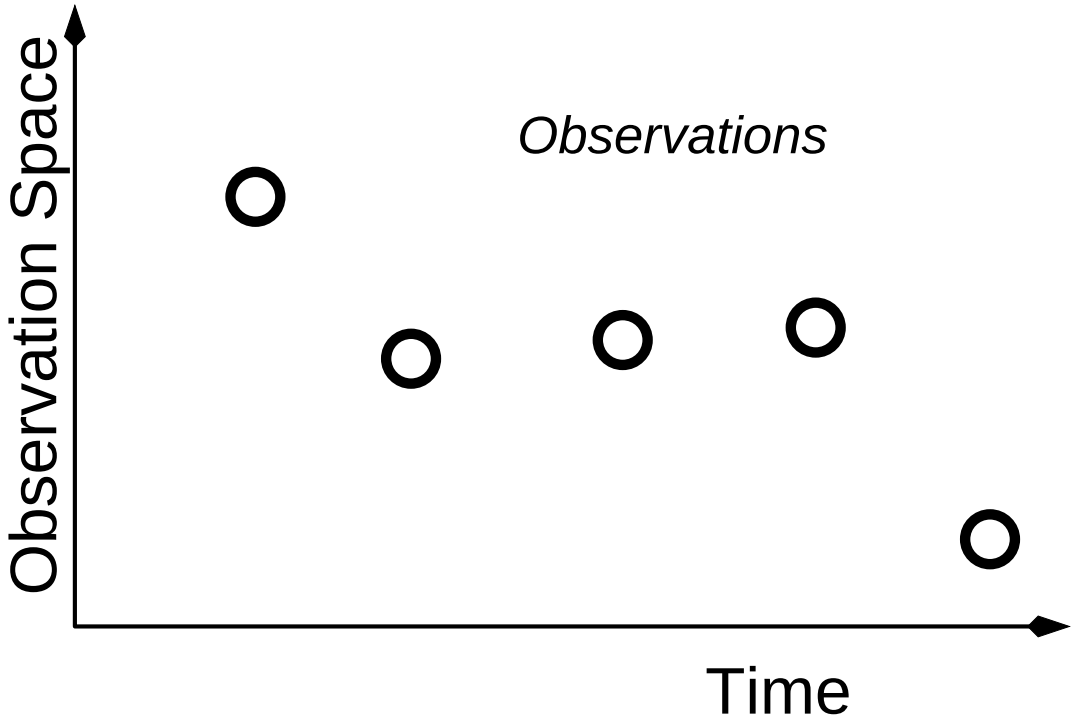
SENSITIVITY AND OUT–OF–SAMPLE ERROR IN DATA ASSIMILATION

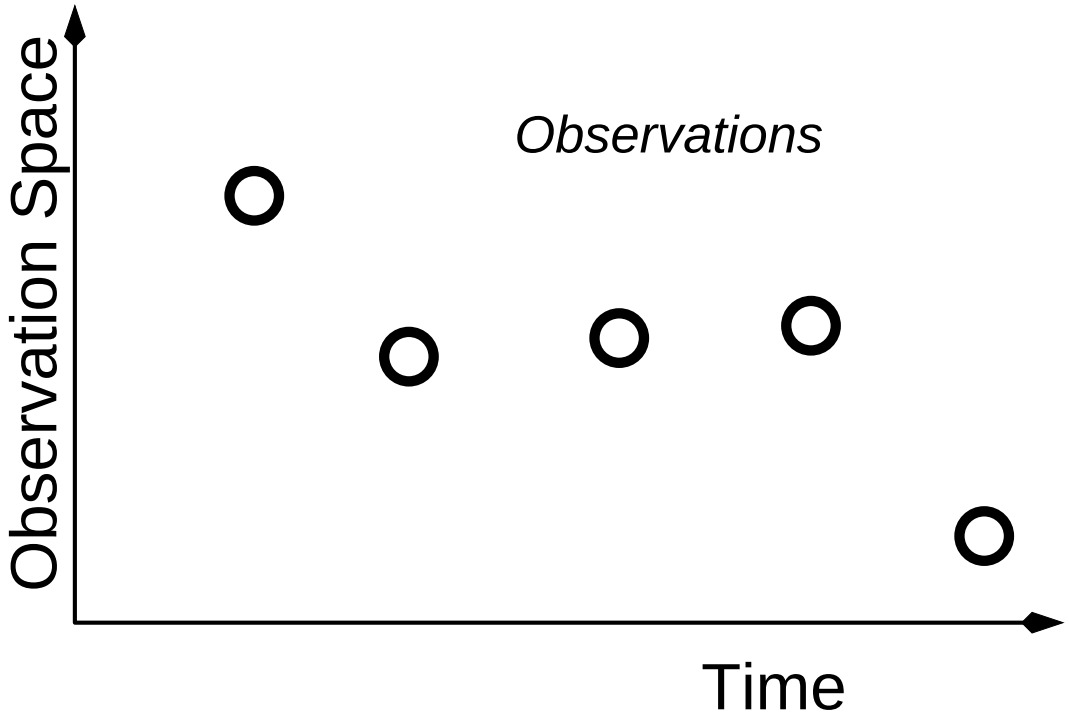
Jochen Bröcker, Ivan G. Szendro

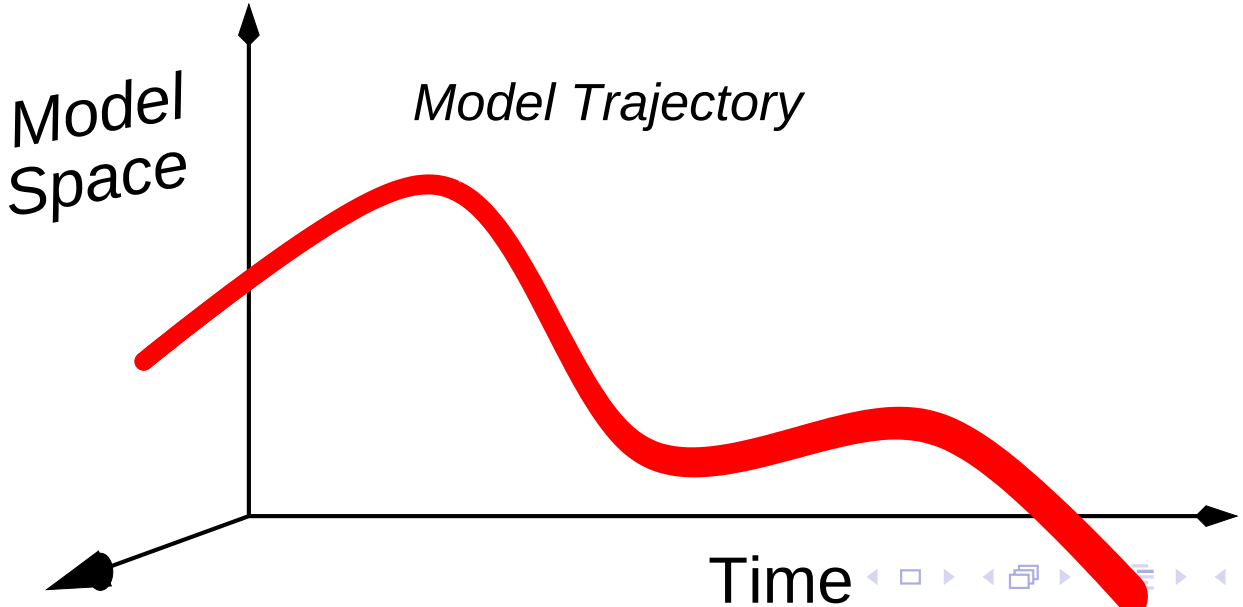
Ackn: H. Kantz, M. Niemann, G. del Magno

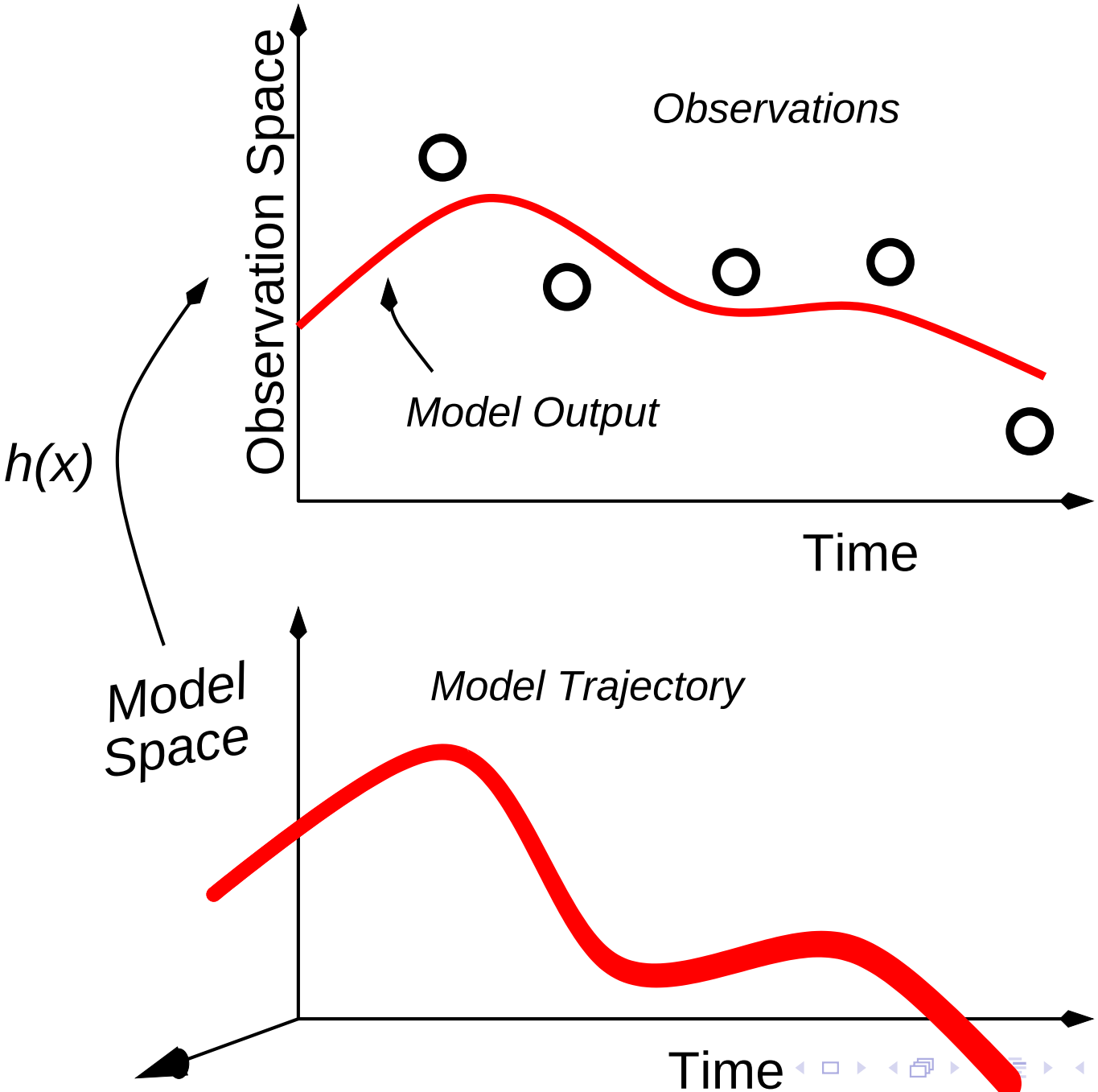
MPI für Physik komplexer Systeme, Dresden, Germany

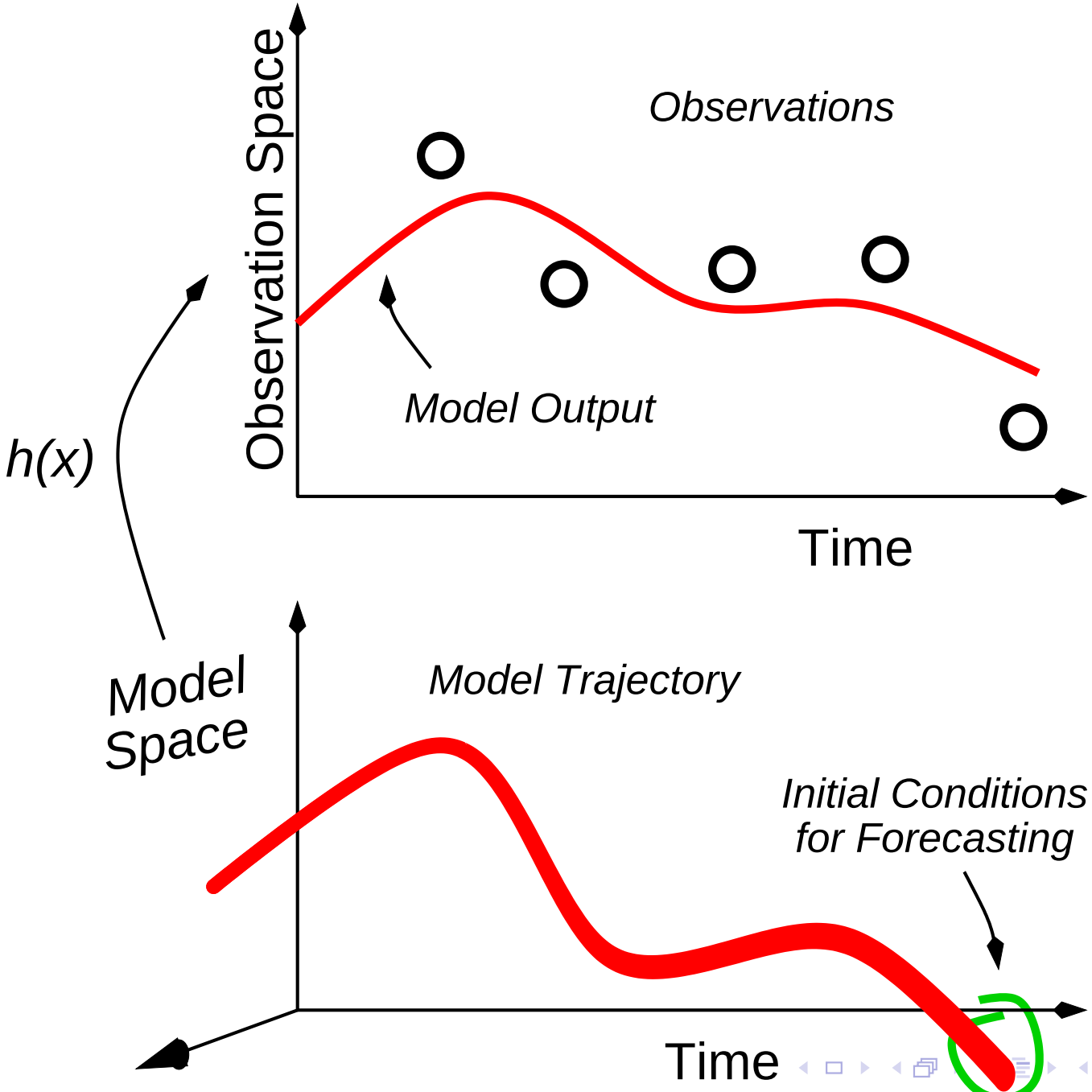
Warwick 2011


Observation Space
Observations
Time


Observation Space
Observations
Time


Model Space
Model Trajectory
Time


Observation Space
Observations
Model Output
Time
h(x)
Model Space
Model Trajectory
Time


Observation Space
Observations
Model Output
Time
h(x)
Model Space
Model Trajectory
Initial Conditions for Forecasting
Time

The problem of DA

Given:

- Observations $\eta_t, t \in [t_1, t_2]$
- Model with output $y_t = h(x_t) \qquad \dot{x}_t = f(x_t)$

Objective: Find a trajectory $\{x_t\}$ so that

$$\eta_t \cong y_t = h(x_t) \qquad \text{and} \qquad \dot{x}_t \cong f(x_t)$$

Problem:

Due to model error, there is no exact solution. We have to trade–off between

- Good model trajectories $\dot{x}_t \cong f(x_t)$
- Good tracking of the observations $\eta_t \cong y_t = h(x_t)$

The problem of DA

Given:

- Observations $\eta_t, t \in [t_1, t_2]$
- Model with output $y_t = h(x_t) \qquad \dot{x}_t = f(x_t)$

Objective: Find a trajectory $\{x_t\}$ so that

$$\eta_t \cong y_t = h(x_t) \qquad \text{and} \qquad \dot{x}_t \cong f(x_t)$$

Problem:

Due to model error, there is no exact solution. We have to trade–off between

- Good model trajectories $\dot{x}_t \cong f(x_t)$
- Good tracking of the observations $\eta_t \cong y_t = h(x_t)$

The problem of DA

Given:

- Observations $\eta_t, t \in [t_1, t_2]$
- Model with output $y_t = h(x_t) \qquad \dot{x}_t = f(x_t)$

Objective: Find a trajectory $\{x_t\}$ so that

$$\eta_t \cong y_t = h(x_t) \qquad \text{and} \qquad \dot{x}_t \cong f(x_t)$$

Problem:

Due to model error, there is no exact solution. We have to trade–off between

- Good model trajectories $\dot{x}_t \cong f(x_t)$
- Good tracking of the observations $\eta_t \cong y_t = h(x_t)$

Example: Variational approach

$$\dot{x}_t = f(x_t) + u_t \tag{1}$$

with perturbations u_t to the model.

$$\eta_t = h(x_t) + r_t \tag{2}$$

with perturbations r_t to the output.

With

Tracking Error $A_T = \int {r_t}^2 \, \mathrm{d}t$, Dynamical Error $A_D = \int {u_t}^2 \, \mathrm{d}t$

minimize

$$A_\alpha := \frac{\alpha}{2} A_T + \frac{1-\alpha}{2} A_D, \tag{3}$$

subject to conditions (1,2).

Example: Variational approach

$$\dot{x}_t = f(x_t) + u_t \tag{1}$$

with perturbations u_t to the model.

$$\eta_t = h(x_t) + r_t \tag{2}$$

with perturbations r_t to the output.
With

Tracking Error $A_T = \int {r_t}^2 \, \mathrm{d}t$, Dynamical Error $A_D = \int {u_t}^2 \, \mathrm{d}t$

minimize

$$A_\alpha := \frac{\alpha}{2} A_T + \frac{1-\alpha}{2} A_D, \tag{3}$$

subject to conditions (1,2).

Remarks

Remark 1: Error Covariance

In practice, we would use

$$A_T = \frac{1}{2} \int {r_t}^T R^{-1} r_t \, \mathrm{d}t$$

$$A_D = \frac{1}{2} \int {u_t}^T Q^{-1} u_t \, \mathrm{d}t$$

Remarks

Remark 2: Background Error

There is often an additional *Background Error*

$$A_B = \frac{1}{2}(x_{t_1} - x_1)^T B^{-1}(x_{t_1} - x_1)$$

with

$$B = \text{“Background error covariance”}$$
$$x_1 = \text{“first guess for } x_{t_1}\text{”}$$

Remarks

Remark 3: White observational noise

The observations might contain a white noise component, i.e.

$$\mathrm{d}\eta_t = \zeta_t \mathrm{d}t + \rho \mathrm{d}W_t.$$

Then the tracking error is re–defined as

$$A_T = \frac{1}{2\rho^2} \int h(x_t)^2 \, \mathrm{d}t - \frac{1}{\rho} \int h(x_t) \, \mathrm{d}\eta_t$$

Remarks

Remark 3: White observational noise

The observations might contain a white noise component, i.e.

$$\mathrm{d}\eta_t = \zeta_t \mathrm{d}t + \rho \mathrm{d}W_t.$$

Then the tracking error is re–defined as

$$A_T = \frac{1}{2\rho^2} \int h(x_t)^2 \, \mathrm{d}t - \frac{1}{\rho} \int h(x_t) \, \mathrm{d}\eta_t$$

Remarks

Remark 4: Solution amounts to two point BVP

States and co–states (x_t, λ_t) satisfy SDE

$$\begin{aligned} \mathrm{d}\lambda_t &= (F(x_t)\lambda_t + \alpha G(x_t))\,\mathrm{d}t + B(x_t)\mathrm{d}\eta_t \\ \dot{x}_t &= f(x_t) + u_t \\ u_t &= \frac{-1}{1-\alpha} Q\lambda_t, \end{aligned}$$

and there are boundary conditions, for example

$$\lambda_{t_1} = \lambda_{t_2} = 0$$

for free ends.

Remarks

Remark 4: Solution amounts to two point BVP

States and co–states (x_t, λ_t) satisfy SDE

$$\begin{aligned} \mathrm{d}\lambda_t &= (F(x_t)\lambda_t + \alpha G(x_t))\,\mathrm{d}t + B(x_t)\mathrm{d}\eta_t \\ \dot{x}_t &= f(x_t) + u_t \\ u_t &= \frac{-1}{1-\alpha} Q\lambda_t, \end{aligned}$$

and there are boundary conditions, for example

$$\lambda_{t_1} = \lambda_{t_2} = 0$$

for free ends.

Remarks

Remark 5: Statistical interpretation

$$A_\alpha = \frac{\alpha}{2}\int {r_t}^2 \, \mathrm{d}t + \frac{1-\alpha}{2}\int {u_t}^2 \, \mathrm{d}t$$

$$\dot{x}_t = f(x_t) + u_t, \qquad \eta_t = h(x_t) + r_t$$

Question

Does the trajectory $(\hat{x}_t, \hat{u}_t), t \in [t_1, t_2]$ minimising the functional A_α have a statistical interpretation?

Remarks

Remark 5: Statistical interpretation

Assume the observations come from

$$\mathrm{d}X_t = f(X_t)\mathrm{d}t + \sigma \mathrm{d}W_t'$$
$$\mathrm{d}\eta_t = h(X_t)\mathrm{d}t + \rho \mathrm{d}W_t.$$

Then it is possible to define the

$$\begin{aligned} &\text{a posteriori} \\ &= \text{“}p(x_t, t \in [t_1, t_2] | \eta_t, t \in [t_1, t_2])\text{”} \\ &= \text{Onsager–Machlup–Functional} \\ &\propto \exp(\beta A_\alpha + \text{further terms}), \end{aligned}$$

Conditions [Zeitouni and Dembo(1987)]:

- $R = \rho^T \rho, Q = \sigma^T \sigma$
- The mappings f, h have to be known exactly
- All model error is attributed to W, W'.

Remarks

Remark 5: Statistical interpretation

Assume the observations come from

$$\mathrm{d}X_t = f(X_t)\mathrm{d}t + \sigma \mathrm{d}W_t'$$
$$\mathrm{d}\eta_t = h(X_t)\mathrm{d}t + \rho \mathrm{d}W_t.$$

Then it is possible to define the

$$\begin{aligned} &\text{a posteriori} \\ &= \text{“}p(x_t, t \in [t_1, t_2] | \eta_t, t \in [t_1, t_2])\text{''} \\ &= \text{Onsager–Machlup–Functional} \\ &\propto \exp(\beta A_\alpha + \text{further terms}), \end{aligned}$$

Conditions [Zeitouni and Dembo(1987)]:

- $R = \rho^T \rho, Q = \sigma^T \sigma$
- The mappings f, h have to be known exactly
- All model error is attributed to W, W'.

Remarks

Remark 5: Statistical interpretation

Assume the observations come from

$$\mathrm{d}X_t = f(X_t)\mathrm{d}t + \sigma \mathrm{d}W_t'$$
$$\mathrm{d}\eta_t = h(X_t)\mathrm{d}t + \rho \mathrm{d}W_t.$$

Then it is possible to define the

$$\begin{aligned} &\text{a posteriori} \\ &= \text{“}p(x_t, t \in [t_1, t_2] | \eta_t, t \in [t_1, t_2])\text{''} \\ &= \text{Onsager–Machlup–Functional} \\ &\propto \exp(\beta A_\alpha + \text{further terms}), \end{aligned}$$

Conditions [Zeitouni and Dembo(1987)]:

- $R = \rho^T \rho, Q = \sigma^T \sigma$
- The mappings f, h have to be known exactly
- All model error is attributed to W, W'.

Remarks

Remark 5: Statistical interpretation

Therefore, A_α is not equal to the (negative log) Onsager–Machlup functional.
The optimal orbit $\hat{x}_t, t \in [t_1, t_2]$ is in general *not* equal to the maximum–aposteriori estimator.

Exception

$\mathrm{div}(f) = \mathrm{const}$, and R, Q are not state–dependent
$\Rightarrow A_\alpha = -\log$ Onsager–Machlup functional.

Remarks

Remark 5: Statistical interpretation

Therefore, A_α is not equal to the (negative log) Onsager–Machlup functional.
The optimal orbit $\hat{x}_t, t \in [t_1, t_2]$ is in general *not* equal to the maximum–aposteriori estimator.

Exception

$\mathrm{div}(f) = \mathrm{const}$, and R, Q are not state–dependent
$\Rightarrow A_\alpha = -\log$ Onsager–Machlup functional.

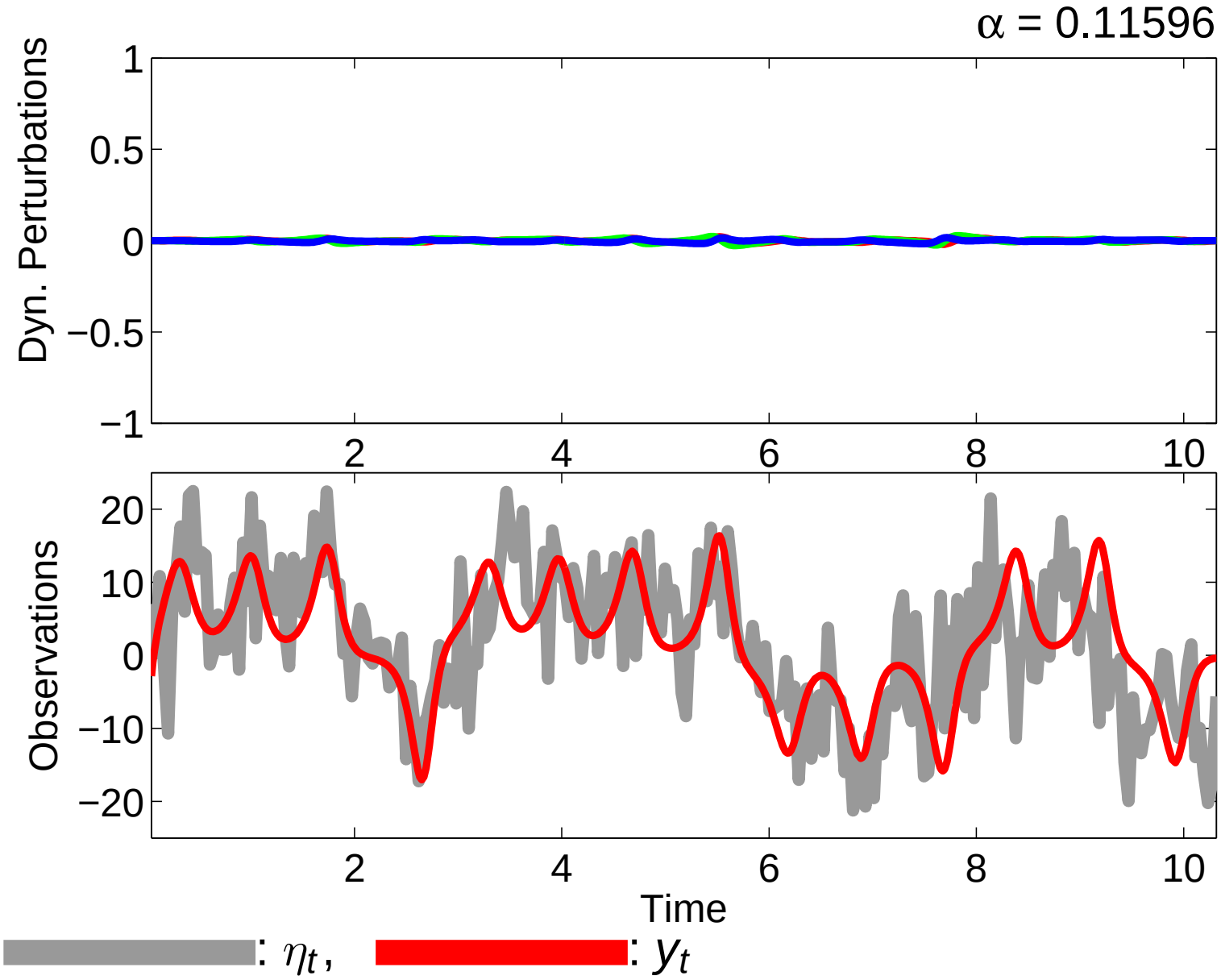
Numerical Example

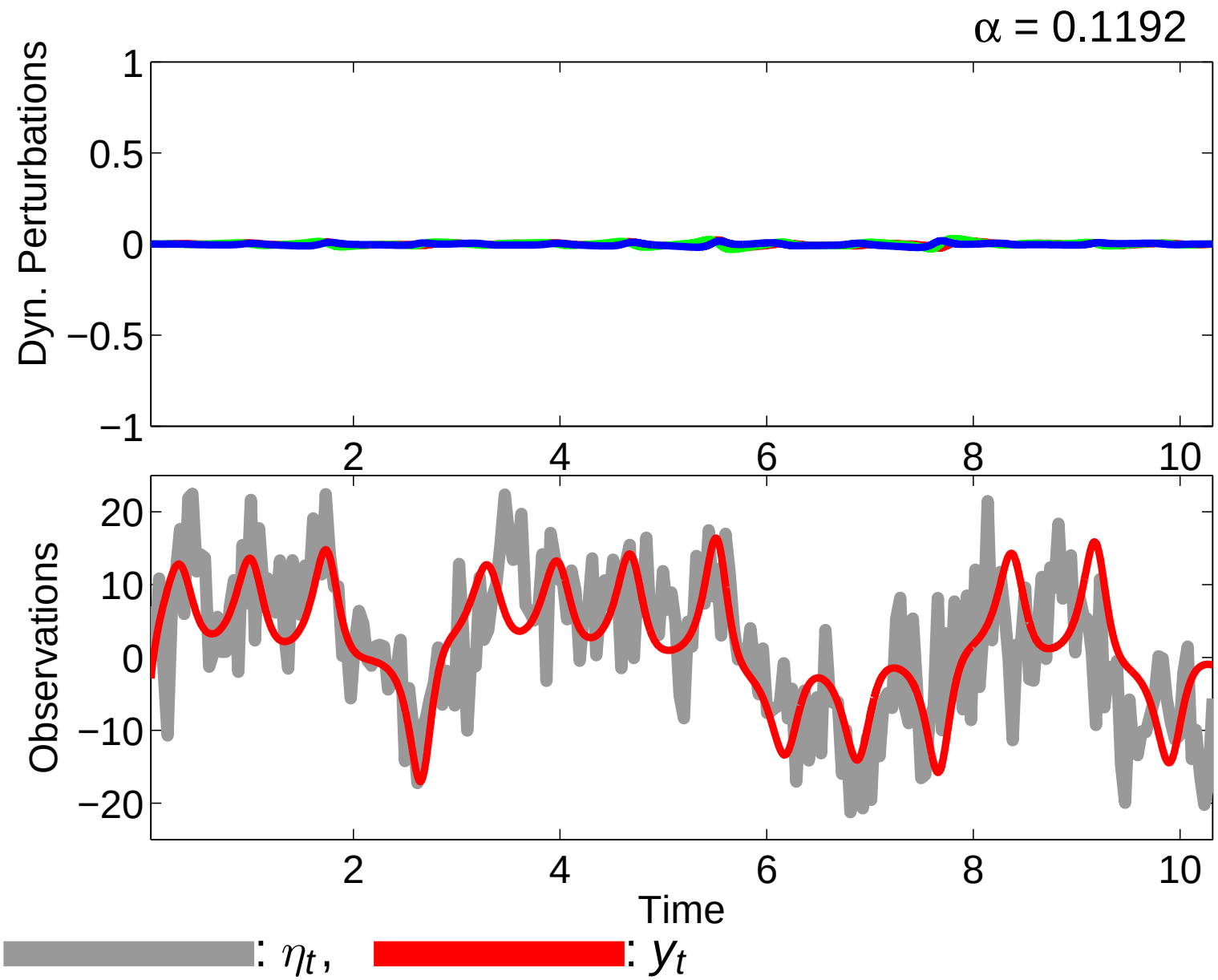
Lorenz'63

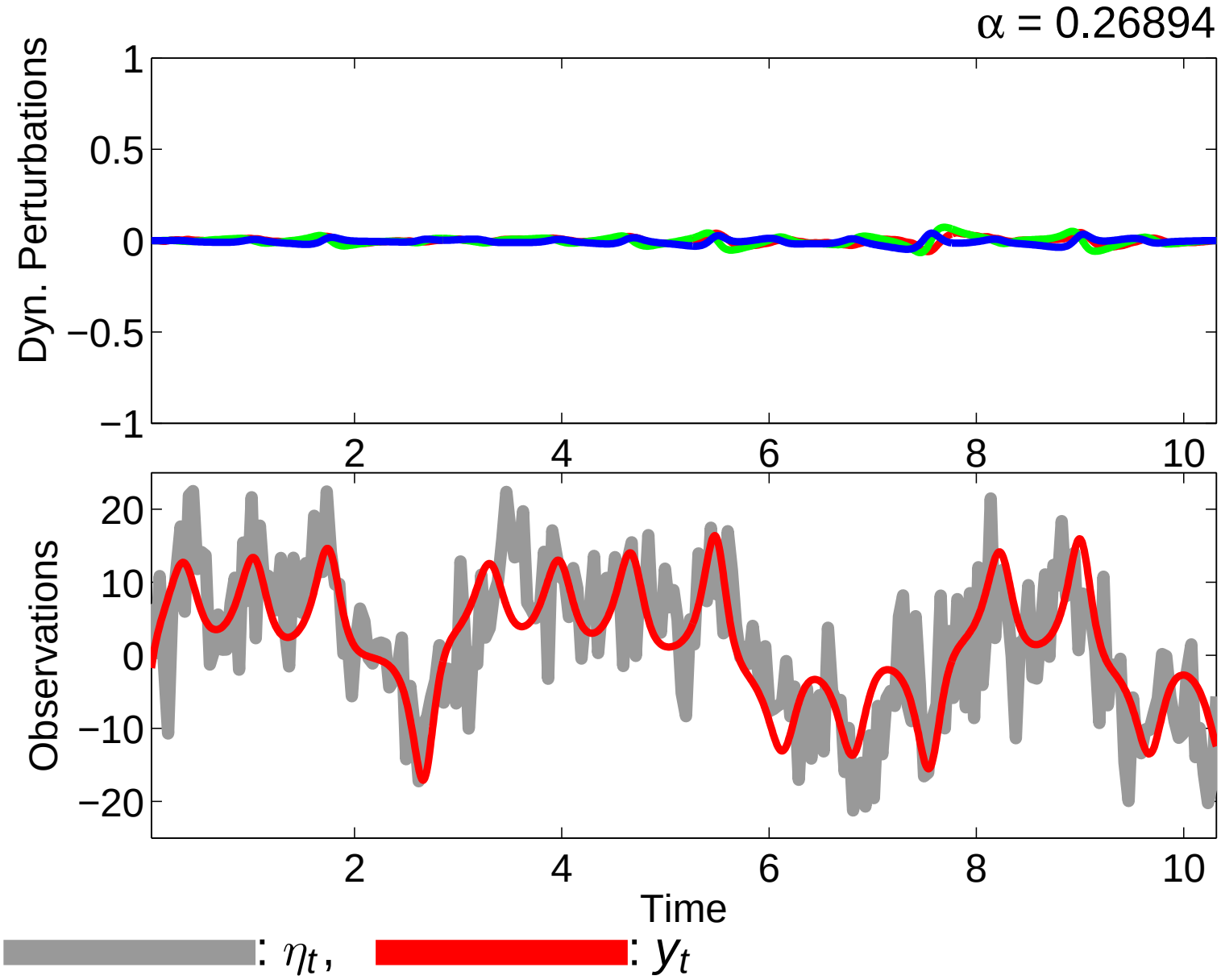
$$A_\alpha = \frac{\alpha}{2}\int {r_t}^2 \, \mathrm{d}t + \frac{1-\alpha}{2}\int {u_t}^2 \, \mathrm{d}t$$

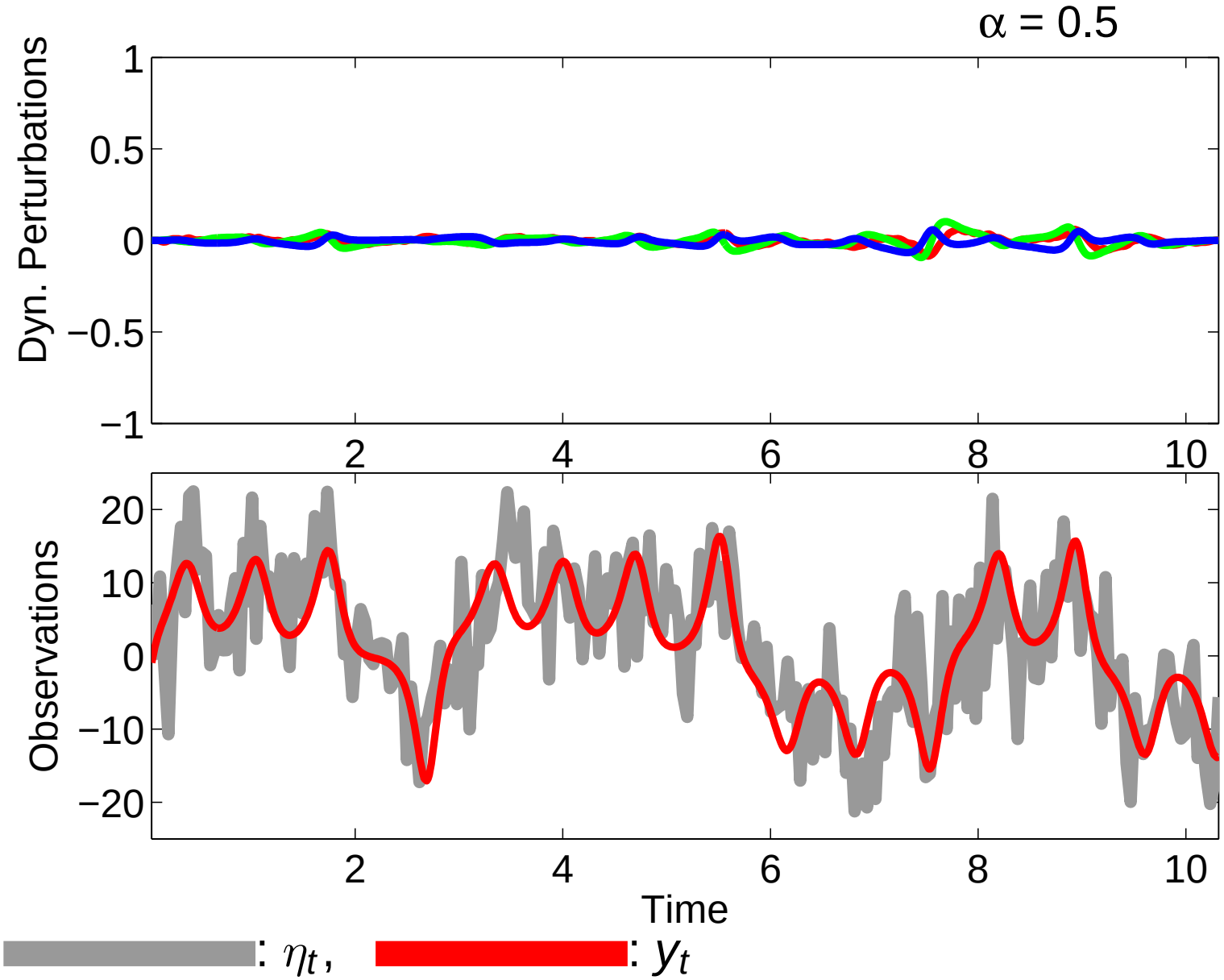
$$\dot{x}_t = f(x_t) + u_t, \qquad \eta_t = h(x_t) + r_t$$

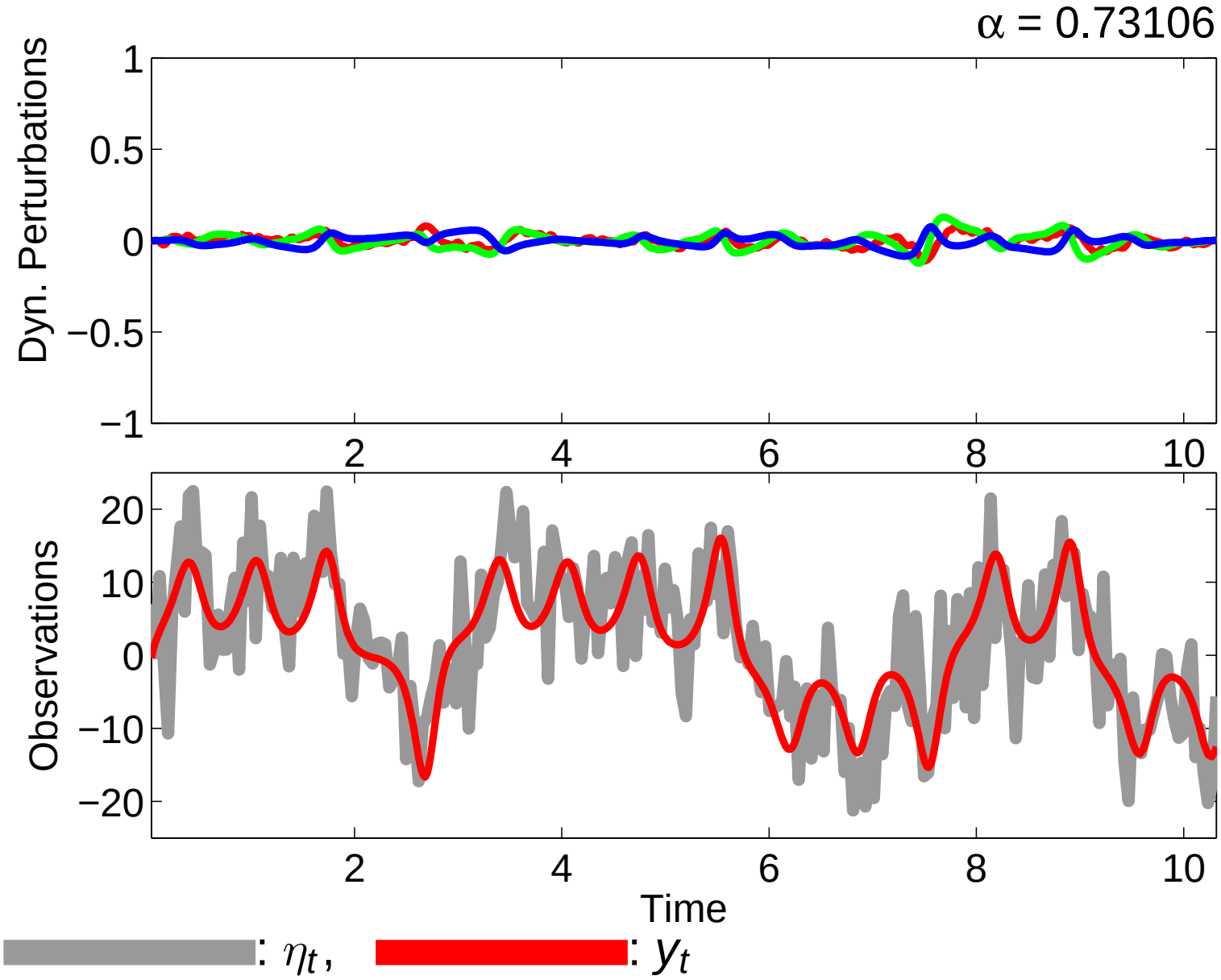
Lorenz'63: Three dimensional chaotic dynamics. Model and "Truth" were different, i.e. there was model error (and observational error).

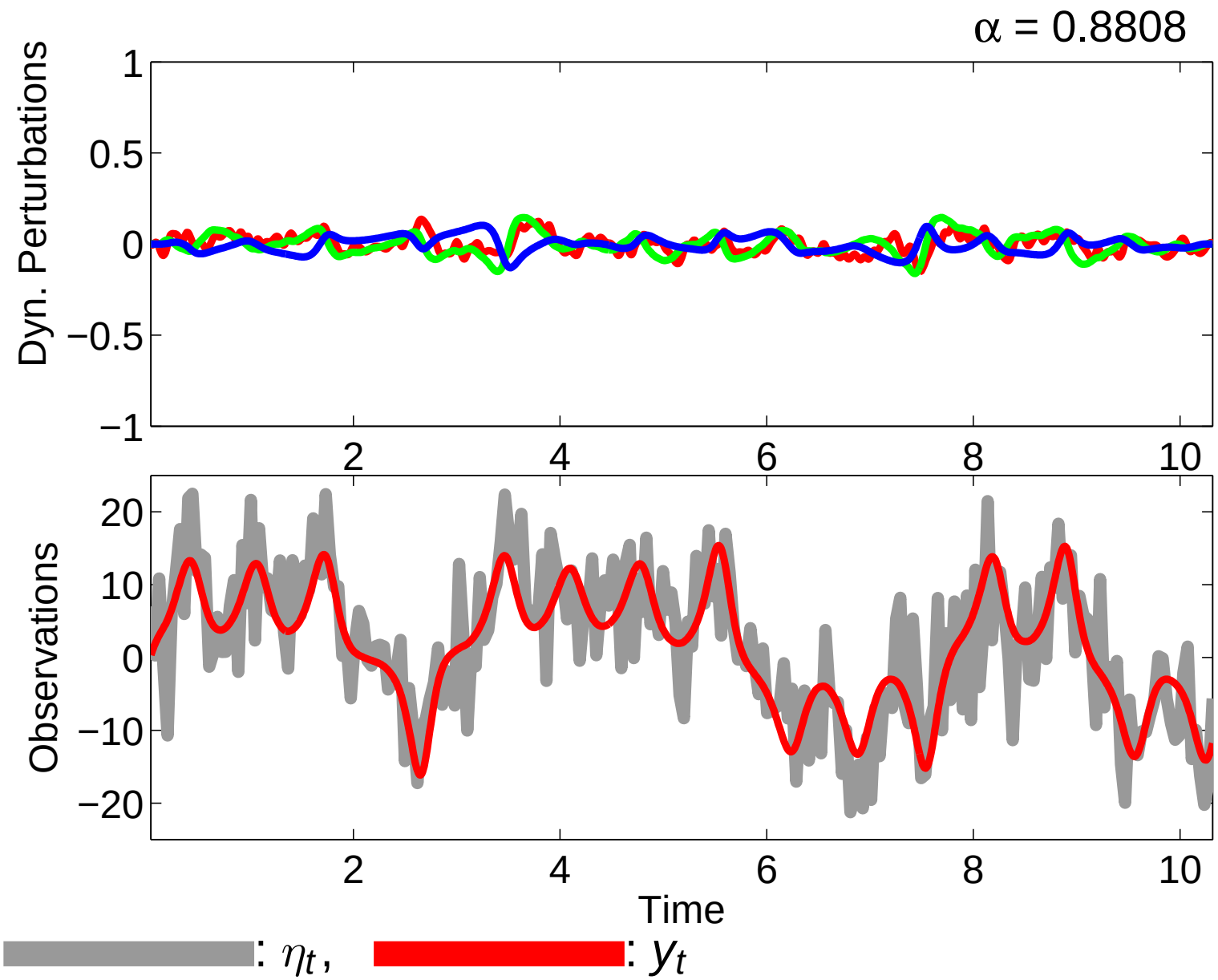

α = 0.11596
Dyn. Perturbations
1
0.5
0
−0.5
−1
2
4
6
8
10
Observations
20
10
0
−10
−20
2
4
6
8
10
Time
: η_t ,
: y_t

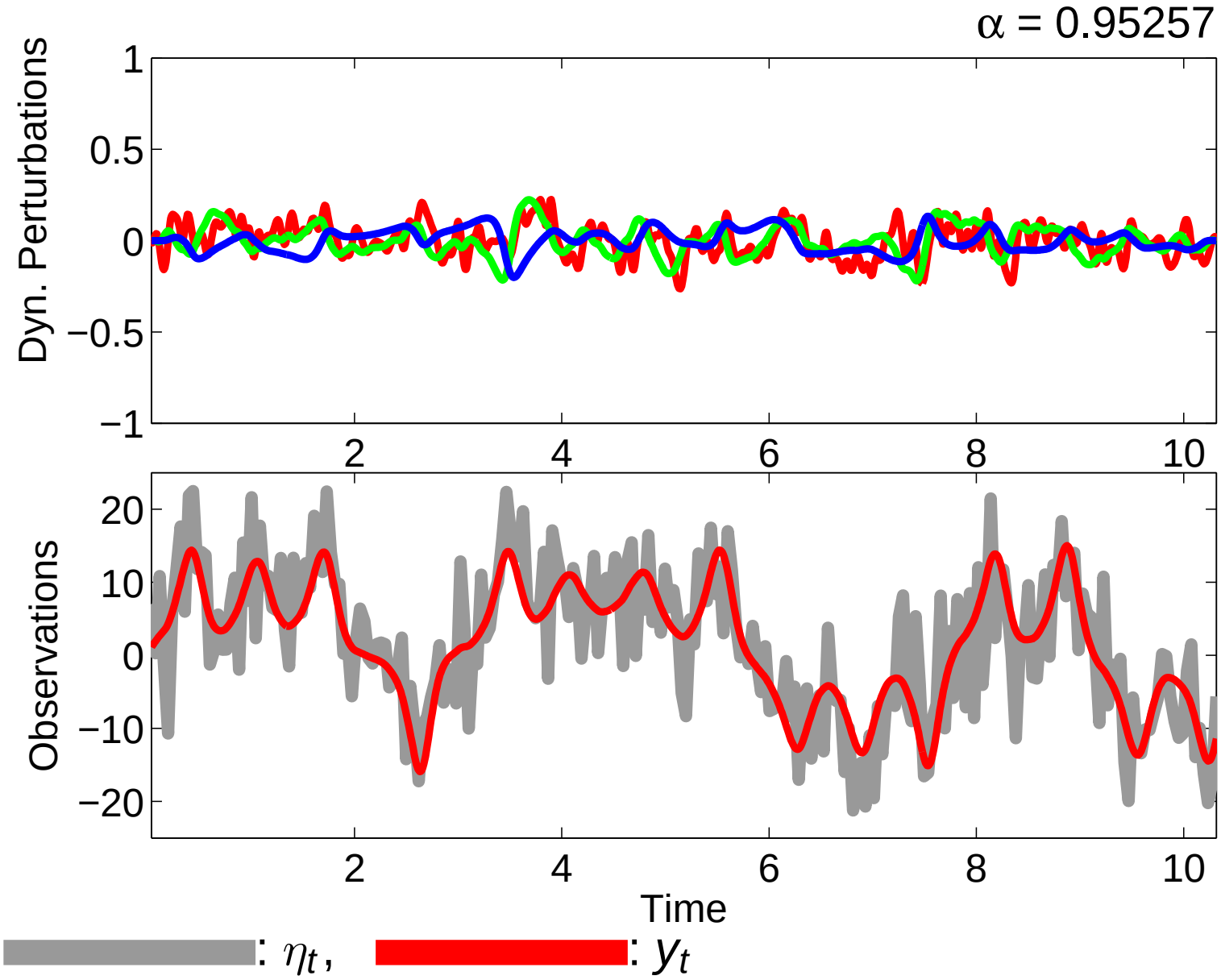

α = 0.1192
Dyn. Perturbations
1
0.5
0
−0.5
−1
2
4
6
8
10
Observations
20
10
0
−10
−20
2
4
6
8
10
Time
: η_t ,
: y_t

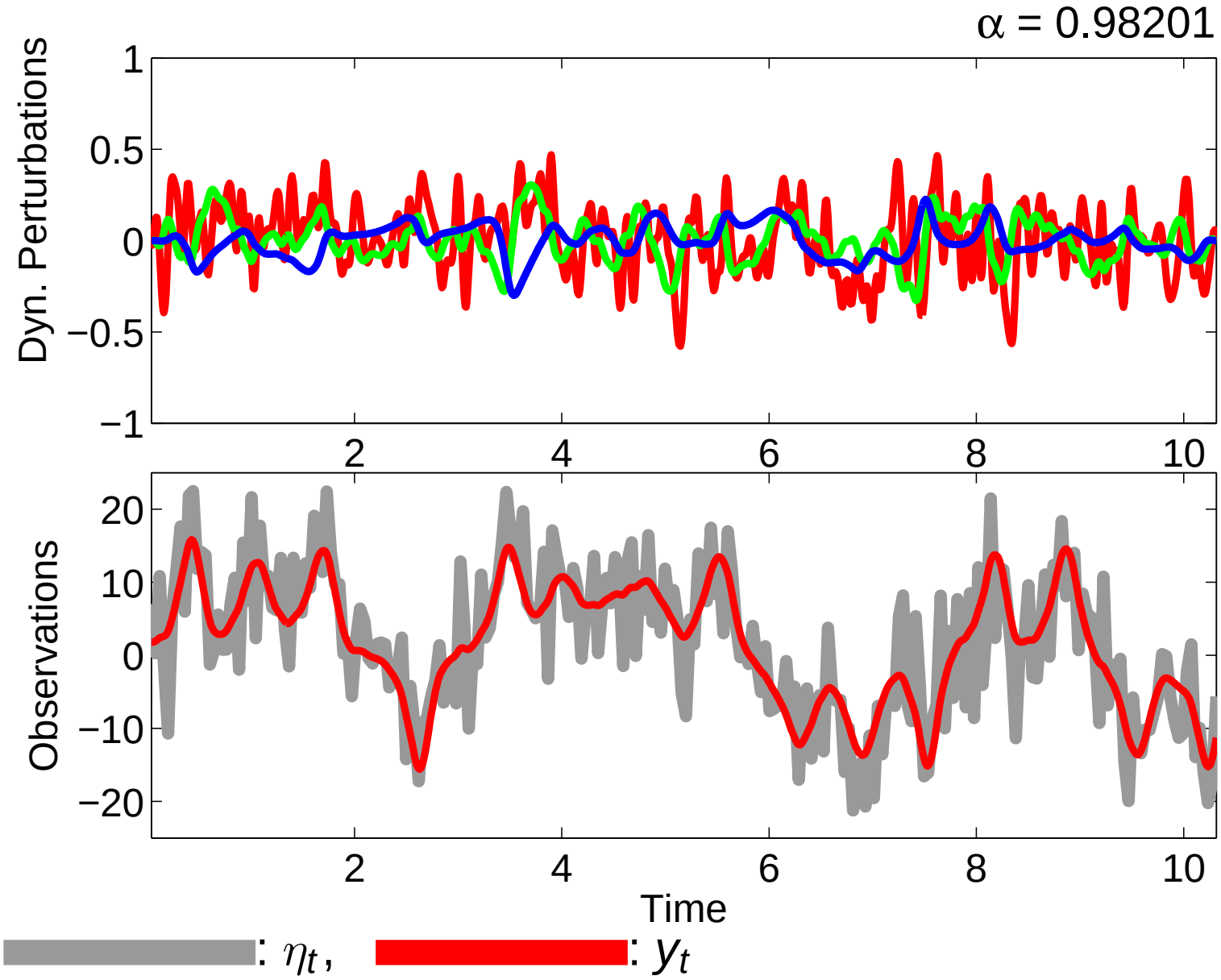

α = 0.26894
Dyn. Perturbations
1
0.5
0
−0.5
−1
2
4
6
8
10
Observations
20
10
0
−10
−20
2
4
6
8
10
Time
: η_t ,
: y_t

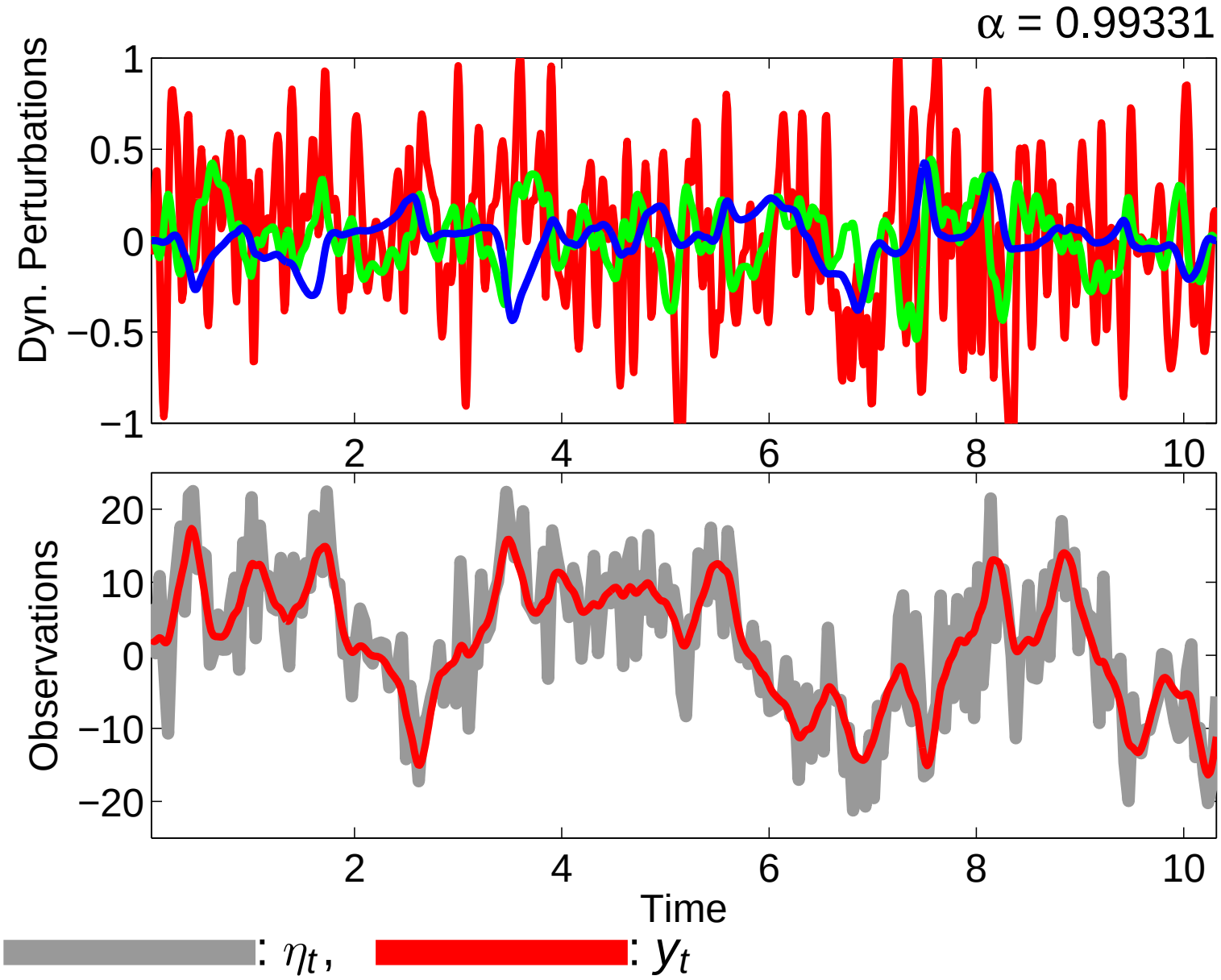

α = 0.5
Dyn. Perturbations
1
0.5
0
−0.5
−1
2
4
6
8
10
Observations
20
10
0
−10
−20
2
4
6
8
10
Time
: η_t ,
: y_t

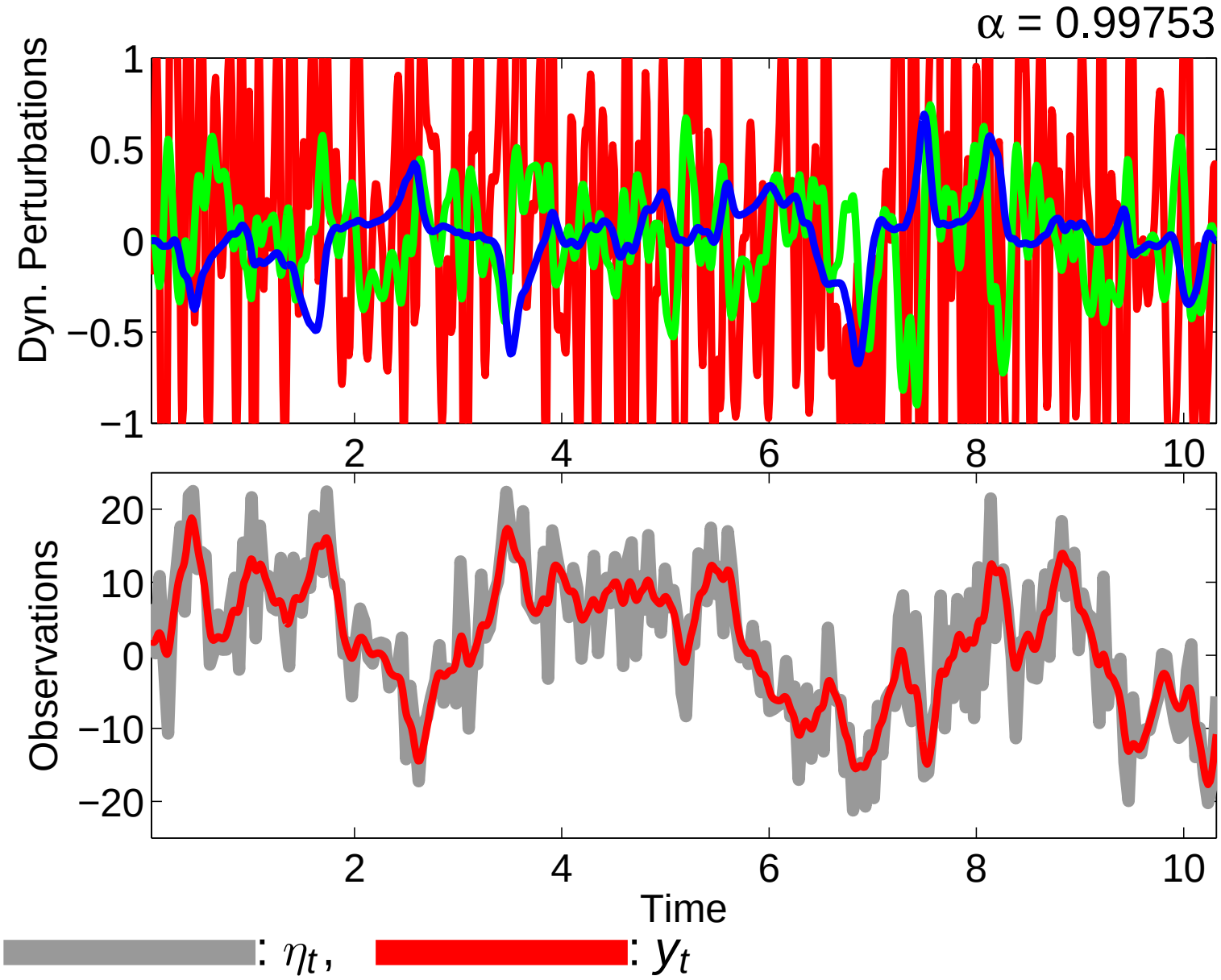

α = 0.73106
Dyn. Perturbations
1
0.5
0
−0.5
−1
2
4
6
8
10
Observations
20
10
0
−10
−20
2
4
6
8
10
Time
: η_t ,
: y_t

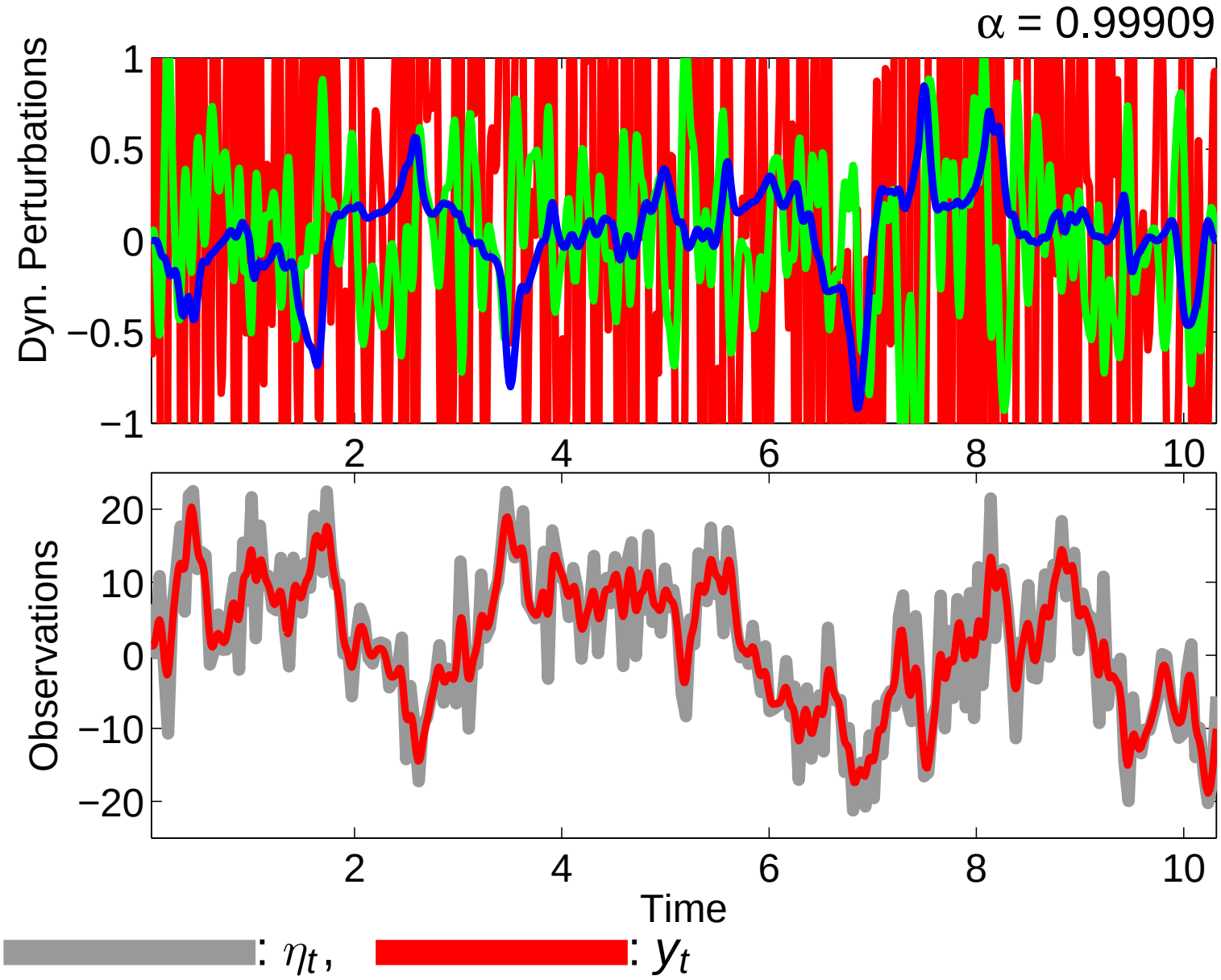

α = 0.8808
Dyn. Perturbations
1
0.5
0
−0.5
−1
2
4
6
8
10
Observations
20
10
0
−10
−20
2
4
6
8
10
Time
: η_t ,
: y_t

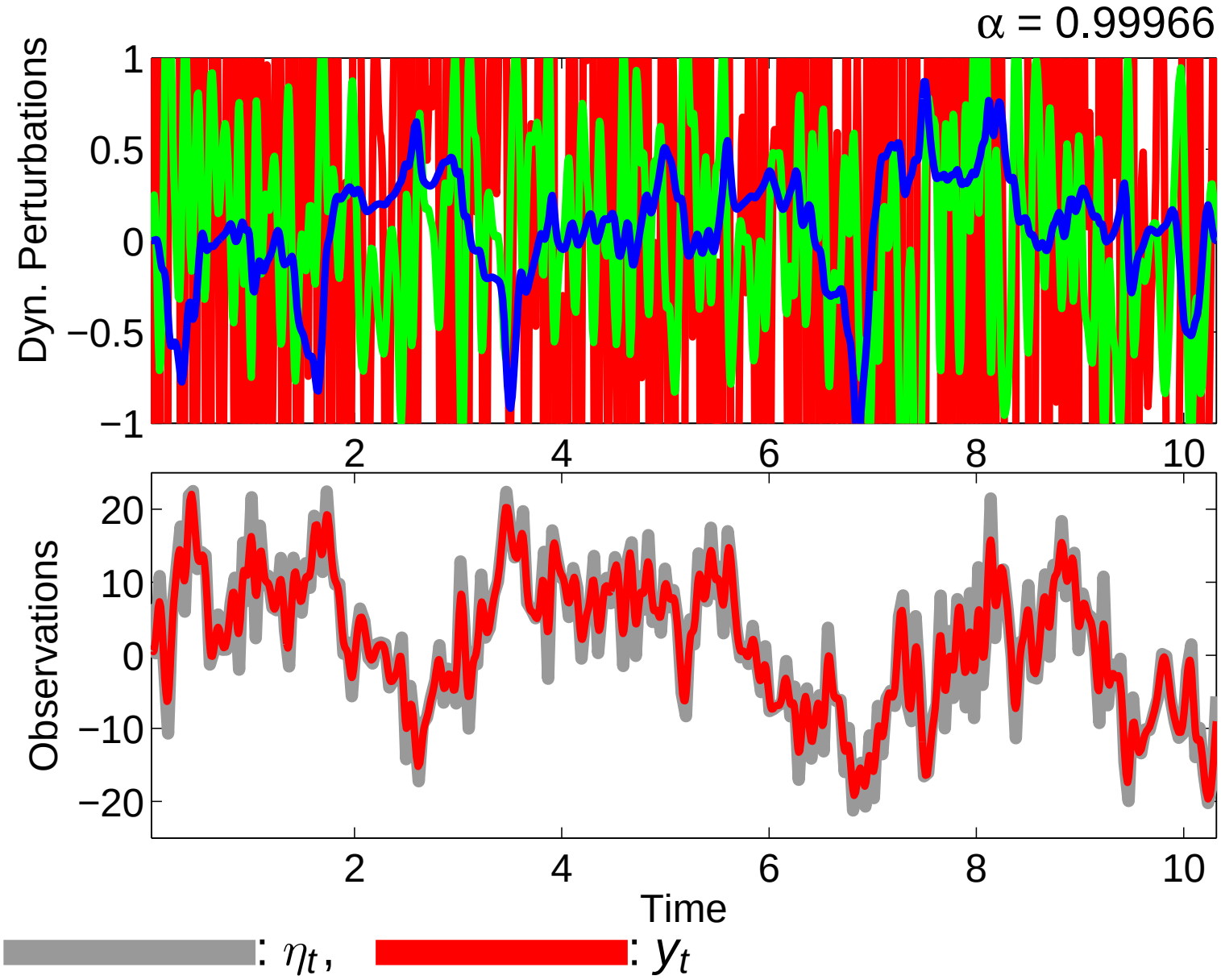

α = 0.95257
Dyn. Perturbations
1
0.5
0
−0.5
−1
2
4
6
8
10
Observations
20
10
0
−10
−20
2
4
6
8
10
Time
: η_t ,
: y_t

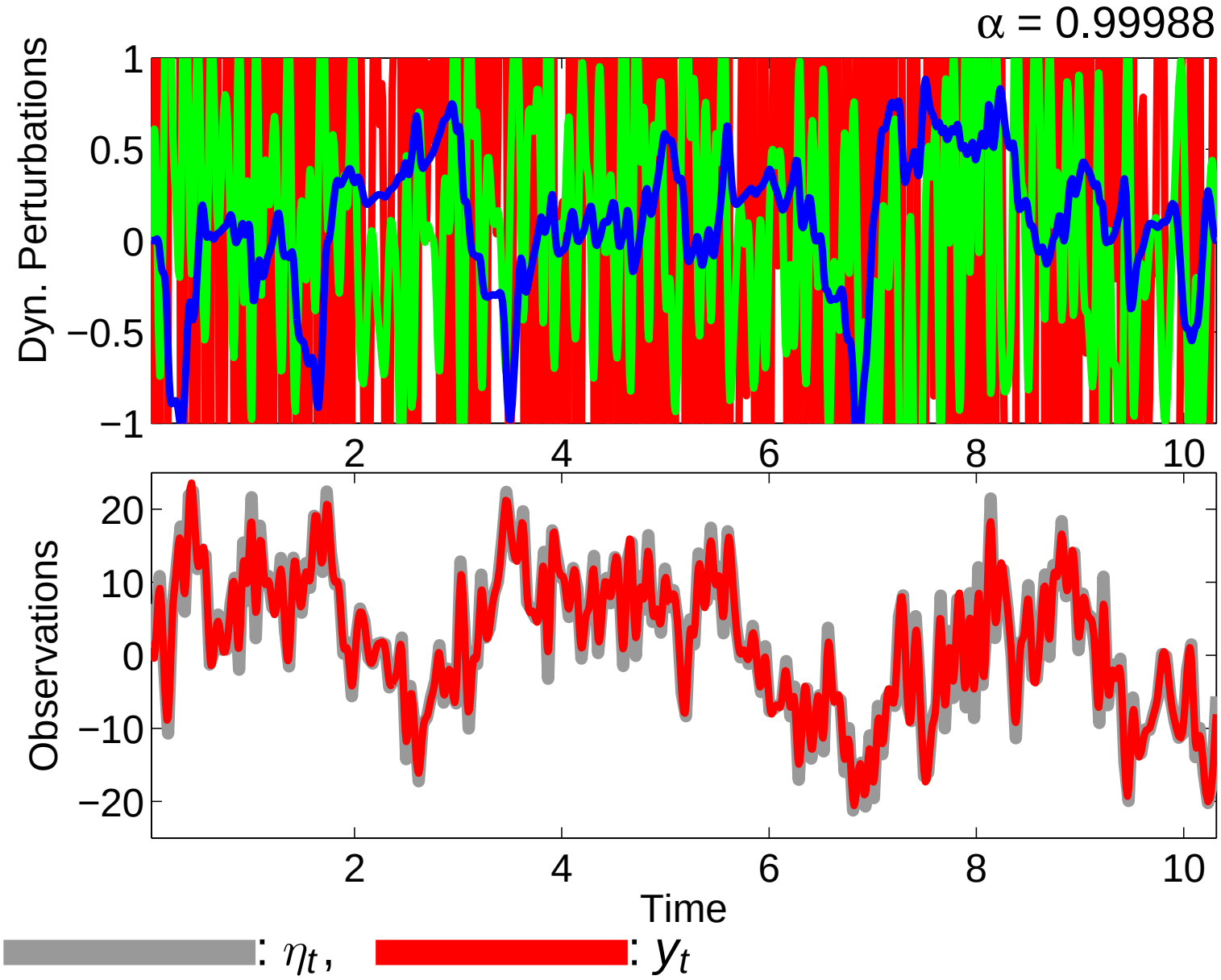

α = 0.98201
Dyn. Perturbations
1
0.5
0
−0.5
−1
2
4
6
8
10
Observations
20
10
0
−10
−20
2
4
6
8
10
Time
: η_t ,
: y_t

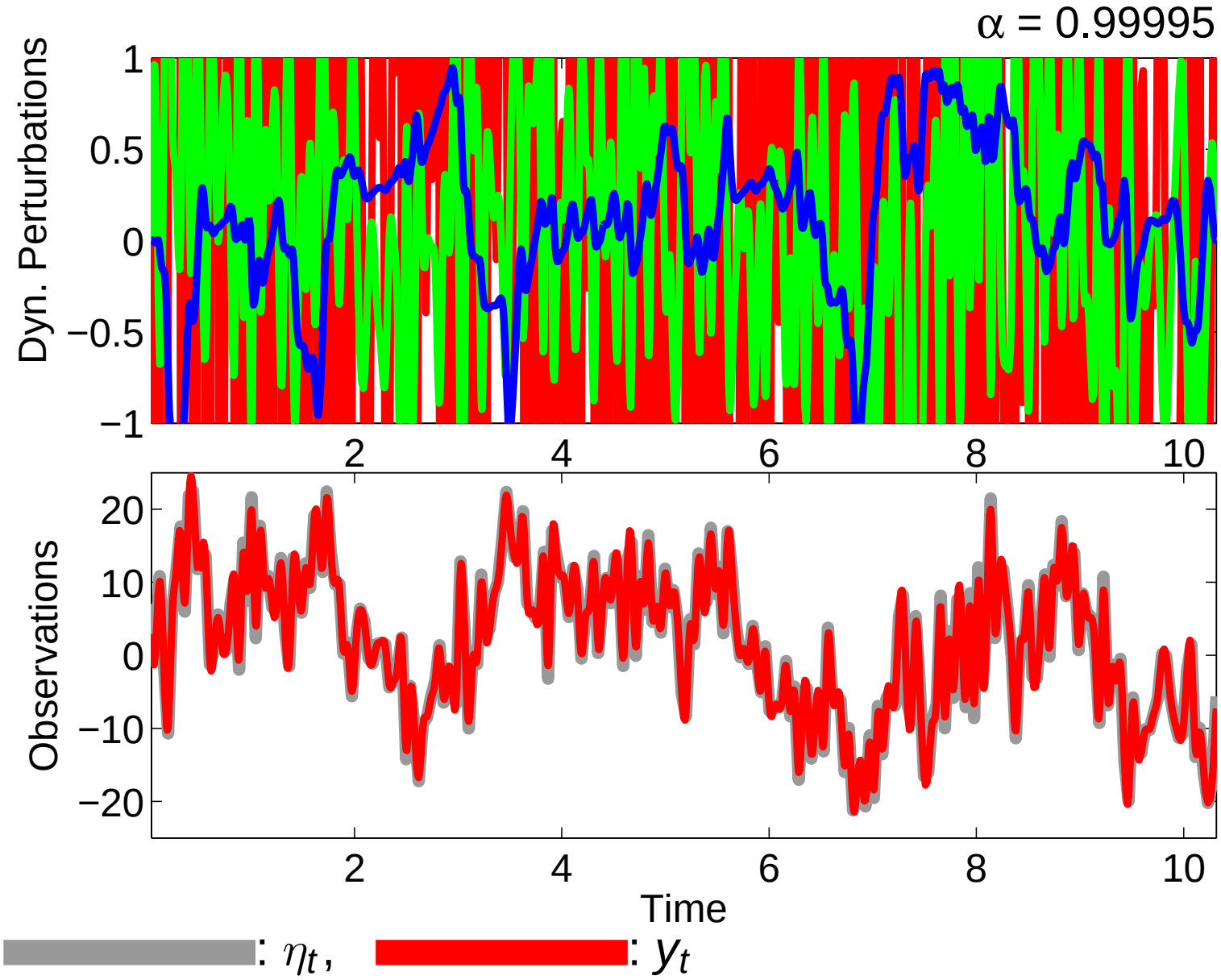

α = 0.99331
Dyn. Perturbations
1
0.5
0
−0.5
−1
2
4
6
8
10
Observations
20
10
0
−10
−20
2
4
6
8
10
Time
: η_t ,
: y_t

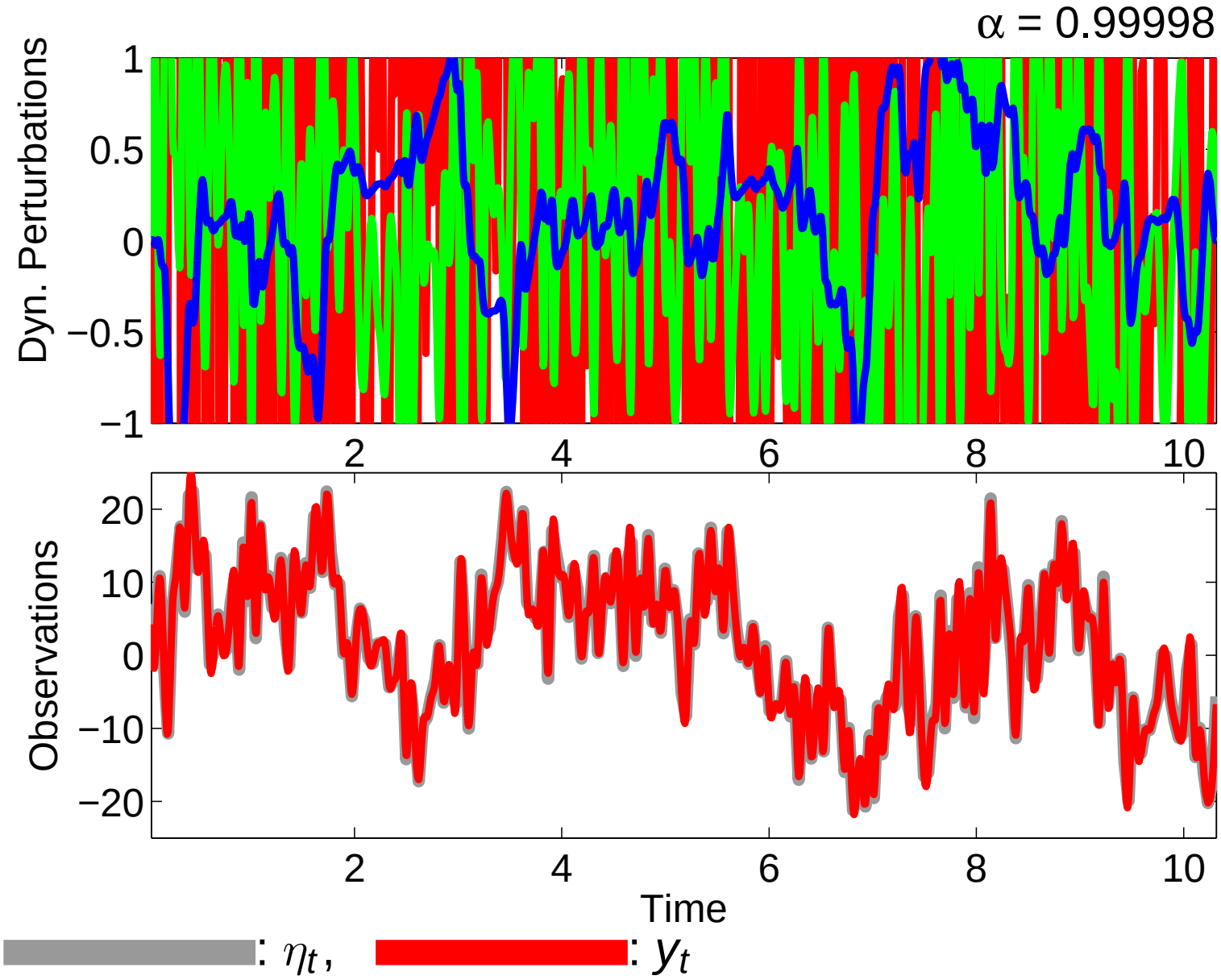

α = 0.99753
Dyn. Perturbations
1
0.5
0
−0.5
−1
2
4
6
8
10
Observations
20
10
0
−10
−20
2
4
6
8
10
Time
: η_t ,
: y_t


α = 0.99909
Dyn. Perturbations
1
0.5
0
−0.5
−1
2
4
6
8
10
Observations
20
10
0
−10
−20
2
4
6
8
10
Time
: η_t ,
: y_t

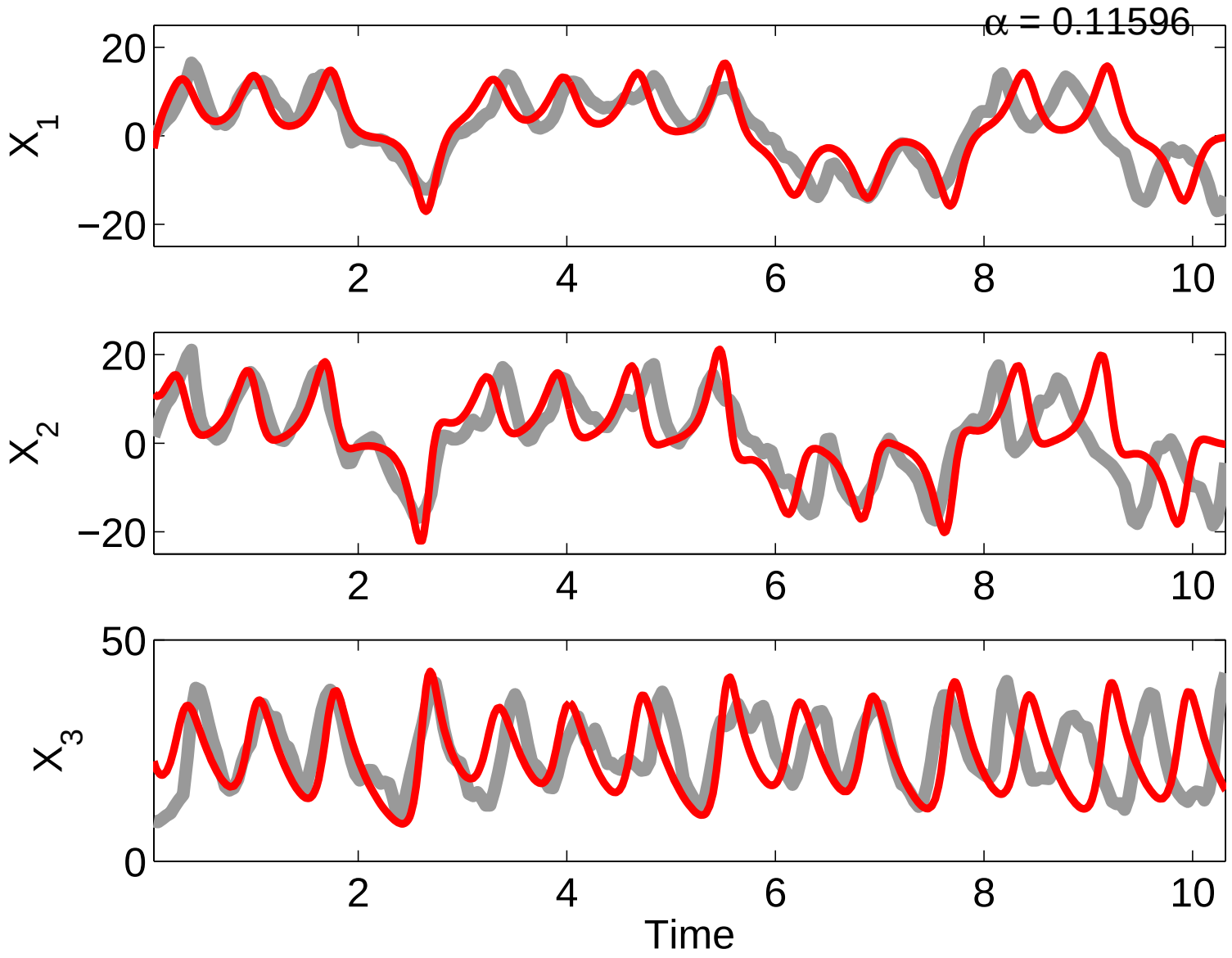

α = 0.99966
Dyn. Perturbations
1
0.5
0
−0.5
−1
2
4
6
8
10
Observations
20
10
0
−10
−20
2
4
6
8
10
Time
: η_t ,
: y_t

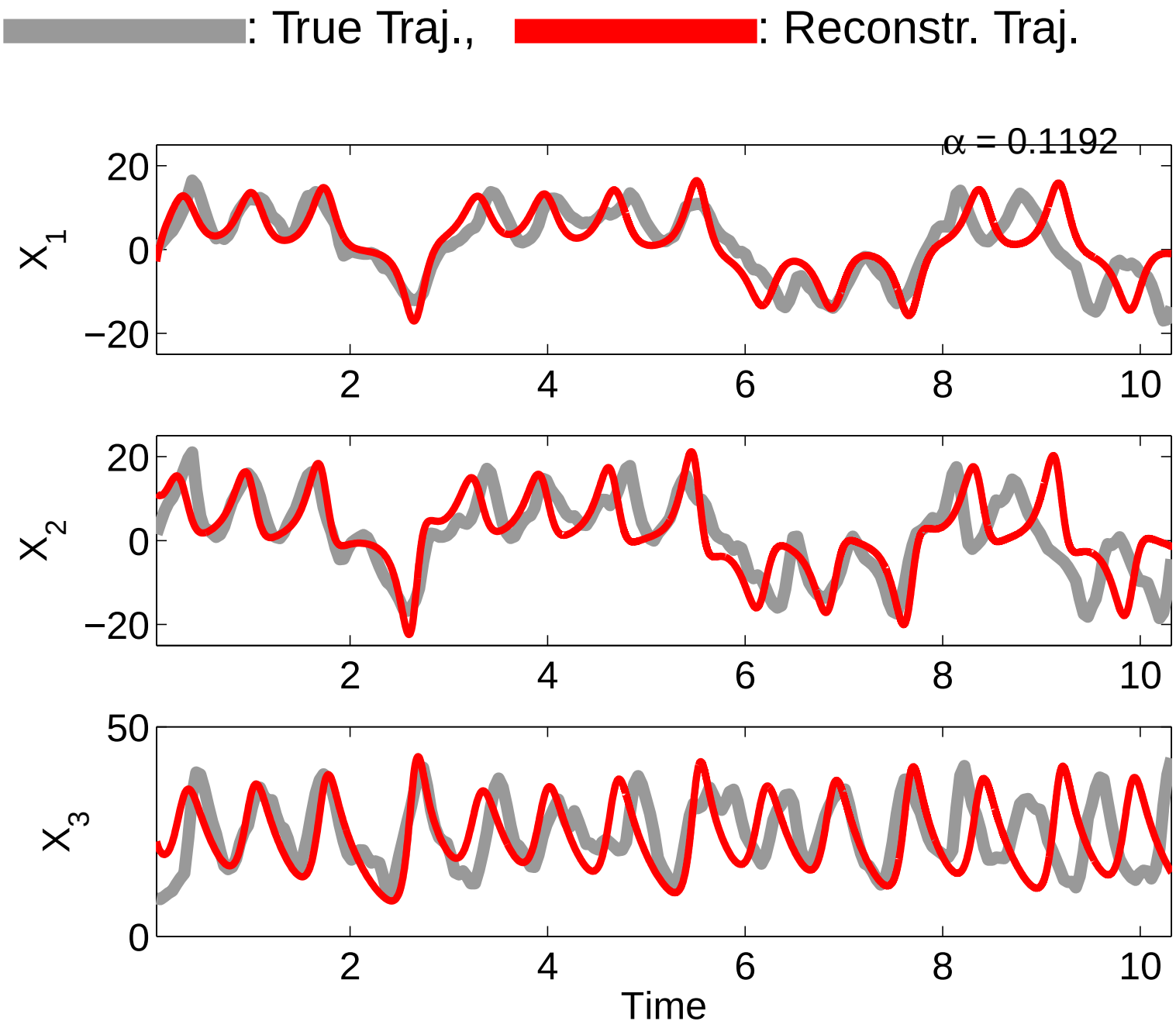

α = 0.99988
Dyn. Perturbations
1
0.5
0
−0.5
−1
2
4
6
8
10
Observations
20
10
0
−10
−20
2
4
6
8
10
Time
: η_t ,
: y_t

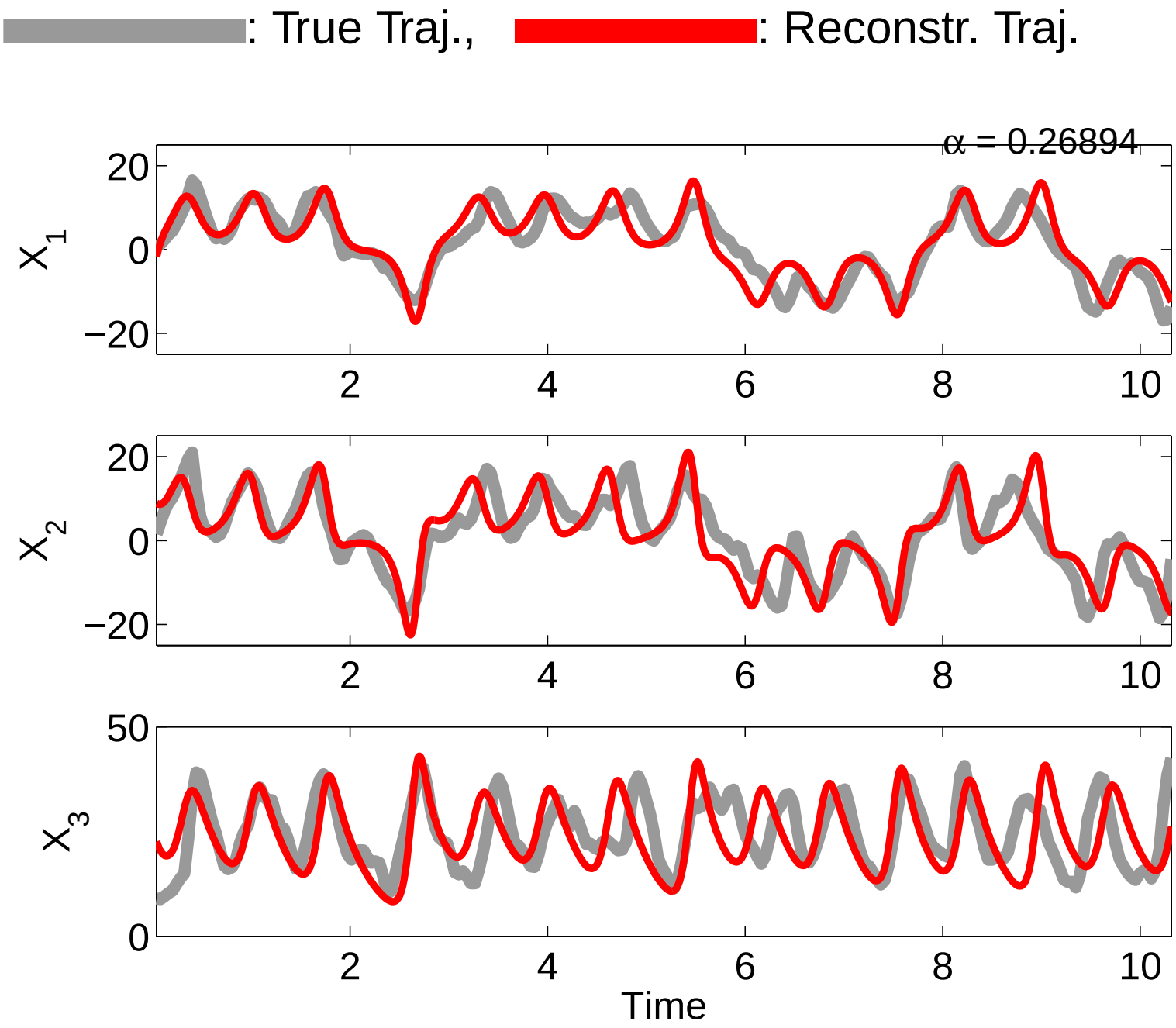

α = 0.99995
Dyn. Perturbations
1
0.5
0
−0.5
−1
2
4
6
8
10
Observations
20
10
0
−10
−20
2
4
6
8
10
Time
: η_t ,
: y_t

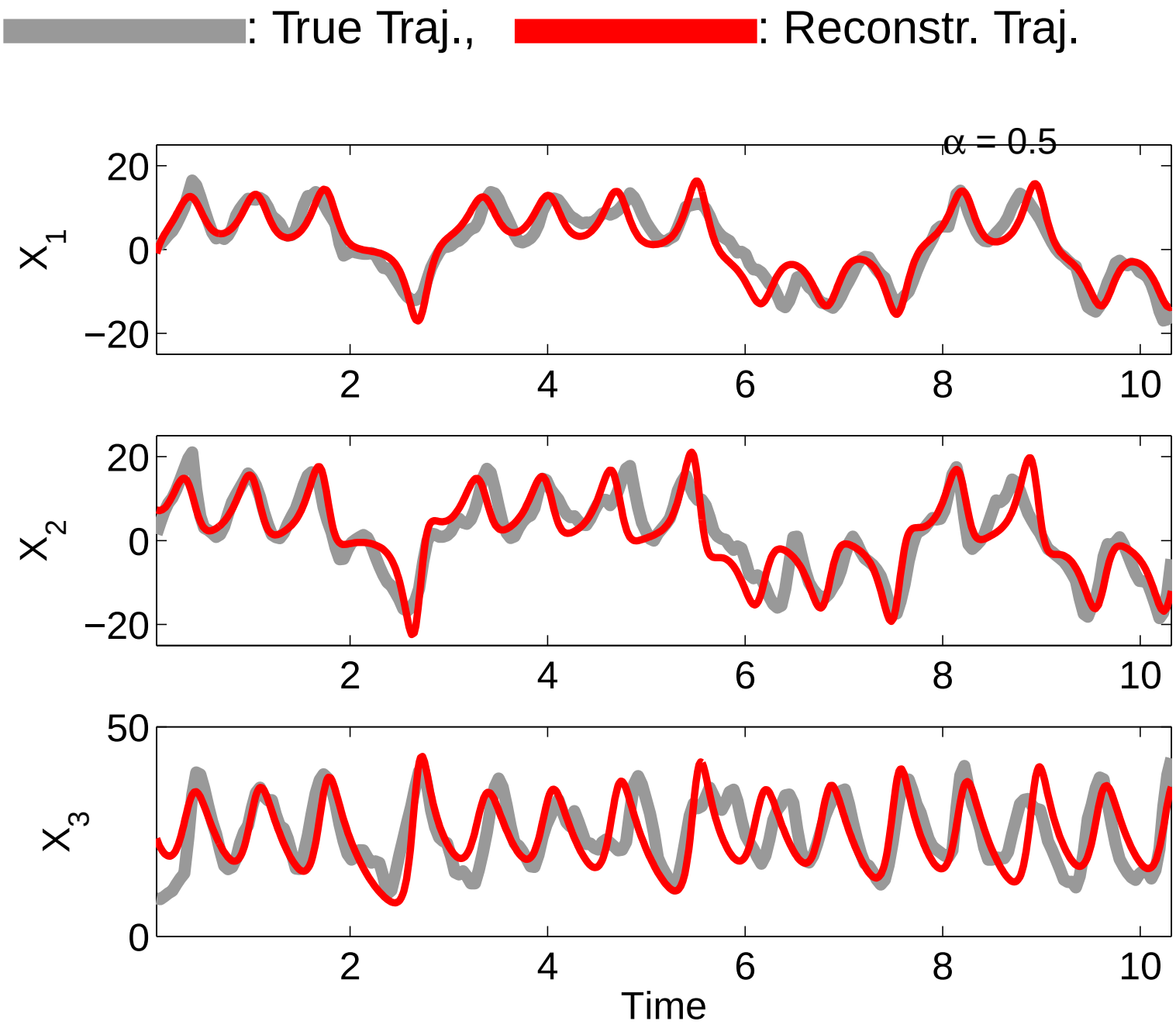

α = 0.99998
Dyn. Perturbations
1
0.5
0
−0.5
−1
2
4
6
8
10
Observations
20
10
0
−10
−20
2
4
6
8
10
Time
: η_t ,
: y_t

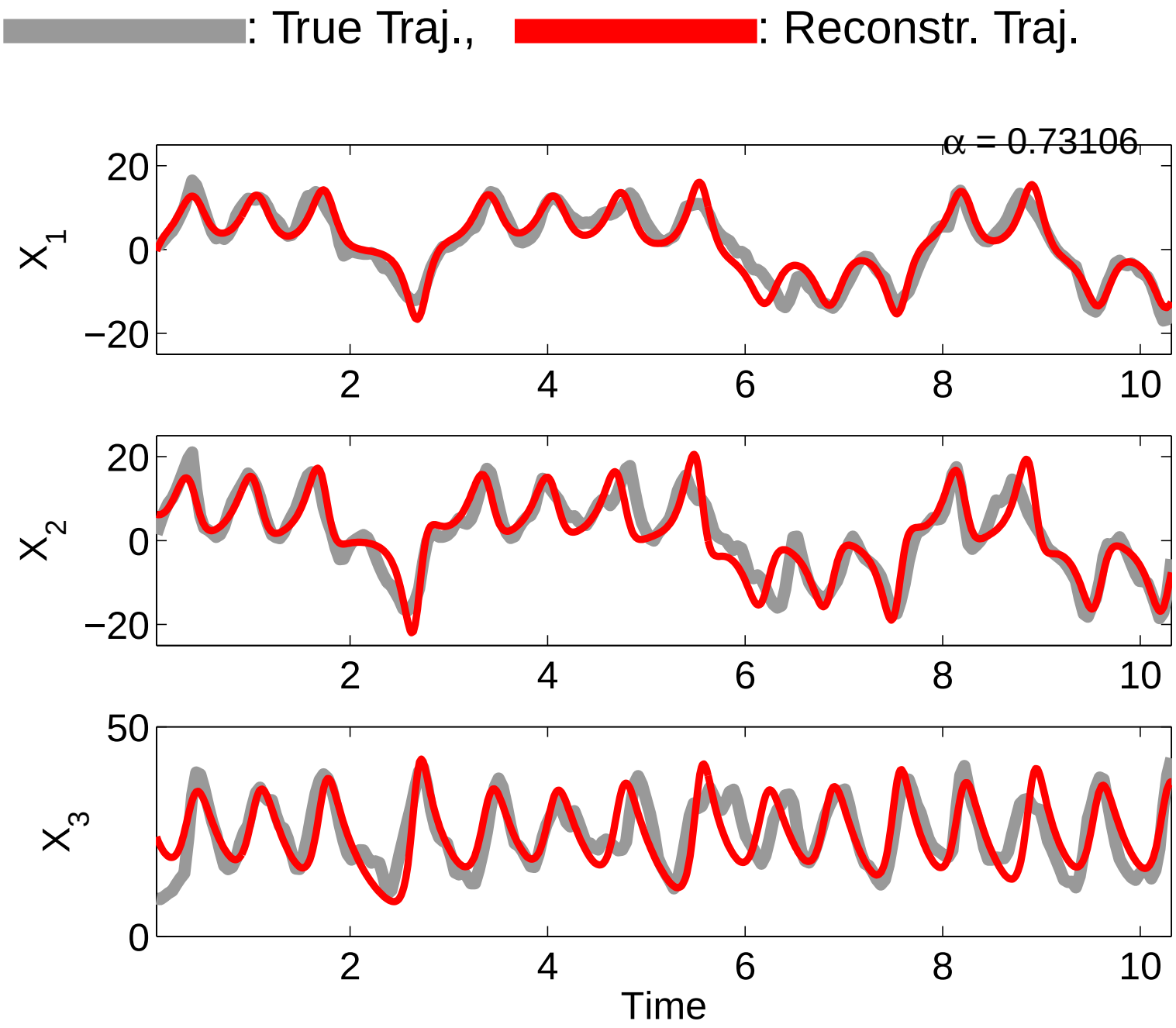
: True Traj., : Reconstr. Traj.

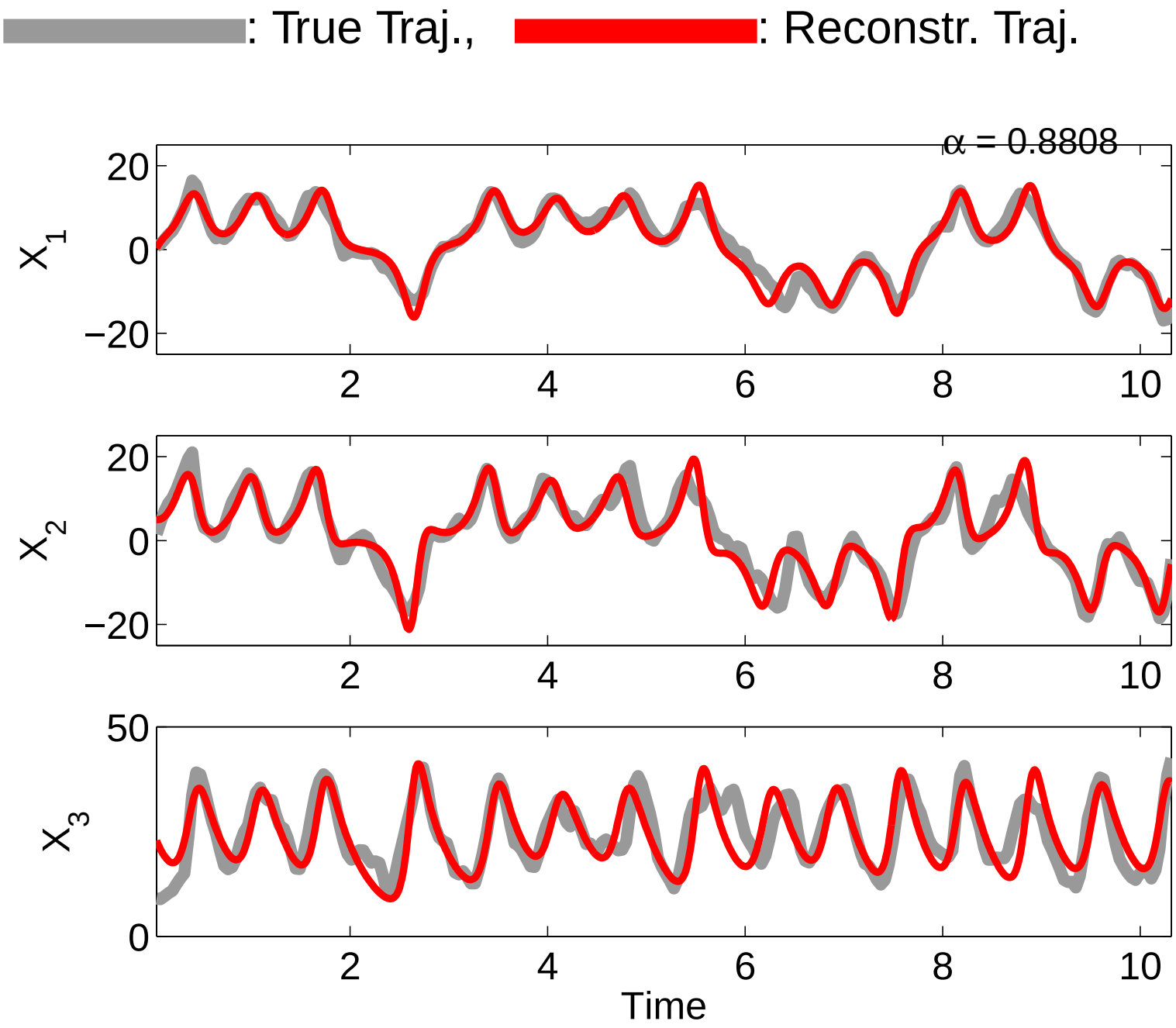


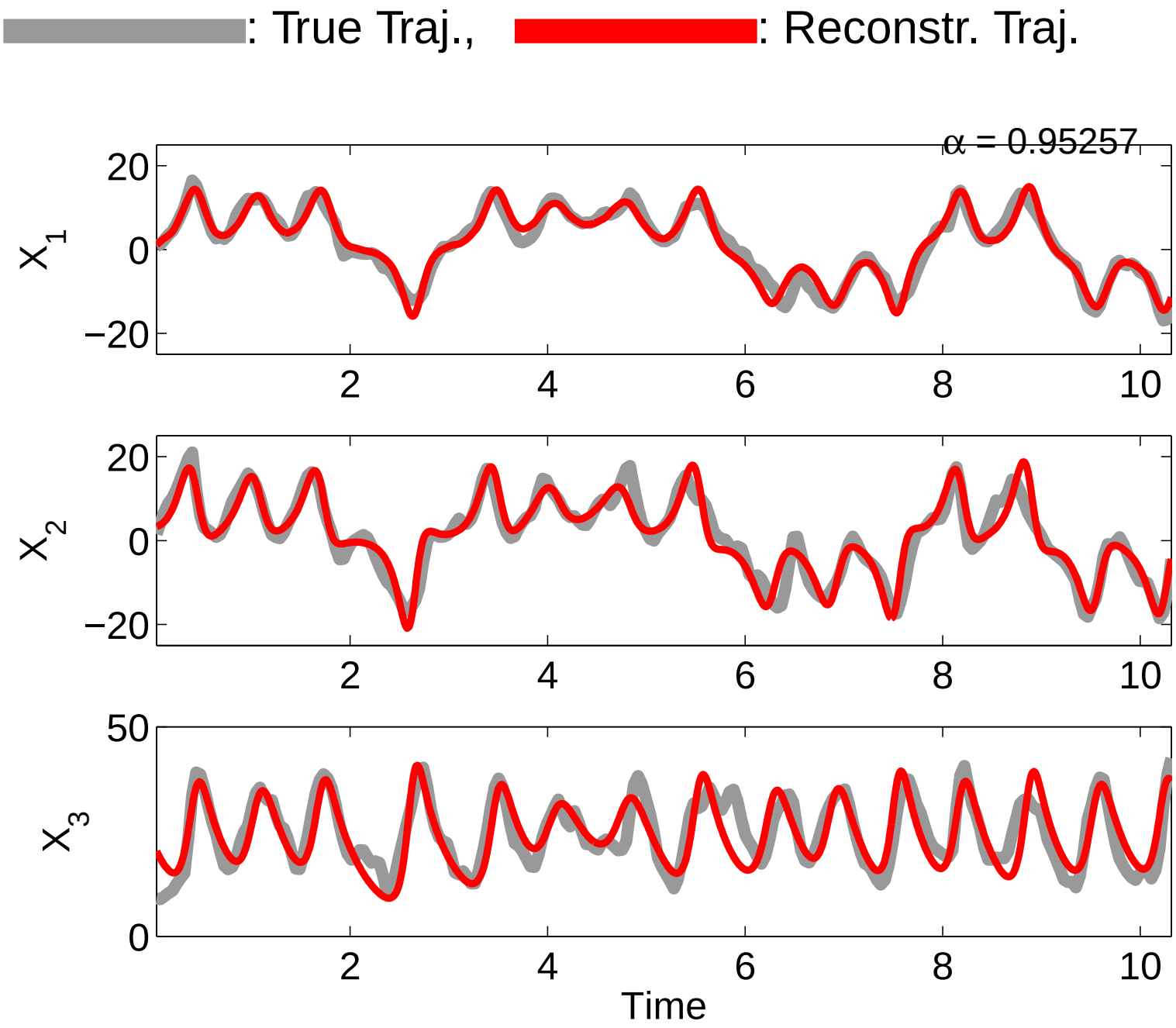

: True Traj., : Reconstr. Traj.
α = 0.1192
X1
X2
X3
20
0
−20
50
0
2
4
6
8
10
Time

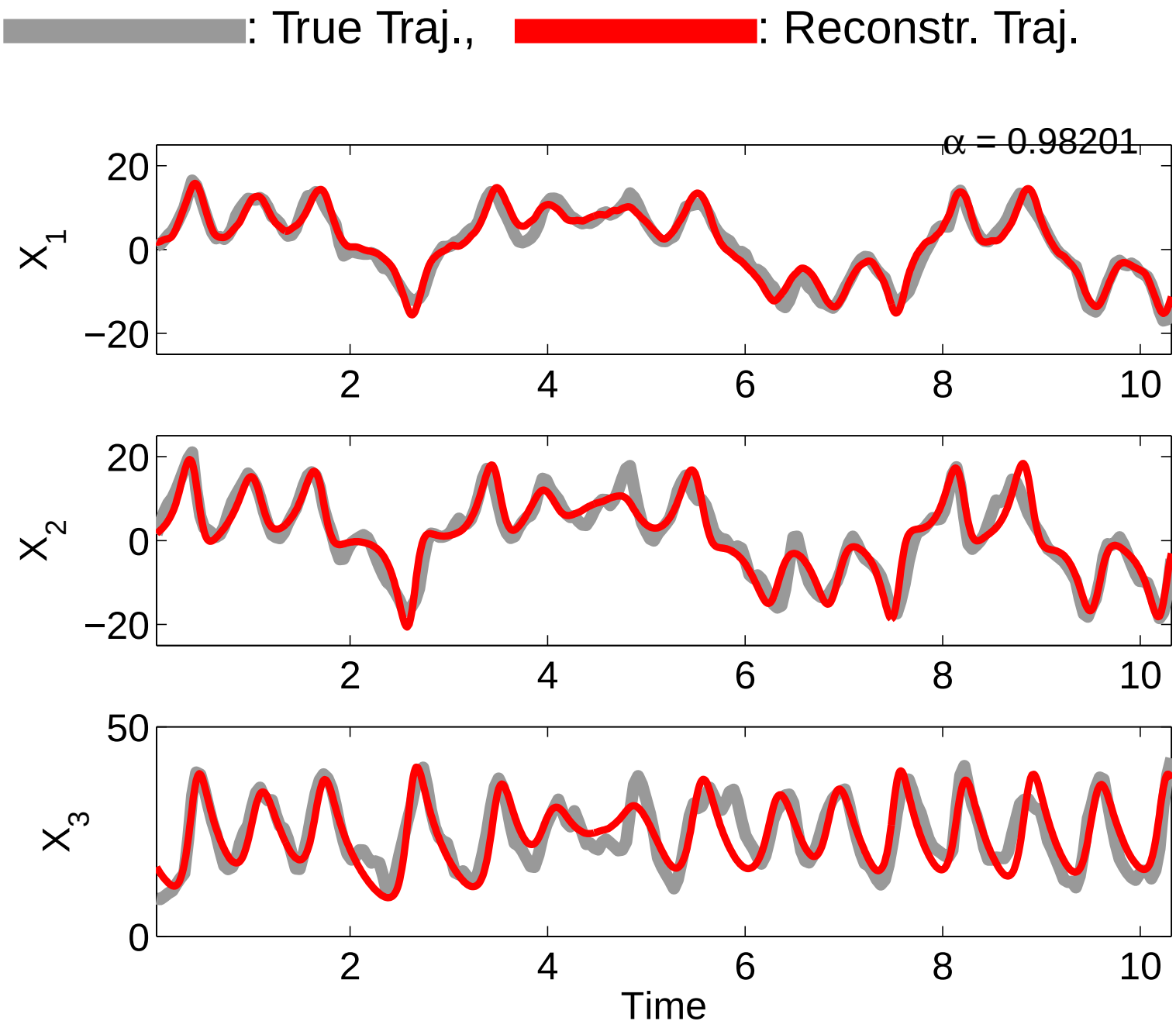

: True Traj., : Reconstr. Traj.
α = 0.26894
X1
X2
X3
20
0
−20
50
0
2
4
6
8
10
Time

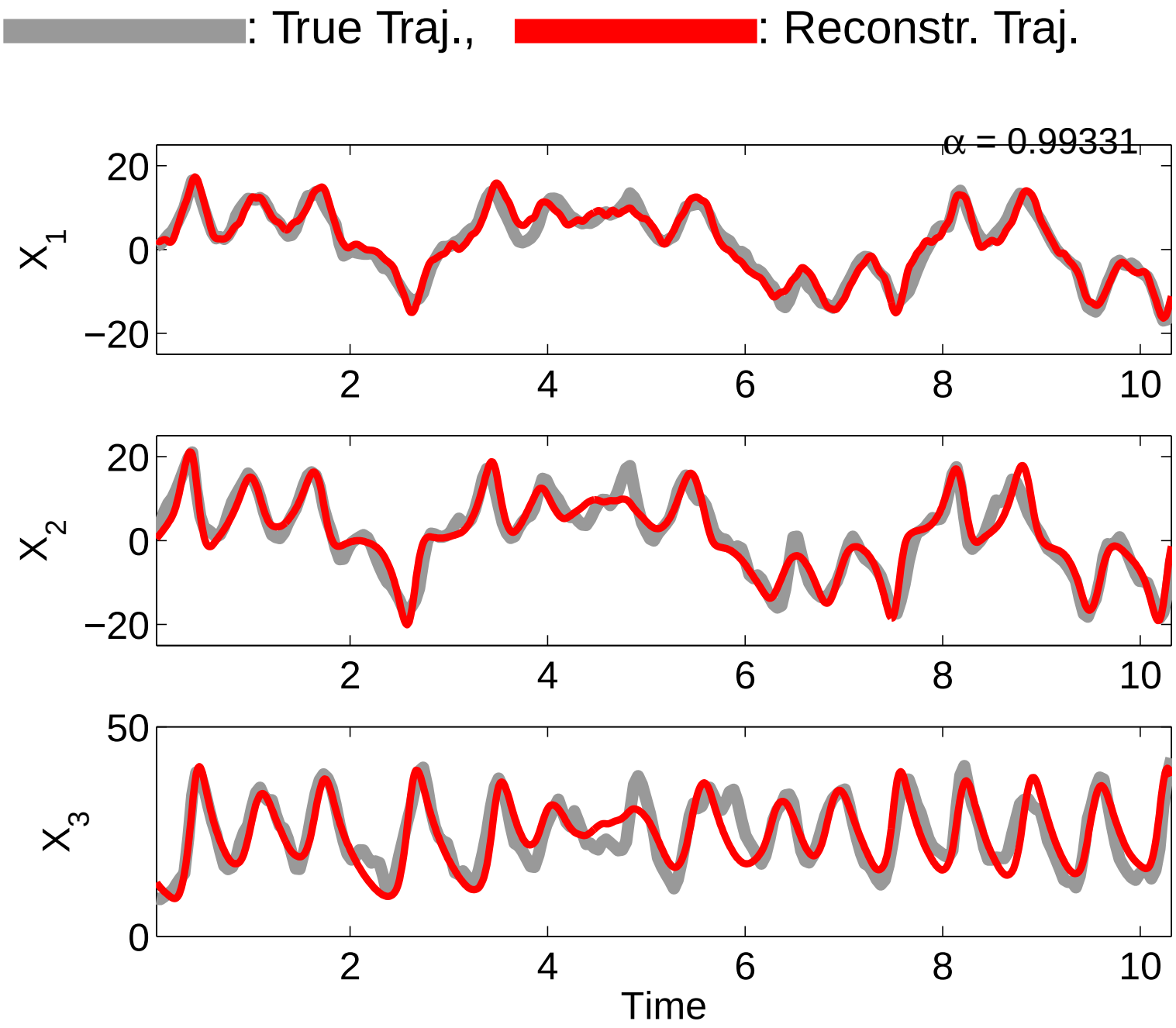

: True Traj., : Reconstr. Traj.
α = 0.5
X1
X2
X3
20
0
−20
50
0
2
4
6
8
10
Time

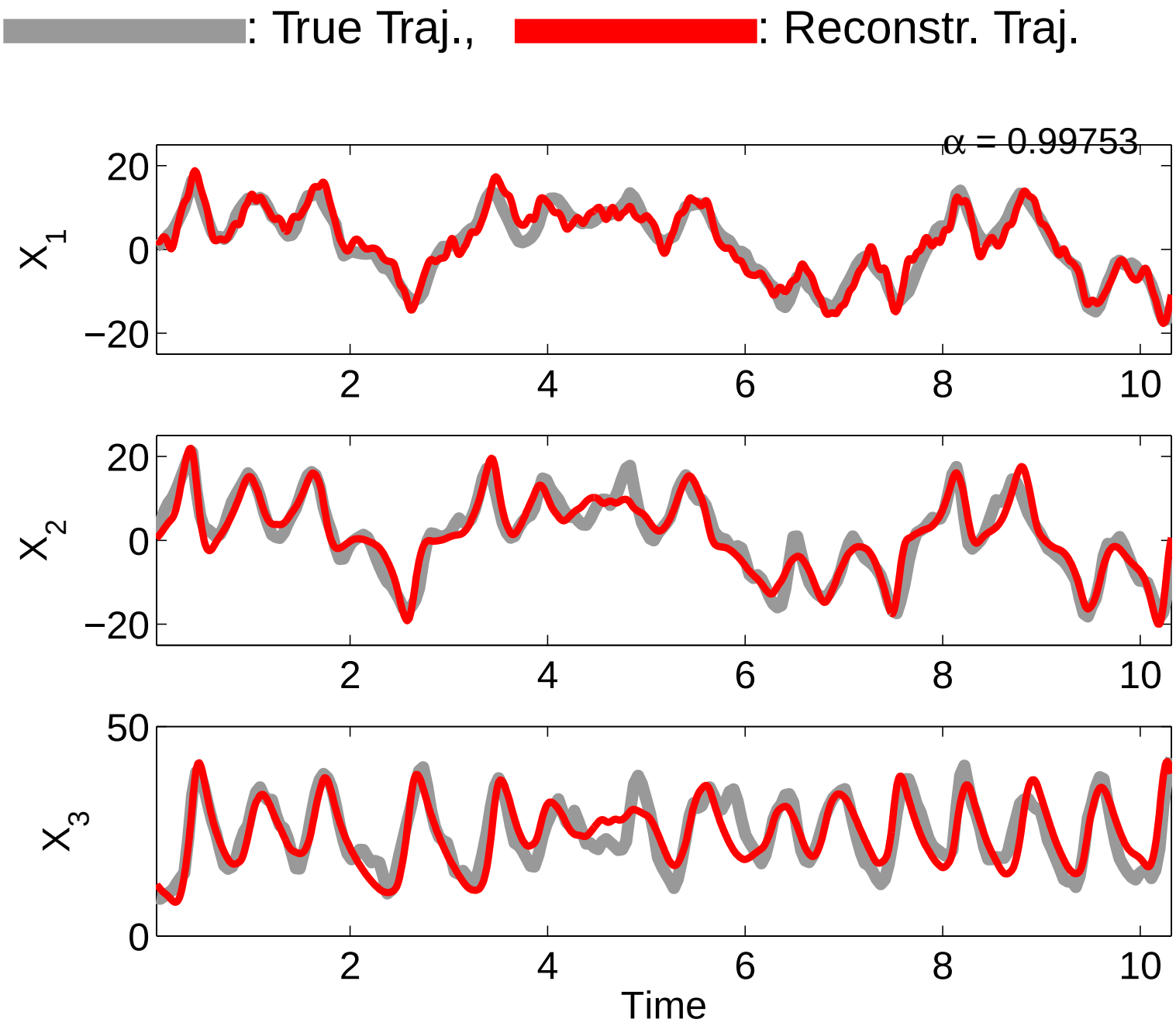

: True Traj.,
: Reconstr. Traj.
α = 0.73106
X1
X2
X3
20
0
−20
50
0
2
4
6
8
10
Time

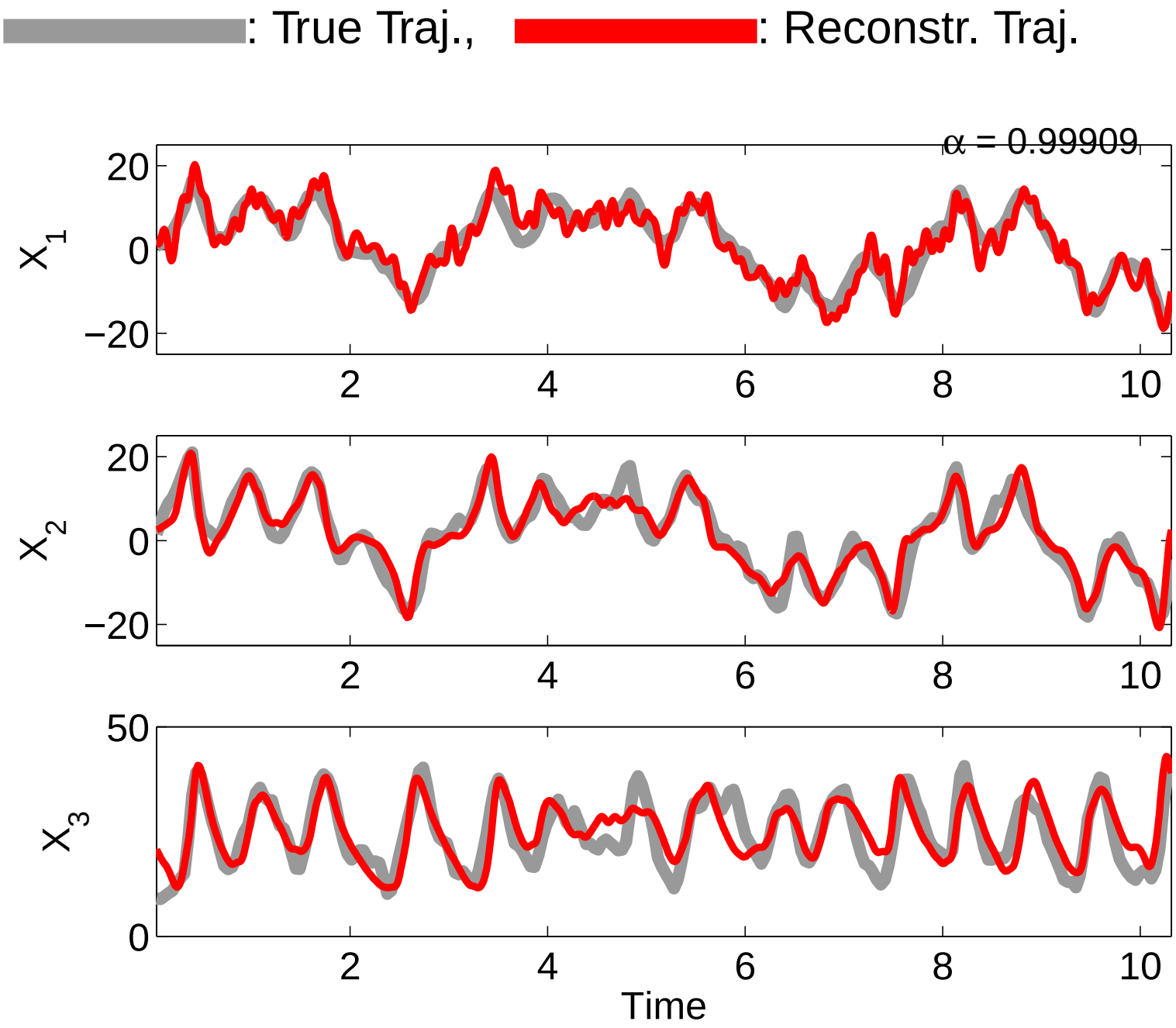

: True Traj., : Reconstr. Traj.
α = 0.8808
X1
X2
X3
20
0
−20
50
0
2
4
6
8
10
Time

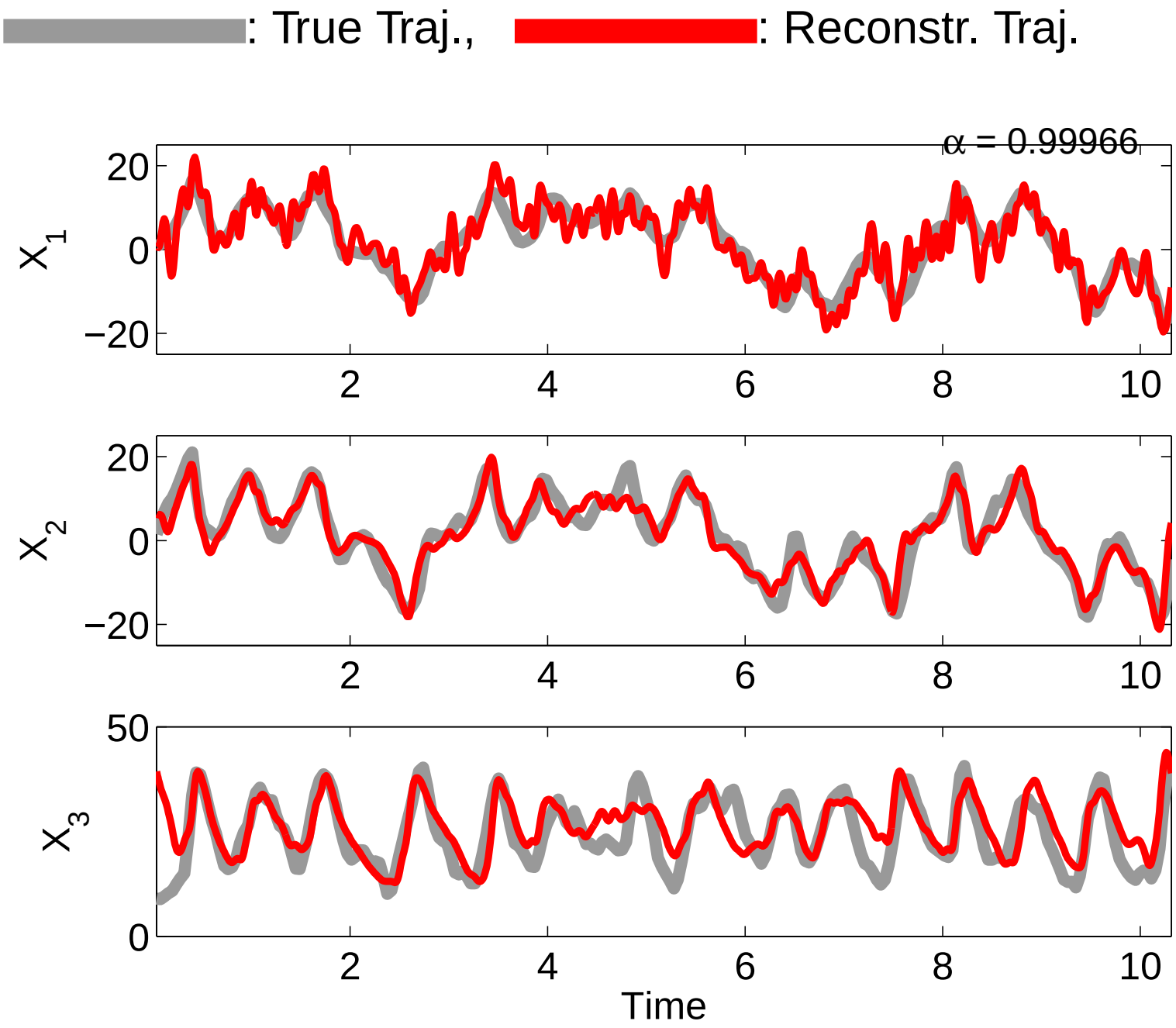

: True Traj., : Reconstr. Traj.
α = 0.95257
X1
X2
X3
20
0
−20
50
0
2
4
6
8
10
Time

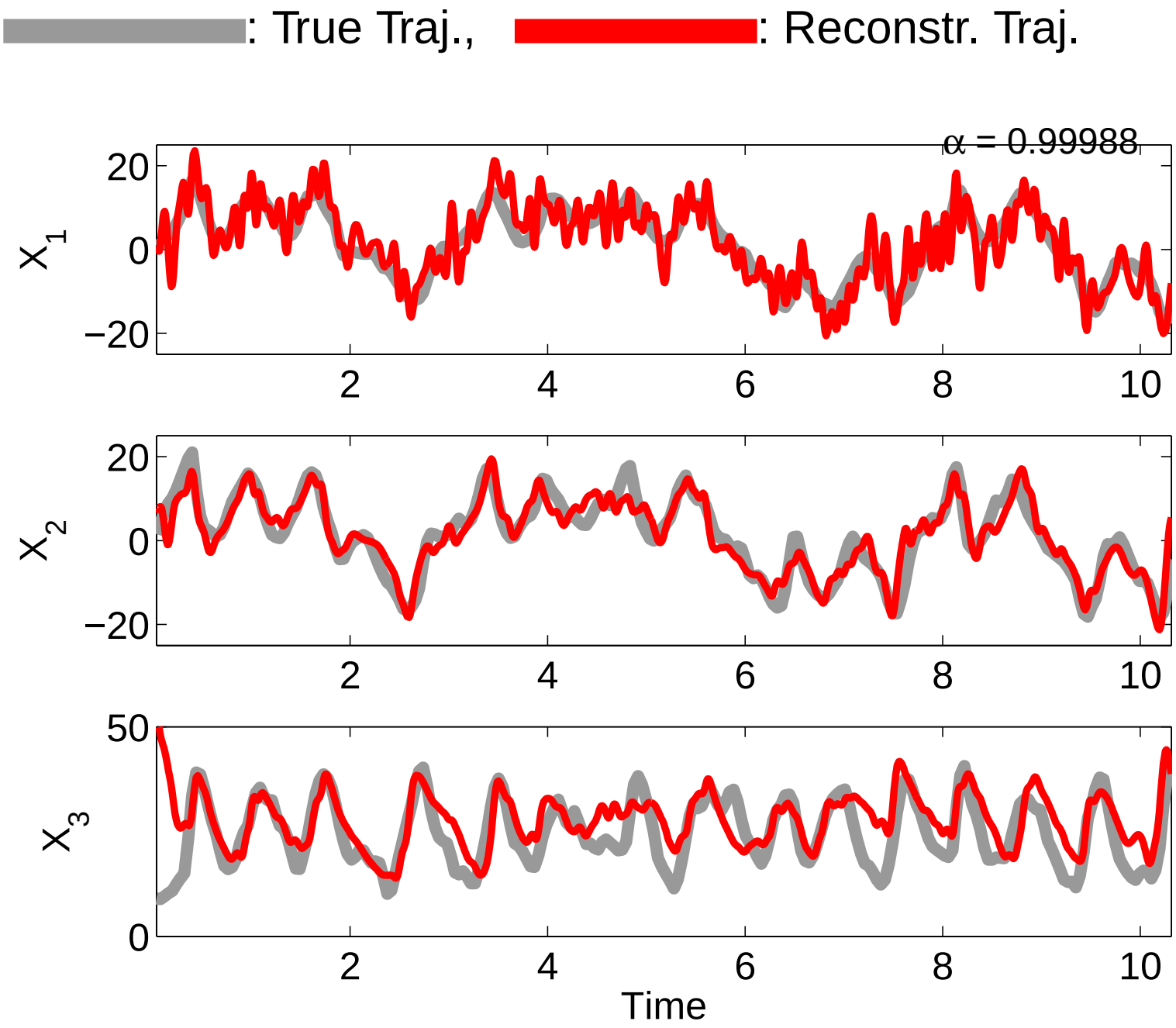

: True Traj., : Reconstr. Traj.
α = 0.98201
X1
X2
X3
20
0
−20
50
0
2
4
6
8
10
Time

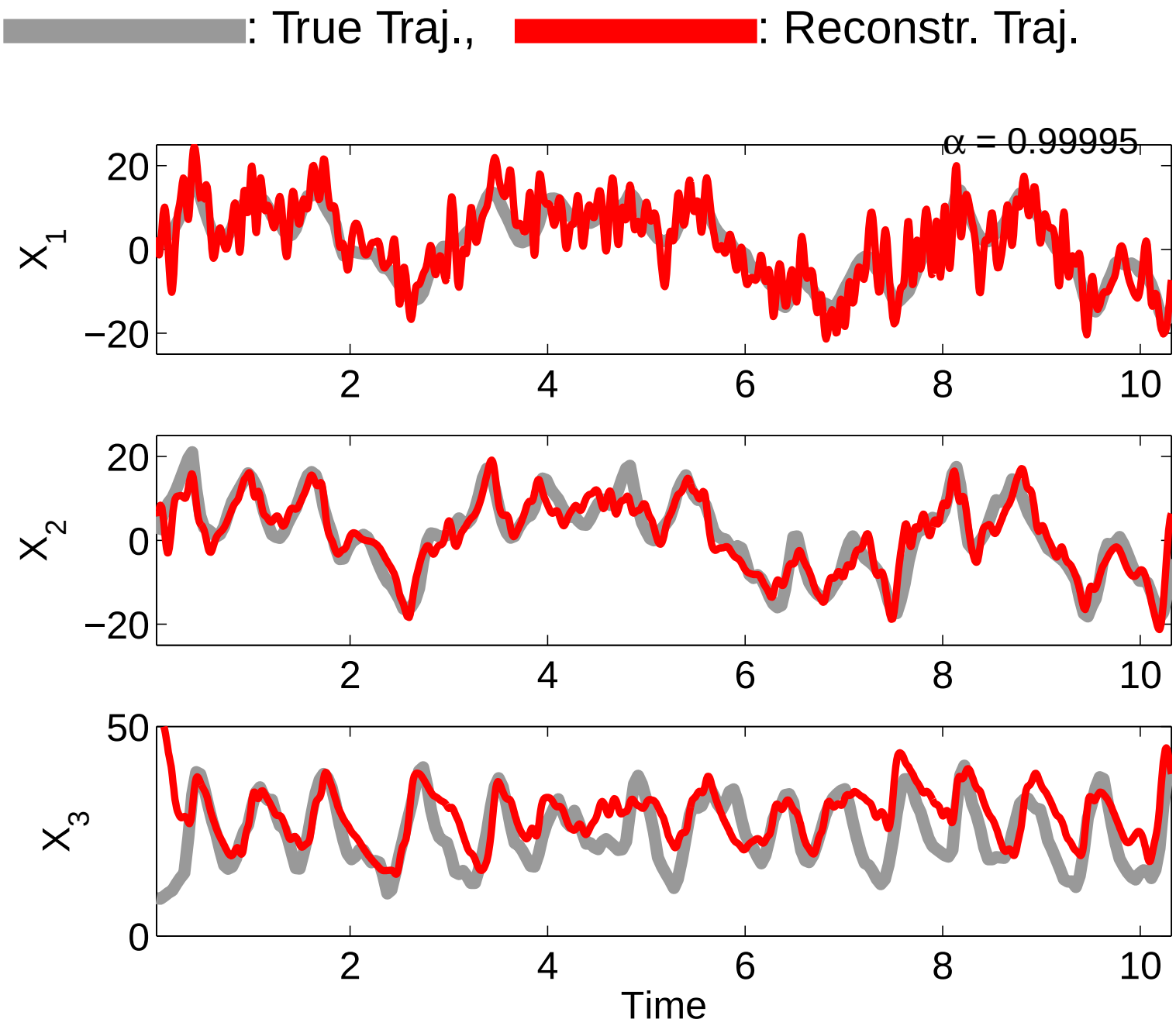

: True Traj.,
: Reconstr. Traj.
α = 0.99331
X1
X2
X3
20
0
−20
50
0
2
4
6
8
10
Time


: True Traj., : Reconstr. Traj.
α = 0.99753
X1
X2
X3
20
0
−20
50
0
2
4
6
8
10
Time

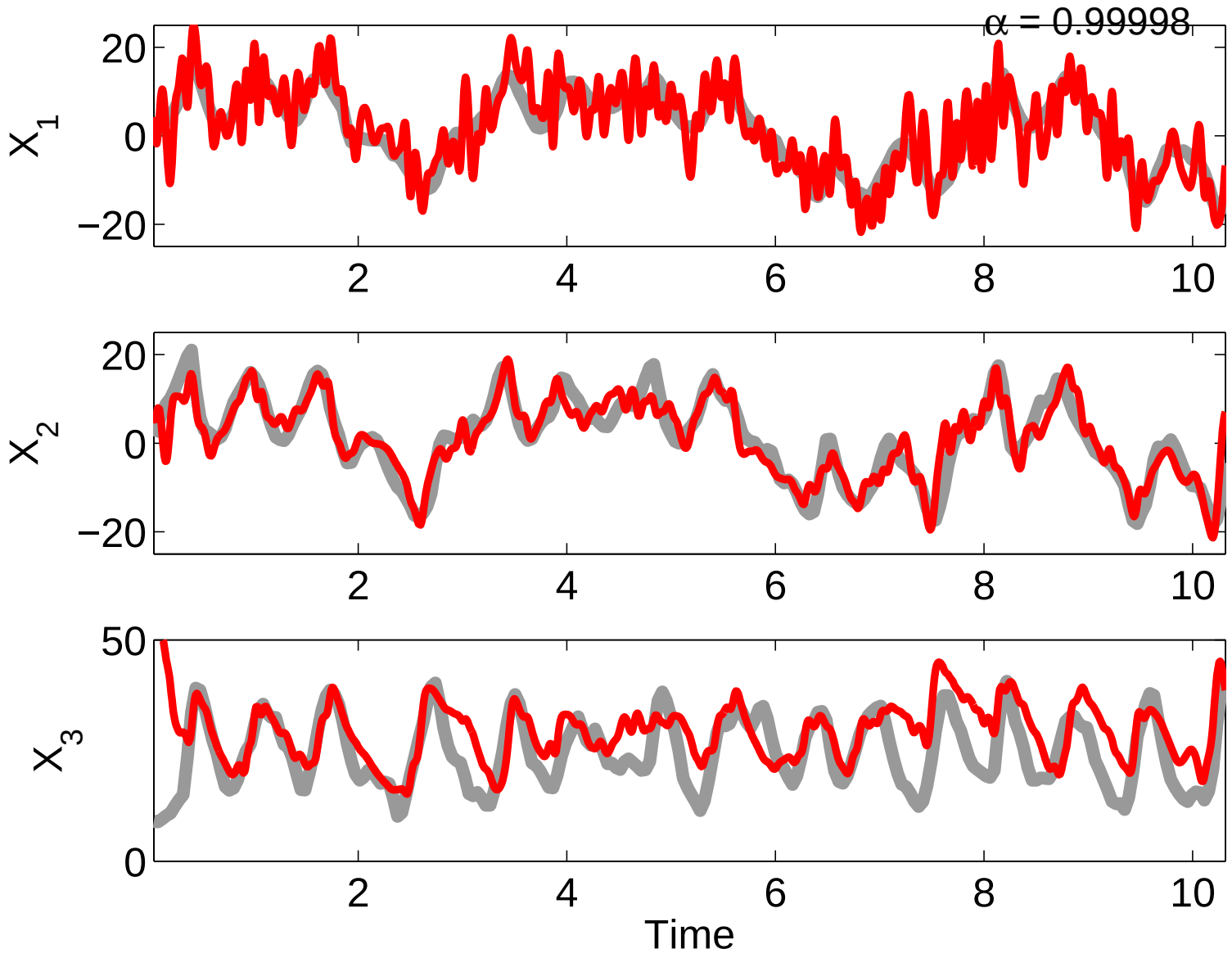

: True Traj., : Reconstr. Traj.
α = 0.99909
X1
X2
X3
20
0
−20
50
0
2
4
6
8
10
Time


: True Traj., : Reconstr. Traj.
α = 0.99966
X1
X2
X3
20
0
−20
50
0
2
4
6
8
10
Time


: True Traj., : Reconstr. Traj.
α = 0.99988
X1
X2
X3
20
0
−20
50
0
2
4
6
8
10
Time


: True Traj., : Reconstr. Traj.
α = 0.99995
X1
X2
X3
20
0
−20
50
0
2
4
6
8
10
Time

: True Traj., : Reconstr. Traj.



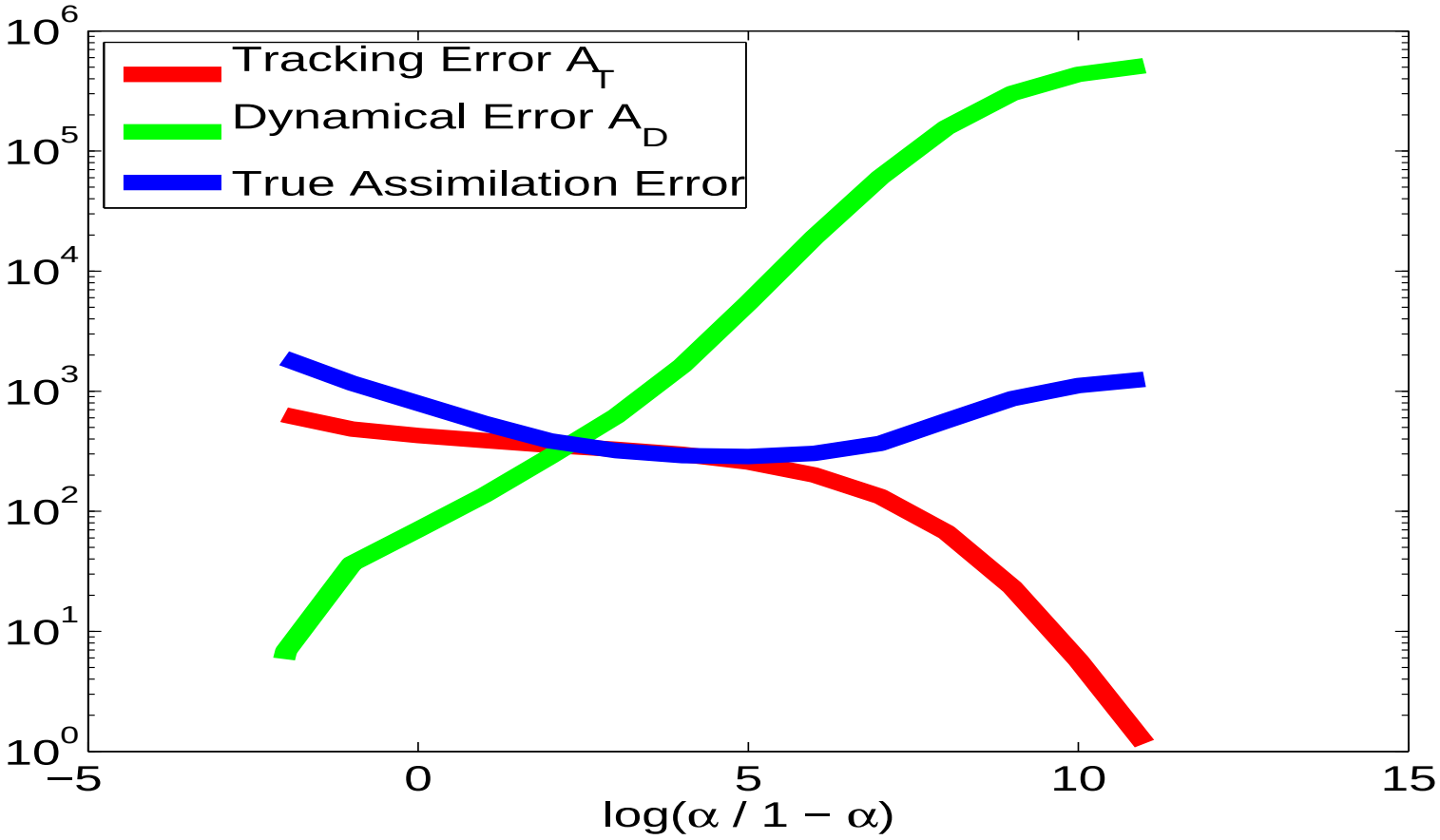
Example: Errors

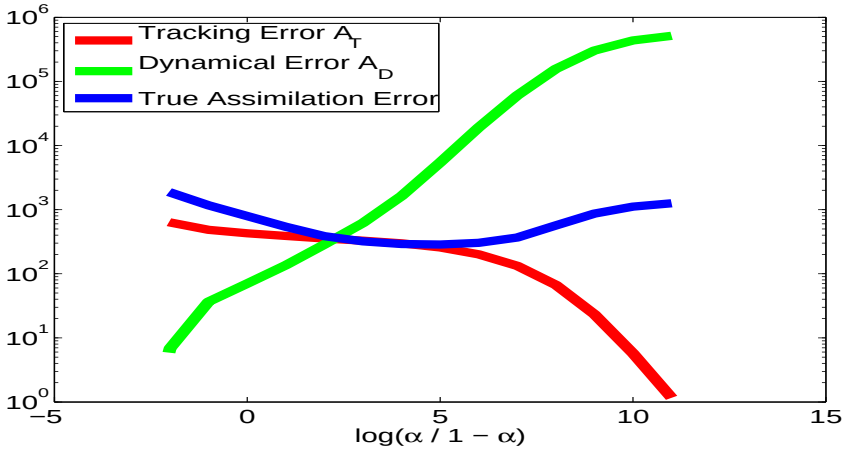
Similar results for Lorenz'63, Lorenz'96, and the Barotropic Vorticity Equation

: Tracking Error $A_T = \int (\eta_t - y_t)^2 \, dt$

: Dynamical Error $A_D = \int u_t^2 \, dt$,

: Assimilation Error $A_S = \int (x_t - X_t)^2 \, dt$.

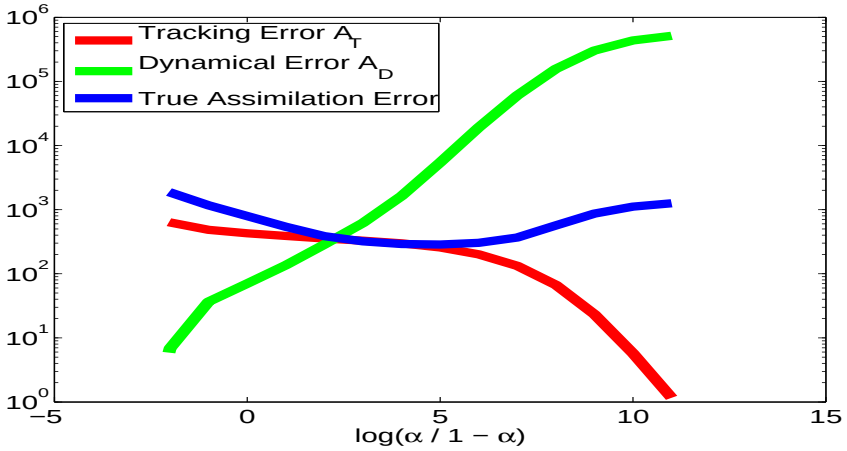




Question

How do we choose α (or any other regularisation parameter)?

A small assimilation error is not an operational criterion since in reality, there is no X_t!
We need a proxy for the assimilation error A_S.



Question

How do we choose α (or any other regularisation parameter)?

A small assimilation error is not an operational criterion since in reality, there is no X_t!
We need a proxy for the assimilation error A_S.

Alternative: the out–of–sample error

Assumption:

$$\underbrace{\eta_t}_{\text{Observation}} = \underbrace{\zeta_t}_{\text{Real Signal}} + \underbrace{r_t}_{\text{Noise}}$$

Imagine we could generate *new* observations

$$\eta_t' = \zeta_t + r_t'$$

with r_t and r_t' having the same statistical properties but being uncorrelated.

Alternative: the out–of–sample error

Assumption:

$$\underbrace{\eta_t}_{\text{Observation}} = \underbrace{\zeta_t}_{\text{Real Signal}} + \underbrace{r_t}_{\text{Noise}}$$

Imagine we could generate *new* observations

$$\eta_t' = \zeta_t + r_t'$$

with r_t and r_t' having the same statistical properties but being uncorrelated.

The out–of–sample error

Motivation

Robustness of data assimilation with respect to the noise.
Output y_t should be close to η'_t as well.

The *Out–Of–Sample* error

$$E_{\text{oos}} = \int \mathbb{E}(\eta'_t - y_t)^2 \mathrm{d}t.$$

expected to behave like the assimilation error of the observed degrees of freedom.

The out–of–sample error

Motivation

Robustness of data assimilation with respect to the noise.
Output y_t should be close to η'_t as well.

The *Out–Of–Sample* error

$$E_{\text{oos}} = \int \mathbb{E}(\eta'_t - y_t)^2 \mathrm{d}t.$$

expected to behave like the assimilation error of the observed degrees of freedom.

The out–of–sample error

Motivation

Robustness of data assimilation with respect to the noise.
Output y_t should be close to η'_t as well.

The *Out–Of–Sample* error

$$E_{\text{oos}} = \int \mathbb{E}(\eta'_t - y_t)^2 \mathrm{d}t.$$

expected to behave like the assimilation error of the observed degrees of freedom.

The sensitivity

Proposition:

$$E_{\text{oos}} = \int \mathbb{E}(\eta'_t - y_t)^2 \mathrm{d}t = \underbrace{\int \mathbb{E}(\eta_t - y_t)^2 \mathrm{d}t}_{\text{Tracking Err. } \mathbb{E}(A_T)} + 2S$$

with

$$S = \int \mathrm{Cov}\,[y_t, r_t]\,\mathrm{d}t$$

the *Sensitivity*.

(Alternatively, we could call S/ρ^2 the sensitivity.)

The sensitivity

Proposition:

$$E_{\text{oos}} = \int \mathbb{E}(\eta'_t - y_t)^2 \mathrm{d}t = \underbrace{\int \mathbb{E}(\eta_t - y_t)^2 \mathrm{d}t}_{\text{Tracking Err. } \mathbb{E}(A_T)} + 2S$$

with

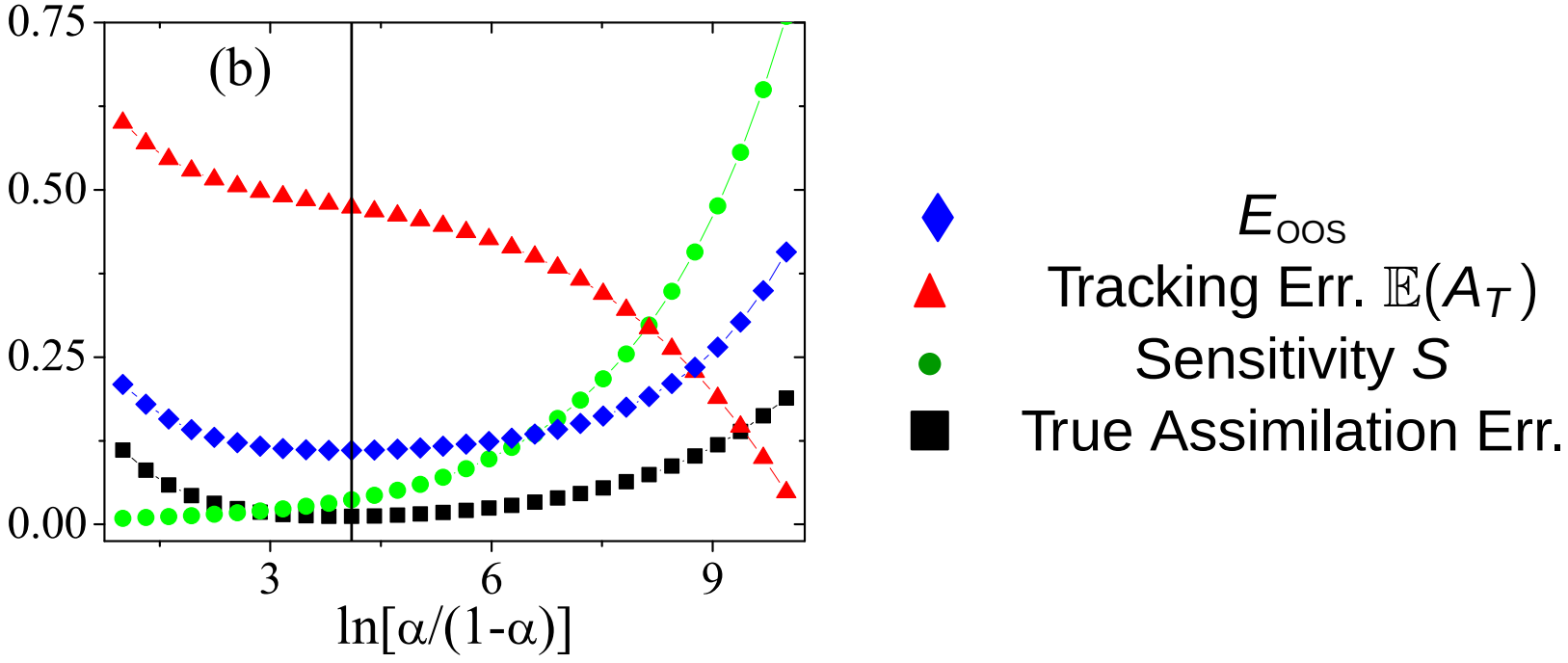
$$S = \int \mathrm{Cov}\,[y_t, r_t]\,\mathrm{d}t$$

the *Sensitivity*.
(Alternatively, we could call S/ρ^2 the sensitivity.)

Variational Data Assimilation

Results for Lorenz'96 [B., Szendro(2011)]

$$E_{\text{OOS}} = \text{Tracking Err.} + 2\text{Sensitivity}$$



Still needed:

A good guess of the observational noise.

We have translated the problem to a new one:
Make a decision what part of the observations you want to model, and what part you don't want to model.
In the presented examples, this was done by comparing the spectrum of the observations with that of the model.

Still needed:

A good guess of the observational noise.

We have translated the problem to a new one:
Make a decision what part of the observations you want to model, and what part you don't want to model.
In the presented examples, this was done by comparing the spectrum of the observations with that of the model.

Still needed:

A good guess of the observational noise.

We have translated the problem to a new one:
Make a decision what part of the observations you want to model, and what part you don't want to model.
In the presented examples, this was done by comparing the spectrum of the observations with that of the model.

Conclusions

- There is a nontrivial trade–off between tracking error and dynamical error (unless the model is perfect).
- A minimum E_{OOS} provides a self–consistent criterion to set the weighting α, or more generally the sensitivity of data assimilation algorithms.
- Approx. to the Out–Of–Sample error (the sensitivity) give reasonable results in the studied examples.

Conclusions

- There is a nontrivial trade–off between tracking error and dynamical error (unless the model is perfect).
- A minimum E_{OOS} provides a self–consistent criterion to set the weighting α, or more generally the sensitivity of data assimilation algorithms.
- Approx. to the Out–Of–Sample error (the sensitivity) give reasonable results in the studied examples.

Conclusions

- There is a nontrivial trade–off between tracking error and dynamical error (unless the model is perfect).
- A minimum E_{OOS} provides a self–consistent criterion to set the weighting α, or more generally the sensitivity of data assimilation algorithms.
- Approx. to the Out–Of–Sample error (the sensitivity) give reasonable results in the studied examples.

Questions

Alternative approaches to determine regularisation parameters should be applicable, too. How do they compare?

For Further Reading I

J.B. and Ivan G. Szendro.
Sensitivity and out–of–sample error in continuous time data assimilation.
Quarterly Journal of the Royal Meteorological Society, 2011 (accepted).

O. Zeitouni and A. Dembo.
A maximum a posteriori estimator for trajectories of diffusion processes.
Stochastics, 20(3):221, 1987.