0.32
0.30
10^0
10^1
ETKF tuned
Speed-scale ratio (c) — both for truth and DA

Benchmarks

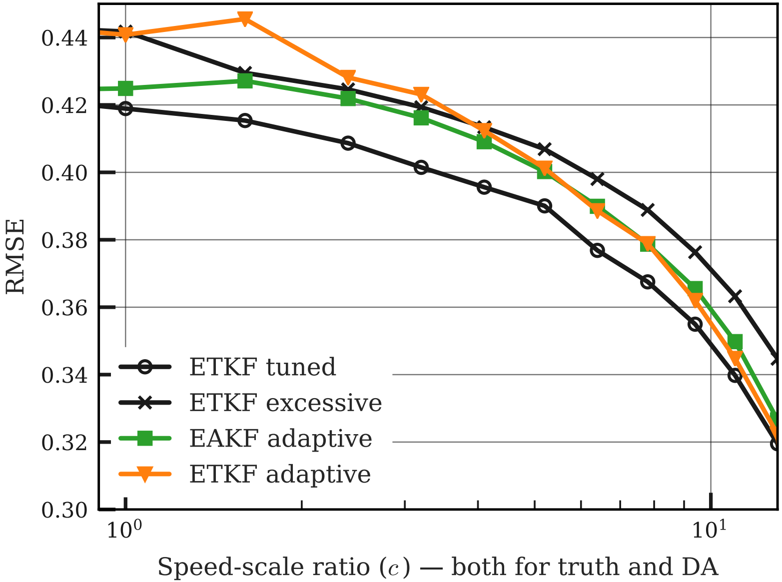

0.44
0.42
0.40
0.38
0.36
0.34
0.32
0.30
RMSE
ETKF tuned
ETKF excessive
10^0
10^1
Speed-scale ratio (c) — both for truth and DA

Benchmarks



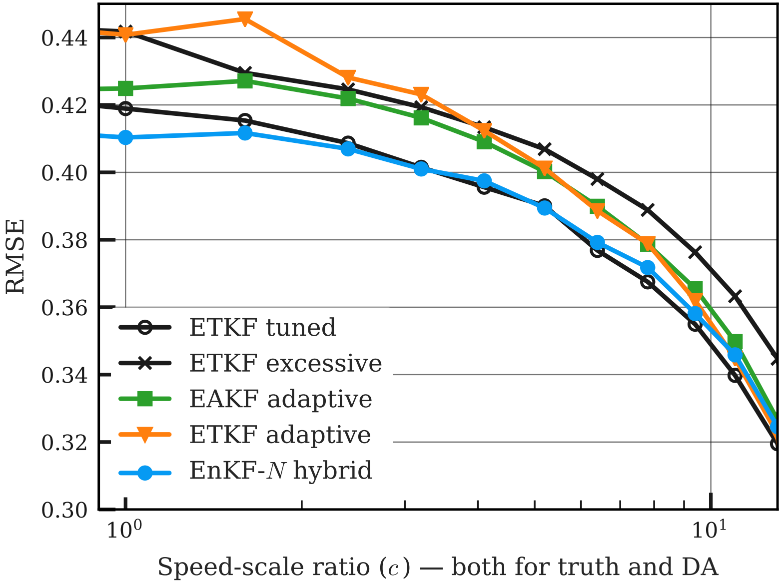
0.44
0.42
0.40
0.38
0.36
0.34
0.32
0.30
RMSE
10^0
10^1
Speed-scale ratio (c) — both for truth and DA
ETKF tuned
ETKF excessive
EAKF adaptive

Benchmarks



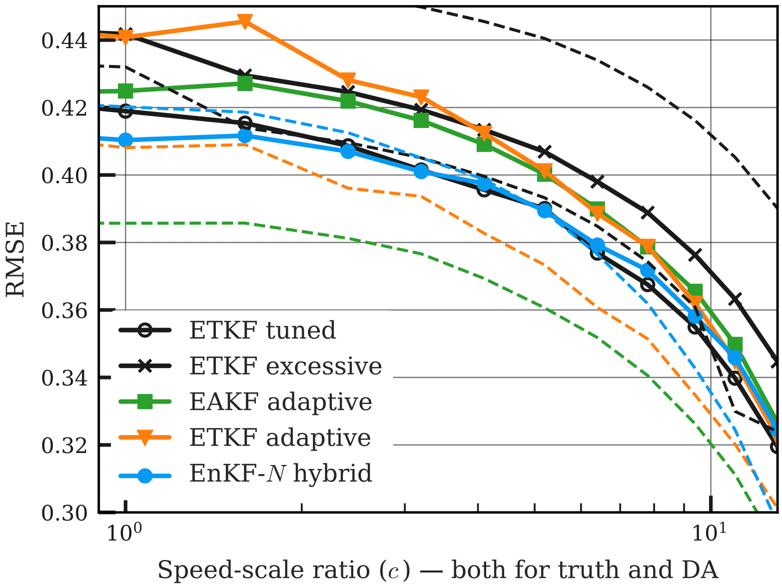
RMSE
0.44
0.42
0.40
0.38
0.36
0.34
0.32
0.30
10^0
10^1
Speed-scale ratio (c) — both for truth and DA
ETKF tuned
ETKF excessive
EAKF adaptive
ETKF adaptive

Benchmarks



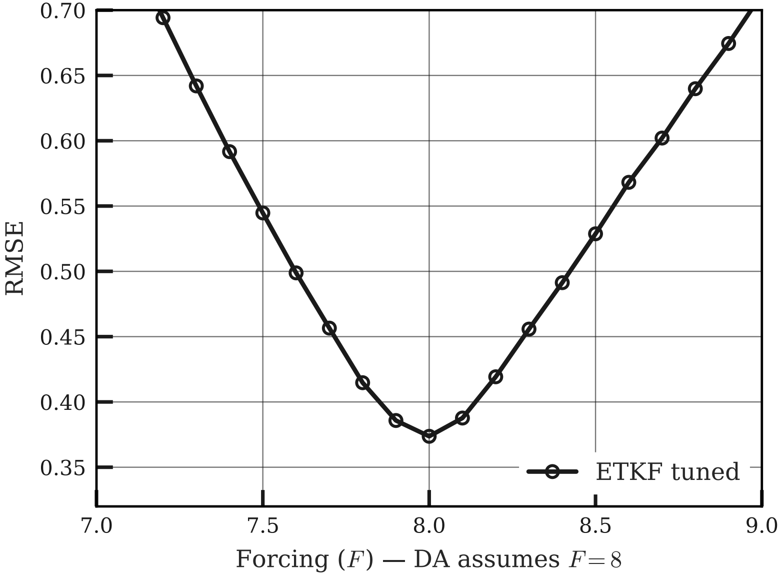
RMSE
0.44
0.42
0.40
0.38
0.36
0.34
0.32
0.30
10^0
10^1
Speed-scale ratio (c) — both for truth and DA
ETKF tuned
ETKF excessive
EAKF adaptive
ETKF adaptive
EnKF-N hybrid

Benchmarks

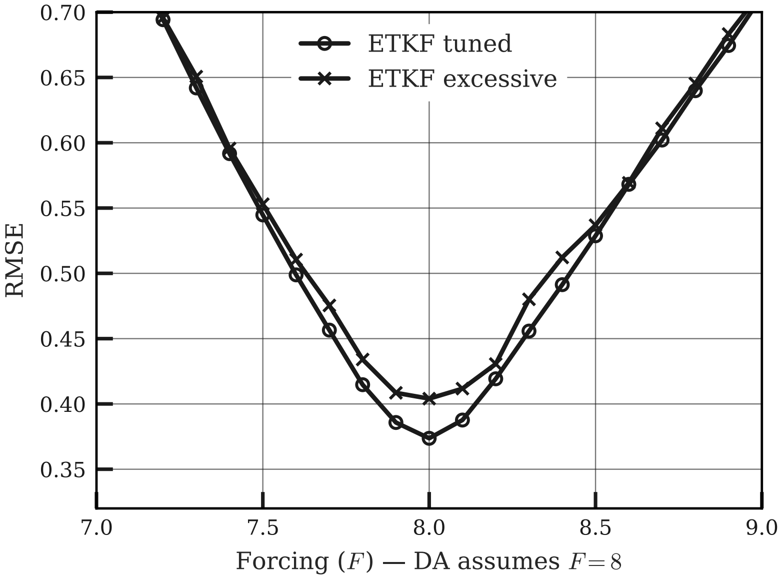


0.44
0.42
0.40
0.38
0.36
0.34
0.32
0.30
RMSE
10^0
10^1
Speed-scale ratio (c) — both for truth and DA
ETKF tuned
ETKF excessive
EAKF adaptive
ETKF adaptive
EnKF-N hybrid

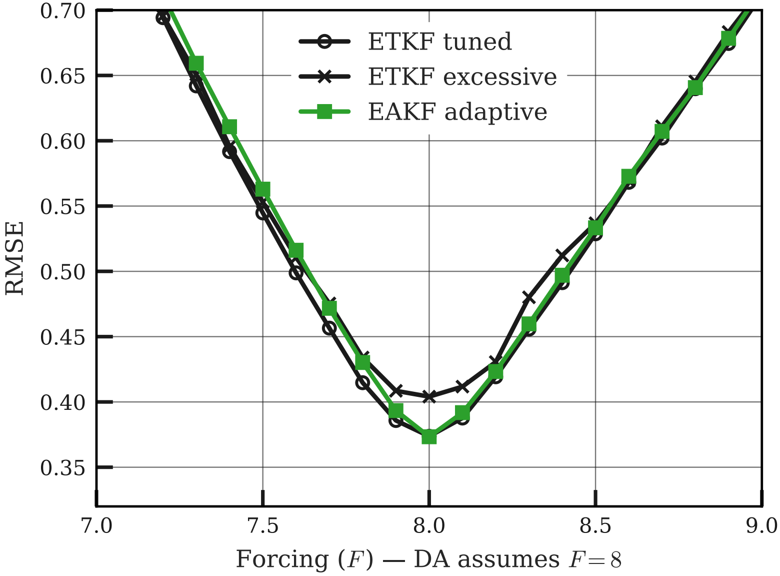
Benchmarks (single-layer)



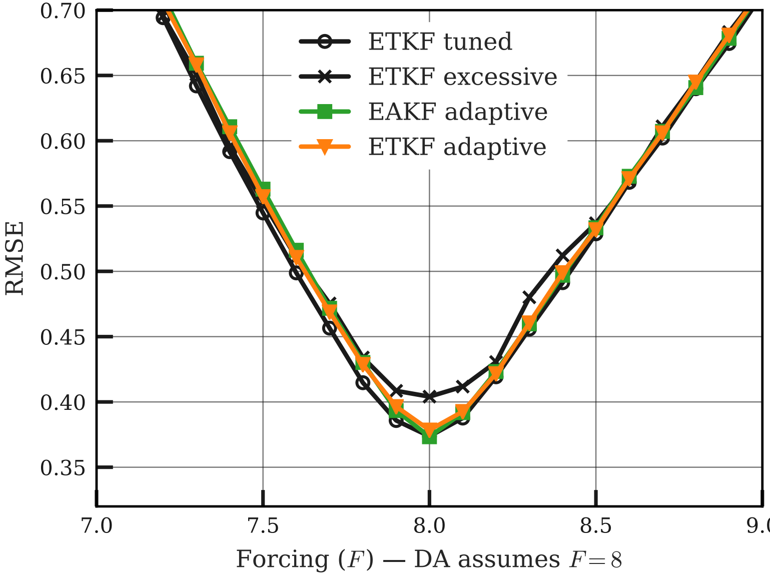
Benchmarks (single-layer)



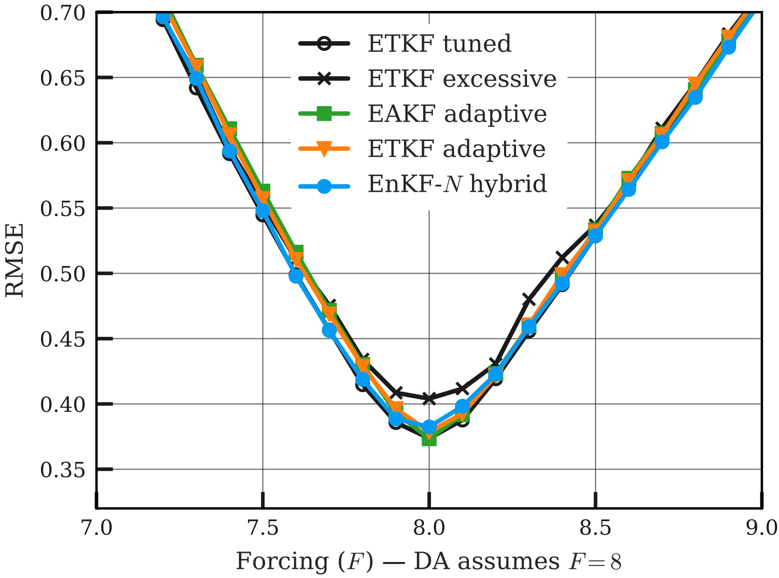
Benchmarks (single-layer)



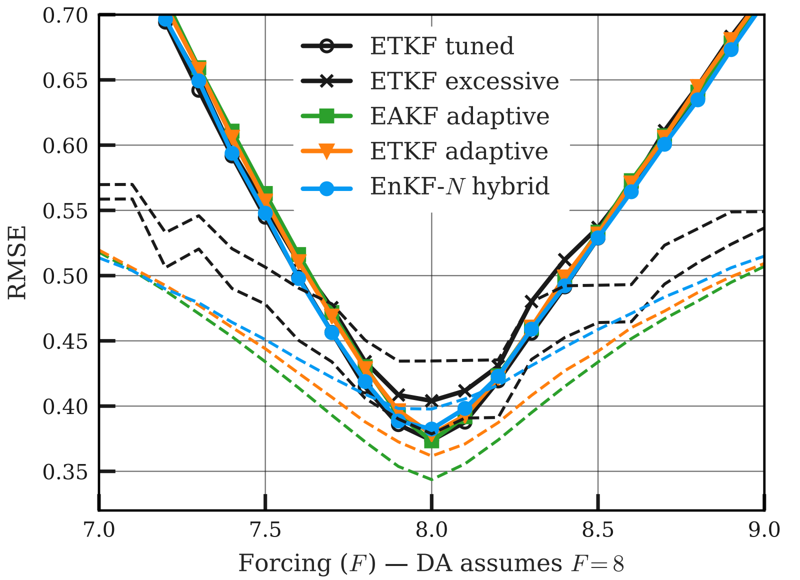
Benchmarks (single-layer)



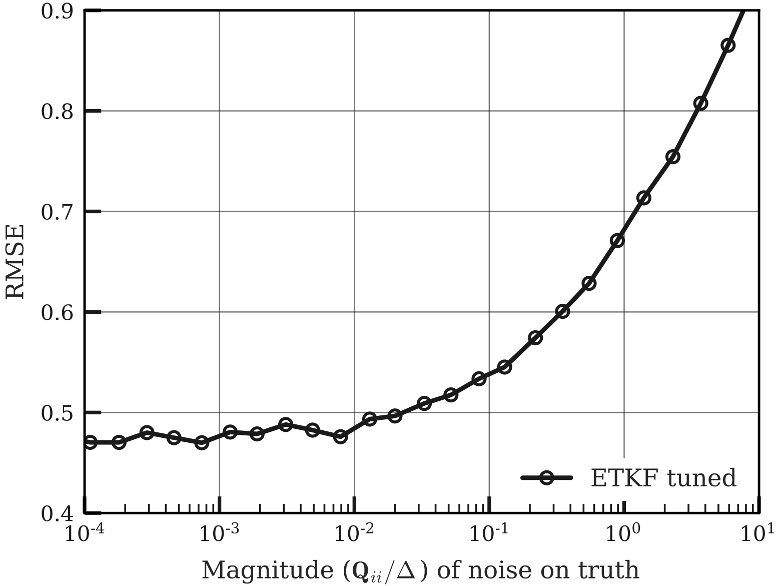
Benchmarks (single-layer)



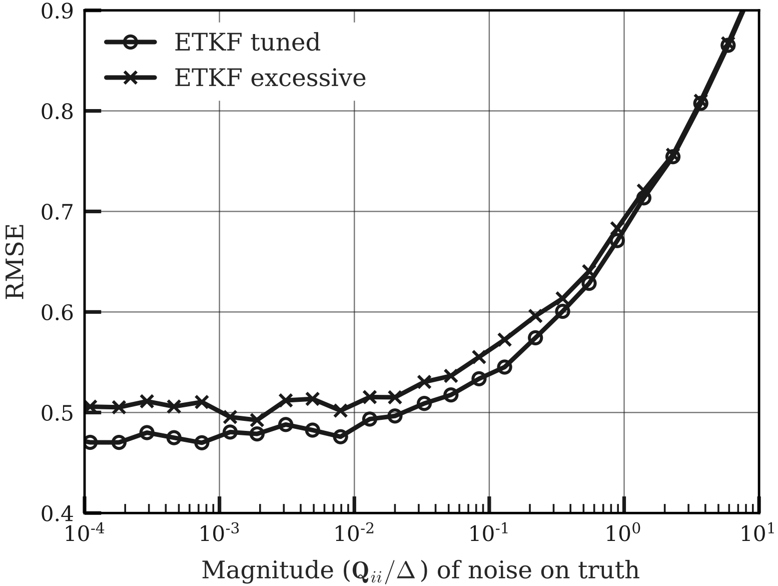
Benchmarks (single-layer)



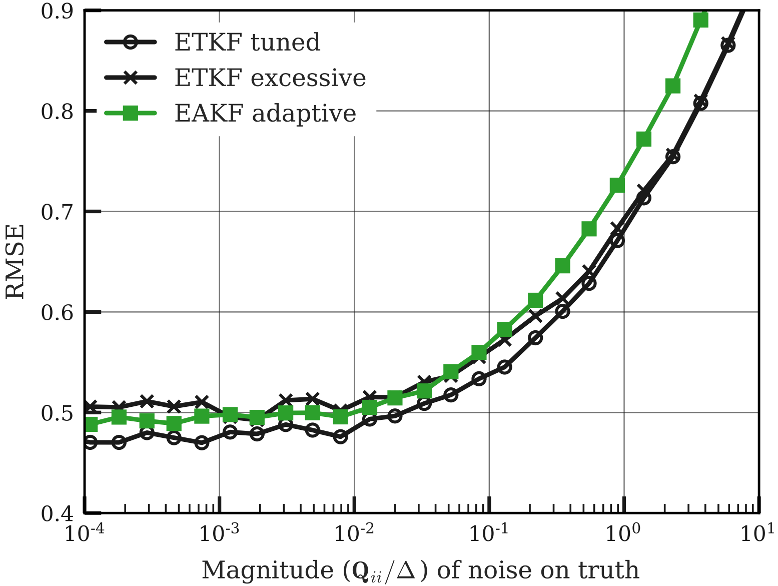
Benchmarks (Lorenz-63) – $N = 3$



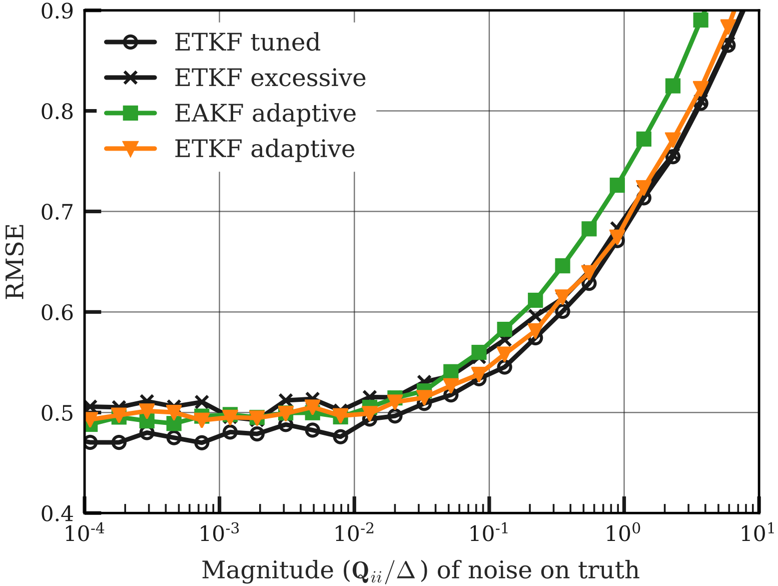
Benchmarks (Lorenz-63) – $N = 3$



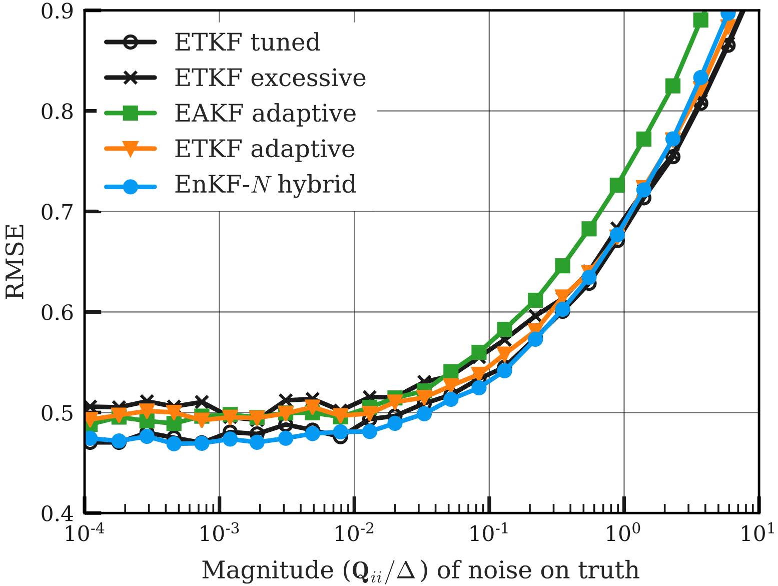
Benchmarks (Lorenz-63) – $N = 3$



Benchmarks (Lorenz-63) – $N = 3$



Benchmarks (Lorenz-63) – $N = 3$



Benchmarks (Lorenz-63) – $N = 3$