Conditional Gaussian Nonlinear Systems

Quanling Deng

Yau Mathematical Sciences Center
Tsinghua University

Introduction to Data Assimilation: L13-L14

Main reference

Nan Chen and Andrew J Majda. "Conditional Gaussian systems for multiscale nonlinear stochastic systems: Prediction, state estimation and uncertainty quantification." Entropy 20.7 (2018): 509.

Nan Chen and Andrew J. Majda. "Filtering nonlinear turbulent dynamical systems through conditional Gaussian statistics." Monthly Weather Review 144.12 (2016): 4885-4917.

Nan Chen. "Improving the prediction of complex nonlinear turbulent dynamical systems using nonlinear filter, smoother and backward sampling techniques." Research in the Mathematical Sciences 7.3 (2020): 1-39.

Acknowledgement: We gratefully acknowledge that these lecture notes are adapted from Nan Chen's original materials.

Introduction

Nonlinear complex dynamical systems

- ubiquitous in geoscience, engineering, neural and material sciences
- a large dimensional phase space
- strong intermittent instabilities
- extreme and rare events, intermittency, fat-tailed probability density functions (PDFs) and other non-Gaussian features ...

Introduction

Nonlinear complex dynamical systems

- ubiquitous in geoscience, engineering, neural and material sciences
- a large dimensional phase space
- strong intermittent instabilities
- extreme and rare events, intermittency, fat-tailed probability density functions (PDFs) and other non-Gaussian features ...

Key applied math/science issues

- mathematical structural properties and qualitative features
- short- and long-range forecasting
- uncertainty quantification
- state estimation, filtering or data assimilation
- model error

I. General Mathematical Framework

Nonlinear Conditional Gaussian Systems

Despite the fully nonlinearity in many multiscale turbulent dynamical systems and the non-Gaussian features in both the marginal and joint PDFs, these systems have conditional Gaussian structures.

Nonlinear Conditional Gaussian Systems

Despite the fully nonlinearity in many multiscale turbulent dynamical systems and the non-Gaussian features in both the marginal and joint PDFs, these systems have conditional Gaussian structures.

The general nonlinear conditional Gaussian systems (Chen & Majda, 2018 *Entropy*, 2016 *MWR*),

$$d\mathbf{u_I} = [\mathbf{A}_0(t,\mathbf{u_I}) + \mathbf{A}_1(t,\mathbf{u_I})\mathbf{u_{II}}]dt + \mathbf{\Sigma_I}(t,\mathbf{u_I})d\mathbf{W_I}(t) \tag{1a}$$

$$d\mathbf{u_{II}} = [\mathbf{a}_0(t,\mathbf{u_I}) + \mathbf{a}_1(t,\mathbf{u_I})\mathbf{u_{II}}]dt + \mathbf{\Sigma_{II}}(t,\mathbf{u_I})d\mathbf{W_{II}}(t) \tag{1b}$$

Once $\mathbf{u_I}(s)$ for $s \leq t$ is given, $\mathbf{u_{II}}(t)$ conditioned on $\mathbf{u_I}(s)$ becomes a Gaussian process,

$$p\big(\mathbf{u_{II}}(t)|\mathbf{u_I}(s \leq t)\big) \sim \mathcal{N}(\bar{\mathbf{u}}_{\mathbf{II}}(t), \mathbf{R_{II}}(t)). \tag{2}$$

Nonlinear Conditional Gaussian Systems

Despite the fully nonlinearity in many multiscale turbulent dynamical systems and the non-Gaussian features in both the marginal and joint PDFs, these systems have conditional Gaussian structures.

The general nonlinear conditional Gaussian systems (Chen & Majda, 2018 *Entropy*, 2016 *MWR*),

$$d\mathbf{u}_{\mathbf{I}} = [\mathbf{A}_0(t,\mathbf{u}_{\mathbf{I}}) + \mathbf{A}_1(t,\mathbf{u}_{\mathbf{I}})\mathbf{u}_{\mathbf{II}}]dt + \mathbf{\Sigma}_{\mathbf{I}}(t,\mathbf{u}_{\mathbf{I}})d\mathbf{W}_{\mathbf{I}}(t) \tag{1a}$$

$$d\mathbf{u}_{\mathbf{II}} = [\mathbf{a}_0(t,\mathbf{u}_{\mathbf{I}}) + \mathbf{a}_1(t,\mathbf{u}_{\mathbf{I}})\mathbf{u}_{\mathbf{II}}]dt + \mathbf{\Sigma}_{\mathbf{II}}(t,\mathbf{u}_{\mathbf{I}})d\mathbf{W}_{\mathbf{II}}(t) \tag{1b}$$

Once $\mathbf{u}_{\mathbf{I}}(s)$ for $s \leq t$ is given, $\mathbf{u}_{\mathbf{II}}(t)$ conditioned on $\mathbf{u}_{\mathbf{I}}(s)$ becomes a Gaussian process,

$$p\big(\mathbf{u}_{\mathbf{II}}(t)|\mathbf{u}_{\mathbf{I}}(s \leq t)\big) \sim \mathcal{N}(\bar{\mathbf{u}}_{\mathbf{II}}(t), \mathbf{R}_{\mathbf{II}}(t)). \tag{2}$$

- Despite the conditional Gaussianity, the coupled system (1) remains highly nonlinear and is able to capture the non-Gaussian features as in nature.
- The conditional Gaussian distribution in (2) has closed analytic form:

$$d\bar{\mathbf{u}}_{\mathbf{II}}(t) =[\mathbf{a}_0(t,\mathbf{u}_{\mathbf{I}}) + \mathbf{a}_1(t,\mathbf{u}_{\mathbf{I}})\bar{\mathbf{u}}_{\mathbf{II}}]dt + (\mathbf{R}_{\mathbf{II}}\mathbf{A}_1^*(t,\mathbf{u}_{\mathbf{I}}))(\mathbf{\Sigma}_{\mathbf{I}}\mathbf{\Sigma}_{\mathbf{I}}^*)^{-1}(t,\mathbf{u}_{\mathbf{I}})\,[d\mathbf{u}_{\mathbf{I}} - (\mathbf{A}_0(t,\mathbf{u}_{\mathbf{I}}) + \mathbf{A}_1(t,\mathbf{u}_{\mathbf{I}})\bar{\mathbf{u}}_{\mathbf{II}})dt]\,,$$

$$d\mathbf{R}_{\mathbf{II}}(t) =\Big\{\mathbf{a}_1(t,\mathbf{u}_{\mathbf{I}})\mathbf{R}_{\mathbf{II}} + \mathbf{R}_{\mathbf{II}}\mathbf{a}_1^*(t,\mathbf{u}_{\mathbf{I}}) + (\mathbf{\Sigma}_{\mathbf{II}}\mathbf{\Sigma}_{\mathbf{II}}^*)(t,\mathbf{u}_{\mathbf{I}}) - (\mathbf{R}_{\mathbf{II}}\mathbf{A}_1^*(t,\mathbf{u}_{\mathbf{I}}))(\mathbf{\Sigma}_{\mathbf{I}}\mathbf{\Sigma}_{\mathbf{I}}^*)^{-1}(t,\mathbf{u}_{\mathbf{I}})(\mathbf{R}_{\mathbf{II}}\mathbf{A}_1^*(t,\mathbf{u}_{\mathbf{I}}))^*\Big\}dt.$$

- This allows the development of both rigourous mathematical theories and efficient numerical algorithms for these complex turbulent dynamical systems.

Special case: the Kalman-Bucy filter.

A special case of the general conditional Gaussian framework is the so-called Kalman-Bucy filter, which is a continuous time version of the Kalman filter and it deals with the linear coupled systems,

$$d\mathbf{u_I} = [\mathbf{A}_0(t) + \mathbf{A}_1(t)\mathbf{u_{II}}]dt + \mathbf{\Sigma_I}(t)d\mathbf{W_I}(t)$$
$$d\mathbf{u_{II}} = [\mathbf{a}_0(t) + \mathbf{a}_1(t)\mathbf{u_{II}}]dt + \mathbf{\Sigma_{II}}(t)d\mathbf{W_{II}}(t)$$

As an analog to the continuous time conditional Gaussian systems, the general form of the discrete conditional Gaussian nonlinear models is as follows,

$$\mathbf{u_I}(t_{j+1}) = \mathbf{A_0}(\mathbf{u_I}(t_j), t_j) + \mathbf{A_1}(\mathbf{u_I}(t_j), t_j)\mathbf{u_{II}}(t_j) + \mathbf{B_1}(\mathbf{u_I}(t_j), t_j)\boldsymbol{\varepsilon}_\mathbf{1}(t_{j+1}) + \mathbf{B_2}(\mathbf{u_I}(t_j), t_j)\boldsymbol{\varepsilon}_\mathbf{2}(t_{j+1}), \tag{3a}$$

$$\mathbf{u_{II}}(t_{j+1}) = \mathbf{a_0}(\mathbf{u_I}(t_j), t_j) + \mathbf{a_1}(\mathbf{u_I}(t_j), t_j)\mathbf{u_{II}}(t_j) + \mathbf{b_1}(\mathbf{u_I}(t_j), t_j)\boldsymbol{\varepsilon}_\mathbf{1}(t_{j+1}) + \mathbf{b_2}(\mathbf{u_I}(t_j), t_j)\boldsymbol{\varepsilon}_\mathbf{2}(t_{j+1}), \tag{3b}$$

As an analog to the continuous time conditional Gaussian systems, the general form of the discrete conditional Gaussian nonlinear models is as follows,

$$\begin{aligned}\mathbf{u_I}(t_{j+1}) = \mathbf{A_0}(\mathbf{u_I}(t_j), t_j) + \mathbf{A_1}(\mathbf{u_I}(t_j), t_j)\mathbf{u_{II}}(t_j) \\ + \mathbf{B_1}(\mathbf{u_I}(t_j), t_j)\varepsilon_\mathbf{1}(t_{j+1}) + \mathbf{B_2}(\mathbf{u_I}(t_j), t_j)\varepsilon_\mathbf{2}(t_{j+1}),\end{aligned} \tag{3a}$$

$$\begin{aligned}\mathbf{u_{II}}(t_{j+1}) = \mathbf{a_0}(\mathbf{u_I}(t_j), t_j) + \mathbf{a_1}(\mathbf{u_I}(t_j), t_j)\mathbf{u_{II}}(t_j) \\ + \mathbf{b_1}(\mathbf{u_I}(t_j), t_j)\varepsilon_\mathbf{1}(t_{j+1}) + \mathbf{b_2}(\mathbf{u_I}(t_j), t_j)\varepsilon_\mathbf{2}(t_{j+1}),\end{aligned} \tag{3b}$$

Assume a sequence of the observed variable $\mathbf{u_I}$, namely $\{\mathbf{u_I}(t_0), \mathbf{u_I}(t_1), \ldots, \mathbf{u_I}(t_{j+1})\}$, is available. Then the distribution of $\mathbf{u_{II}}(t_{j+1})$ conditioned on this given observed sequence is conditional Gaussian,

$$p(\mathbf{u_{II}}(t_{j+1})|\mathbf{u_I}(s), s \leq t_{j+1}) \sim \mathcal{N}(\boldsymbol{\mu}(t_{j+1}), \mathbf{R}(t_{j+1})). \tag{4}$$

As an analog to the continuous time conditional Gaussian systems, the general form of the discrete conditional Gaussian nonlinear models is as follows,

$$\begin{aligned}\mathbf{u_I}(t_{j+1}) = \mathbf{A_0}(\mathbf{u_I}(t_j), t_j) + \mathbf{A_1}(\mathbf{u_I}(t_j), t_j)\mathbf{u_{II}}(t_j)\\ + \mathbf{B_1}(\mathbf{u_I}(t_j), t_j)\boldsymbol{\varepsilon}_\mathbf{1}(t_{j+1}) + \mathbf{B_2}(\mathbf{u_I}(t_j), t_j)\boldsymbol{\varepsilon}_\mathbf{2}(t_{j+1}),\end{aligned} \tag{3a}$$

$$\begin{aligned}\mathbf{u_{II}}(t_{j+1}) = \mathbf{a_0}(\mathbf{u_I}(t_j), t_j) + \mathbf{a_1}(\mathbf{u_I}(t_j), t_j)\mathbf{u_{II}}(t_j)\\ + \mathbf{b_1}(\mathbf{u_I}(t_j), t_j)\boldsymbol{\varepsilon}_\mathbf{1}(t_{j+1}) + \mathbf{b_2}(\mathbf{u_I}(t_j), t_j)\boldsymbol{\varepsilon}_\mathbf{2}(t_{j+1}),\end{aligned} \tag{3b}$$

Assume a sequence of the observed variable $\mathbf{u_I}$, namely $\{\mathbf{u_I}(t_0), \mathbf{u_I}(t_1), \ldots, \mathbf{u_I}(t_{j+1})\}$, is available. Then the distribution of $\mathbf{u_{II}}(t_{j+1})$ conditioned on this given observed sequence is conditional Gaussian,

$$p(\mathbf{u_{II}}(t_{j+1})|\mathbf{u_I}(s), s \leq t_{j+1}) \sim \mathcal{N}(\boldsymbol{\mu}(t_{j+1}), \mathbf{R}(t_{j+1})). \tag{4}$$

The time evolutions of the conditional mean $\boldsymbol{\mu}(t_{j+1})$ and conditional covariance $\mathbf{R}(t_{j+1})$ are given by the following explicit formulae,

$$\begin{aligned}\boldsymbol{\mu}(t_{j+1}) = \mathbf{a_0} + \mathbf{a_1}\boldsymbol{\mu}(t_j) + (\mathbf{b}\circ\mathbf{B} + \mathbf{a_1}\mathbf{R}(t_i)\mathbf{A_1^*})\times\\ (\mathbf{B}\circ\mathbf{B} + \mathbf{A_1}\mathbf{R}(t_i)\mathbf{A_1^*})^{-1}(\mathbf{u_I}(t_{i+1}) - \mathbf{A_0} - \mathbf{A_1}\boldsymbol{\mu}(t_i)),\end{aligned} \tag{5a}$$

$$\begin{aligned}\mathbf{R}(t_{j+1}) = \mathbf{a_1}\mathbf{R}(t_i)\mathbf{a_1^*} + \mathbf{b}\circ\mathbf{b} - (\mathbf{b}\circ\mathbf{B} + \mathbf{a_1}\mathbf{R}(t_i)\mathbf{A_1^*})\times\\ (\mathbf{B}\circ\mathbf{B} + \mathbf{A_1}\mathbf{R}(t_i)\mathbf{A_1^*})^{-1}(\mathbf{b}\circ\mathbf{B} + \mathbf{a_1}\mathbf{R}(t_i)\mathbf{A_1^*})^*,\end{aligned} \tag{5b}$$

where

$$\mathbf{b}\circ\mathbf{b} = \mathbf{b_1}\mathbf{b_1^*} + \mathbf{b_2}\mathbf{b_2^*}, \quad \mathbf{b}\circ\mathbf{B} = \mathbf{b_1}\mathbf{B_1^*} + \mathbf{b_2}\mathbf{B_2^*}, \quad \mathbf{B}\circ\mathbf{B} = \mathbf{B_1}\mathbf{B_1^*} + \mathbf{B_2}\mathbf{B_2^*}.$$

Lemma

Let the Gaussian random variables be

$$\mathbf{x} = \begin{pmatrix} \mathbf{x_1} \\ \mathbf{x_2} \end{pmatrix},$$

with mean $\boldsymbol{\mu}$ *and covariance* $\mathbf{R}$*,*

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu_1} \\ \boldsymbol{\mu_2} \end{pmatrix}, \qquad \mathbf{R} = \begin{pmatrix} \mathbf{R_{11}} & \mathbf{R_{12}} \\ \mathbf{R_{21}} & \mathbf{R_{22}} \end{pmatrix}.$$

The conditional distribution

$$p(\mathbf{x_1}|\mathbf{x_2}) \sim \mathcal{N}(\overline{\boldsymbol{\mu}}, \overline{\mathbf{R}}),$$

where

$$\begin{aligned} \overline{\boldsymbol{\mu}} &= \boldsymbol{\mu_1} + \mathbf{R_{12}}\mathbf{R_{22}^{-1}}(\mathbf{x_2} - \boldsymbol{\mu}_2), \\ \overline{\mathbf{R}} &= \mathbf{R_{11}} - \mathbf{R_{12}}\mathbf{R_{22}^{-1}}\mathbf{R_{21}}. \end{aligned} \tag{6}$$

$$\mathbf{u_I}(t_{j+1}) = \mathbf{A_0}(\mathbf{u_I}(t_j), t_j) + \mathbf{A_1}(\mathbf{u_I}(t_j), t_j)\mathbf{u_{II}}(t_j) + \mathbf{B_1}(\mathbf{u_I}(t_j), t_j)\boldsymbol{\varepsilon_1}(t_{j+1}) + \mathbf{B_2}(\mathbf{u_I}(t_j), t_j)\boldsymbol{\varepsilon_2}(t_{j+1}), \tag{3a}$$

$$\mathbf{u_{II}}(t_{j+1}) = \mathbf{a_0}(\mathbf{u_I}(t_j), t_j) + \mathbf{a_1}(\mathbf{u_I}(t_j), t_j)\mathbf{u_{II}}(t_j) + \mathbf{b_1}(\mathbf{u_I}(t_j), t_j)\boldsymbol{\varepsilon_1}(t_{j+1}) + \mathbf{b_2}(\mathbf{u_I}(t_j), t_j)\boldsymbol{\varepsilon_2}(t_{j+1}), \tag{3b}$$

Proof.

Consider the joint distribution $p(\mathbf{u_I}^{j+1}, \mathbf{u_{II}}^{j+1} | \mathbf{u_I}^s, s \leq j)$. In light of (3a),

$$p(\mathbf{u_I}^{j+1} | \mathbf{u_I}^s, s \leq j) \sim \mathcal{N}(\mathbf{A}_0^j + \mathbf{A}_1^j \boldsymbol{\mu}^j, \mathbf{A}_1^j \mathbf{R}^j (\mathbf{A}_1^j)^* + \mathbf{B}^j \circ \mathbf{B}^j). \tag{7}$$

Using the same argument, running the model (3b) forward yields

$$p(\mathbf{u_{II}}^{j+1} | \mathbf{X}^s, s \leq j) \sim \mathcal{N}(\mathbf{a}_0^j + \mathbf{a}_1^j \boldsymbol{\mu}^j, \mathbf{a}_1^j \mathbf{R}^j (\mathbf{a}_1^j)^* + \mathbf{b}^j \circ \mathbf{b}^j). \tag{8}$$

$$\mathbf{u_I}(t_{j+1}) = \mathbf{A_0}(\mathbf{u_I}(t_j), t_j) + \mathbf{A_1}(\mathbf{u_I}(t_j), t_j)\mathbf{u_{II}}(t_j) + \mathbf{B_1}(\mathbf{u_I}(t_j), t_j)\varepsilon_1(t_{j+1}) + \mathbf{B_2}(\mathbf{u_I}(t_j), t_j)\varepsilon_2(t_{j+1}), \tag{3a}$$

$$\mathbf{u_{II}}(t_{j+1}) = \mathbf{a_0}(\mathbf{u_I}(t_j), t_j) + \mathbf{a_1}(\mathbf{u_I}(t_j), t_j)\mathbf{u_{II}}(t_j) + \mathbf{b_1}(\mathbf{u_I}(t_j), t_j)\varepsilon_1(t_{j+1}) + \mathbf{b_2}(\mathbf{u_I}(t_j), t_j)\varepsilon_2(t_{j+1}), \tag{3b}$$

Proof.

Consider the joint distribution $p(\mathbf{u_I}^{j+1}, \mathbf{u_{II}}^{j+1}|\mathbf{u_I}^{s}, s \leq j)$. In light of (3a),

$$p(\mathbf{u_I}^{j+1}|\mathbf{u_I}^{s}, s \leq j) \sim \mathcal{N}(\mathbf{A}_0^j + \mathbf{A}_1^j\boldsymbol{\mu}^j, \mathbf{A}_1^j\mathbf{R}^j(\mathbf{A}_1^j)^* + \mathbf{B}^j \circ \mathbf{B}^j). \tag{7}$$

Using the same argument, running the model (3b) forward yields

$$p(\mathbf{u_{II}}^{j+1}|\mathbf{X}^{s}, s \leq j) \sim \mathcal{N}(\mathbf{a}_0^j + \mathbf{a}_1^j\boldsymbol{\mu}^j, \mathbf{a}_1^j\mathbf{R}^j(\mathbf{a}_1^j)^* + \mathbf{b}^j \circ \mathbf{b}^j). \tag{8}$$

The cross-covariance can be derived by removing the mean in (3a) and multiplying the resulting equation by $(\mathbf{u_{II}}^{\prime,j+1})^*$, where $(\mathbf{u_{II}}^{\prime,j+1})^*$ is $(\mathbf{u_{II}}^{j+1})^*$ subtracting its mean,

$$\langle \mathbf{u_I}^{\prime,j+1}(\mathbf{u_{II}}^{\prime,j+1})^* \rangle = \mathbf{A}_1^j\mathbf{R}^j(\mathbf{a}_1^j)^* + (\mathbf{b}^j \circ \mathbf{B}^j)^*. \tag{9}$$

Collecting (7), (8) and (9) leads to

$$\begin{aligned}&p(\mathbf{u_I}^{j+1}, \mathbf{u_{II}}^{j+1}|\mathbf{u_I}^{s}, s \leq j)\\ &\sim \mathcal{N}\left(\begin{pmatrix} \mathbf{A}_0^j + \mathbf{A}_1^j\boldsymbol{\mu}^j \\ \mathbf{a}_0^j + \mathbf{a}_1^j\boldsymbol{\mu}^j \end{pmatrix}, \begin{pmatrix} \mathbf{A}_1^j\mathbf{R}^j(\mathbf{A}_1^j)^* + \mathbf{B}^j \circ \mathbf{B}^j & \mathbf{A}_1^j\mathbf{R}^j(\mathbf{a}_1^j)^* + (\mathbf{b}^j \circ \mathbf{B}^j)^* \\ \mathbf{a}_1^j\mathbf{R}^j(\mathbf{A}_1^j)^* + \mathbf{b}^j \circ \mathbf{B}^j & \mathbf{a}_1^j\mathbf{R}^j(\mathbf{a}_1^j)^* + \mathbf{b}^j \circ \mathbf{b}^j \end{pmatrix}\right)\end{aligned}$$

$$p(\mathbf{u}_{\mathbf{I}}^{j+1}, \mathbf{u}_{\mathbf{II}}^{j+1} | \mathbf{u}_{\mathbf{I}}^{s}, s \leq j)$$
$$\sim \mathcal{N}\left(\left(\begin{array}{c} \mathbf{A}_0^j + \mathbf{A}_1^j \boldsymbol{\mu}^j \\ \mathbf{a}_0^j + \mathbf{a}_1^j \boldsymbol{\mu}^j \end{array}\right), \left(\begin{array}{cc} \mathbf{A}_1^j \mathbf{R}^j (\mathbf{A}_1^j)^* + \mathbf{B}^j \circ \mathbf{B}^j & \mathbf{A}_1^j \mathbf{R}^j (\mathbf{a}_1^j)^* + (\mathbf{b}^j \circ \mathbf{B}^j)^* \\ \mathbf{a}_1^j \mathbf{R}^j (\mathbf{A}_1^j)^* + \mathbf{b}^j \circ \mathbf{B}^j & \mathbf{a}_1^j \mathbf{R}^j (\mathbf{a}_1^j)^* + \mathbf{b}^j \circ \mathbf{b}^j \end{array}\right)\right)$$

Then making use of (6) in the Lemma finishes the proof,

$$\boldsymbol{\mu}(t_{j+1}) = \mathbf{a_0} + \mathbf{a_1}\boldsymbol{\mu}(t_j) + (\mathbf{b} \circ \mathbf{B} + \mathbf{a_1}\mathbf{R}(t_i)\mathbf{A}_1^*)(\mathbf{B} \circ \mathbf{B} + \mathbf{A_1}\mathbf{R}(t_i)\mathbf{A}_1^*)^{-1}(\mathbf{u_I}(t_{i+1}) - \mathbf{A_0} - \mathbf{A_1}\boldsymbol{\mu}(t_i)),$$
$$\mathbf{R}(t_{j+1}) = \mathbf{a_1}\mathbf{R}(t_i)\mathbf{a}_1^* + \mathbf{b} \circ \mathbf{b} - (\mathbf{b} \circ \mathbf{B} + \mathbf{a_1}\mathbf{R}(t_i)\mathbf{A}_1^*)(\mathbf{B} \circ \mathbf{B} + \mathbf{A_1}\mathbf{R}(t_i)\mathbf{A}_1^*)^{-1}(\mathbf{b} \circ \mathbf{B} + \mathbf{a_1}\mathbf{R}(t_i)\mathbf{A}_1^*)^*.$$

Lemma

Let the Gaussian random variables be

$$\mathbf{x} = \left(\begin{array}{c} \mathbf{x_1} \\ \mathbf{x_2} \end{array}\right),$$

with mean $\boldsymbol{\mu}$ and covariance $\mathbf{R}$,

$$\boldsymbol{\mu} = \left(\begin{array}{c} \boldsymbol{\mu_1} \\ \boldsymbol{\mu_2} \end{array}\right), \qquad \mathbf{R} = \left(\begin{array}{cc} \mathbf{R_{11}} & \mathbf{R_{12}} \\ \mathbf{R_{21}} & \mathbf{R_{22}} \end{array}\right).$$

The conditional distribution

$$p(\mathbf{x_1}|\mathbf{x_2}) \sim \mathcal{N}(\overline{\boldsymbol{\mu}}, \overline{\mathbf{R}}),$$

where

$$\overline{\boldsymbol{\mu}} = \boldsymbol{\mu_1} + \mathbf{R_{12}}\mathbf{R}_{\mathbf{22}}^{-1}(\mathbf{x_2} - \boldsymbol{\mu}_2),$$
$$\overline{\mathbf{R}} = \mathbf{R_{11}} - \mathbf{R_{12}}\mathbf{R}_{\mathbf{22}}^{-1}\mathbf{R_{21}}. \qquad (6)$$

Special case: the Kalman filter.

$$\mathbf{u_I}(t_{j+1}) = \mathbf{G}(t_j)\mathbf{u_{II}}(t_{j+1}) + \mathbf{B_2}(t_j)\varepsilon_\mathbf{2}(t_{j+1}), \tag{10a}$$

$$\mathbf{u_{II}}(t_{j+1}) = \mathbf{a_0}(t_j) + \mathbf{a_1}(t_j)\mathbf{u_{II}}(t_j) + \mathbf{b_1}(t_j)\varepsilon_\mathbf{1}(t_{j+1}), \tag{10b}$$

Special case: the Kalman filter.

$$\mathbf{u_I}(t_{j+1}) = \mathbf{G}(t_j)\mathbf{u_{II}}(t_{j+1}) + \mathbf{B_2}(t_j)\varepsilon_\mathbf{2}(t_{j+1}), \tag{10a}$$

$$\mathbf{u_{II}}(t_{j+1}) = \mathbf{a_0}(t_j) + \mathbf{a_1}(t_j)\mathbf{u_{II}}(t_j) + \mathbf{b_1}(t_j)\varepsilon_\mathbf{1}(t_{j+1}), \tag{10b}$$

Replacing $\mathbf{u_{II}}(t_{j+1})$ on the right hand side of (10a) by the equation (10b) and the resulting coupled system reads,

$$\mathbf{u_I}(t_{j+1}) = \mathbf{A_0}(t_j) + \mathbf{A_1}(t_j)\mathbf{u_{II}}(t_j) + \mathbf{B_1}(t_j)\varepsilon_\mathbf{1}(t_{j+1}) + \mathbf{B_2}(t_j)\varepsilon_\mathbf{2}(t_{j+1}), \tag{11a}$$

$$\mathbf{u_{II}}(t_{j+1}) = \mathbf{a_0}(t_j) + \mathbf{a_1}(t_j)\mathbf{u_{II}}(t_j) + \mathbf{b_1}(t_j)\varepsilon_\mathbf{1}(t_{j+1}), \tag{11b}$$

where in (11a)

$$\mathbf{A_0}(t_j) = \mathbf{G}(t_j)\mathbf{a_0}(t_j), \qquad \mathbf{A_1}(t_j) = \mathbf{G}(t_j)\mathbf{a_1}(t_j) \qquad \text{and} \qquad \mathbf{B_1}(t_j) = \mathbf{G}(t_j)\mathbf{b_1}(t_j). \tag{12}$$

Data v.s. Model

Vast amount of observational data are available for understanding nature.

However ... many purely data-driven statistical methods fail to capture the key features of nature and suffer from the curse of dimensionality.

Data v.s. Model

Vast amount of observational data are available for understanding nature.
However ... many purely data-driven statistical methods fail to capture the key features of nature and suffer from the curse of dimensionality.

Data-driven modeling framework

- **Combining model with data becomes necessary.**
- Systematic physics-constrained data-driven models have been developed.

Data v.s. Model

Vast amount of observational data are available for understanding nature.
However ... many purely data-driven statistical methods fail to capture the key features of nature and suffer from the curse of dimensionality.

Data-driven modeling framework

- **Combining model with data becomes necessary.**
- Systematic physics-constrained data-driven models have been developed.

The importance of models and data assimilation

- Only **partial** and **noisy observations** are available!
- Models are combined with the available observations to
  - estimate the states of the unobserved variables, and
  - reduce the noise in the observed variables.

Data v.s. Model

Vast amount of observational data are available for understanding nature.
However ... many purely data-driven statistical methods fail to capture the key features of nature and suffer from the curse of dimensionality.

Data-driven modeling framework

- **Combining model with data becomes necessary.**
- Systematic physics-constrained data-driven models have been developed.

The importance of models and data assimilation

- Only **partial** and **noisy observations** are available!
- Models are combined with the available observations to
  - estimate the states of the unobserved variables, and
  - reduce the noise in the observed variables.

Efficient data assimilation strategies with solvable conditional statistics

- Major challenges come from strong nonlinearity and large system dimension.
- Effective multiscale data assimilation with **suitable** approximate forecast models
  - Large scale: **fully non-Gaussian**,
  - Small scale: **conditional Gaussian** to the large scale.

  e.g., stochastic superparameterization (Majda & Grooms, 2014 *JCP*), blended particle filter (Majda, Qi & Sapsis, 2014, *PNAS*).

Multiscale conditional Gaussian with stochastic mode reduction strategy.

Let's start with a general nonlinear deterministic model with quadratic nonlinearity,

$$d\mathbf{u} = \left[(\mathbf{L} + \mathbf{D})\mathbf{u} + \mathbf{B}(\mathbf{u}, \mathbf{u}) + \mathbf{F}(t)\right] dt,$$

Here the state variables $\mathbf{u} = (\mathbf{u}_{\mathbf{I}}, \mathbf{u}_{\mathbf{II}})$ has multiscale features:

- $\mathbf{u}_{\mathbf{I}}$ denotes the resolved variables that evolve slowly in time (e.g., climate variables) while
- $\mathbf{u}_{\mathbf{II}}$ are unresolved or unobserved fast variables (e.g., weather variables).

Multiscale conditional Gaussian with stochastic mode reduction strategy.

Let's start with a general nonlinear deterministic model with quadratic nonlinearity,

$$d\mathbf{u} = \big[(\mathbf{L}+\mathbf{D})\mathbf{u} + \mathbf{B}(\mathbf{u},\mathbf{u}) + \mathbf{F}(t)\big]dt,$$

Here the state variables $\mathbf{u} = (\mathbf{u}_\mathbf{I}, \mathbf{u}_\mathbf{II})$ has multiscale features:

- $\mathbf{u}_\mathbf{I}$ denotes the resolved variables that evolve slowly in time (e.g., climate variables) while
- $\mathbf{u}_\mathbf{II}$ are unresolved or unobserved fast variables (e.g., weather variables).

The above system can be written down into more detailed forms:

$$\begin{aligned}
d\mathbf{u}_\mathbf{I} = \Big[&(\mathbf{L}_{11}+\mathbf{D}_{11})\mathbf{u}_\mathbf{I} + (\mathbf{L}_{12}+\mathbf{D}_{12})\mathbf{u}_\mathbf{II} + \mathbf{B}^1_{11}(\mathbf{u}_\mathbf{I},\mathbf{u}_\mathbf{I}) \\
&+ \mathbf{B}^1_{12}(\mathbf{u}_\mathbf{I},\mathbf{u}_\mathbf{II}) + \mathbf{B}^1_{22}(\mathbf{u}_\mathbf{II},\mathbf{u}_\mathbf{II}) + \mathbf{F}_1(t)\Big]dt, \\
d\mathbf{u}_\mathbf{II} = \Big[&(\mathbf{L}_{22}+\mathbf{D}_{22})\mathbf{u}_\mathbf{II} + (\mathbf{L}_{21}+\mathbf{D}_{21})\mathbf{u}_\mathbf{I} + \mathbf{B}^2_{11}(\mathbf{u}_\mathbf{I},\mathbf{u}_\mathbf{I}) \\
&+ \mathbf{B}^2_{12}(\mathbf{u}_\mathbf{I},\mathbf{u}_\mathbf{II}) + \mathbf{B}^2_{22}(\mathbf{u}_\mathbf{II},\mathbf{u}_\mathbf{II}) + \mathbf{F}_2(t)\Big]dt.
\end{aligned}$$

Multiscale conditional Gaussian with stochastic mode reduction strategy.

Let's start with a general nonlinear deterministic model with quadratic nonlinearity,

$$d\mathbf{u} = \big[(\mathbf{L}+\mathbf{D})\mathbf{u} + \mathbf{B}(\mathbf{u},\mathbf{u}) + \mathbf{F}(t)\big]\,dt,$$

Here the state variables $\mathbf{u} = (\mathbf{u_I}, \mathbf{u_{II}})$ has multiscale features:

- $\mathbf{u_I}$ denotes the resolved variables that evolve slowly in time (e.g., climate variables) while
- $\mathbf{u_{II}}$ are unresolved or unobserved fast variables (e.g., weather variables).

The above system can be written down into more detailed forms:

$$\begin{aligned} d\mathbf{u_I} = \Big[&(\mathbf{L}_{11}+\mathbf{D}_{11})\mathbf{u_I} + (\mathbf{L}_{12}+\mathbf{D}_{12})\mathbf{u_{II}} + \mathbf{B}^1_{11}(\mathbf{u_I},\mathbf{u_I}) \\ &+ \mathbf{B}^1_{12}(\mathbf{u_I},\mathbf{u_{II}}) + \mathbf{B}^1_{22}(\mathbf{u_{II}},\mathbf{u_{II}}) + \mathbf{F}_1(t)\Big]\,dt, \\ d\mathbf{u_{II}} = \Big[&(\mathbf{L}_{22}+\mathbf{D}_{22})\mathbf{u_{II}} + (\mathbf{L}_{21}+\mathbf{D}_{21})\mathbf{u_I} + \mathbf{B}^2_{11}(\mathbf{u_I},\mathbf{u_I}) \\ &+ \mathbf{B}^2_{12}(\mathbf{u_I},\mathbf{u_{II}}) + \mathbf{B}^2_{22}(\mathbf{u_{II}},\mathbf{u_{II}}) + \mathbf{F}_2(t)\Big]\,dt. \end{aligned}$$

To make the above mutiscale system fit into the conditional Gaussian framework, two modifications are needed.

1. The quadratic terms involving the interactions between $\mathbf{u_{II}}$ and itself, namely $\mathbf{B}^1_{22}(\mathbf{u_{II}},\mathbf{u_{II}})$ and $\mathbf{B}^2_{22}(\mathbf{u_{II}},\mathbf{u_{II}})$, are not allowed there.
2. Stochastic noise is required at least to the system of $\mathbf{u_I}$.

To fill in these gaps, the most natural way is to apply idea for stochastic mode reduction:

The equations of motion for the unresolved fast modes are modified by representing the nonlinear self-interactions terms between unresolved modes by stochastic terms.

Using ϵ to represent the time scale separation between $\mathbf{u_I}$ and $\mathbf{u_{II}}$, the terms with quadratic nonlinearity of $\mathbf{u_{II}}$ and itself are approximated by

$$
\begin{aligned}
\mathbf{B}_{22}^1(\mathbf{u_{II}}, \mathbf{u_{II}}) &\approx -\frac{\mathbf{\Gamma}_1}{\epsilon}\mathbf{u_{II}} + \frac{\mathbf{\Sigma_I}}{\sqrt{\epsilon}}\mathbf{W_I}, \\
\mathbf{B}_{22}^2(\mathbf{u_{II}}, \mathbf{u_{II}}) &\approx -\frac{\mathbf{\Gamma}_2}{\epsilon}\mathbf{u_{II}} + \frac{\mathbf{\Sigma_{II}}}{\sqrt{\epsilon}}\mathbf{W_{II}}.
\end{aligned}
\tag{13}
$$

Using ϵ to represent the time scale separation between $\mathbf{u_I}$ and $\mathbf{u_{II}}$, the terms with quadratic nonlinearity of $\mathbf{u_{II}}$ and itself are approximated by

$$\begin{aligned} \mathbf{B}_{22}^{1}(\mathbf{u_{II}}, \mathbf{u_{II}}) &\approx -\frac{\mathbf{\Gamma}_1}{\epsilon}\mathbf{u_{II}} + \frac{\mathbf{\Sigma_I}}{\sqrt{\epsilon}}\mathbf{W_I}, \\ \mathbf{B}_{22}^{2}(\mathbf{u_{II}}, \mathbf{u_{II}}) &\approx -\frac{\mathbf{\Gamma}_2}{\epsilon}\mathbf{u_{II}} + \frac{\mathbf{\Sigma_{II}}}{\sqrt{\epsilon}}\mathbf{W_{II}}. \end{aligned} \tag{13}$$

What's the motivation?

The nonlinear self-interacting terms of fast variables $\mathbf{u_{II}}$ are responsible for the chaotic sensitive dependence on small perturbations and do not require a more detailed description if their effect on the coarse-grained dynamics for the climate variables alone is the main objective. On the other hand, the quadratic nonlinear interactions between $\mathbf{u_I}$ and $\mathbf{u_{II}}$ are retained.

Using ϵ to represent the time scale separation between $\mathbf{u_I}$ and $\mathbf{u_{II}}$, the terms with quadratic nonlinearity of $\mathbf{u_{II}}$ and itself are approximated by

$$\begin{aligned} \mathbf{B}^1_{22}(\mathbf{u_{II}}, \mathbf{u_{II}}) &\approx -\frac{\mathbf{\Gamma}_1}{\epsilon}\mathbf{u_{II}} + \frac{\mathbf{\Sigma_I}}{\sqrt{\epsilon}}\mathbf{W_I}, \\ \mathbf{B}^2_{22}(\mathbf{u_{II}}, \mathbf{u_{II}}) &\approx -\frac{\mathbf{\Gamma}_2}{\epsilon}\mathbf{u_{II}} + \frac{\mathbf{\Sigma_{II}}}{\sqrt{\epsilon}}\mathbf{W_{II}}. \end{aligned} \tag{13}$$

What's the motivation?

The nonlinear self-interacting terms of fast variables $\mathbf{u_{II}}$ are responsible for the chaotic sensitive dependence on small perturbations and do not require a more detailed description if their effect on the coarse-grained dynamics for the climate variables alone is the main objective. On the other hand, the quadratic nonlinear interactions between $\mathbf{u_I}$ and $\mathbf{u_{II}}$ are retained.

Therefore,

$$d\mathbf{u_I} = \Big[(\mathbf{L}_{11} + \mathbf{D}_{11})\mathbf{u_I} + (\mathbf{L}'_{12} + \mathbf{D}_{12})\mathbf{u_{II}} + \mathbf{B}^1_{11}(\mathbf{u_I}, \mathbf{u_I}) + \mathbf{B}^1_{12}(\mathbf{u_I}, \mathbf{u_{II}}) + \mathbf{F}_1(t)\Big]dt + \mathbf{\Sigma}'_{\mathbf{I}} d\mathbf{W_I}(t),$$

$$d\mathbf{u_{II}} = \Big[(\mathbf{L}'_{22} + \mathbf{D}_{22})\mathbf{u_{II}} + (\mathbf{L}_{21} + \mathbf{D}_{21})\mathbf{u_I} + \mathbf{B}^2_{11}(\mathbf{u_I}, \mathbf{u_I}) + \mathbf{B}^2_{12}(\mathbf{u_I}, \mathbf{u_{II}}) + \mathbf{F}_2(t)\Big]dt + \mathbf{\Sigma}'_{\mathbf{II}} \mathbf{W_{II}}(t),$$

where $\mathbf{L}'_{12} = \mathbf{L}_{12} - \mathbf{\Gamma}_1/\epsilon$, $\mathbf{L}'_{22} = \mathbf{L}_{22} - \mathbf{\Gamma}_2/\epsilon$, $\mathbf{\Sigma}'_{\mathbf{I}} = \mathbf{\Sigma_I}/\sqrt{\epsilon}$ and $\mathbf{\Sigma}'_{\mathbf{II}} = \mathbf{\Sigma_{II}}/\sqrt{\epsilon}$.

Clearly, this system belongs to the conditional Gaussian framework.

Notably, if the nonlinear terms satisfy $\mathbf{u} \cdot \mathbf{B}(\mathbf{u}, \mathbf{u}) = 0$, then the system becomes a physics-constrained nonlinear model.

Physics constraint.

For a nonlinear system (either deterministic or stochastic),

$$d\mathbf{u} = \big[(\mathbf{L}+\mathbf{D})\mathbf{u} + \mathbf{B}(\mathbf{u},\mathbf{u}) + \mathbf{F}(t)\big]dt + \boldsymbol{\Sigma} d\mathbf{W}(t),$$

physics constraint means $\mathbf{u}\cdot\mathbf{B}(\mathbf{u},\mathbf{u}) = 0$.

- Most of the key nonlinear dynamical features in fluids and turbulence are given by quadratic nonlinear terms.
- Examples: Eulerian equation, Navier-Stokes equation, Boussinesq equation ...
- Nonlinearity: Advection, convection ... $\mathbf{u}\cdot\nabla\mathbf{u}$, $\mathbf{u}\cdot\nabla T$
- Without the physics constraint (at least in the large-scale dynamics), a fluid system usually lacks physical meaning and suffers from finite time blowup of solution.
- Physics constraint “=” conservation of energy in the quadratic nonlinear terms.

Example:

$$\begin{aligned} dv_1 &= \big((-d_1 + v_2)v_1 + f\big)dt + \sigma_1 dW_1 \\ dv_2 &= (-d_2 v_2 - v_1^2)dt + \sigma_2 dW_2 \end{aligned}$$

Here $\mathbf{u} = (v_1, v_2)^T$ and $\mathbf{B}(\mathbf{u}, \mathbf{u}) = (v_2 v_1, -v_1^2)^T$.
The nonlinear part of the system is

$$\begin{aligned} dv_1 &= v_2 v_1 dt \\ dv_2 &= -v_1^2 dt \end{aligned}$$

Example:

$$
\begin{aligned}
dv_1 &= \big((-d_1 + v_2)v_1 + f\big)dt + \sigma_1 dW_1 \\
dv_2 &= (-d_2 v_2 - v_1^2)dt + \sigma_2 dW_2
\end{aligned}
$$

Here $\mathbf{u} = (v_1, v_2)^T$ and $\mathbf{B}(\mathbf{u}, \mathbf{u}) = (v_2 v_1, -v_1^2)^T$.
The nonlinear part of the system is

$$
\begin{aligned}
dv_1 &= v_2 v_1 dt \\
dv_2 &= -v_1^2 dt
\end{aligned}
$$

Multiplying the two equations by v_1 and v_2 respectively,

$$
\begin{aligned}
& & v_1 dv_1 &= v_1 v_2 v_1 dt \\
+ & & v_2 dv_2 &= -v_2 v_1^2 dt \\
\to & & \frac{1}{2} d(v_1^2 + v_2^2) &= 0 = \mathbf{u} \cdot \mathbf{B}(\mathbf{u}, \mathbf{u})
\end{aligned}
$$

Note that $E \equiv \frac{1}{2}(v_1^2 + v_2^2)$ is the most natural representation of energy.

II. A gallery of examples of conditional Gaussian systems

$$
\begin{aligned}
d\mathbf{u}_{\mathbf{I}} &= [\mathbf{A}_0(t,\mathbf{u}_{\mathbf{I}}) + \mathbf{A}_1(t,\mathbf{u}_{\mathbf{I}})\mathbf{u}_{\mathbf{II}}]dt + \boldsymbol{\Sigma}_{\mathbf{I}}(t,\mathbf{u}_{\mathbf{I}})d\mathbf{W}_{\mathbf{I}}(t)\\
d\mathbf{u}_{\mathbf{II}} &= [\mathbf{a}_0(t,\mathbf{u}_{\mathbf{I}}) + \mathbf{a}_1(t,\mathbf{u}_{\mathbf{I}})\mathbf{u}_{\mathbf{II}}]dt + \boldsymbol{\Sigma}_{\mathbf{II}}(t,\mathbf{u}_{\mathbf{I}})d\mathbf{W}_{\mathbf{II}}(t)
\end{aligned}
$$

1. Physics-constrained nonlinear low-order stochastic models

(Majda & Harlim 2012 *Nonlinearity*, Harlim, Mahdi & Majda, 2014 *JCP*)

- the recent development of data driven statistical-dynamical models for the time series of a partial subset of observed variables
- succeed in overcoming both the finite-time blowup and the lack of physical meaning issues in various ad hoc multi-layer regression models
- often require only a short training period
- contain energy-conserving quadratic nonlinear interactions

$$d\mathbf{u} = \big[(\mathbf{L}+\mathbf{D})\mathbf{u} + \mathbf{B}(\mathbf{u},\mathbf{u}) + \mathbf{F}(t)\big]dt + \mathbf{\Sigma}(t,\mathbf{u})d\mathbf{W}(t),$$
$$\text{with} \qquad \mathbf{u}\cdot\mathbf{B}(\mathbf{u},\mathbf{u}) = 0.$$

Denote $\mathbf{u} = (\mathbf{u_I}, \mathbf{u_{II}})$. Many of the physics-constrained nonlinear stochastic models belong to the nonlinear conditional Gaussian framework.

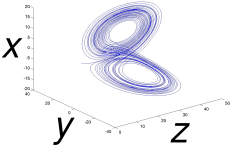
$$
\boxed{\begin{aligned} d\mathbf{u_I} &= [\mathbf{A}_0(t,\mathbf{u_I}) + \mathbf{A}_1(t,\mathbf{u_I})\mathbf{u_{II}}]dt + \mathbf{\Sigma_I}(t,\mathbf{u_I})d\mathbf{W_I}(t) \\ d\mathbf{u_{II}} &= [\mathbf{a}_0(t,\mathbf{u_I}) + \mathbf{a}_1(t,\mathbf{u_I})\mathbf{u_{II}}]dt + \mathbf{\Sigma_{II}}(t,\mathbf{u_I})d\mathbf{W_{II}}(t) \end{aligned}}
$$

Examples.

1) The noisy versions of Lorenz models (L63, L84, two-layer L96 ...)

A noisy Lorenz 63 model

$$
\begin{aligned}
dx &= \sigma(y-x)dt + \sigma_x dW_x, \\
dy &= \big(x(\rho - z) - y\big)dt + \sigma_y dW_y, \\
dz &= (xy - \beta z)dt + \sigma_z dW_z.
\end{aligned}
\qquad
\begin{aligned}
\rho &= 28 \\
\sigma &= 10 \\
\beta &= 8/3
\end{aligned}
$$



A simplified mathematical model for atmospheric convection.

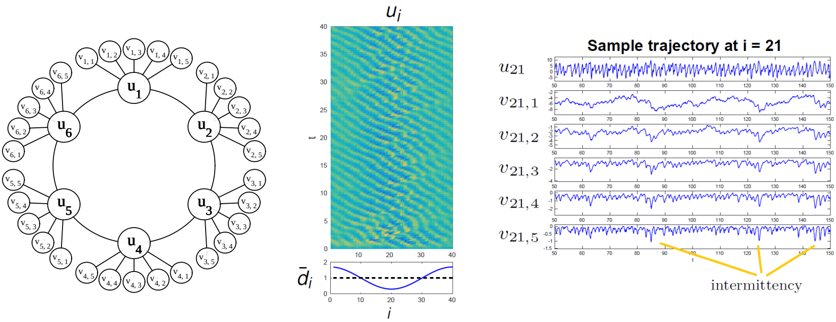
- x is proportional to the rate of convection.
- y to the horizontal temperature variation.
- z to the vertical temperature variation.

A two-layer Lorenz 96 model

$$\frac{du_i}{dt} = u_{i-1}(u_{i+1} - u_{i-2}) + \sum_{j=1}^{J} \gamma_{i,j} u_i v_{i,j} - \bar{d}_i u_i + F + \sigma_u \dot{W}_{u_i}, \qquad i = 1, \ldots, I,$$

$$\frac{dv_{i,j}}{dt} = -d_{v_{i,j}} v_{i,j} - \gamma_j u_i^2 + \sigma_{i,j} \dot{W}_{v_{i,j}}, \quad j = 1, \ldots, J,$$

with $I = 40$ and $J = 5$. The total number of dimension is 240.



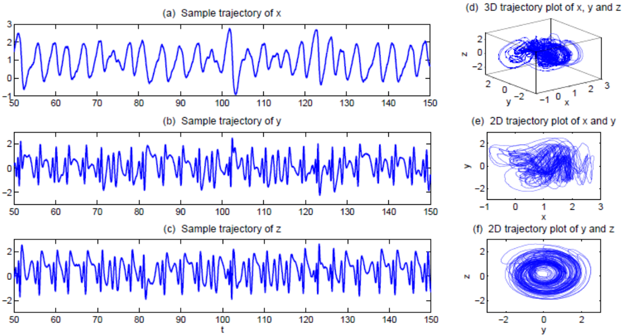
- The first layer can be regarded as a coarse discretization of atmospheric flow on a latitude circle with complicated wave-like and chaotic behavior.
- The second layer includes small-scale fluctuations.

Lorenz 84 model

$$
\begin{aligned}
dx &= \left(-(y^2 + z^2) - a(x - f)\right)dt + \sigma_x dW_x,\\
dy &= \left(-bxz + xy - y + g\right)dt + \sigma_y dW_y,\\
dz &= \left(bxy + xz - z\right)dt + \sigma_z dW_z.
\end{aligned}
$$

This model is an extremely simple analogue of the global atmospheric circulation.

- x represents the intensity of the mid-latitude westerly wind current.
- y and z representing the cosine and sine phases of a chain of vortices superimposed on the zonal flow.
- $x^2 + y^2 + z^2$ is the total scaled energy (kinetic plus potential plus internal).
- These equations can be derived as a Galerkin truncation of the two-layer quasigeostrophic potential vorticity equations in a channel.

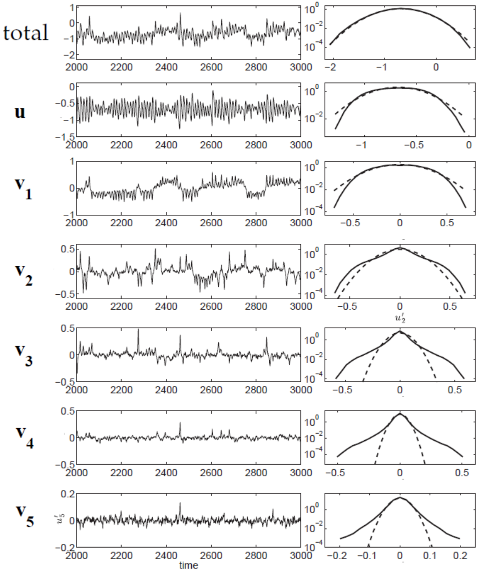


2) Conceptual models for turbulent dynamical systems (Majda & Lee, 2014 *PNSA*)

$$du = \left(-d_u u + \gamma \sum_{k=1}^{K} v_k^2 + F\right) dt,$$

$$dv_k = \left(-d_{v_k} v_k - \gamma u v_k\right) dt + \sigma_{v_k} dW_{v_k},$$

- The large-scale mean flow is usually chaotic but more predictable than the smaller-scale fluctuations.
- The overall single point PDF of the flow field is nearly Gaussian whereas the mean flow pdf is sub-Gaussian.
- The PDFs of the larger-scale fluctuating components of the turbulent field are nearly Gaussian, whereas the smaller-scale fluctuating components are intermittent and have fat-tailed PDFs.



3) A low-order model of Charney-DeVore flows (Olbers 2001)

$$
\begin{aligned}
dx_1 &= \Big(\gamma_1^* x_3 - C(x_1 - x_1^*)\Big)\,dt + \sigma_1 dW_1,\\
dx_4 &= \Big(\gamma_2^* x_6 - C(x_4 - x_4^*) + \epsilon(x_2 x_6 - x_3 x_5)\Big)\,dt + \sigma_4 dW_4,\\
dx_2 &= \Big(-(\alpha_1 x_1 - \beta_1)x_3 - C x_2 - \delta_1 x_4 x_6\Big)\,dt + \sigma_2 dW_2,\\
dx_3 &= \Big((\alpha_1 x_1 - \beta_1)x_2 - \gamma_1 x_1 - C x_3 + \delta_1 x_4 x_5\Big)\,dt + \sigma_3 dW_3,\\
dx_5 &= \Big(-(\alpha_2 x_1 - \beta_2)x_6 - C x_5 - \delta_2 x_4 x_3\Big)\,dt + \sigma_5 dW_5,\\
dx_6 &= \Big((\alpha_2 x_1 - \beta_2)x_5 - \gamma_2 x_4 - C x_6 + \delta_2 x_4 x_2\Big)\,dt + \sigma_6 dW_6.
\end{aligned}
$$

- Charney and DeVore (CDV) made an fundamental contribution for the regime switching behavior of the atmosphere.
- This 6-dimensional low-order model is obtained by a Galerkin projection and truncation of the barotropic vorticity equation on a β-plane channel
- x_1, x_4 represent the zonal flow, x_2, x_3 are the topographic Rossby waves and x_5, x_6 are the Rossby waves.

Derivations.

The barotropic vorticity equation is the following,

$$\frac{\partial}{\partial t}\nabla^2\psi = -J(\psi, \nabla^2\psi + f + \gamma h) - C\nabla^2(\psi - \psi^*). \tag{14}$$

- The domain of longitude and latitude (x, y) are given by $[0, 2\pi] \times [0, \pi b]$.
- The parameter $b = 2B/L$ determines the ratio between the dimensional zonal length L and the meridional width B of the channel.
- The stream function ψ is periodic in x. The meridional boundaries $y = 0$ and $y = \pi$ have the conditions $\partial\psi/\partial x = 0$. In addition, $\int_0^{2\pi}(\partial\psi/\partial y)dx = 0$.
- The Coriolis parameter f generates the beta effect in model.
- Orography enters with h, the orographic height, and is scaled with γ.
- J is the Jacobi operator $J(A, B) = (\partial A/\partial x)(\partial B/\partial y) - (\partial A/\partial y)(\partial B/\partial x)$.
- The damping coefficient C is the newtonian relaxation to the streamfunction profile ψ^*.

Next, the barotropic vorticity equation (14) is projected on a set of basis functions which are eigenfunctions of the Laplace operator ∇^2,

$$\phi_{0m}(y) = \sqrt{2}\cos(my/b), \qquad \phi_{nm}(x, y) = \sqrt{2}e^{inx}\sin(my/b),$$

The 6-dimensional model is obtained by truncating the expansion of the stream function and the topographic height after $|n| = 1$ and $m = 2$.

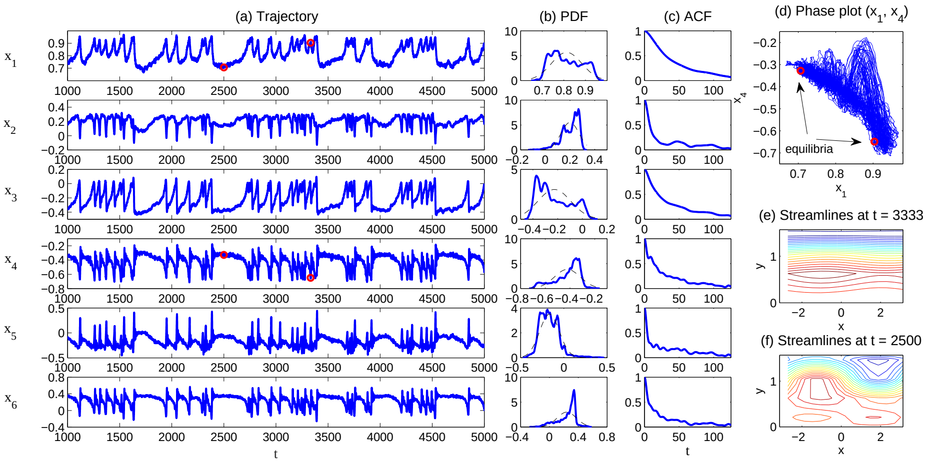
Then the time-dependent complex variables of the stream functions ψ_{01}, ψ_{02}, $\psi_{\pm 11}$, $\psi_{\pm 12}$ are transformed to real variables:

$$x_1 = \frac{1}{b}\psi_{01}, \qquad x_2 = \frac{1}{b\sqrt{2}}(\psi_{11} + \psi_{-11}), \qquad x_3 = \frac{i}{b\sqrt{2}}(\psi_{11} - \psi_{-11}),$$
$$x_4 = \frac{1}{b}\psi_{02}, \qquad x_5 = \frac{1}{b\sqrt{2}}(\psi_{12} + \psi_{-12}), \qquad x_6 = \frac{i}{b\sqrt{2}}(\psi_{12} - \psi_{-12}),$$

while the topography h is chosen to have only the $(1, 1)$ wave profile,

$$h(x, y) = \cos(x)\sin(y/b).$$

These manipulations lead to a 6-dimensional ODE model, where x_1, x_4 represent the zonal flow, x_2, x_3 are the topographic Rossby waves and x_5, x_6 are the Rossby waves.



(a) Trajectory
(b) PDF
(c) ACF
(d) Phase plot (x_1, x_4)
equilibria
(e) Streamlines at t = 3333
(f) Streamlines at t = 2500
t
x
y

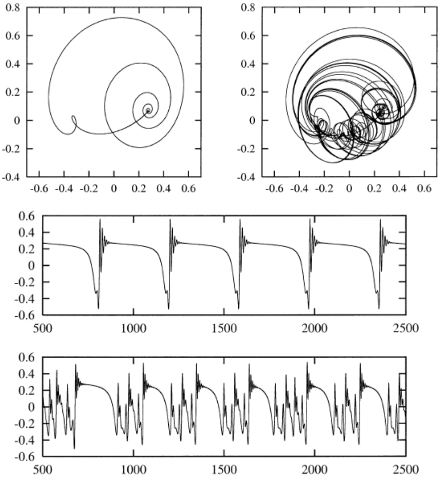


FIG. 2. Results from 2000-day integrations using four-EOF model and CDV model. (upper left) Four-EOF model, M_1 norm, projection onto EOF 1, 2 plane. (upper right) CDV model, same projection. (middle) Four-EOF model, EOF 1 vs time. (bottom) CDV model, EOF 1 vs time.

TABLE 1. EOF variance spectra, using L2 norm M_0 and kinetic energy norm M_1. Shown are the cumulative variances of the CDV model data.

No. of EOF	Cumulative variance, norm M_0	Cumulative variance, norm M_1
1	0.679 54	0.659 68
2	0.934 27	0.946 65
3	0.975 76	0.987 80
4	0.988 49	0.994 22
5	0.996 11	0.998 44
6	1.000 00	1.000 00

TABLE 2. Summary of dynamics of various EOF models.

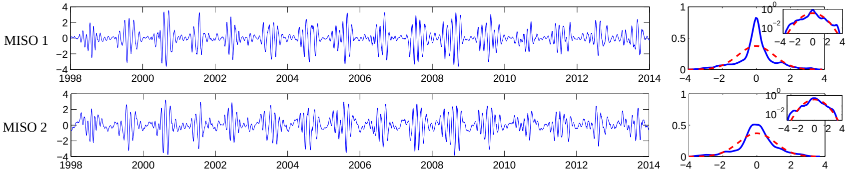
Model	Norm	Dynamics
Original CDV	—	Chaotic, regimes
Five EOFs	M_0	Fixed point
Five EOFs	M_1	Fixed point
Four EOFs	M_0	Fixed point
Four EOFs	M_1	Periodic, regimes
Three EOFs	M_0	Fixed point
Three EOFs	M_1	Periodic, no regimes

Projecting this 6-dimensional model to its leading 5 Empirical Orthogonal Functions (EOFs) explains 99.5% of the variance. However, such a 5-dimensional projected dynamics completely misses the dynamical features in the original model, where the multiple equilibria disappears and the 5-dimensional model cannot reproduce regime transitions.

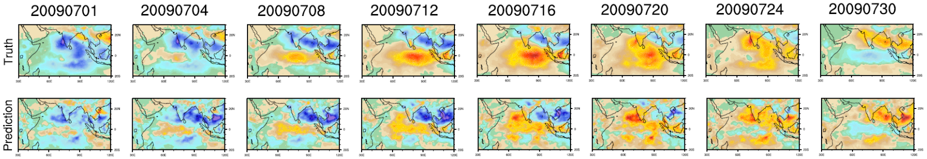
4) Nonlinear stochastic models for predicting intermittent MJO and monsoon indices

(Chen, Majda & Giannakis 2014 *GRL*, Chen, Majda, Sabeerali & Ajayamohan 2018 *J Climate*)

Physics-Constrained Low-Order Stochastic Models



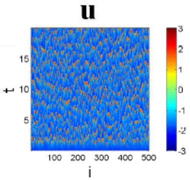
$$d u_1 = (-d_u(t)\, u_1 + \gamma\, v\, u_1 - \omega\, u_2)\, dt + \sigma_u\, dW_{u_1},$$
$$d u_2 = (-d_u(t)\, u_2 + \gamma\, v\, u_2 + \omega\, u_1)\, dt + \sigma_u\, dW_{u_2},$$
$$d v = (-d_v\, v - \gamma\, (u_1^2 + u_2^2))\, dt + \sigma_v\, dW_v,$$
$$d\omega = (-d_\omega \omega + \hat{\omega})\, dt + \sigma_\omega\, dW_\omega,$$



2. Stochastically coupled reaction-diffusion models in neuroscience and ecology

1) Stochastically coupled FitzHugh-Nagumo (FHN) models — a prototype of an excitable system (Lindner et al, 2004 *Physics Report*),

$$\epsilon du_i = \Big(d_u(u_{i+1} + u_{i-1} - 2u_i) + u_i - \frac{1}{3}u_i^3 + m(\bar{u} - u_i) - v_i\Big)dt + \sqrt{\epsilon}\delta_1 dW_{u_i},$$
$$dv_i = \big(u_i + a\big)dt + \delta_2 dW_{v_i}, \qquad i = 1, \dots, N.$$



2) A stochastically coupled SIR epidemic model (Gray et al, 2011 *SIAM JAM*)

$$\text{susceptible} \quad \longrightarrow \quad \text{infectious} \quad \longrightarrow \quad \text{recovered.}$$

$$dS = (\nabla^2 S - \beta SI - \mu_1 S + b)dt + \sigma(S)dW_S,$$
$$dI = (\nabla^2 I + \beta SI - \mu_2 I - \alpha I)dt,$$
$$dR = (\nabla^2 R + \alpha I - \mu_3 R)dt,$$

3) A stochastic version of the predator-prey system (Medvinsky et al, 2002 *SIAM Review*)

4) A nutrient-limited model for avascular cancer growth (Ferreira, Martins & Vilela 2002 *PRE*)

5) ...

3. Large-scale dynamical models in turbulence, fluids and geophysical flows

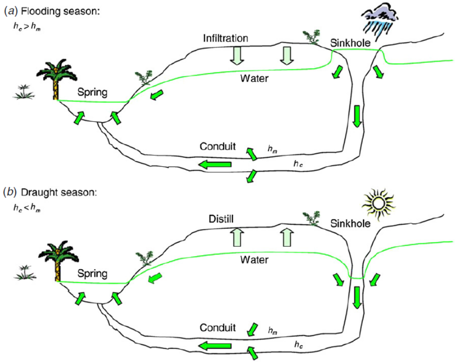
1) The Boussinesq equation — with applications in modeling the Rayleigh-Bénard convection and describing strongly stratified flows as in geophysics (Majda 2003),

$$\nabla \cdot \mathbf{u} = 0,$$

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} = -\frac{1}{\rho_0} \nabla p + \nu \nabla^2 \mathbf{u} - g\alpha T + \mathbf{F}_{\mathbf{u}},$$

$$\frac{\partial T}{\partial t} + \mathbf{u} \cdot \nabla T = \kappa \nabla^2 T + F_T.$$

2) Darcy-Brinkman-Oberbeck-Boussinesq system – convection phenomena in porous media (Kelliher et al, 2011 *Physica D*)



3) The rotating shallow water equations

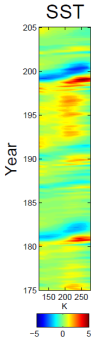
$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} + f\mathbf{u}^{\perp} + g\nabla h = \mathbf{F}_{\mathbf{u}},$$
$$\frac{\partial h}{\partial t} + \mathbf{u} \cdot \nabla h + (H + h)\nabla \cdot \mathbf{u} = F_h,$$

4) The MJO stochastic skeleton model (Thual, Majda & Stechmann, 2014 *JAS*)

5) A coupled El Niño model capturing observed El Niño diversity (Chen & Majda, 2017 *PNAS*)

6) ...

A coupled El Niño model capturing observed El Niño diversity (Chen & Majda, 2017 *PNAS*).



Atmosphere

$$-yv - \partial_x\theta = 0,$$
$$yu - \partial_y\theta = 0,$$
$$-(\partial_x u + \partial_y v) = E_q/(1-\overline{Q})$$

Ocean

$$\partial_\tau U - c_1 YV + c_1\partial_x H = c_1\tau_x,$$
$$YU + \partial_Y H = 0,$$
$$\partial_\tau H + c_1(\partial_x U + \partial_Y V) = 0$$

SST

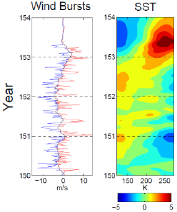
$$\partial_\tau T + \mu\partial_x(UT) = -c_1\zeta E_q + c_1\eta H,$$

Coupling:

$$E_q = \alpha_q T, \qquad \tau_x = \gamma(u + u_p).$$

The wind bursts and easterly mean trade wind are parameterized as

$$u_p = a_p(\tau)s_p(x)\phi_0(y),$$
$$\frac{da_p}{d\tau} = -d_p(a_p - \hat{a}_p) + \sigma_p(T_W)\dot{W}(\tau),$$



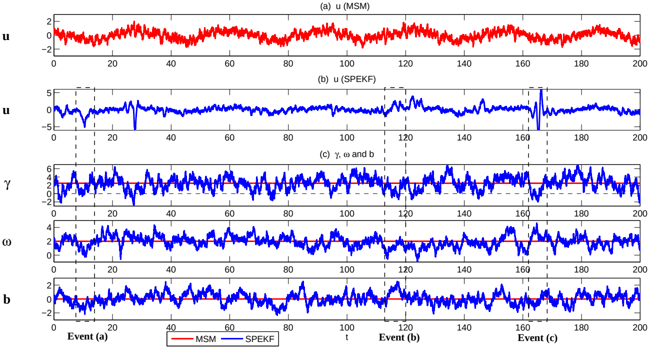
First simple dynamical model capturing

- the observed El Niño diversity,
- the non-Gaussian statistics in different regions across equatorial Pacific, and
- different extreme El Niño events.

4. Other low-order models for filtering and prediction

The stochastic parameterized extended Kalman filter (SPEKF) model — filter and predict the highly nonlinear and intermittent turbulent signals as observed in nature,

$$
\begin{aligned}
du &= \Big((-\gamma + i\omega)u + F(t) + b\Big)dt + \sigma_u dW_u,\\
d\gamma &= -d_\gamma(\gamma - \hat{\gamma})dt + \sigma_\gamma dW_\gamma\\
d\omega &= -d_\omega(\omega - \hat{\omega})dt + \sigma_\omega dW_\omega\\
db &= -d_b(b - \hat{b})dt + \sigma_b dW_b,
\end{aligned}
$$



A good paper for SPEKF's application: Branicki, Michal, Andrew J. Majda, and Kody JH Law. "Accuracy of Some Approximate Gaussian Filters for the Navier–Stokes Equation in the Presence of Model Error." Multiscale Modeling & Simulation 16.4 (2018): 1756-1794.

III. Parameter Estimation Using Data Assimilation

Recall the conditional Gaussian systems,

$$
\begin{aligned}
d\mathbf{u}_{\mathbf{I}} &= [\mathbf{A}_0(t, \mathbf{u}_{\mathbf{I}}) + \mathbf{A}_1(t, \mathbf{u}_{\mathbf{I}})\mathbf{u}_{\mathbf{II}}]dt + \boldsymbol{\Sigma}_{\mathbf{I}}(t, \mathbf{u}_{\mathbf{I}})d\mathbf{W}_{\mathbf{I}}(t) \\
d\mathbf{u}_{\mathbf{II}} &= [\mathbf{a}_0(t, \mathbf{u}_{\mathbf{I}}) + \mathbf{a}_1(t, \mathbf{u}_{\mathbf{I}})\mathbf{u}_{\mathbf{II}}]dt + \boldsymbol{\Sigma}_{\mathbf{II}}(t, \mathbf{u}_{\mathbf{I}})d\mathbf{W}_{\mathbf{II}}(t)
\end{aligned}
$$

The conditional Gaussian system also provides a framework for parameter estimation. In fact, $\mathbf{u}_{\mathbf{II}}$ can be written as

$$
\mathbf{u}_{\mathbf{II}} = (\widetilde{\mathbf{u}}_{\mathbf{II}}, \boldsymbol{\Lambda}),
$$

where $\mathbf{u}_{\mathbf{II}}$ in $\mathbb{R}^{\widetilde{N}_2}$ is physical process variables and $\boldsymbol{\Lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_p) \in \mathbb{R}^{N_{2,p}}$ denotes the model parameters. Here $N_2 = \widetilde{N}_2 + N_{2,p}$. Rewriting the conditional Gaussian system (1) in terms of $\mathbf{u}_{\mathbf{II}} = (\widetilde{\mathbf{u}}_{\mathbf{II}}, \boldsymbol{\Lambda})$ yields

$$
\begin{aligned}
d\mathbf{u}_{\mathbf{I}} &= [\mathbf{A}_0(t, \mathbf{u}_{\mathbf{I}}) + \mathbf{A}_1(t, \mathbf{u}_{\mathbf{I}})\widetilde{\mathbf{u}}_{\mathbf{II}} + \mathbf{A}_{1,\lambda}(t, \mathbf{u}_{\mathbf{I}})\boldsymbol{\Lambda}]dt + \boldsymbol{\Sigma}_{\mathbf{I}}(t, \mathbf{u}_{\mathbf{I}})d\mathbf{W}_{\mathbf{I}}(t), \\
d\mathbf{u}_{\mathbf{II}} &= [\mathbf{a}_0(t, \mathbf{u}_{\mathbf{I}}) + \mathbf{a}_1(t, \mathbf{u}_{\mathbf{I}})\widetilde{\mathbf{u}}_{\mathbf{II}} + \mathbf{a}_{1,\lambda}(t, \mathbf{u}_{\mathbf{I}})\boldsymbol{\Lambda}]dt + \boldsymbol{\Sigma}_{\mathbf{II}}(t, \mathbf{u}_{\mathbf{I}})d\mathbf{W}_{\mathbf{II}}(t).
\end{aligned}
$$

Consider the parameter estimation in the following simple setup:

$$d\mathbf{u_I} = [\mathbf{A}_0(t, \mathbf{u_I}) + \mathbf{A}_{1,\lambda}(t, \mathbf{u_I})\mathbf{\Lambda}^*]dt + \mathbf{\Sigma_I}(t, \mathbf{u_I})d\mathbf{W_I}(t).$$

Given the observed trajectory $\mathbf{u_I}$, our goal is to estimate the parameter $\mathbf{\Lambda}^*$.

Note: There are different parameter estimation methods.

1. Direct parameter estimation algorithm.

Since $\mathbf{\Lambda}$ are constant parameters, it is natural to augment the dynamics with the following relationship,

$$d\mathbf{u_I} = [\mathbf{A}_0(t, \mathbf{u_I}) + \mathbf{A}_{1,\lambda}(t, \mathbf{u_I})\mathbf{\Lambda}]dt + \mathbf{\Sigma_I}(t, \mathbf{u_I})d\mathbf{W_I}(t), \tag{15a}$$

$$d\mathbf{\Lambda} = 0, \tag{15b}$$

1. Direct parameter estimation algorithm.

Since $\mathbf{\Lambda}$ are constant parameters, it is natural to augment the dynamics with the following relationship,

$$d\mathbf{u_I} = [\mathbf{A}_0(t, \mathbf{u_I}) + \mathbf{A}_{1,\lambda}(t, \mathbf{u_I})\mathbf{\Lambda}]dt + \mathbf{\Sigma_I}(t, \mathbf{u_I})d\mathbf{W_I}(t), \tag{15a}$$

$$d\mathbf{\Lambda} = 0, \tag{15b}$$

where an initial uncertainty for the parameter $\mathbf{\Lambda}$ is assigned.

1. Direct parameter estimation algorithm.

Since $\mathbf{\Lambda}$ are constant parameters, it is natural to augment the dynamics with the following relationship,

$$d\mathbf{u_I} = [\mathbf{A}_0(t,\mathbf{u_I}) + \mathbf{A}_{1,\lambda}(t,\mathbf{u_I})\mathbf{\Lambda}]dt + \mathbf{\Sigma_I}(t,\mathbf{u_I})d\mathbf{W_I}(t), \tag{15a}$$

$$d\mathbf{\Lambda} = 0, \tag{15b}$$

where an initial uncertainty for the parameter $\mathbf{\Lambda}$ is assigned.

The time evolutions of the mean $\overline{\mathbf{u}}_{\mathbf{II}}$ and covariance $\mathbf{R_{II}}$ of the estimate of $\mathbf{\Lambda}$ are given by

$$d\overline{\mathbf{u}}_{\mathbf{II}}(t) =(\mathbf{R_{II}}\mathbf{A}_1^*(t,\mathbf{u_I}))(\mathbf{\Sigma_I}\mathbf{\Sigma_I}^*)^{-1}(t,\mathbf{u_I})\,[d\mathbf{u_I} - (\mathbf{A}_0(t,\mathbf{u_I}) + \mathbf{A}_1(t,\mathbf{u_I})\overline{\mathbf{u}}_{\mathbf{II}})dt]\,, \tag{16a}$$

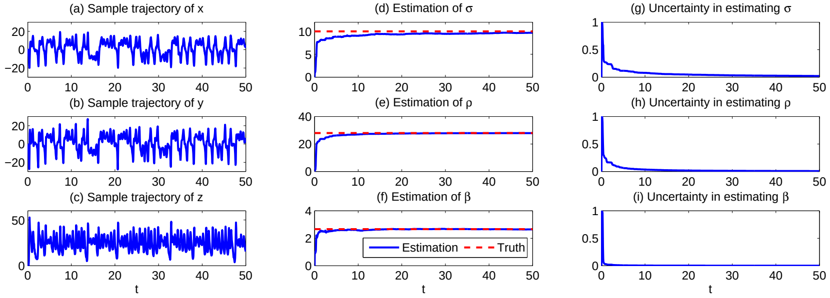
$$d\mathbf{R_{II}}(t) = -\,(\mathbf{R_{II}}\mathbf{A}_1^*(t,\mathbf{u_I}))(\mathbf{\Sigma_I}\mathbf{\Sigma_I}^*)^{-1}(t,\mathbf{u_I})(\mathbf{R_{II}}\mathbf{A}_1^*(t,\mathbf{u_I}))^*dt. \tag{16b}$$

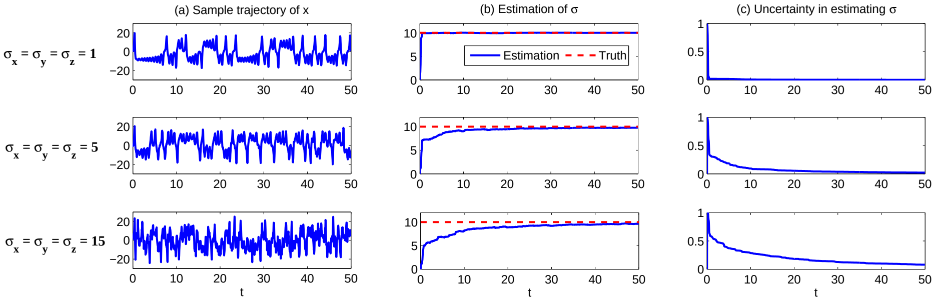
The formula in (16b) indicates that $\mathbf{R_{II}} = 0$ is a solution, plugging which into (16a) results in $\overline{\mathbf{u}}_{\mathbf{II}} = \mathbf{\Lambda}^*$. This means by knowing the perfect model the estimated parameters in (15)–(16) under certain conditions will converge to the truth.

As a simple test example, consider estimating the three parameters σ, ρ and γ in the noisy L-63 model with $\rho = 28, \sigma = 10, \beta = 8/3$.

$$
\begin{aligned}
dx &= \sigma(y - x)dt + \sigma_x dW_x, \\
dy &= \big(x(\rho - z) - y\big)dt + \sigma_y dW_y, \\
dz &= (xy - \beta z)dt + \sigma_z dW_z, \\
d\sigma &= 0, \\
d\rho &= 0, \\
d\beta &= 0,
\end{aligned}
$$

with $\sigma_x = \sigma_y = \sigma_z = 5$.




(a) Sample trajectory of x
(b) Estimation of σ
(c) Uncertainty in estimating σ
σx = σy = σz = 1
σx = σy = σz = 5
σx = σy = σz = 15
Estimation
Truth
t
t
t

2. Parameter estimation using stochastic parameterized equation.

A new approach of the augmented system can be formed in the following way:

$$d\mathbf{u_I} = [\mathbf{A}_0(t, \mathbf{u_I}) + \mathbf{A}_{1,\lambda}(t, \mathbf{u_I})\mathbf{\Lambda}]dt + \mathbf{\Sigma_I}(t, \mathbf{u_I})d\mathbf{W_I}(t), \tag{17a}$$

$$d\mathbf{\Lambda} = [\mathbf{c}_1\mathbf{\Lambda} + \mathbf{c}_2]dt + \boldsymbol{\sigma}_{\mathbf{\Lambda}}d\mathbf{W}_{\mathbf{\Lambda}}(t). \tag{17b}$$

Here, $\mathbf{c}_1$ is a negative-definite diagonal matrix, $\mathbf{c}_2$ is a constant vector and $\boldsymbol{\sigma}_{\mathbf{\Lambda}}$ is a diagonal noise matrix.

- The stochastic parameterized equations in (17b) serve as the prior information of the parameter estimation.

2. Parameter estimation using stochastic parameterized equation.

A new approach of the augmented system can be formed in the following way:

$$d\mathbf{u_I} = [\mathbf{A}_0(t, \mathbf{u_I}) + \mathbf{A}_{1,\lambda}(t, \mathbf{u_I})\mathbf{\Lambda}]dt + \mathbf{\Sigma_I}(t, \mathbf{u_I})d\mathbf{W_I}(t), \tag{17a}$$

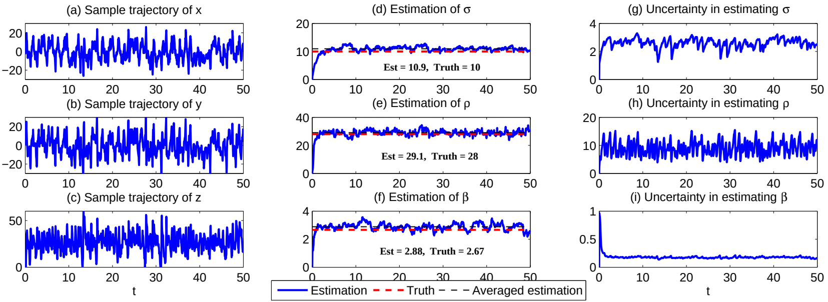
$$d\mathbf{\Lambda} = [\mathbf{c}_1\mathbf{\Lambda} + \mathbf{c}_2]dt + \boldsymbol{\sigma}_{\mathbf{\Lambda}}d\mathbf{W}_{\mathbf{\Lambda}}(t). \tag{17b}$$

Here, $\mathbf{c}_1$ is a negative-definite diagonal matrix, $\mathbf{c}_2$ is a constant vector and $\boldsymbol{\sigma}_{\mathbf{\Lambda}}$ is a diagonal noise matrix.

- The stochastic parameterized equations in (17b) serve as the prior information of the parameter estimation.
- Although certain model error will be introduced in the stochastic parameterized equations due to the appearance of $\mathbf{c}_1$, $\mathbf{c}_2$ and $\boldsymbol{\sigma}_{\mathbf{\Lambda}}$, it has shown that the convergence rate will be greatly accelerated.
- In fact, in linear models, rigorous analysis reveals that the convergence rate using stochastic parameterized equations (17) is exponential while that using the direct method (17) is only algebraic.

Now we apply the parameter estimation using stochastic parameterized equations (17) for the noisy L-63 model with a large noise $\sigma_x = \sigma_y = \sigma_z = 15$. The augmented system reads,

$$\begin{aligned}
dx &= \sigma(y - x)dt + \sigma_x dW_x,\\
dy &= \big(x(\rho - z) - y\big)dt + \sigma_y dW_y,\\
dz &= (xy - \beta z)dt + \sigma_z dW_z,\\
d\sigma &= -d_\sigma(\sigma - \hat{\sigma})dt + \sigma_\sigma dW_\sigma,\\
d\rho &= -d_\rho(\rho - \hat{\rho})dt + \sigma_\rho dW_\rho,\\
d\beta &= -d_\beta(\beta - \hat{\beta})dt + \sigma_\beta dW_\beta.
\end{aligned}$$



IV. Hybrid Data Assimilation Revisited

Recall the multiscale data assimilation framework in the previous lectures,

$$p^f(\mathbf{u}) = p^f(\overline{\mathbf{u}}, \mathbf{u}') \approx p^f(\overline{\mathbf{u}}) p^f_G(\mathbf{u}'|\overline{\mathbf{u}})$$

where

- $p^f(\overline{\mathbf{u}})$ is solved via particle filter
- $p^f_G(\mathbf{u}'|\overline{\mathbf{u}})$ is solved via ensemble Kalman filter

For conditional Gaussian system,

$$\begin{aligned} d\mathbf{u}_{\mathbf{I}} &= [\mathbf{A}_0(t,\mathbf{u}_{\mathbf{I}}) + \mathbf{A}_1(t,\mathbf{u}_{\mathbf{I}})\mathbf{u}_{\mathbf{II}}]dt + \mathbf{\Sigma}_{\mathbf{I}}(t,\mathbf{u}_{\mathbf{I}})d\mathbf{W}_{\mathbf{I}}(t) \\ d\mathbf{u}_{\mathbf{II}} &= [\mathbf{a}_0(t,\mathbf{u}_{\mathbf{I}}) + \mathbf{a}_1(t,\mathbf{u}_{\mathbf{I}})\mathbf{u}_{\mathbf{II}}]dt + \mathbf{\Sigma}_{\mathbf{II}}(t,\mathbf{u}_{\mathbf{I}})d\mathbf{W}_{\mathbf{II}}(t) \end{aligned}$$

The forecast joint PDF $p^f(\mathbf{u}) = p^f(\mathbf{u}_{\mathbf{I}}, \mathbf{u}_{\mathbf{II}}) = p^f(\mathbf{u}_{\mathbf{I}})p^f(\mathbf{u}_{\mathbf{II}}|\mathbf{u}_{\mathbf{I}})$

- No approximation is here.
- $p^f(\mathbf{u}_{\mathbf{I}})$ is solved via particle filter
- $p^f(\mathbf{u}_{\mathbf{II}}|\mathbf{u}_{\mathbf{I}})$ is solved via closed analytic formulae of the conditional Gaussian framework.

- The conditional Gaussian nonlinear models can be used as approximate models for many natural phenomena.
- The framework has quite a few salient features, allowing rigorous mathematical analysis and efficient numerical algorithms.
- A few selected topics of the conditional Gaussian nonlinear models will be presented in the following lectures.