Efficient Nonlinear Optimal Smoothing and Sampling Algorithms for Complex Turbulent Nonlinear Dynamical Systems with Partial Observations

Instructor: Nan Chen

Department of Mathematics
University of Wisconsin-Madison

Graduate Course, MATH 801, Spring 2020
MWF 8:50AM – 9:40AM, B239 Van Vleck Hall
Main reference

Filtering and Smoothing

Both filtering and smoothing are techniques for state estimation.

Assume \( \{ u_0, u_1, \ldots, u_T \} \) is a time series of the state variable and \( \{ v_0, v_1, \ldots, v_T \} \) are the corresponding noisy observations. The model also contains uncertainty.

**Filtering**

Model + Obs \((v_0, \ldots, v_t)\)

\[ p(u_t|v_s, 0 \leq s \leq t) \]

**Smoothing**

Model + Obs \((v_0, \ldots, v_t, \ldots, v_T)\)

\[ p(u_t|v_s, 0 \leq s \leq T) \]
Comparison of Filtering and Smoothing

Filtering
Model + Obs \((v_0, ..., v_t)\)
\(p(u_t|v_s, 0 \leq s \leq t)\)

Smoothing
Model + Obs \((v_0, ..., v_t, ..., v_T)\)
\(p(u_t|v_s, 0 \leq s \leq T)\)

<table>
<thead>
<tr>
<th>Filtering</th>
<th>Smoothing</th>
</tr>
</thead>
<tbody>
<tr>
<td>using the information only in the past</td>
<td>using the entire available information</td>
</tr>
<tr>
<td>cheaper but less accurate</td>
<td>more accurate but more expensive</td>
</tr>
<tr>
<td>online</td>
<td>offline</td>
</tr>
<tr>
<td>the precursor of real-time prediction</td>
<td>applying for post-processing of data</td>
</tr>
</tbody>
</table>
Linear model and Gaussian noise: Kalman filter and Kalman smoother (or RTS smoother).

In general nonlinear and/or non-Gaussian situations, particle or ensemble methods are required to approximate the state estimates.

- Ensemble Kalman filter: only involving Gaussian statistics
- Particle filter: non-Gaussian statistics in low-dimensional system

Smoothing involves a backward pass and the particle methods are expensive and hard to implement for nonlinear systems.

![Diagram of Filtering and Smoothing](image)
Motivation of Developing an Efficient Nonlinear Optimal Smoother

- Improve the state estimation, especially for non-Gaussian and extreme events.
- The smoother estimates can be used to extract the temporal information and thus to sample hidden model trajectories. But filtering fails to do so in an unbiased way.
In light of the conditional Gaussian model structure, the nonlinear optimal smoother and the optimal sampling strategy can be solved via closed analytic formulae and therefore they are computationally efficient and the methods are applicable to high-dimensional nonlinear turbulent dynamical systems.

1. The nonlinear optimal smoother estimates provide a more accurate description of the non-Gaussian features, including the timing, duration and amplitudes of the hidden extreme events.
In light of the conditional Gaussian model structure, the nonlinear optimal smoother and the optimal sampling strategy can be solved via closed analytic formulae and therefore they are computationally efficient and the methods are applicable to high-dimensional nonlinear turbulent dynamical systems.

1. The nonlinear optimal smoother estimates provide a more accurate description of the non-Gaussian features, including the timing, duration and amplitudes of the hidden extreme events.

2. The optimal sampling technique is extremely useful in obtaining both the statistical and path-wise properties of the unobserved or unresolved variables.
In light of the conditional Gaussian model structure, the nonlinear optimal smoother and the optimal sampling strategy can be solved via *closed analytic formulae* and therefore they are *computationally efficient* and the methods are *applicable to high-dimensional nonlinear turbulent dynamical systems*.

1. The nonlinear optimal smoother estimates provide a *more accurate description of the non-Gaussian features*, including the timing, duration and amplitudes of the *hidden extreme events*.

2. The optimal sampling technique is extremely useful in obtaining *both the statistical and path-wise properties of the unobserved or unresolved variables*.

3. It provides extra dynamical information of the hidden processes which *cannot be fully described by the filter and smoother estimates*, such as the temporal correlation of the underlying systems, e.g. the ACF.
4. The dynamical and statistical information provided by the resulting sampled trajectories can be adopted to **systematically improve the approximate models and the stochastic parameterizations of unresolved variables**, which advance the reduction of the error and uncertainty in real-time forecasts.
4. The dynamical and statistical information provided by the resulting sampled trajectories can be adopted to **systematically improve the approximate models and the stochastic parameterizations of unresolved variables**, which advance the reduction of the error and uncertainty in real-time forecasts.

5. This optimal sampling strategy can then be used to **create a sufficient number of data** in an unbiased fashion for both the observed and hidden variables, which facilitates the statistical description of various significant non-Gaussian features.
4. The dynamical and statistical information provided by the resulting sampled trajectories can be adopted to **systematically improve the approximate models and the stochastic parameterizations of unresolved variables**, which advance the reduction of the error and uncertainty in real-time forecasts.

5. This optimal sampling strategy can then be used to **create a sufficient number of data** in an unbiased fashion for both the observed and hidden variables, which facilitates the statistical description of various significant non-Gaussian features.

6. The nonlinear smoother plays an important role in **parameter estimation of complex turbulent systems with only partial observations** based on an expectation-maximization algorithm.
The Conditional Gaussian Nonlinear System

\[
\begin{align*}
\text{d}X(t) &= \left[ A_0(X, t) + A_1(X, t)Y(t) \right] \text{d}t + B_1(X, t) \text{d}W_1(t) + B_2(X, t) \text{d}W_2(t), \quad (1a) \\
\text{d}Y(t) &= \left[ a_0(X, t) + a_1(X, t)Y(t) \right] \text{d}t + b_1(X, t) \text{d}W_1(t) + b_2(X, t) \text{d}W_2(t). \quad (1b)
\end{align*}
\]

**Theorem (Optimal Filter)**

*Given one realization of the time series \( X(s) \) for \( s \in [0, t] \), the conditional distribution

\[
p(Y(t)|X(s), s \leq t) \sim \mathcal{N}(\mu_f(t), R_f(t))
\]

is Gaussian, where the conditional mean \( \mu_f \) and the conditional covariance \( R_f \) are given by the following explicit formulae

\[
\begin{align*}
\text{d}\mu_f &= (a_0 + a_1 \mu_f) \text{d}t + (b \circ B + R_f A_1^\ast)(B \circ B^\ast)^{-1}(\text{d}X - (A_0 + A_1 \mu_f) \text{d}t), \quad (3a) \\
\text{d}R_f &= (a_1 R_f + R_f a_1^\ast + b \circ b - (b \circ B + R_f A_1^\ast)(B \circ B)^{-1}(b \circ B + A_1 R_f)) \text{d}t, \quad (3b)
\end{align*}
\]

with

\[
\begin{align*}
b \circ b &= b_1 b_1^\ast + b_2 b_2^\ast, \\
b \circ B &= b_1 B_1^\ast + b_2 B_2^\ast, \\
B \circ B &= B_1 B_1^\ast + B_2 B_2^\ast.
\end{align*}
\]
\[ d\mathbf{X}(t) = \left[ A_0(\mathbf{X}, t) + A_1(\mathbf{X}, t)\mathbf{Y}(t) \right] dt + B_1(\mathbf{X}, t) dW_1(t) + B_2(\mathbf{X}, t) dW_2(t), \]

\[ d\mathbf{Y}(t) = \left[ a_0(\mathbf{X}, t) + a_1(\mathbf{X}, t)\mathbf{Y}(t) \right] dt + b_1(\mathbf{X}, t) dW_1(t) + b_2(\mathbf{X}, t) dW_2(t). \]

**Theorem (Optimal Nonlinear Smoother)**

Given one realization of the observed variable \( \mathbf{X}(s) \) for \( s \in [0, T] \), the optimal smoother estimate \( p(\mathbf{Y}(t)|\mathbf{X}(s), 0 \leq s \leq T) \) is conditional Gaussian,

\[ p(\mathbf{Y}(t)|\mathbf{X}(s), 0 \leq s \leq T) \sim \mathcal{N}(\mu_s(t), R_s(t)), \tag{4} \]

where the conditional mean \( \mu_s(t) \) and conditional covariance \( R_s(t) \) of the smoother at time \( s \) satisfy the following equations,

\[ d\hat{\mu}_s = \left( -a_0 - a_1\mu_s + (b \circ b)R_f^{-1}(\mu_f - \mu_s) \right) dt, \tag{5a} \]

\[ d\hat{R}_s = -((a_1 + (b \circ b)R_f^{-1})R_s + R_s(a_1^* + (b \circ b)R_f^{-1}) - b \circ b) dt. \tag{5b} \]

In (5), the terms of the left hand side are understood as

\[ d\hat{\mu}_s = \lim_{\Delta t \to 0} \mu_s(t) - \mu_s(t + \Delta t) \]

\[ d\hat{R}_s = \lim_{\Delta t \to 0} R_s(t) - R_s(t + \Delta t). \]

The starting value of the nonlinear smoother \( (\mu_s(T), R_s(T)) \) is the same as the filter estimate at the endpoint \( (\mu_f(T), R_f(T)) \).
\[ \begin{align*}
\frac{dX(t)}{dt} &= \left[ A_0(X, t) + A_1(X, t)Y(t) \right] dt + B_1(X, t) dW_1(t) + B_2(X, t) dW_2(t), \\
\frac{dY(t)}{dt} &= \left[ a_0(X, t) + a_1(X, t)Y(t) \right] dt + b_1(X, t) dW_1(t) + b_2(X, t) dW_2(t).
\end{align*} \]

**Theorem (Optimal Backward Sampling Formula)**

Conditioned on one realization of the observed variable \( X(s) \) for \( s \in [0, T] \), the optimal strategy of sampling the trajectories associated with the unobserved variable \( Y \) satisfies the following explicit formula,

\[ \frac{d\hat{Y}}{dt} = \left( -a_0 - a_1 Y \right) dt + \left( b \circ b \right) R_f^{-1} (\mu_f - Y) dt + b_1 dW_{Y,1} + b_2 dW_{Y,2}, \]

where \( \mu_f(t) \) and \( R_f(t) \) are the conditional mean and conditional covariance from the filter estimates in (3), and \( W_{Y,1} \) and \( W_{Y,2} \) are independent white noise sources. In (6), the left hand side is understood as

\[ \frac{d\hat{Y}}{dt} = \lim_{\Delta t \to 0} Y(t) - Y(t + \Delta t). \]

The formula (6) starts from \( t = T \) and it is run backwards towards \( t = 0 \). Therefore, it is named as a backward sampling formula. The initial value of \( Y \) in (6) is drawn from the conditional Gaussian distribution \( \mathcal{N}(\mu_f(T), R_f(T)) \).
\[ \text{d} \mathbf{X}(t) = \left[ A_0(\mathbf{X}, t) + A_1(\mathbf{X}, t)\mathbf{Y}(t) \right] \text{d}t + B_1(\mathbf{X}, t) \text{d}W_1(t) + B_2(\mathbf{X}, t) \text{d}W_2(t), \]
\[ \text{d} \mathbf{Y}(t) = \left[ a_0(\mathbf{X}, t) + a_1(\mathbf{X}, t)\mathbf{Y}(t) \right] \text{d}t + b_1(\mathbf{X}, t) \text{d}W_1(t) + b_2(\mathbf{X}, t) \text{d}W_2(t). \]

\[ \text{d} \hat{\mathbf{Y}} = ( -a_0 - a_1 \mathbf{Y}) \text{d}t + (b \circ b)R_f^{-1}(\mu_f - \mathbf{Y}) \text{d}t + b_1 \text{d}W_{Y,1} + b_2 \text{d}W_{Y,2}, \]

- Comparing with the true underlying dynamics of \( \mathbf{Y} \), the backward sampling equation involves an extra term \((b \circ b)R_f^{-1}(\mu_f - \mathbf{Y}) \text{d}t\).
\[
\begin{align*}
\mathrm{d}X(t) &= \left[ A_0(X, t) + A_1(X, t)Y(t) \right] \mathrm{d}t + B_1(X, t) \mathrm{d}W_1(t) + B_2(X, t) \mathrm{d}W_2(t), \\
\mathrm{d}Y(t) &= \left[ a_0(X, t) + a_1(X, t)Y(t) \right] \mathrm{d}t + b_1(X, t) \mathrm{d}W_1(t) + b_2(X, t) \mathrm{d}W_2(t).
\end{align*}
\]

\[
\mathrm{d}Y = (-a_0 - a_1 Y) \mathrm{d}t + (b \circ b) R_f^{-1}(\mu_f - Y) \mathrm{d}t + b_1 \mathrm{d}W_{Y,1} + b_2 \mathrm{d}W_{Y,2},
\]

- Comparing with the true underlying dynamics of \( Y \), the backward sampling equation involves an extra term \((b \circ b) R_f^{-1}(\mu_f - Y) \mathrm{d}t\).
- This correction term plays an important role as a forcing and it drives the sampled trajectory to meander around the filter mean state \( \mu_f \).
\[
\begin{align*}
\text{d}X(t) &= \left[ A_0(X, t) + A_1(X, t)Y(t) \right] \text{d}t + B_1(X, t) \text{d}W_1(t) + B_2(X, t) \text{d}W_2(t), \\
\text{d}Y(t) &= \left[ a_0(X, t) + a_1(X, t)Y(t) \right] \text{d}t + b_1(X, t) \text{d}W_1(t) + b_2(X, t) \text{d}W_2(t).
\end{align*}
\]

\[
\text{d}\tilde{Y} = ( -a_0 - a_1 Y) \text{d}t + (b \circ b) R_f^{-1}(\mu_f - Y) \text{d}t + b_1 \text{d}W_{Y,1} + b_2 \text{d}W_{Y,2},
\]

- Comparing with the true underlying dynamics of \( Y \), the backward sampling equation involves an extra term \((b \circ b) R_f^{-1}(\mu_f - Y) \text{d}t\).
- This correction term plays an important role as a forcing and it drives the sampled trajectory to meander around the filter mean state \( \mu_f \).
- Yet, due to the memory of the process, the system response of the forcing has a delayed effect. The sampled trajectory \( Y \) actually fluctuates around the smoother mean state \( \mu_s \), which is a desirable feature since the optimal smoother estimate makes use of the entire observational information and is thus unbiased.
\[\begin{align*}
\frac{d\mathbf{X}(t)}{dt} &= \left[ A_0(\mathbf{X}, t) + A_1(\mathbf{X}, t)\mathbf{Y}(t) \right] dt + B_1(\mathbf{X}, t) dW_1(t) + B_2(\mathbf{X}, t) dW_2(t), \\
\frac{d\mathbf{Y}(t)}{dt} &= \left[ a_0(\mathbf{X}, t) + a_1(\mathbf{X}, t)\mathbf{Y}(t) \right] dt + b_1(\mathbf{X}, t) dW_1(t) + b_2(\mathbf{X}, t) dW_2(t).
\end{align*}\]

\[\frac{d\widehat{\mathbf{Y}}}{dt} = (-a_0 - a_1\mathbf{Y}) dt + (b \circ b) R_f^{-1} (\mu_f - \mathbf{Y}) dt + b_1 dW_{Y,1} + b_2 dW_{Y,2},\]

- Comparing with the true underlying dynamics of \(\mathbf{Y}\), the backward sampling equation involves an extra term \((b \circ b) R_f^{-1} (\mu_f - \mathbf{Y}) dt\).
- This correction term plays an important role as a forcing and it drives the sampled trajectory to meander around the filter mean state \(\mu_f\).
- Yet, due to the memory of the process, the system response of the forcing has a delayed effect. The sampled trajectory \(\mathbf{Y}\) actually fluctuates around the smoother mean state \(\mu_s\), which is a desirable feature since the optimal smoother estimate makes use of the entire observational information and is thus unbiased.
- One important feature of the backward sampling equation: it retains the dynamical structures of the true underlying dynamics of \(\mathbf{Y}\). Therefore the temporal autocorrelation function (ACF) and higher order temporal correlations associated with the underlying nonlinear systems can be accurately recovered using the sampled trajectories.
\[
\begin{align*}
\dot{X}(t) &= \left[ A_0(X, t) + A_1(X, t)Y(t) \right]dt + B_1(X, t)\,dW_1(t) + B_2(X, t)\,dW_2(t), \\
\dot{Y}(t) &= \left[ a_0(X, t) + a_1(X, t)Y(t) \right]dt + b_1(X, t)\,dW_1(t) + b_2(X, t)\,dW_2(t).
\end{align*}
\]

\[
\dot{\hat{Y}} = (-a_0 - a_1Y)\,dt + (b \circ b)R_f^{-1}(\mu_f - Y)\,dt + b_1\,dW_{Y,1} + b_2\,dW_{Y,2},
\]

- If the noise strength in the \( Y \) process is fixed, namely \( b \circ b \) is a constant matrix, then the amplitude of the filter covariance estimate \( R_f \) is positively correlated with the noise level of the observational process \( X \).
\[ d\mathbf{X}(t) = \left[ A_0(\mathbf{X}, t) + A_1(\mathbf{X}, t)\mathbf{Y}(t) \right] dt + B_1(\mathbf{X}, t) dW_1(t) + B_2(\mathbf{X}, t) dW_2(t), \]
\[ d\mathbf{Y}(t) = \left[ a_0(\mathbf{X}, t) + a_1(\mathbf{X}, t)\mathbf{Y}(t) \right] dt + b_1(\mathbf{X}, t) dW_1(t) + b_2(\mathbf{X}, t) dW_2(t). \]

\[ d\mathbf{\hat{Y}} = (-a_0 - a_1\mathbf{Y}) dt + (b \circ b) R_f^{-1} (\mu_f - \mathbf{Y}) dt + b_1 dW_{\mathbf{Y},1} + b_2 dW_{\mathbf{Y},2}, \]

- If the noise strength in the \( \mathbf{Y} \) process is fixed, namely \( b \circ b \) is a constant matrix, then the amplitude of the filter covariance estimate \( R_f \) is positively correlated with the noise level of the observational process \( \mathbf{X} \).

- A low noise level in \( \mathbf{X} \) implies a small uncertainty in \( R_f \), which leads to a large weight towards the correction term.
\[
\begin{align*}
\frac{d\mathbf{X}(t)}{dt} &= \left[A_0(\mathbf{X}, t) + A_1(\mathbf{X}, t) Y(t)\right] \, dt + B_1(\mathbf{X}, t) \, dW_1(t) + B_2(\mathbf{X}, t) \, dW_2(t), \\
\frac{d\mathbf{Y}(t)}{dt} &= \left[a_0(\mathbf{X}, t) + a_1(\mathbf{X}, t) Y(t)\right] \, dt + b_1(\mathbf{X}, t) \, dW_1(t) + b_2(\mathbf{X}, t) \, dW_2(t). 
\end{align*}
\]

\[
\frac{d\mathbf{\hat{Y}}}{dt} = \left(-a_0 - a_1 \mathbf{Y}\right) \, dt + (b \circ b) R_f^{-1} (\mu_f - \mathbf{Y}) \, dt + b_1 \, dW_{\mathbf{Y},1} + b_2 \, dW_{\mathbf{Y},2}.
\]

- If the noise strength in the \(\mathbf{Y}\) process is fixed, namely \(b \circ b\) is a constant matrix, then the amplitude of the filter covariance estimate \(R_f\) is positively correlated with the noise level of the observational process \(\mathbf{X}\).

- A low noise level in \(\mathbf{X}\) implies a small uncertainty in \(R_f\), which leads to a large weight towards the correction term.

- In addition, the filter mean estimate \(\mu_f\) in such a situation is largely determined by the observations.
\[
\begin{align*}
\mathrm{d}X(t) &= \left[A_0(X, t) + A_1(X, t)Y(t)\right] \mathrm{d}t + B_1(X, t) \mathrm{d}W_1(t) + B_2(X, t) \mathrm{d}W_2(t), \\
\mathrm{d}Y(t) &= \left[a_0(X, t) + a_1(X, t)Y(t)\right] \mathrm{d}t + b_1(X, t) \mathrm{d}W_1(t) + b_2(X, t) \mathrm{d}W_2(t).
\end{align*}
\]

\[
\mathrm{d} \hat{Y} = \left(-a_0 - a_1Y\right) \mathrm{d}t + (b \circ b)R^{-1}_f (\mu_f - Y) \mathrm{d}t + b_1 \mathrm{d}W_{Y,1} + b_2 \mathrm{d}W_{Y,2}.
\]

- If the noise strength in the \(Y\) process is fixed, namely \(b \circ b\) is a constant matrix, then the amplitude of the filter covariance estimate \(R_f\) is positively correlated with the noise level of the observational process \(X\).

- A low noise level in \(X\) implies a small uncertainty in \(R_f\), which leads to a large weight towards the correction term.

- In addition, the filter mean estimate \(\mu_f\) in such a situation is largely determined by the observations.

- As a consequence, the observations play a primary role in creating the sampled trajectories.
Theorem (Transient PDF)
Assume there are $L$ independent trajectories of the observed variable $X(t)$ from $t = 0$ to $t = T$, denoted by $X_l(t)$ with $l = 1, \ldots, L$. The PDF of $Y$ at time instant $t$ is given by

$$p(Y(t)) = \lim_{L \to \infty} \frac{1}{L} \sum_{l=1}^{L} p(Y(t) | X_l(0 \leq s \leq T)),$$

where for each $l$, $p(Y(t) | X_l(0 \leq s \leq T)) \sim \mathcal{N}(\mu_{l,s}, R_{l,s})$, with $\mu_{l,s}$ and $R_{l,s}$ being the mean and covariance computed from the nonlinear smoother.
Theorem (Transient PDF)
Assume there are \( L \) independent trajectories of the observed variable \( X(t) \) from \( t = 0 \) to \( t = T \), denoted by \( X_l(t) \) with \( l = 1, \ldots, L \). The PDF of \( Y \) at time instant \( t \) is given by

\[
p(Y(t)) = \lim_{L \to \infty} \frac{1}{L} \sum_{l=1}^{L} p(Y(t)|X_l(0 \leq s \leq T)),
\]

where for each \( l \), \( p(Y(t)|X_l(0 \leq s \leq T)) \sim \mathcal{N}(\mu_{l,s}, R_{l,s}) \), with \( \mu_{l,s} \) and \( R_{l,s} \) being the mean and covariance computed from the nonlinear smoother.

Corollary (Equilibrium PDF)
Assume a long trajectory of \( X \) from \( t = 0 \) to \( t = T \) is available. The equilibrium PDF of \( Y \), denoted by \( p(Y_\infty) \) is given by

\[
p(Y_\infty) = \lim_{I \to \infty} \frac{1}{I} \sum_{i=1}^{I} p(Y(t_i)|X(0 \leq s \leq T)),
\]

where all \( t_i \) with \( 0 \leq t_1 \leq \ldots \leq t_I \leq T \) are distributed between \( t = 0 \) and \( t = T \) with equal distance.
Theorem (Transient PDF)
Assume there are $L$ independent trajectories of the observed variable $X(t)$ from $t = 0$ to $t = T$, denoted by $X_l(t)$ with $l = 1, \ldots, L$. The PDF of $Y$ at time instant $t$ is given by

$$ p(Y(t)) = \lim_{L \to \infty} \frac{1}{L} \sum_{l=1}^{L} p(Y(t) | X_l(0 \leq s \leq T)),$$

(7)

where for each $l$, $p(Y(t) | X_l(0 \leq s \leq T)) \sim N(\mu_{l,s}, R_{l,s})$, with $\mu_{l,s}$ and $R_{l,s}$ being the mean and covariance computed from the nonlinear smoother.

Corollary (Equilibrium PDF)
Assume a long trajectory of $X$ from $t = 0$ to $t = T$ is available. The equilibrium PDF of $Y$, denoted by $p(Y_\infty)$ is given by

$$ p(Y_\infty) = \lim_{I \to \infty} \frac{1}{I} \sum_{i=1}^{I} p(Y(t_i) | X(0 \leq s \leq T)),$$

(8)

where all $t_i$ with $0 \leq t_1 \leq \ldots \leq t_I \leq T$ are distributed between $t = 0$ and $t = T$ with equal distance.

Corollary (Equilibrium PDF; An Alternative Method)
Under the same condition as Corollary (5), an alternative way of solving the equilibrium PDF of $Y$ is by collecting all the points in the trajectory calculated from the backward sampling equation (6).
Recovering the path-wise and statistical information of hidden variables

A perfect model test.

\[
du = \left(-du + cv\right)u + F_u \; dt + \sigma_u \; dW_u, \\
\dv = \left(-dv v - cu^2\right) dt + \sigma_v \; dW_v, \\
\]

\[ (9) \]
A nonlinear model test in the presence of model error.

Perfect model:

\[ d\gamma = \left( a\gamma + b\gamma^2 + c\gamma^3 + f\right) dt + (A\gamma + B\gamma) dW_\gamma,1 + \sigma_\gamma dW_\gamma,2. \]  \hspace{1cm} (10b)

The approximate model:

\[ d\gamma = -d_\gamma (\gamma - \hat{\gamma}) dt + \sigma_\gamma dW_\gamma. \]  \hspace{1cm} (11b)
Detecting Hidden Extreme Events Using Filtering and Smoothing

A nonlinear dyad model.

\[ du = ((-du + cv)u + Fu) \, dt + \sigma_u \, dW_u, \]
\[ dv = (-dv \, v - cu^2) \, dt + \sigma_v \, dW_v, \]  

\[(12)\]
A four-dimensional stochastic climate model with multiscale features.

\[
\begin{align*}
\, d\,x_1 &= \left(-x_2(L_{12} + a_1 x_1 + a_2 x_2) + d_1 x_1 + F_1 + L_{13} y_1 + b_{123} x_2 y_1\right) \, dt + \sigma_{x_1} \, dW_{x_1}, \\
\, d\,x_2 &= \left(+x_1(L_{12} + a_1 x_1 + a_2 x_2) + d_2 x_2 + F_2 + L_{24} y_2 + b_{213} x_1 y_1\right) \, dt + \sigma_{x_2} \, dW_{x_2}, \\
\, d\,y_1 &= \left(-L_{13} x_1 + b_{312} x_1 x_2 + F_3 - \frac{\gamma_1}{\epsilon} y_1\right) \, dt + \frac{\sigma_{y_1}}{\sqrt{\epsilon}} \, dW_{y_1}, \\
\, d\,y_2 &= \left(-L_{24} x_2 + F_4 - \frac{\gamma_2}{\epsilon} y_2\right) \, dt + \frac{\sigma_{y_2}}{\sqrt{\epsilon}} \, dW_{y_2},
\end{align*}
\]

where \( b_{123} + b_{213} + b_{312} = 0 \) and \( \epsilon = 1 \) here.
A four-dimensional stochastic climate model with multiscale features.

\[
\begin{align*}
\text{d} x_1 &= \left(-x_2(L_{12} + a_1 x_1 + a_2 x_2) + d_1 x_1 + F_1 + L_{13} y_1 + b_{123} x_2 y_1\right) \text{d} t + \sigma_{x_1} \text{d} W_{x_1}, \\
\text{d} x_2 &= \left(+x_1(L_{12} + a_1 x_1 + a_2 x_2) + d_2 x_2 + F_2 + L_{24} y_2 + b_{213} x_1 y_1\right) \text{d} t + \sigma_{x_2} \text{d} W_{x_2}, \\
\text{d} y_1 &= \left(-L_{13} x_1 + b_{312} x_1 x_2 + F_3 - \frac{\gamma_1}{\epsilon} y_1\right) \text{d} t + \frac{\sigma_{y_1}}{\sqrt{\epsilon}} \text{d} W_{y_1}, \\
\text{d} y_2 &= \left(-L_{24} x_2 + F_4 - \frac{\gamma_2}{\epsilon} y_2\right) \text{d} t + \frac{\sigma_{y_2}}{\sqrt{\epsilon}} \text{d} W_{y_2},
\end{align*}
\]

where \(b_{123} + b_{213} + b_{312} = 0\) and \(\epsilon = 0.1\) here.
Recovering non-Gaussian statistics of the observed variables using only short training period

A perfect model test.

\[ d\mathbf{u} = ((-d_u + cv)\mathbf{u} + F_u)\,dt + \sigma_u\,dW_u, \]
\[ d\mathbf{v} = (-d_v\mathbf{v} - cu^2)\,dt + \sigma_v\,dW_v, \]

(a) True signal of \( u \)

(b) Truth and recovered \( v \) from smoothing

(c) Truth and recovered \( v \) from filtering

(d) Comparison of the recovered PDFs of \( u \)

(e) PDFs of \( u \) in log scale

(f) Trajectories of \( u \)
A nonlinear model test in the presence of model error.

Perfect model:

\[
    \begin{align*}
        du &= \left( (-du + cv)u + F_u \right) dt + \sigma_u \, dW_u, \\
        dv &= (-dv v - cu^2) \, dt + \sigma_v \, dW_v,
    \end{align*}
\]

(14)

Approximate model:

\[
    \begin{align*}
        du &= \left( (-du + cv)u + F_u \right) dt + \sigma_u \, dW_u, \\
        dv &= (-dv v - cu^2 + F_v) \, dt + \sigma_v \, dW_v,
    \end{align*}
\]

(15)

\[ F_v = -2. \]
Improving the stochastic parameterizations

- **Perfect model:**
  \[
  du = (-d_u u + \gamma v) \, dt + \sigma_u \, dW_u, \tag{16a}
  \]
  \[
  dv = -d_v (v - \hat{v}) \, dt + \sigma_v(v) \, dW_v, \tag{16b}
  \]
  \[
  \sigma_v(v) = \exp\left(-\frac{|v| - m_c}{v_c}\right). \tag{17}
  \]

- **Approximate model,**
  \[
  du = (-d_u u + \gamma v) \, dt + \sigma_u \, dW_u, \tag{18a}
  \]
  \[
  dv = -d_v^M (v - \hat{v}^M) \, dt + \sigma_v^M \, dW_v, \tag{18b}
  \]

- **An improved approximate model is as follows,**
  \[
  du = (-d_u u + \gamma v) \, dt + \sigma_u \, dW_u, \tag{19a}
  \]
  \[
  dv = -\lambda^M (v - m^M) \, dt + \sigma_v^M (v) \, dW_v, \tag{19b}
  \]
**Theorem**

Given the decorrelation time $\tau$, the mean value $m$ and non-Gaussian PDF $p(x)$, there is an unique stochastic differential equation that satisfies these conditions,

$$dx(t) = -\lambda(x(t) - m) \, dt + \sigma(x) \, dW(t),$$

(20)

where $\lambda = 1/\tau$ and the multiplicative noise coefficient $\sigma(x)$ is given by

$$\sigma^2(x) = \frac{2}{p(x)} \{-\lambda \Phi(x)\}, \quad \text{with} \quad \Phi(x) = \int_b^x (y - m)p(y) \, dy.$$