Data Assimilation as Synchronization of Model Forecasts to Observations

Christopher M. Danforth
Department of Mathematics and Statistics
University of Vermont
July 29, 2008

Thanks to ...

University of Maryland

Eugenia Kalnay, Atmospheric Science
James A. Yorke, Math and Physics
Robert F. Cahalan, NASA GFSC
Takemasa Miyoshi, JMO
Chaos group (Hunt, Kalnay, Kostelich, Ott, Patil, Sauer, Szunyogh, Yorke)

Bates College

Chip Ross, Math
Mark Semon, Physics
George Ruff, Physics

University of Vermont

Darren Hitt, Mechanical Engineering
Floyd Vilmont, Lab

Students

Kameron Harris, Undergrad
Nicholas Allgaier, Graduate
El Hassan Ridouane, Postdoc

Crossing the Atlantic

Christiaan Huygens (1629-1695)

Christiaan Huygens (1629-1695)

Huygens Pendulum Clocks (1660)

Huygens Pendulum Clocks (1660)

British Royal Society: "Occasion was taken here by some of the members to doubt the exactness of the motion of these watches at sea, since so slight and almost insensible motion was able to cause an alteration in their going. "

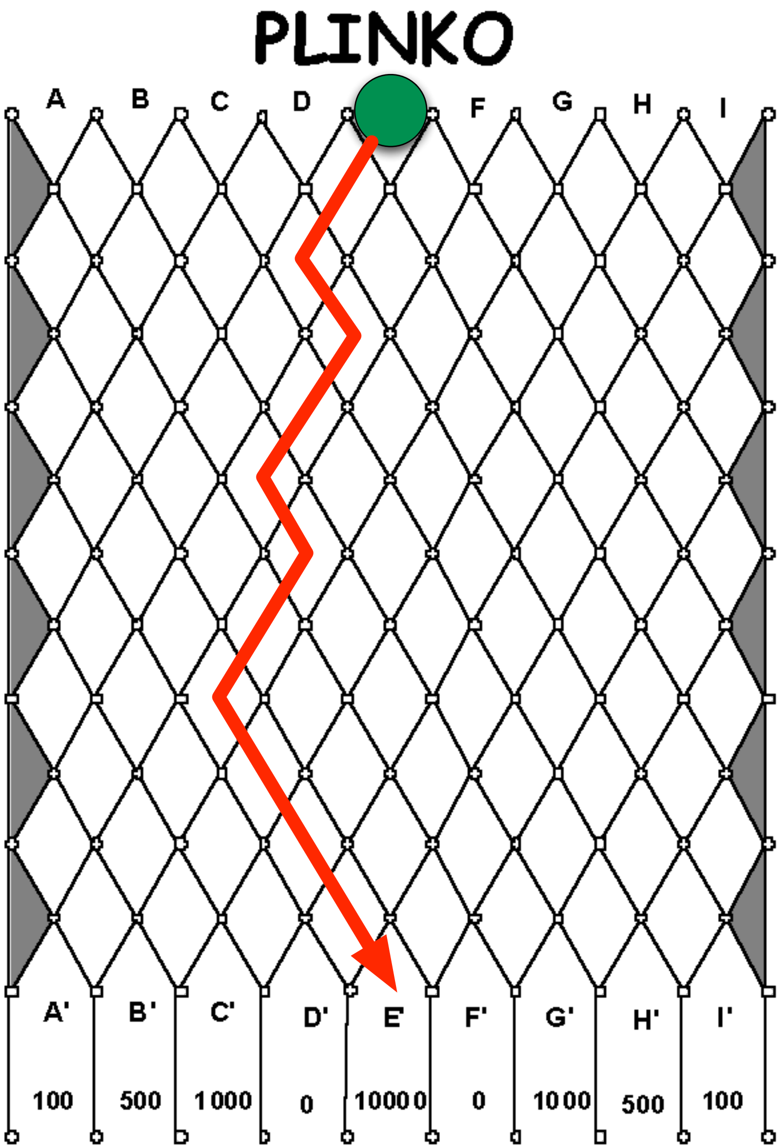
Nonlinear Systems

Chaos: 'When the present determines the future, but the approximate present does not approximately determine the future.' –Lorenz

Nonlinear Systems

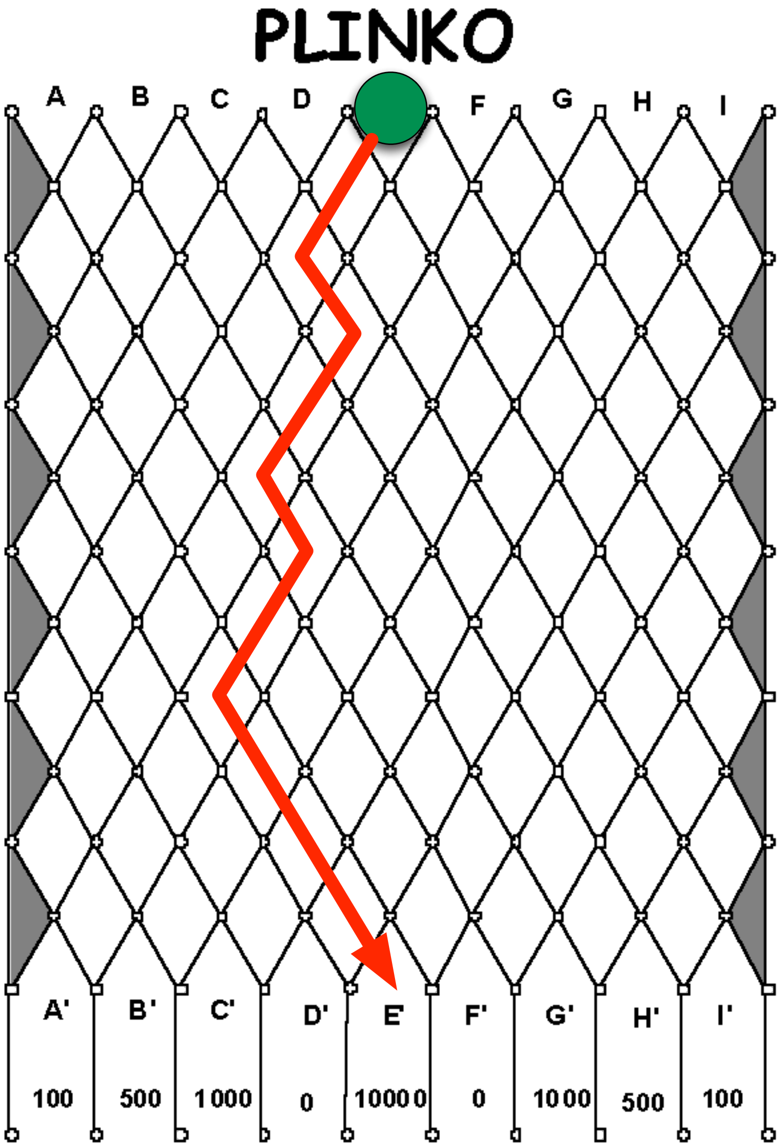
Chaos: ‘When the present determines the future, but the approximate present does not approximately determine the future.’ –Lorenz

This Game is Completely Deterministic

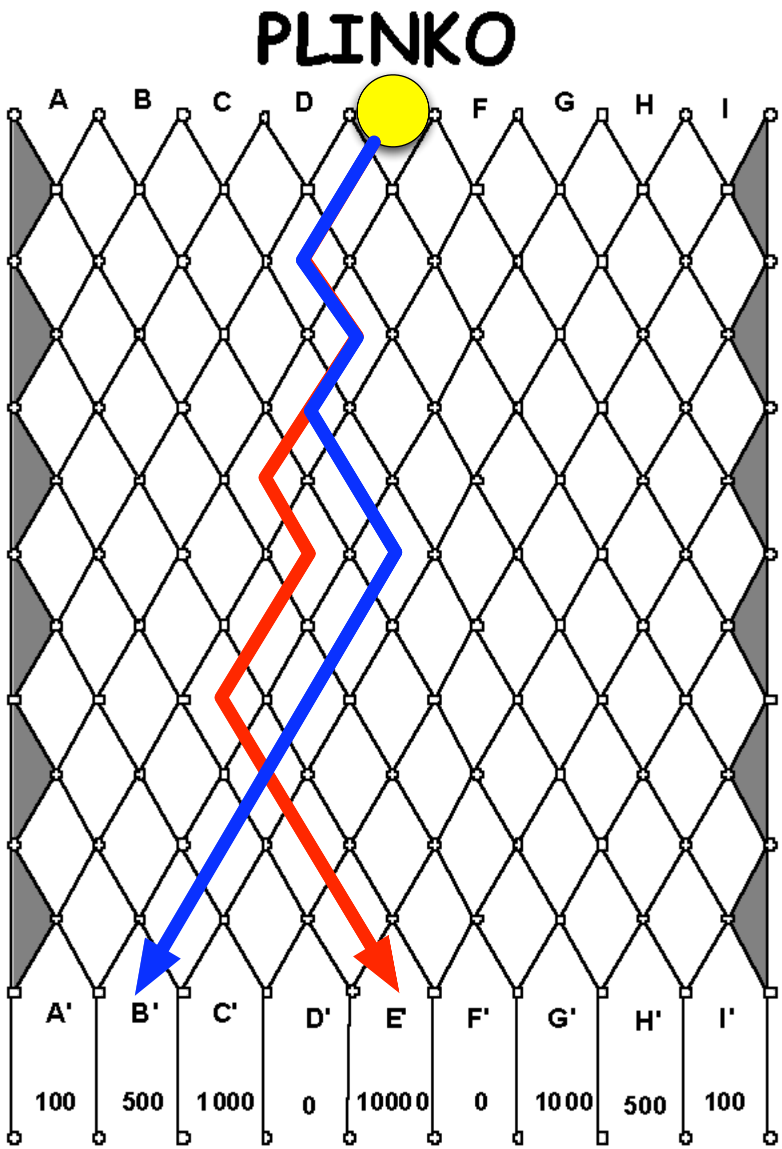


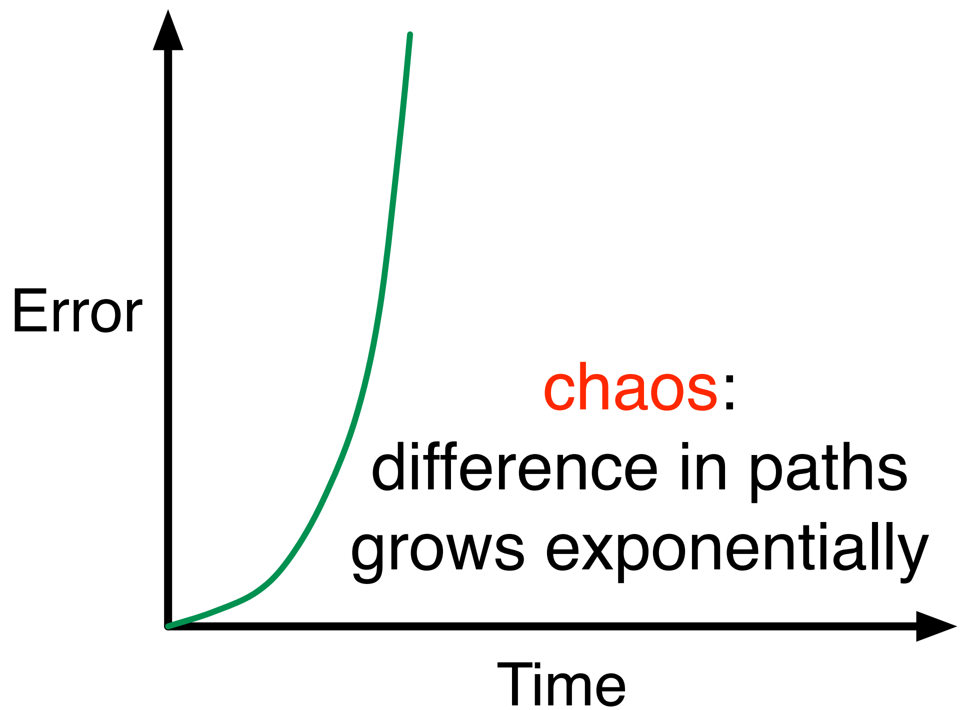
This Game is Completely Deterministic

"You can't fool television viewers with dancing girls and flashing lights." –Bob Barker



Linear vs. Nonlinear





Why is Chaos a Problem?

Double Pendulum

- 4 model variables (position and velocity of each arm)
- model approximates mother nature's rules

(Loading Movie)

Why is Chaos a Problem?

Double Pendulum

- 4 model variables (position and velocity of each arm)
- model approximates mother nature's rules
- prediction time w/supercomputer?

Why is Chaos a Problem?

Double Pendulum

- 4 model variables (position and velocity of each arm)
- model approximates mother nature's rules
- prediction time w/supercomputer $<$ 10 seconds!

Why is Chaos a Problem?

Double Pendulum

- 4 model variables (position and velocity of each arm)
- model approximates mother nature's rules
- prediction time w/supercomputer $<$ 10 seconds!

Earth's atmosphere

- $1,000,000,000$ model variables (need values for all 1 billion!)
  - 7 values per location (3D wind, temp, pressure, humidity, ozone)
  - 1,500,000 locations on surface
  - 100 vertical layers up to the edge of space

Why is Chaos a Problem?

Double Pendulum

- 4 model variables (position and velocity of each arm)
- model approximates mother nature's rules
- prediction time w/supercomputer $<$ 10 seconds!

Earth's atmosphere

- $1,000,000,000$ model variables (need values for all 1 billion!)
  - 7 values per location (3D wind, temp, pressure, humidity, ozone)
  - 1,500,000 locations on surface
  - 100 vertical layers up to the edge of space
- prediction time w/supercomputer ?

A Mathematician's Research Goal:

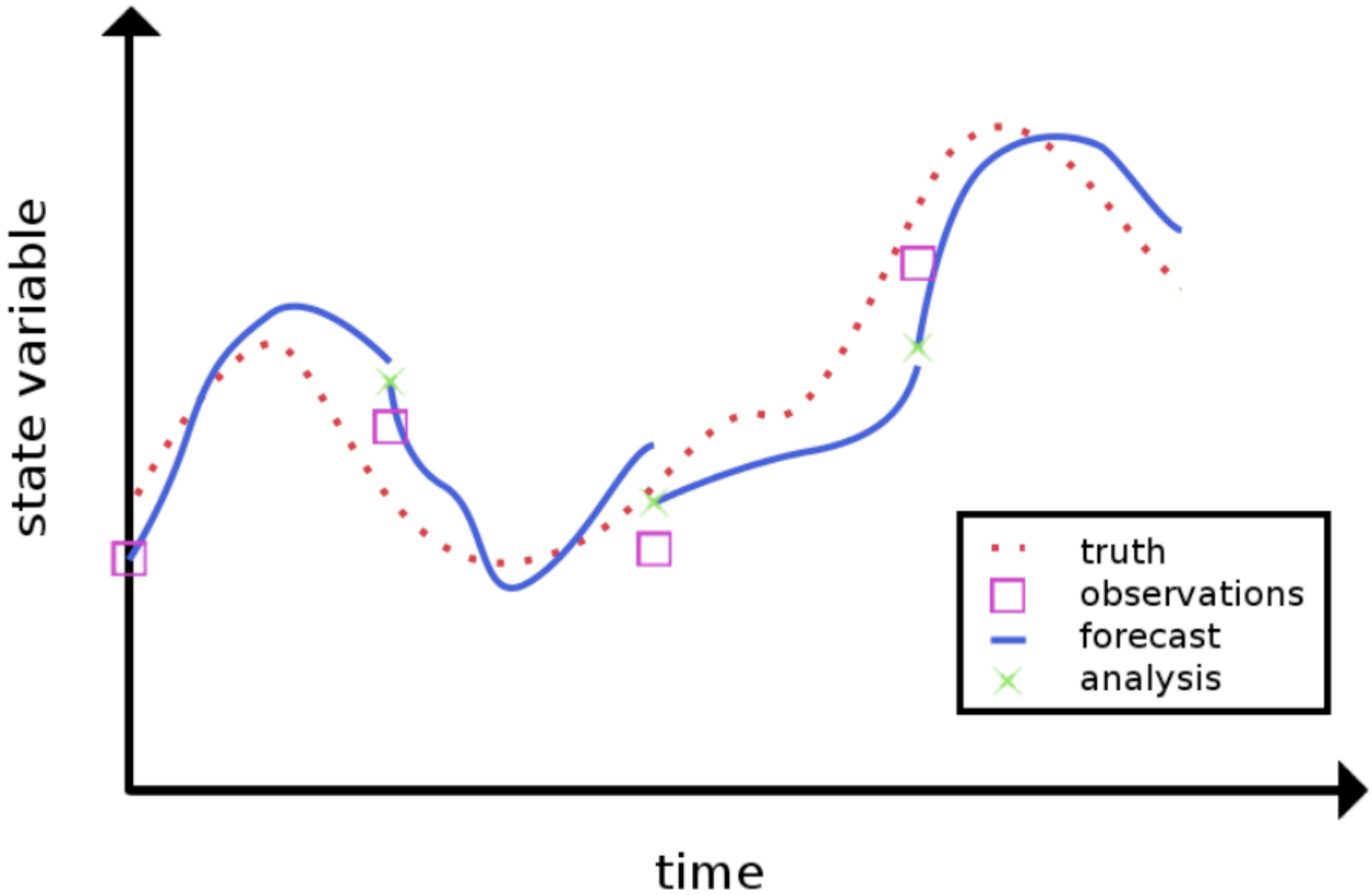
Generate better forecasts *without* improving the initial conditions *or* the model physics

A Mathematician's Research Goal:

Generate better forecasts *without* improving the initial conditions *or* the model physics

by altering the method by which predictions are generated.

Data Assimilation Cartoon

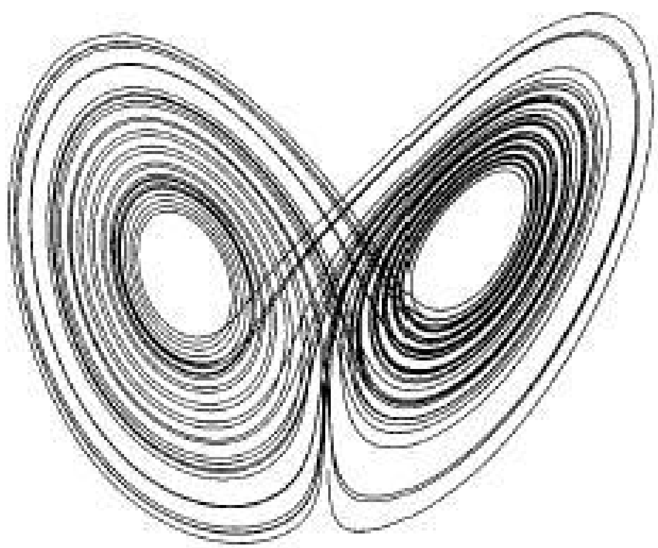


Three Experiments

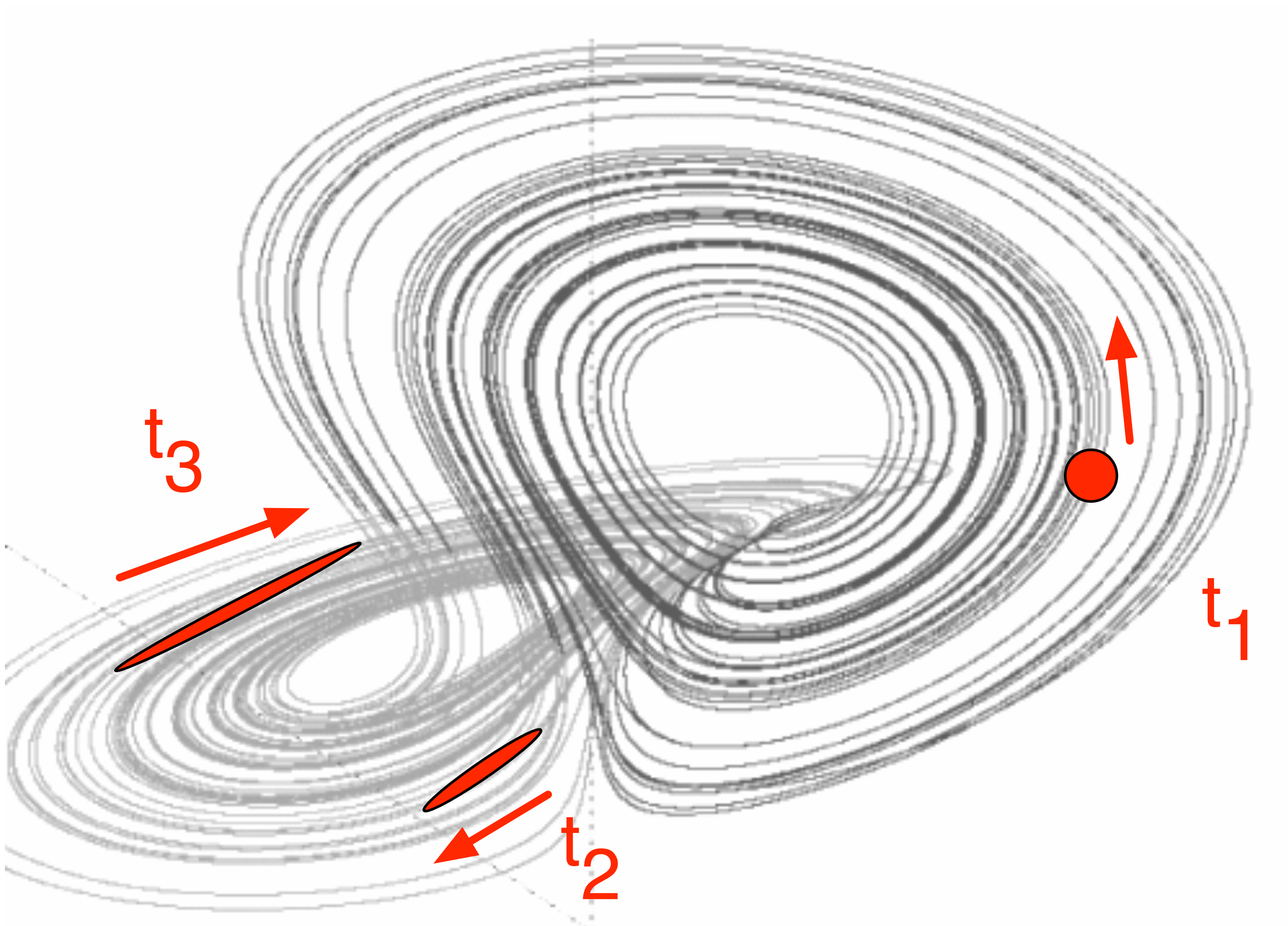
- An experimental analog to Lorenz's 1963 model
- Stalking observations with a numerical trajectory
- *Online* empirical correction of model error

Lorenz (1963)

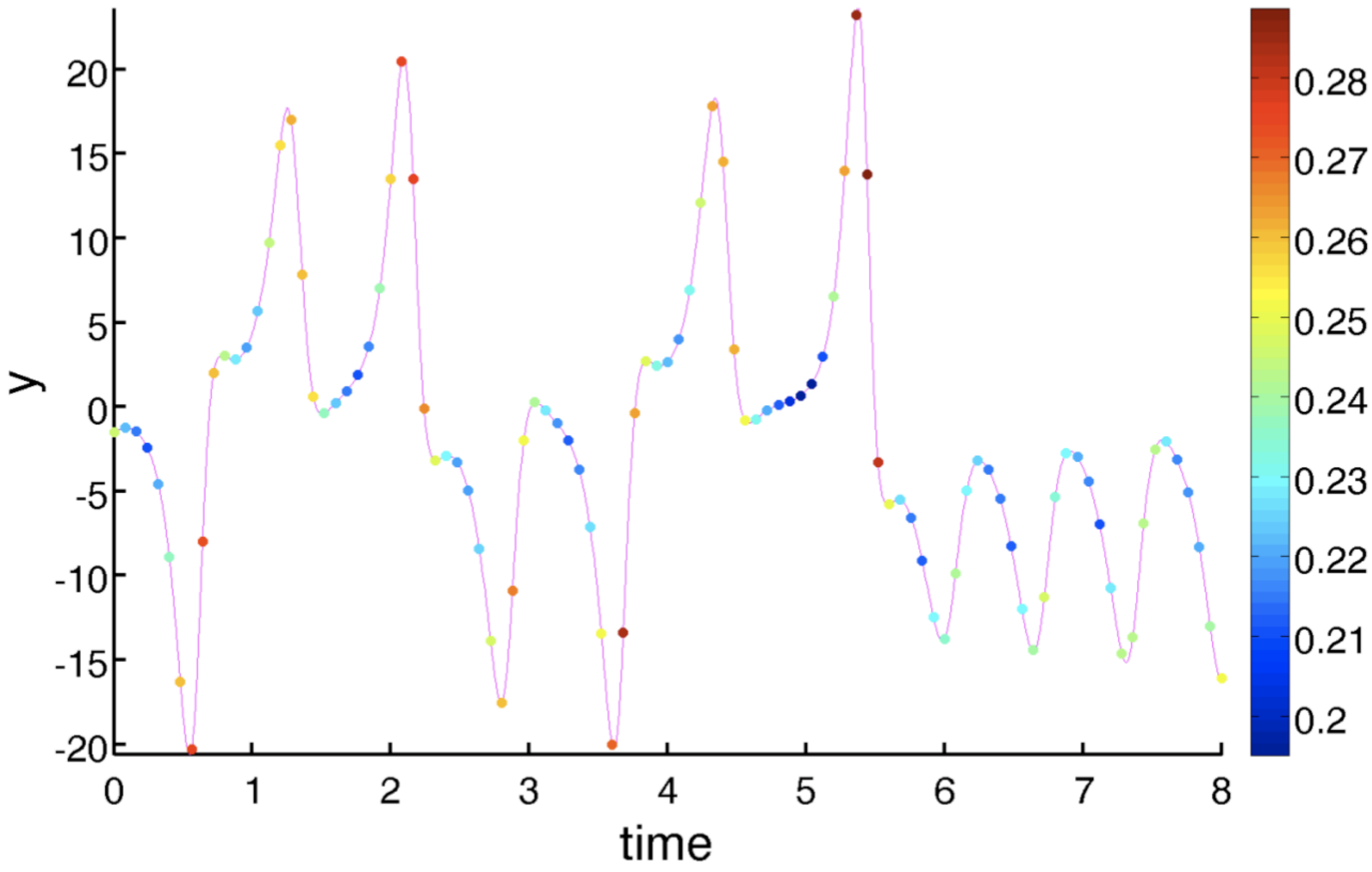
$$\frac{dx}{dt} = \sigma(y - x)$$
$$\frac{dy}{dt} = \rho x - y - xz$$
$$\frac{dz}{dt} = xy - \beta z$$



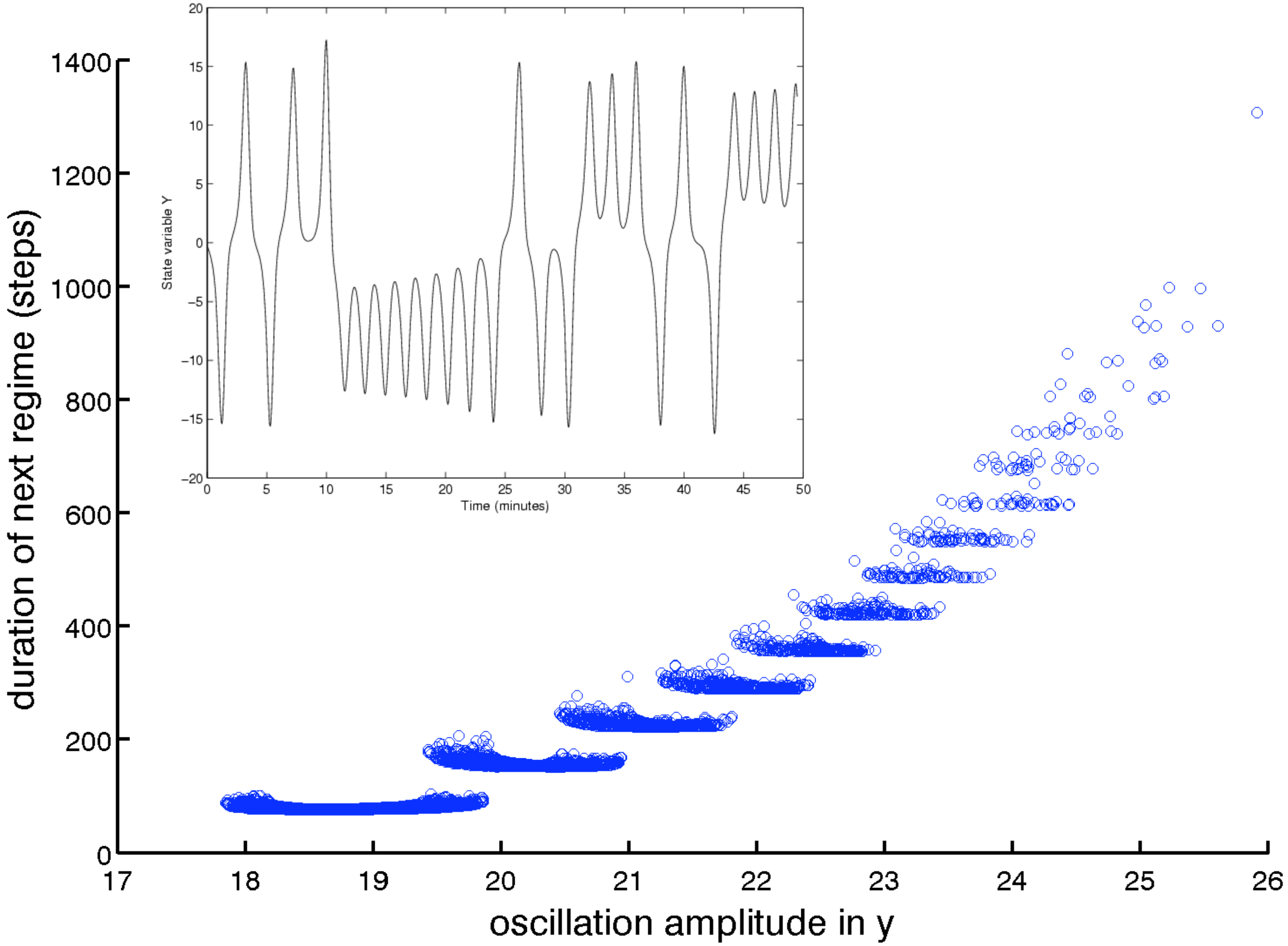
Breeding, Growth of Perturbations (Toth and Kalnay 1993)



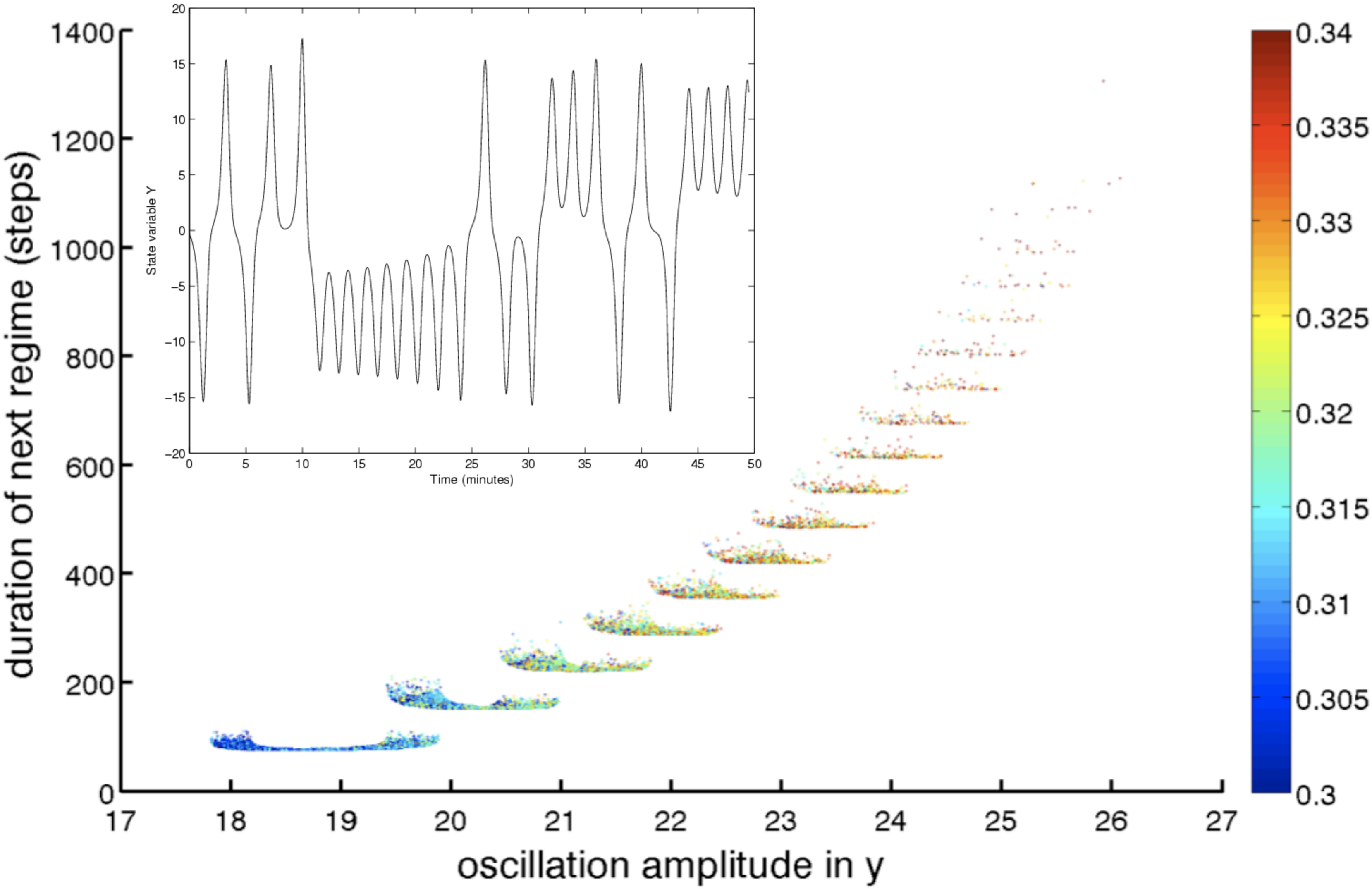
Breeding, Growth rate of Perturbations (Yang et al. 2006)



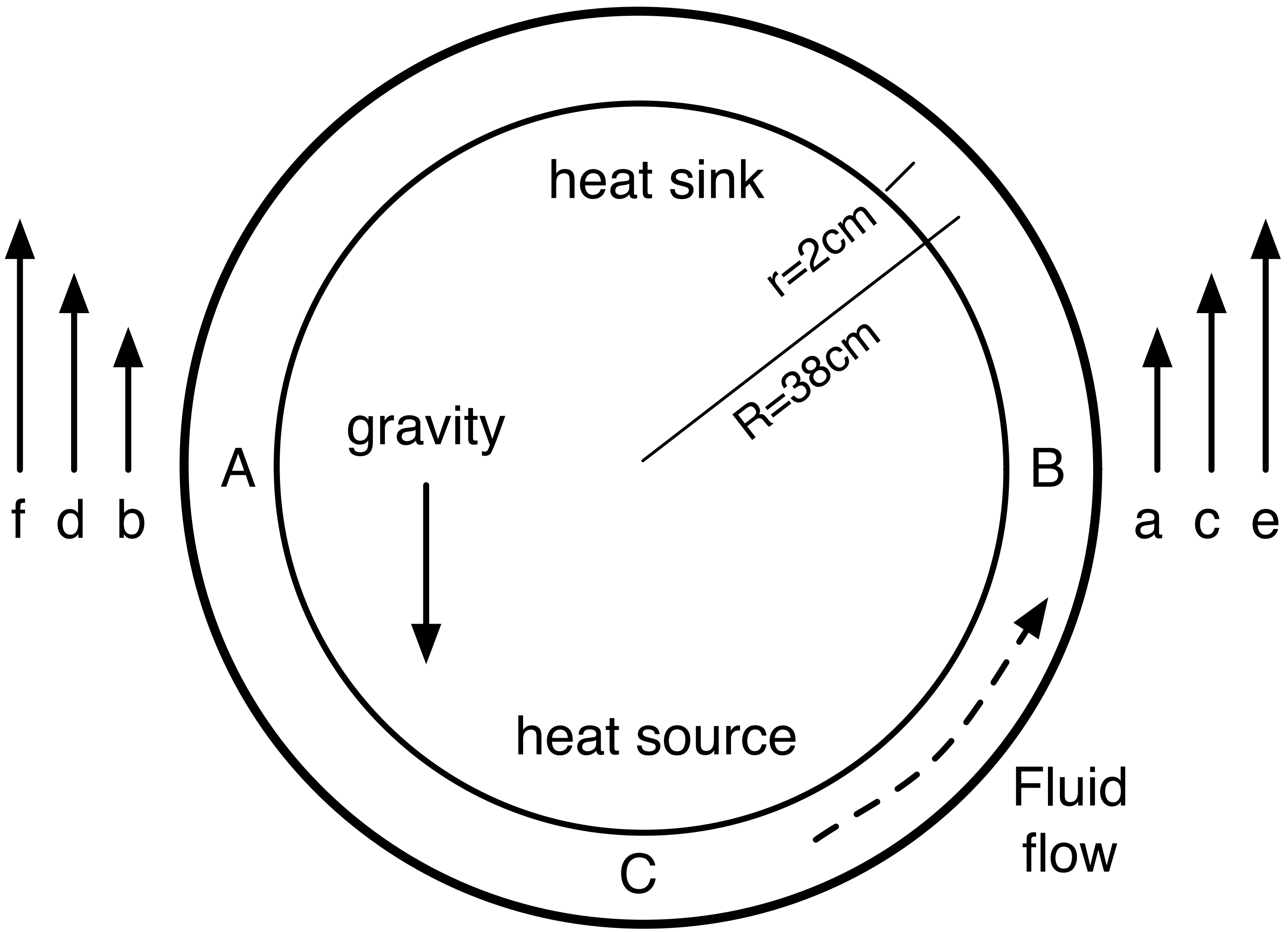
Duration of Regimes (Flow Reversals)



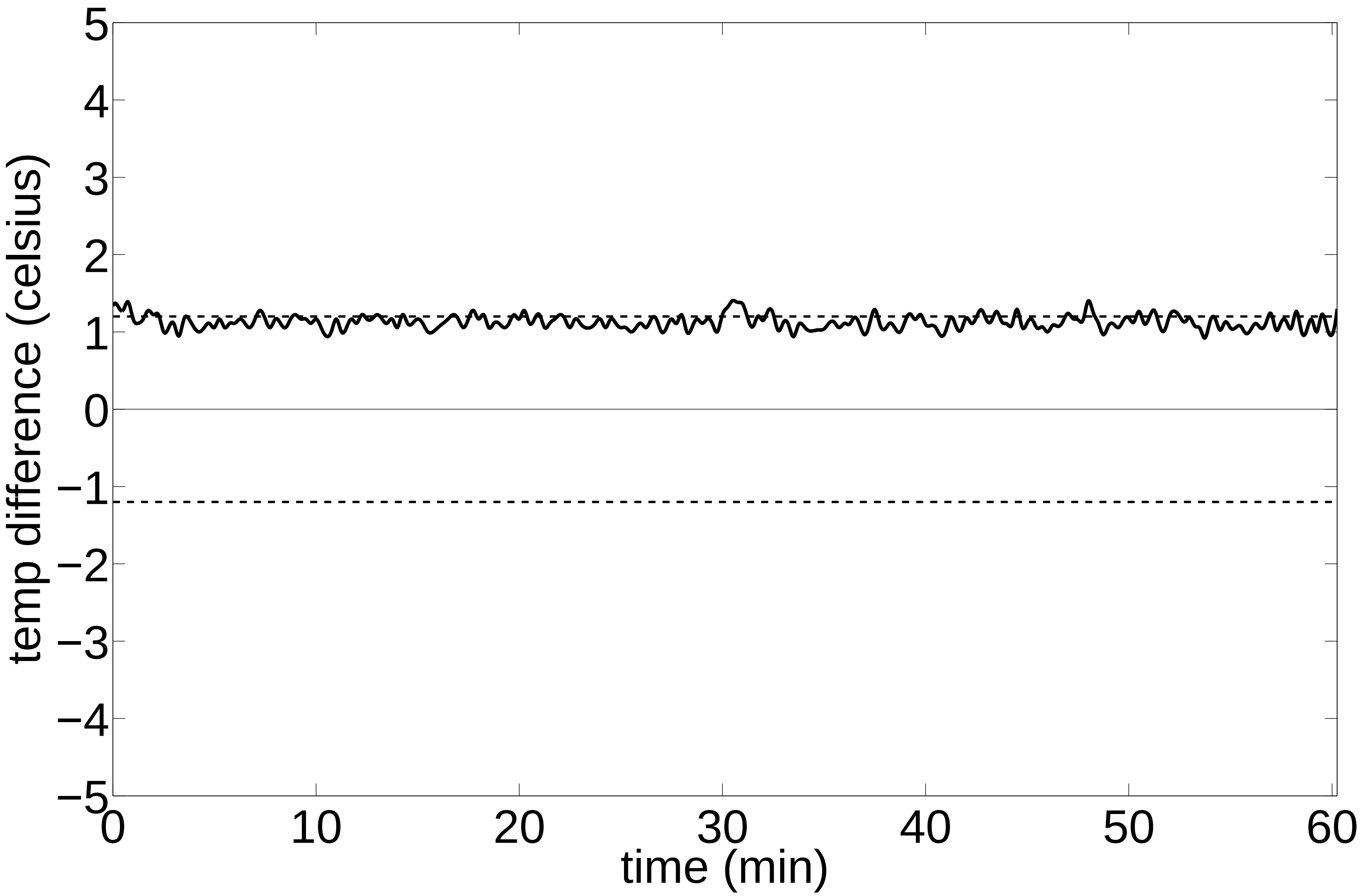
Duration of Regimes (Flow Reversals)



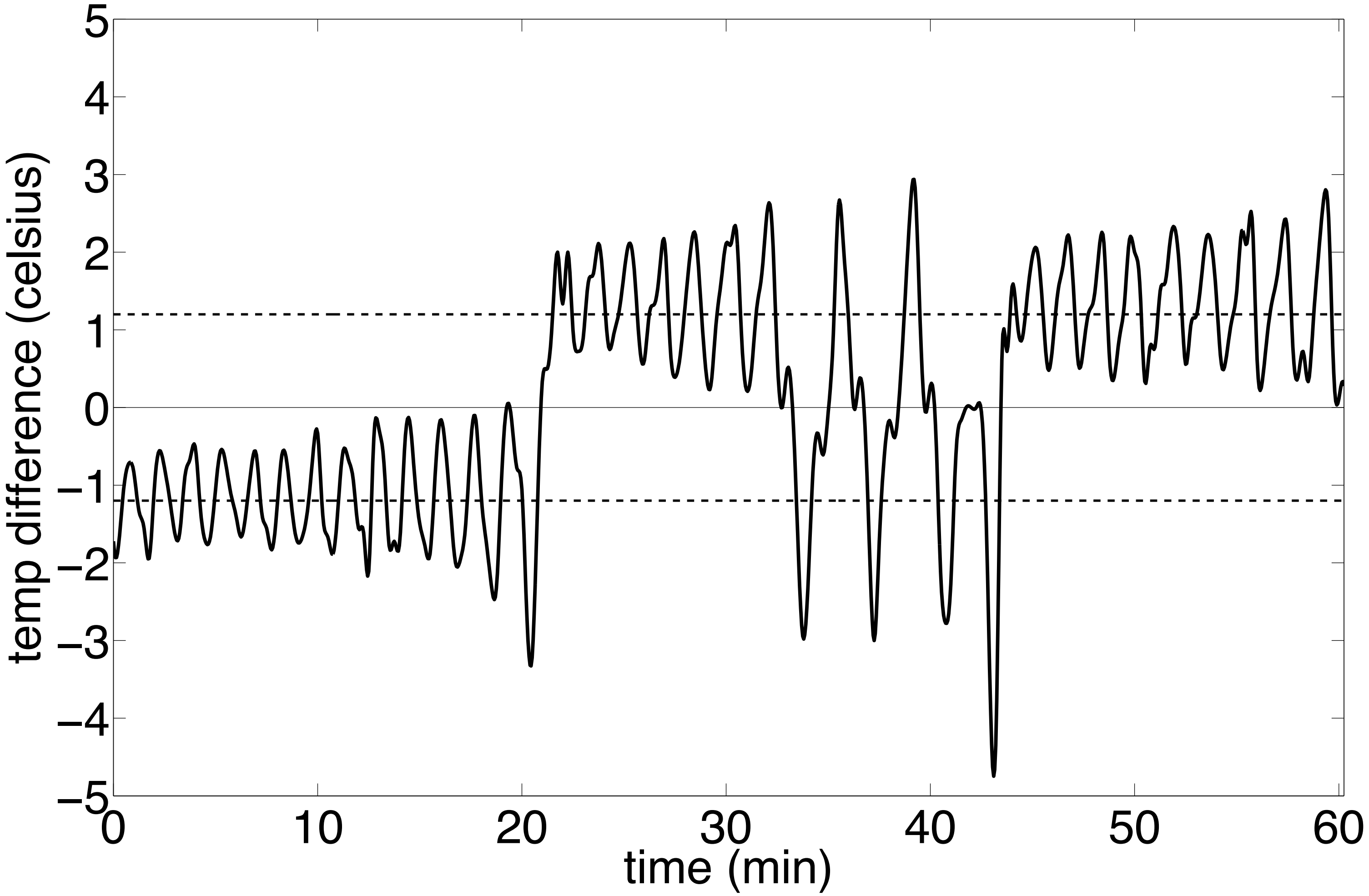
Thermal Convection Loop (An Experiment!)



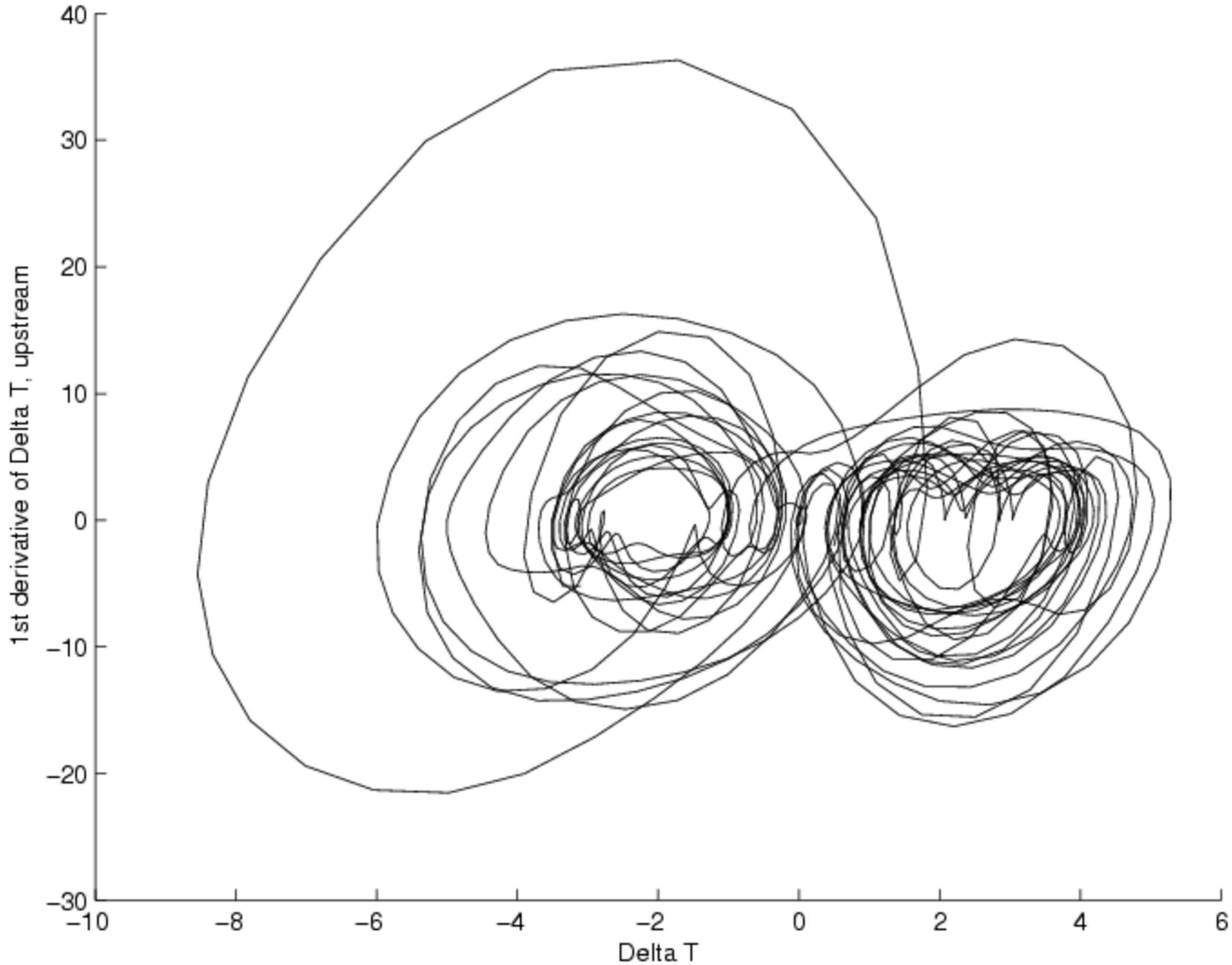
Typical Observations of Delta Temp (@9 - 3 o'clock) $\approx y$
From my Undergraduate Thesis: Stable Convection



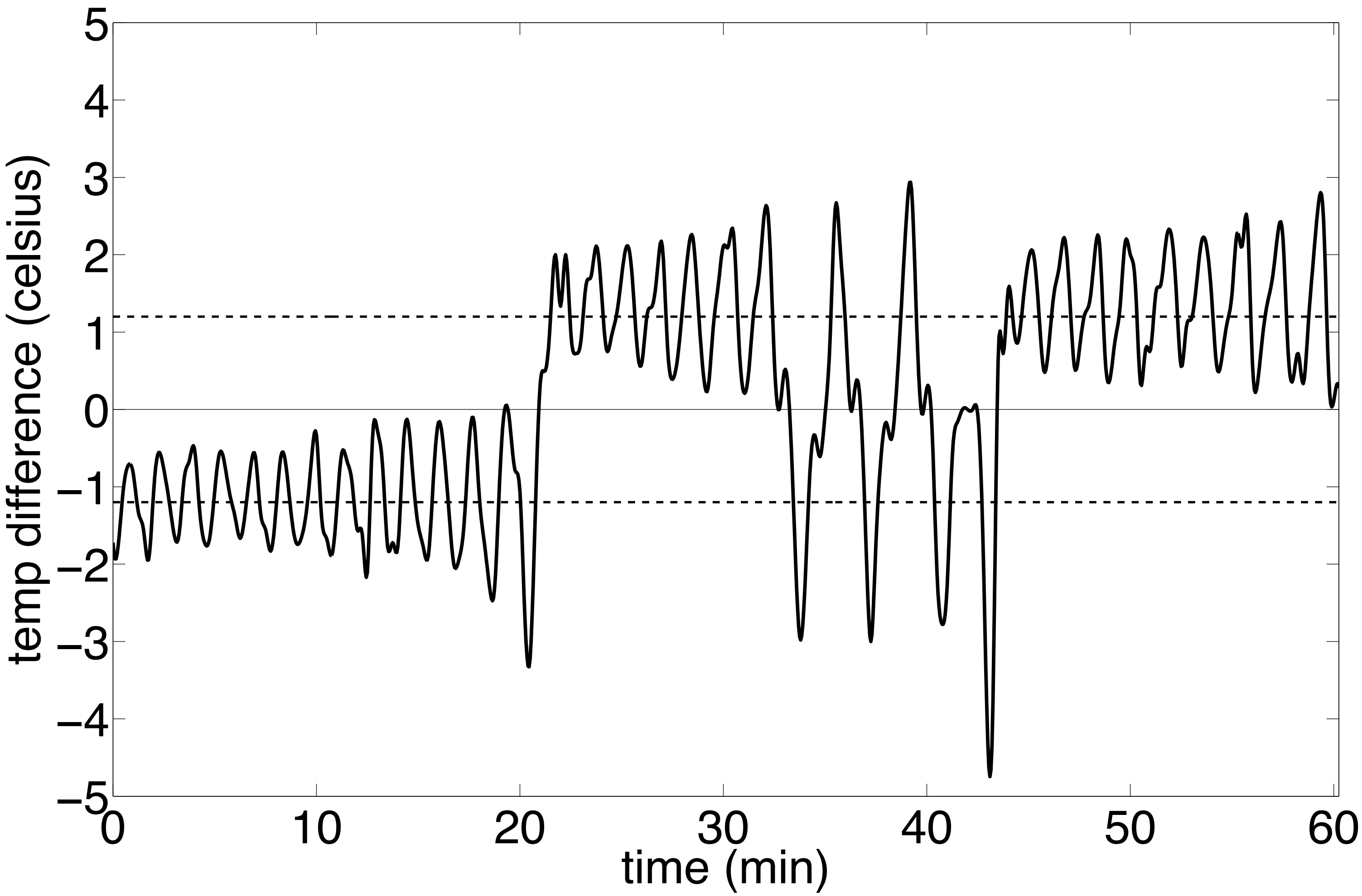
Typical Observations of Delta Temp (@9 - 3 o'clock) $\approx y$ From my Undergraduate Thesis: Chaotic Convection



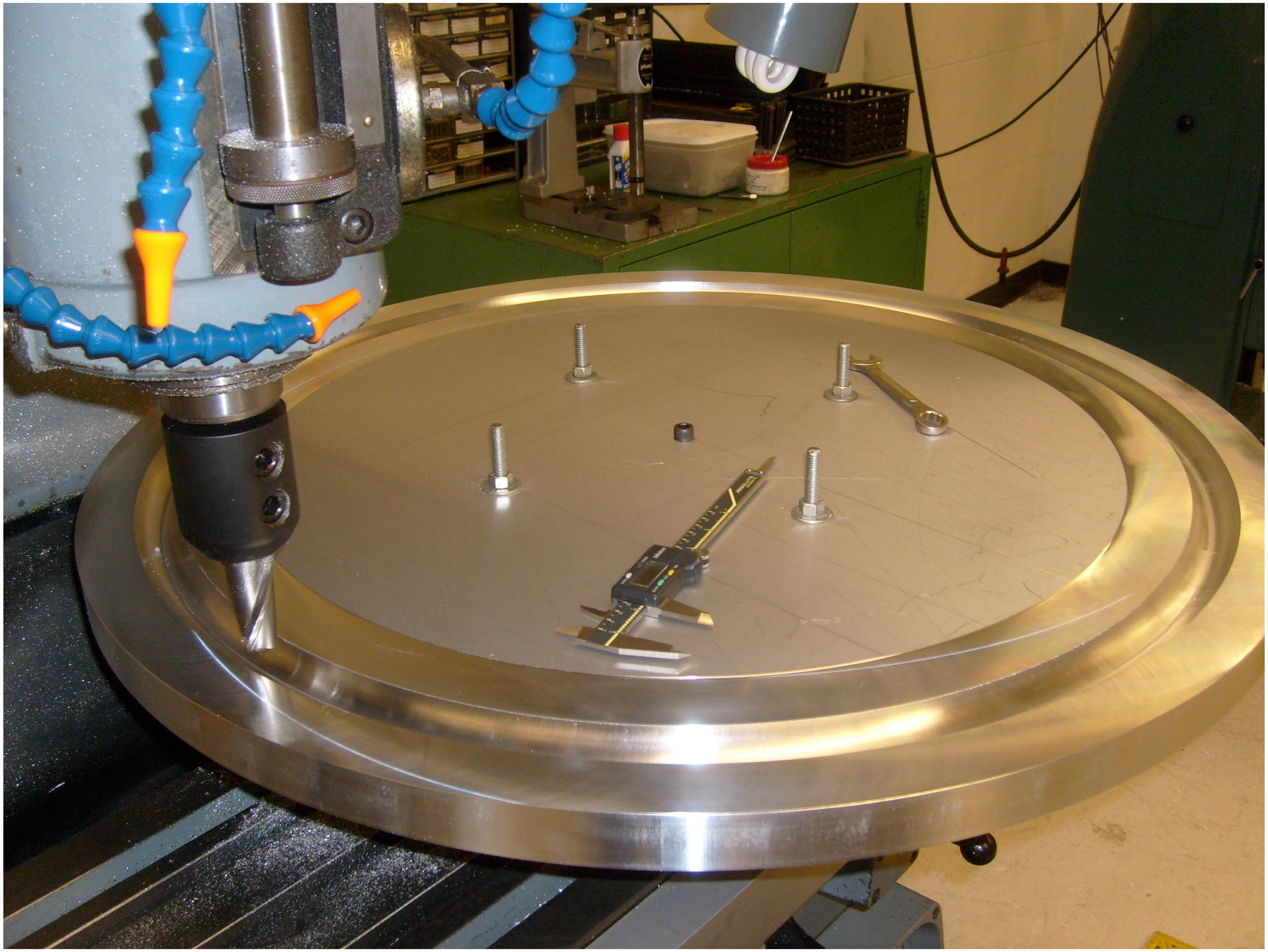
Reconstructing an Attractor



Only!!! Observations of Delta Temp (@9 - 3 o'clock) $\approx y$ From my Undergraduate Thesis: Melted Experiment



Kameron's Undergraduate Thesis (a classier operation...)



Kameron's Undergraduate Thesis (a classier operation...)

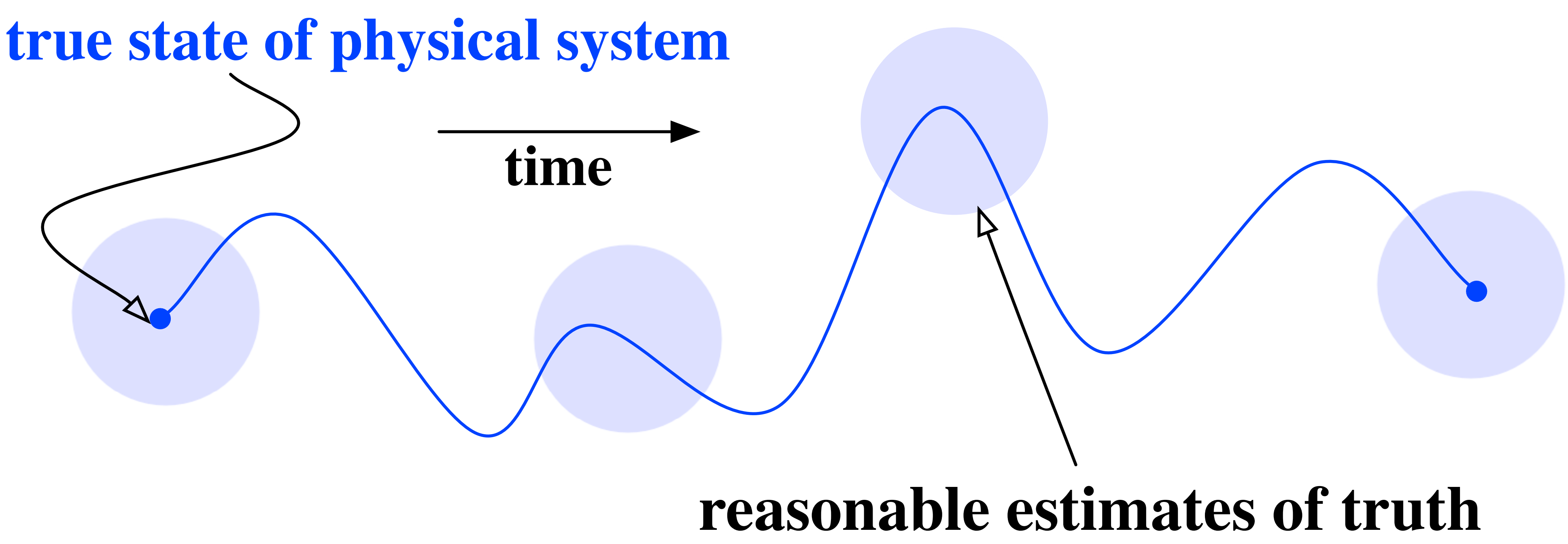
- Experiment will be used as a testbed for improving data assimilation and ensemble forecasting using simple (Lorenz 3-D) and sophisticated (CFD 10^6-D) models.
- We can control the climate (i.e. visit specified regions of state space experimentally)!

Three Experiments

- An experimental analog to Lorenz's 1963 model
- Stalking observations with a numerical trajectory
- *Online* empirical correction of model error

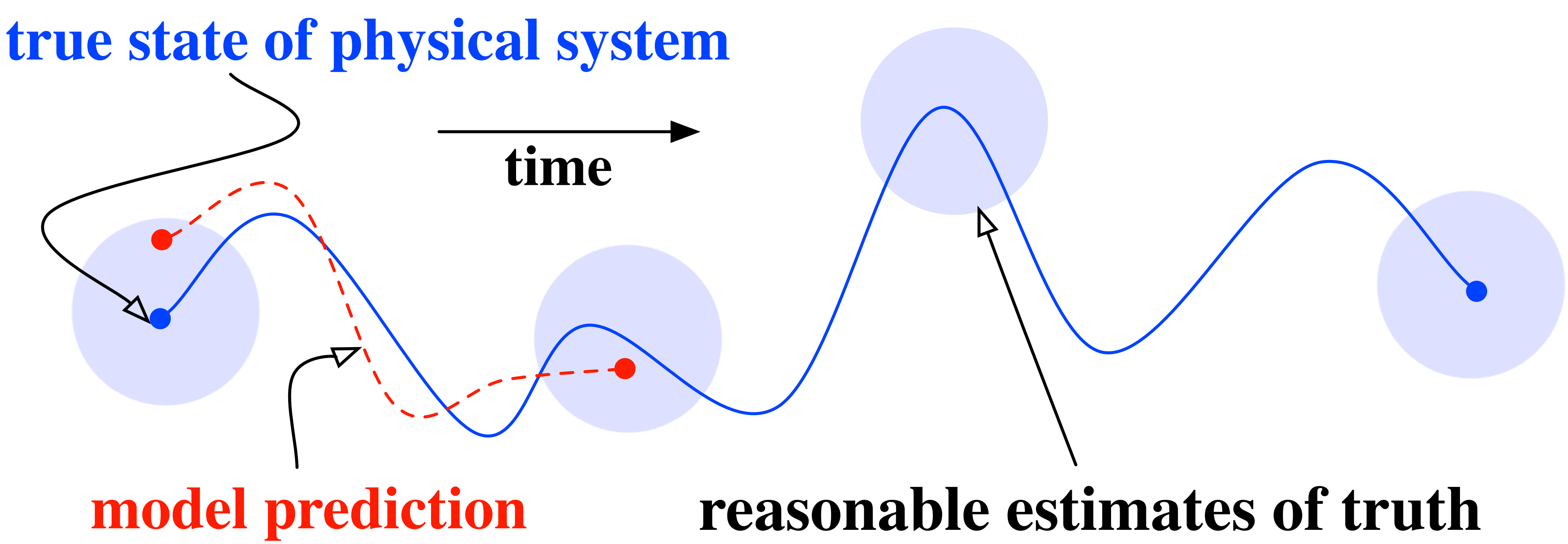
A Mathematical Question about Prediction:

How long can we shadow a trajectory of the physical system H with a trajectory of the model L?



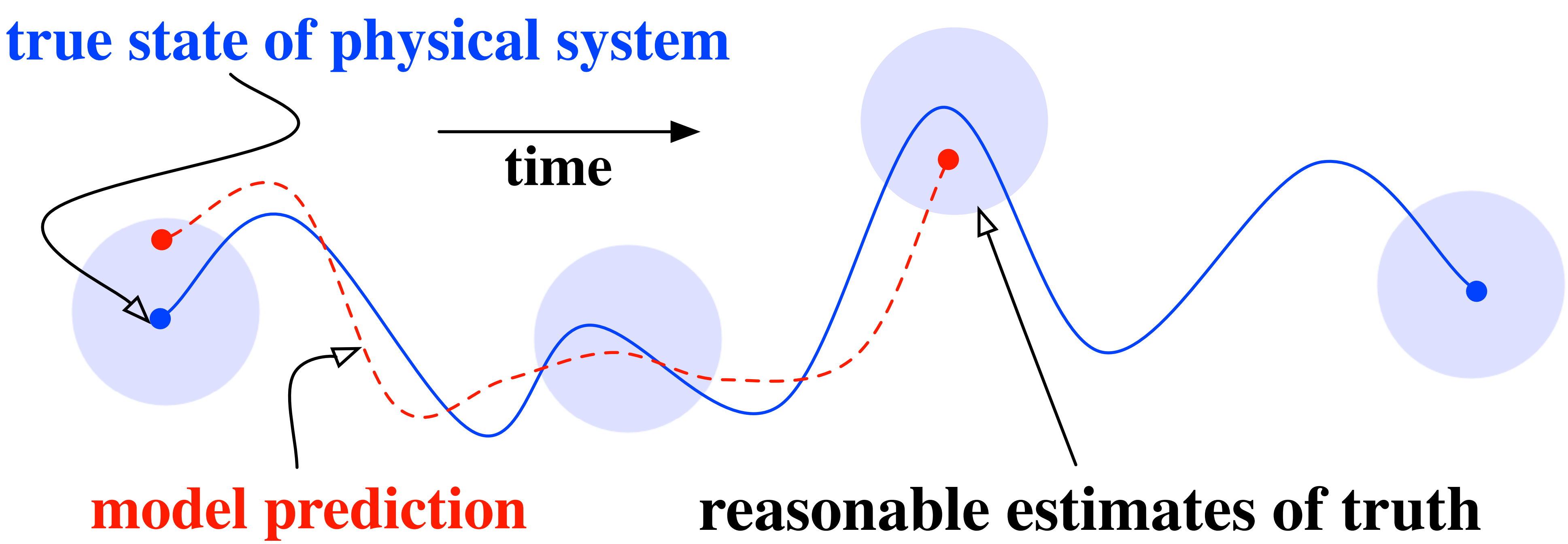
A Mathematical Question about Prediction:

How long can we shadow a trajectory of the physical system H with a trajectory of the model L?



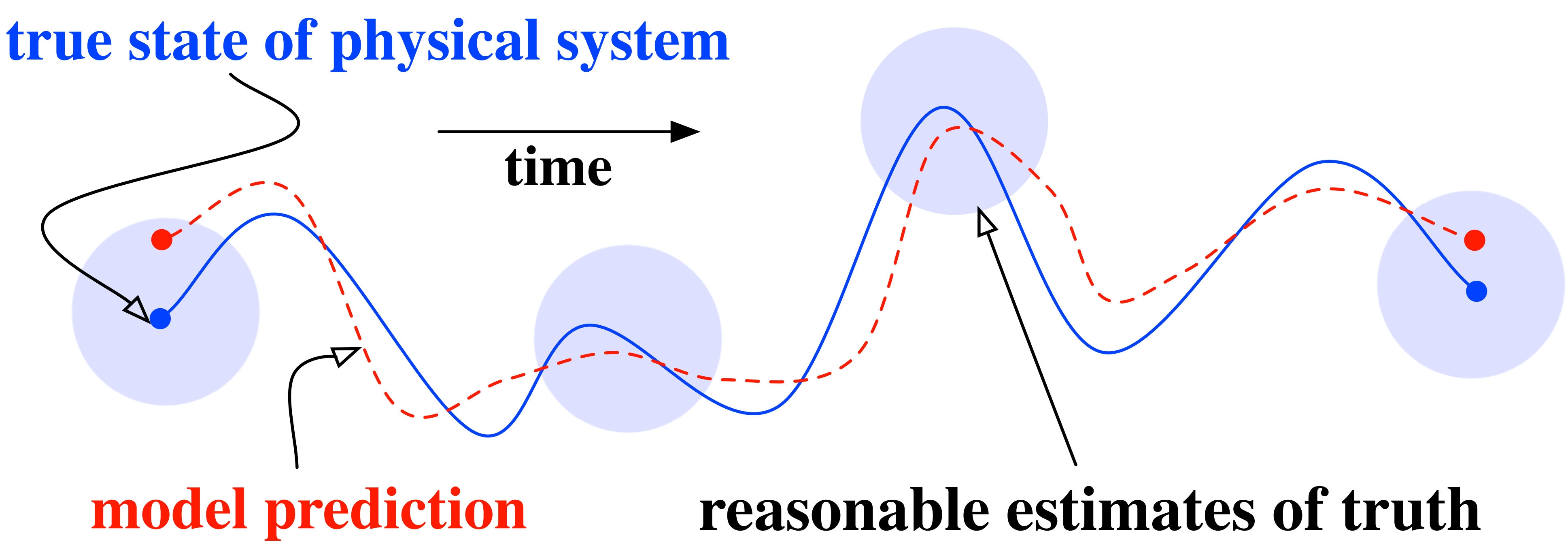
A Mathematical Question about Prediction:

How long can we shadow a trajectory of the physical system H with a trajectory of the model L?

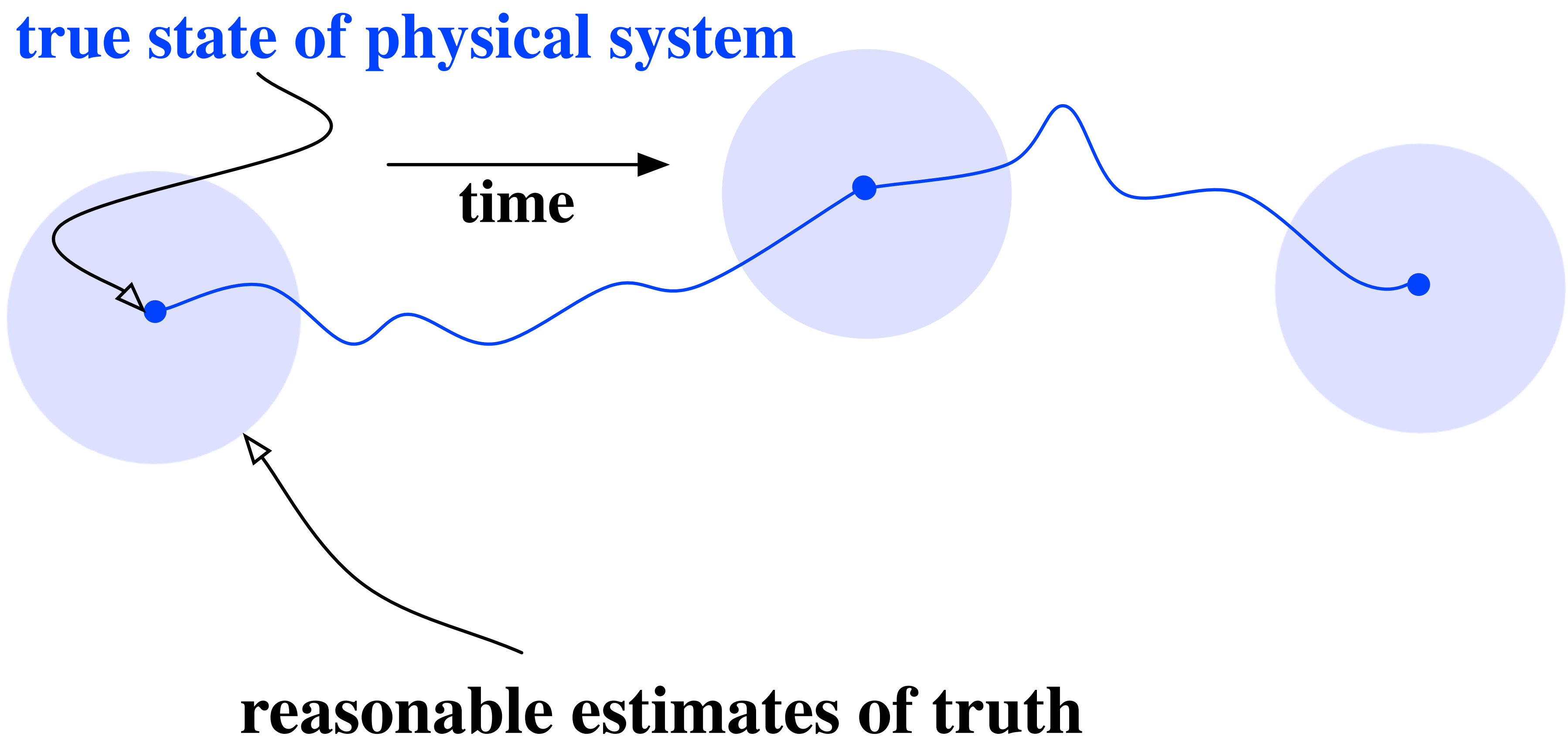


A Mathematical Question about Prediction:

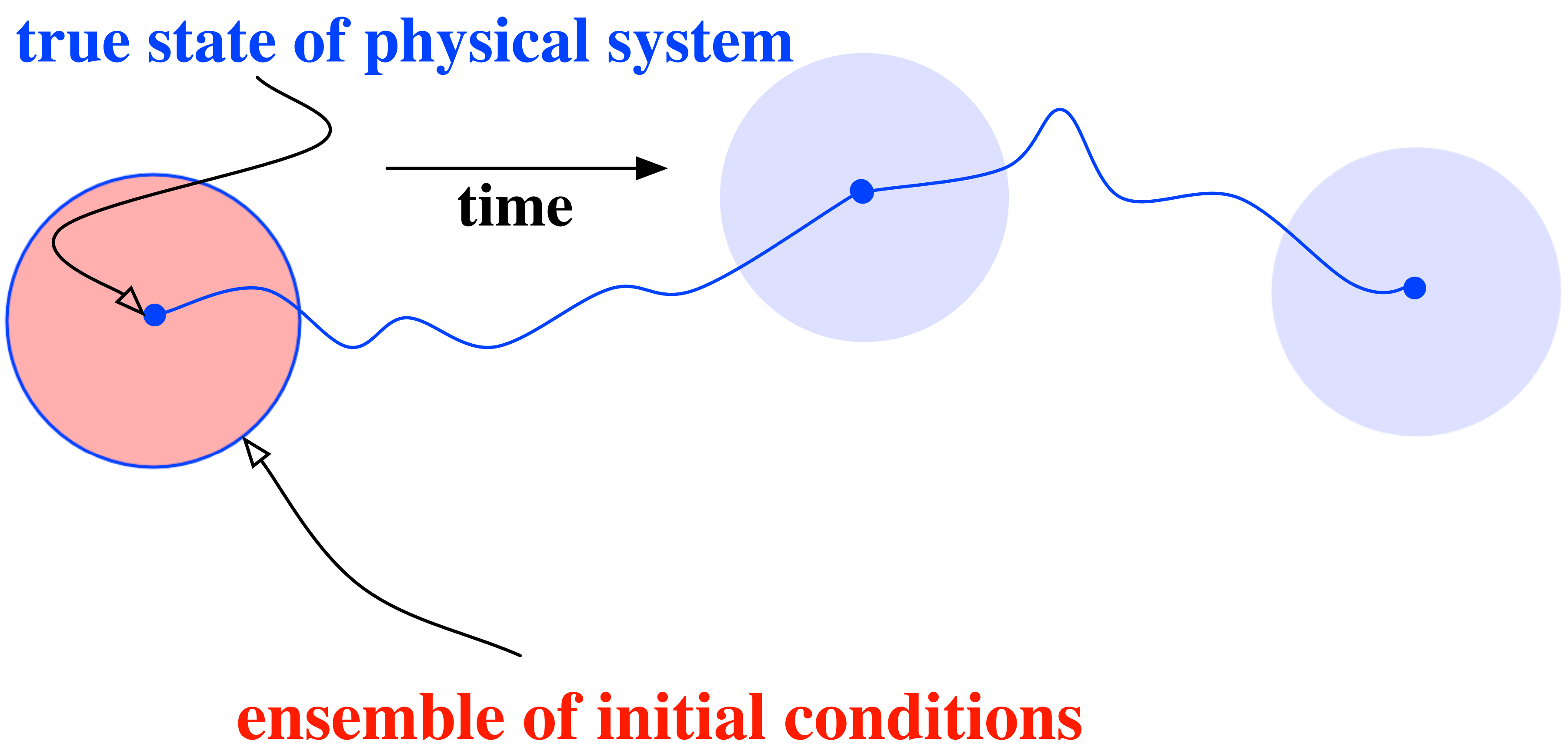
How long can we shadow a trajectory of the physical system **H** with a trajectory of the model **L**?



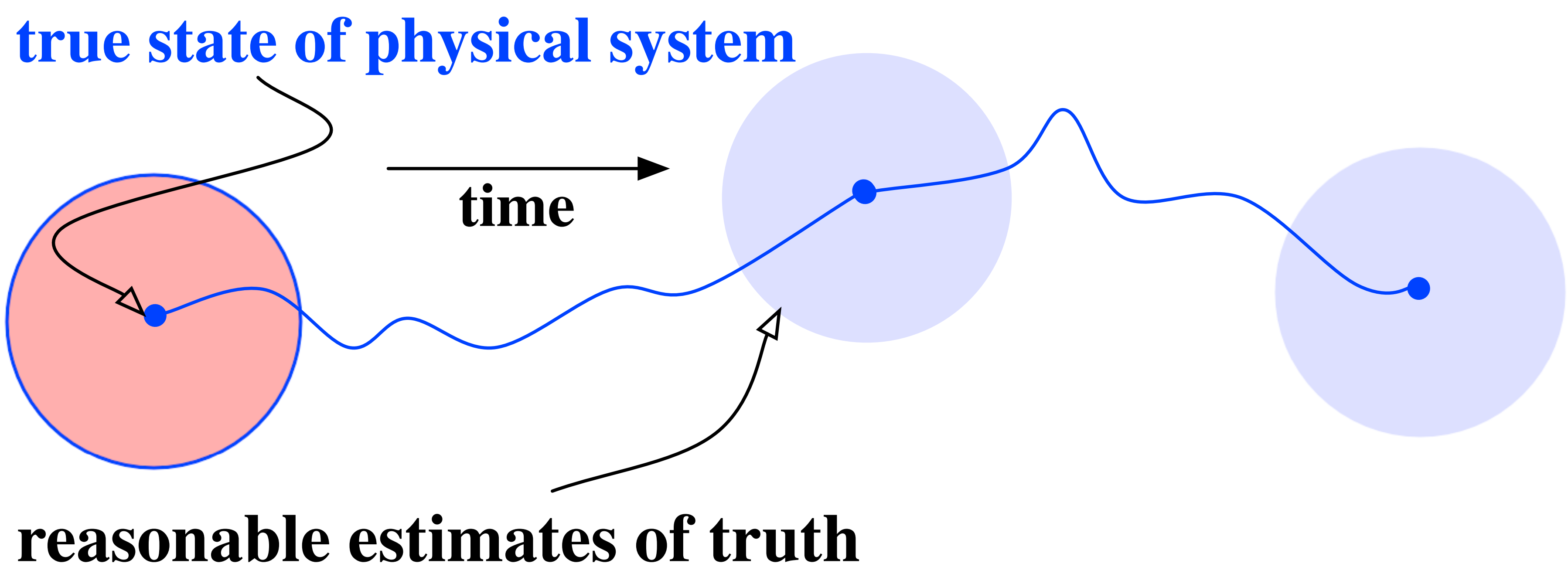
Shadowing a Physical System with an Ensemble



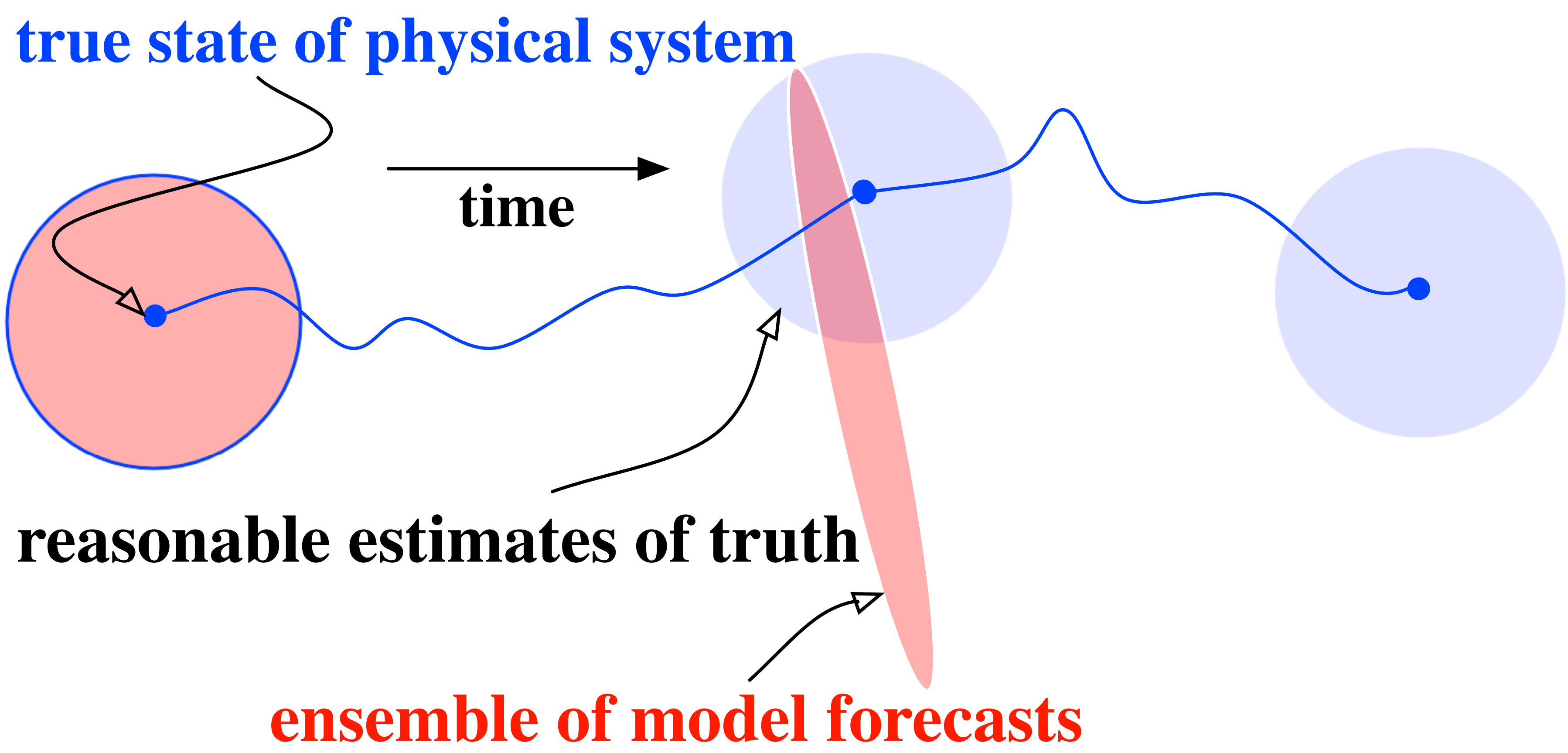
Shadowing a Physical System with an Ensemble



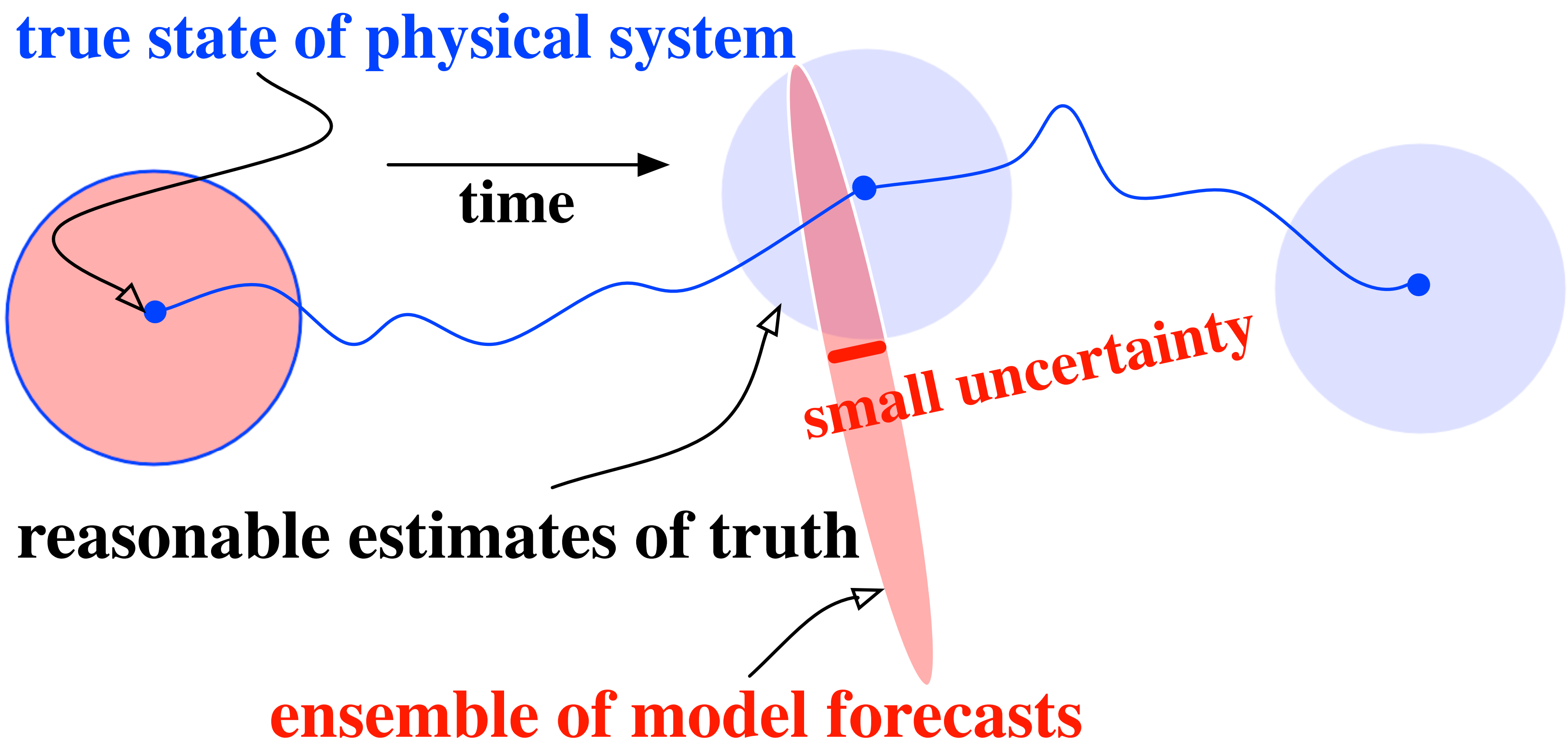
Shadowing a Physical System with an Ensemble



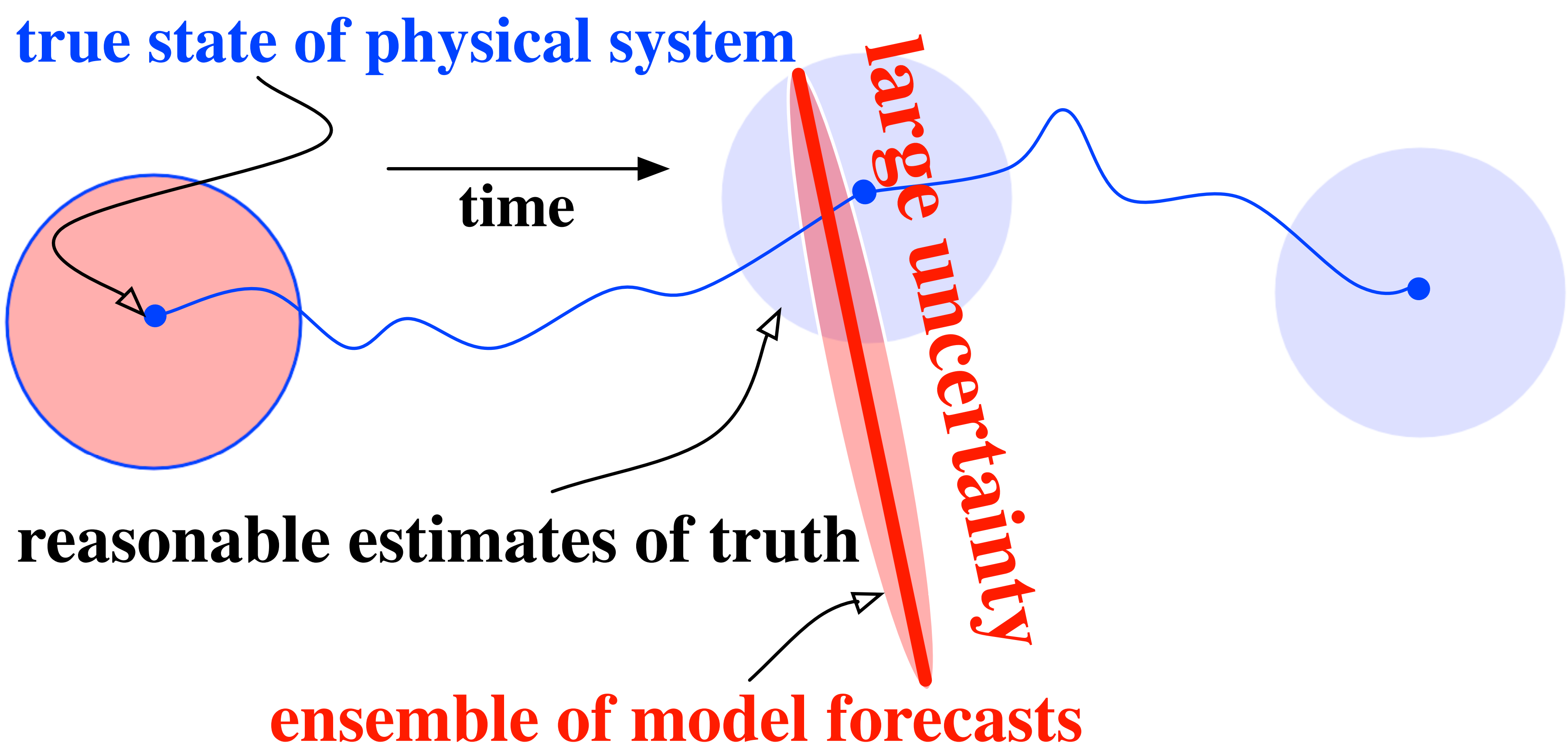
Shadowing a Physical System with an Ensemble



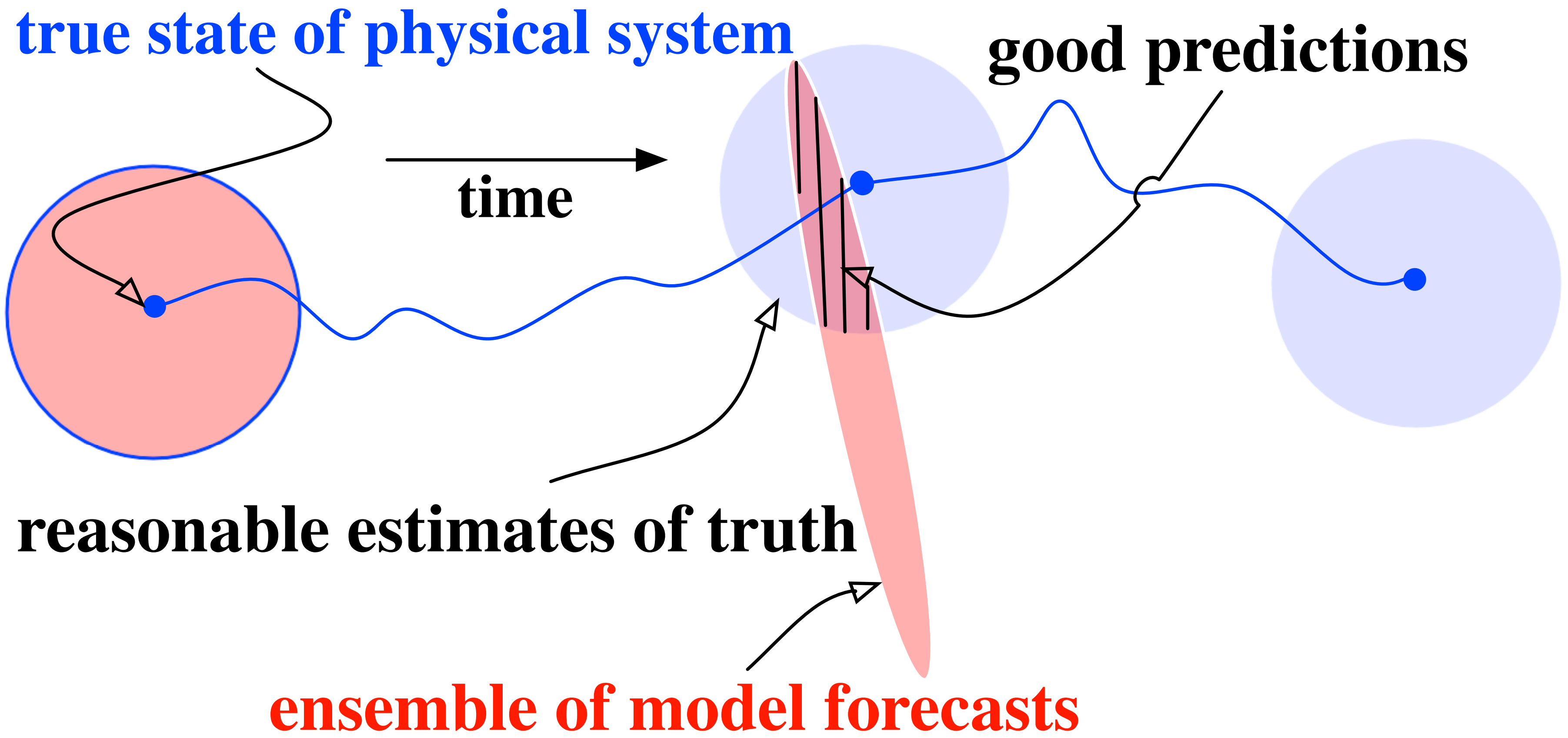
Shadowing a Physical System with an Ensemble



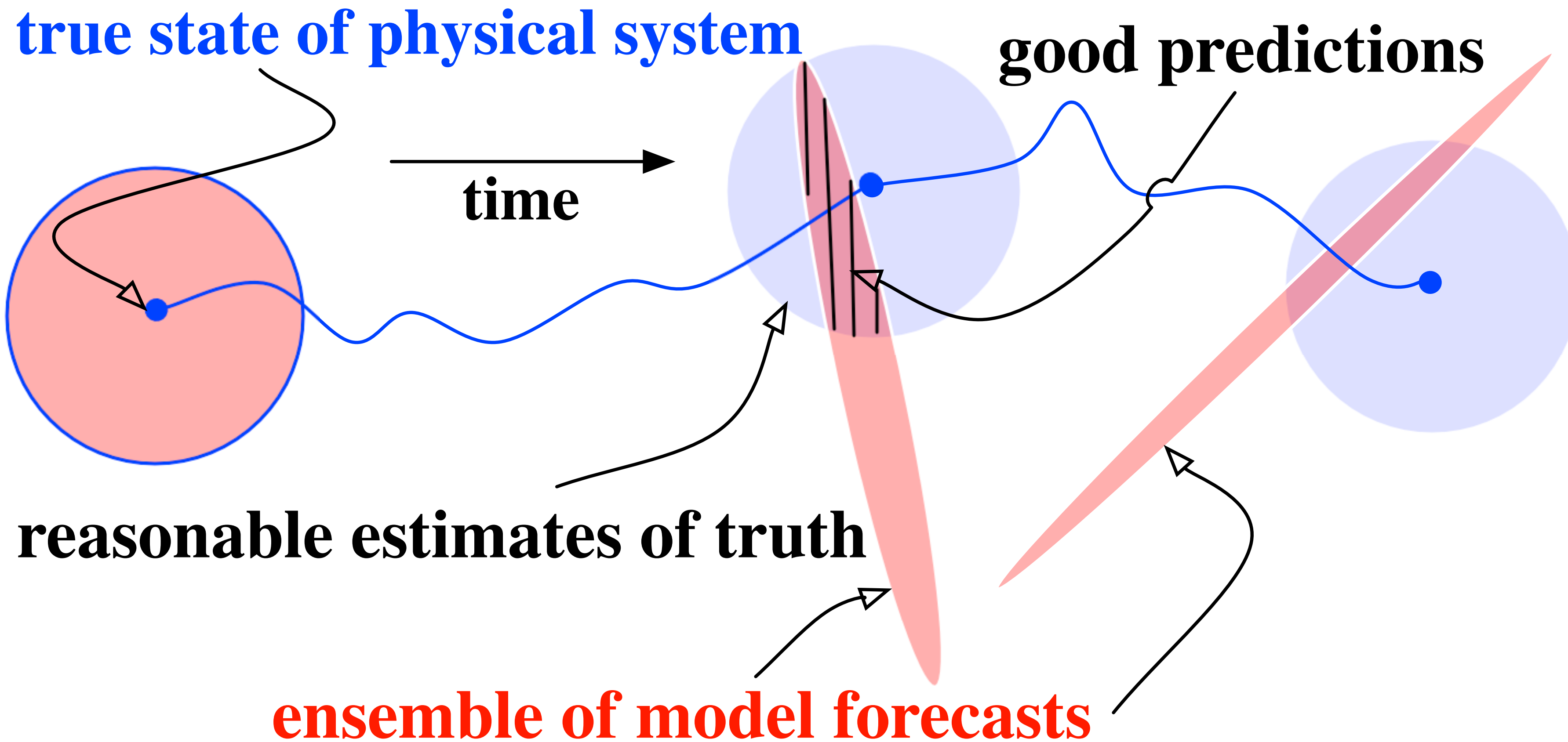
Shadowing a Physical System with an Ensemble



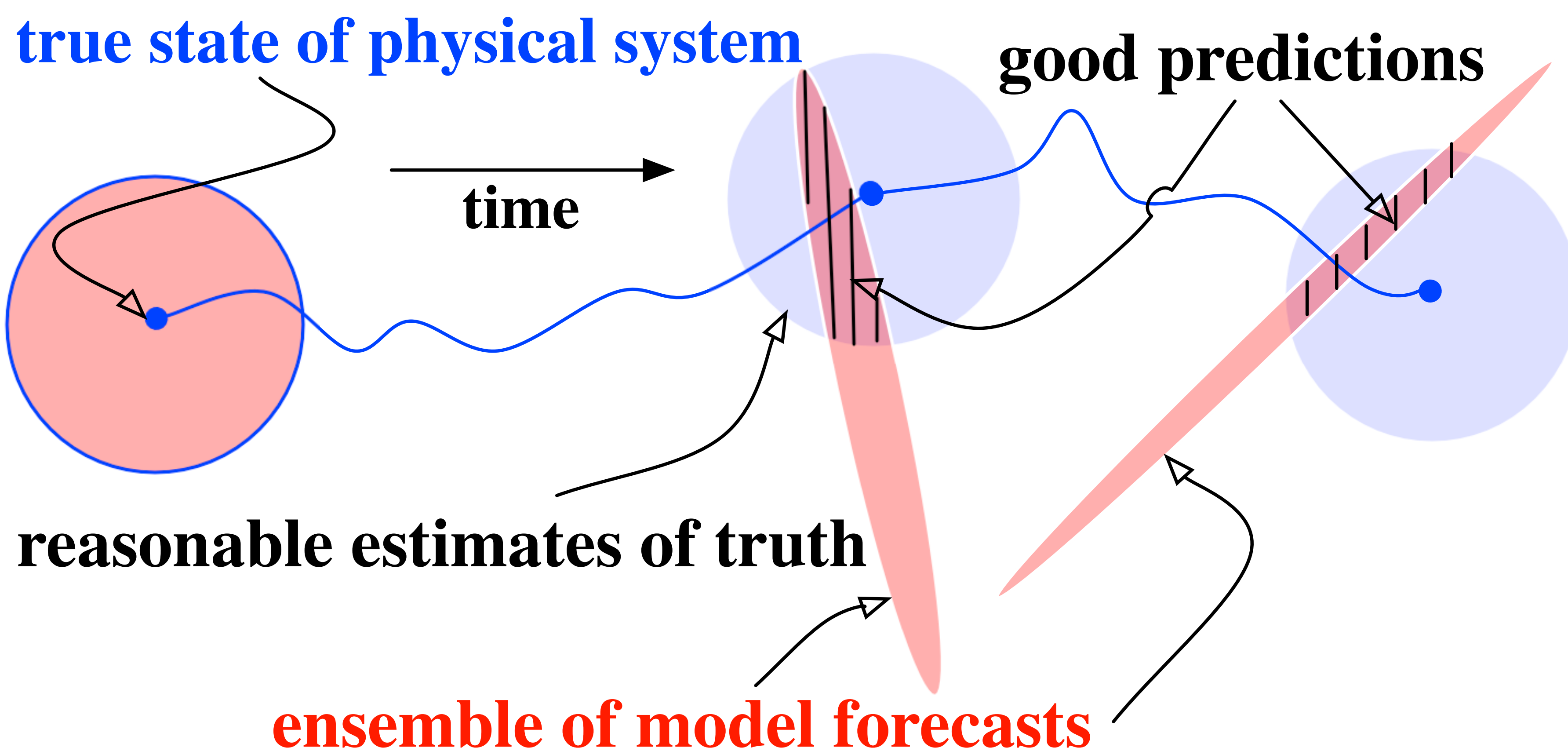
Shadowing a Physical System with an Ensemble



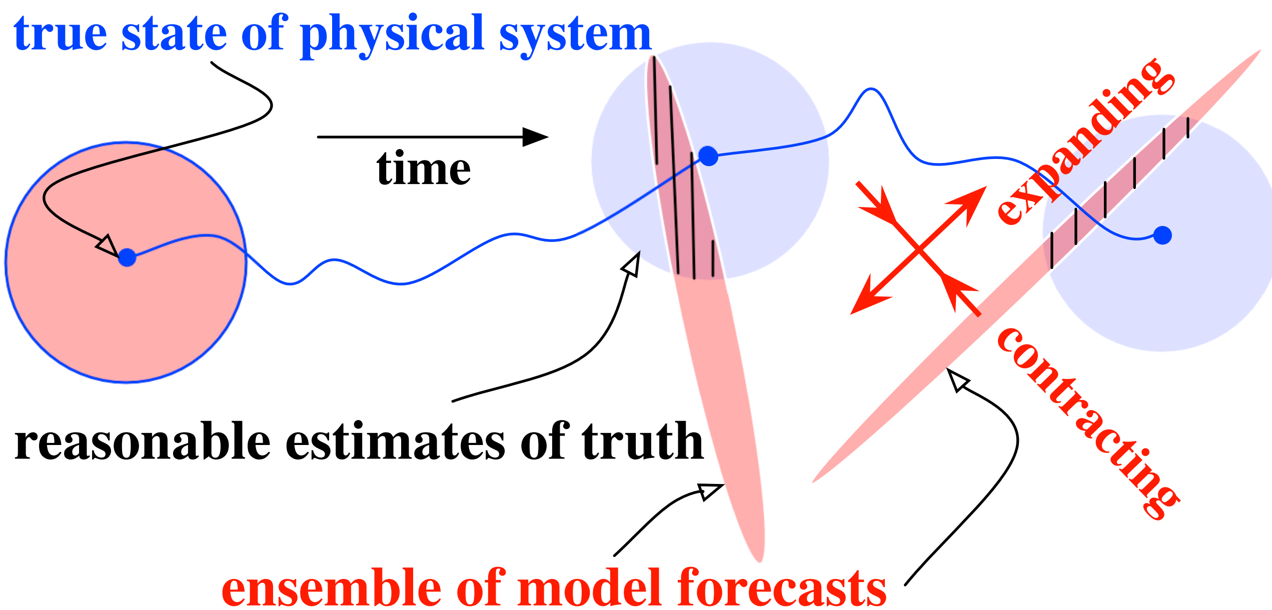
Shadowing a Physical System with an Ensemble



Shadowing a Physical System with an Ensemble

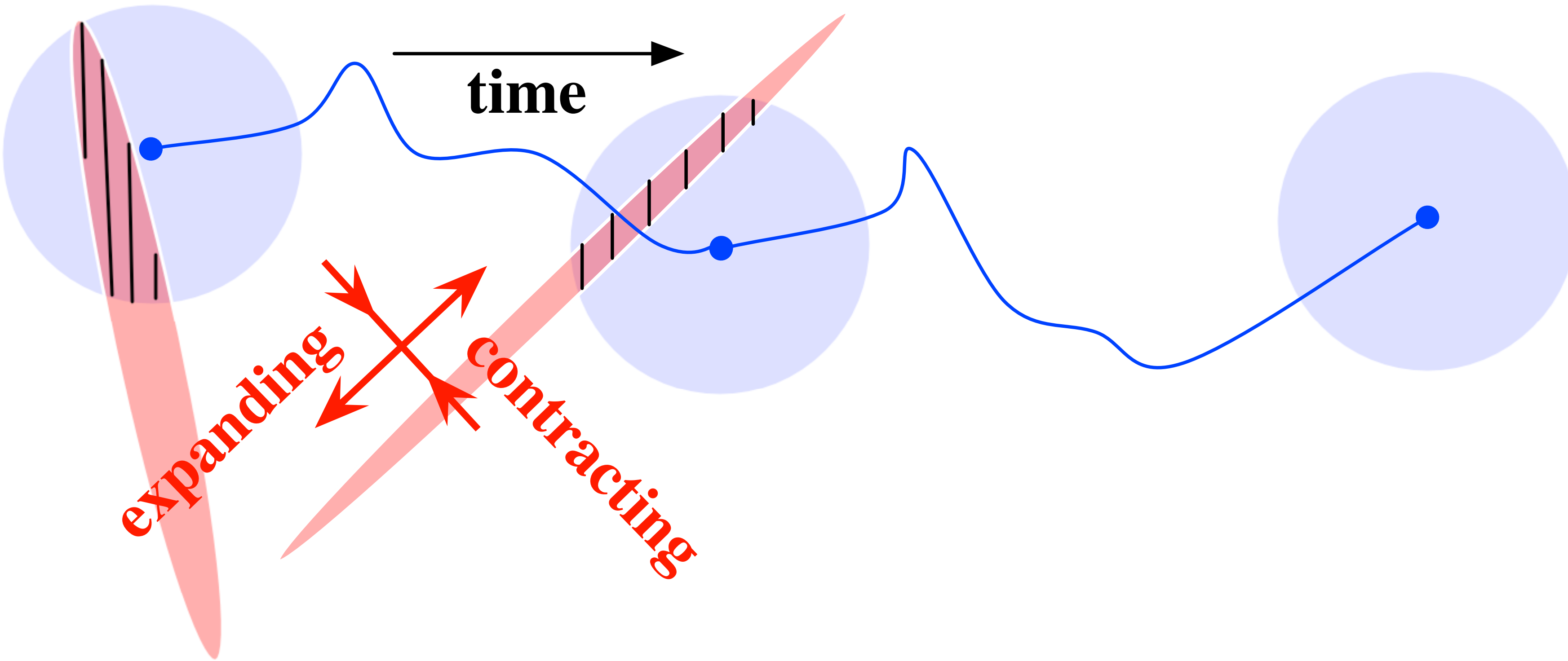


Shadowing a *Hyperbolic* System

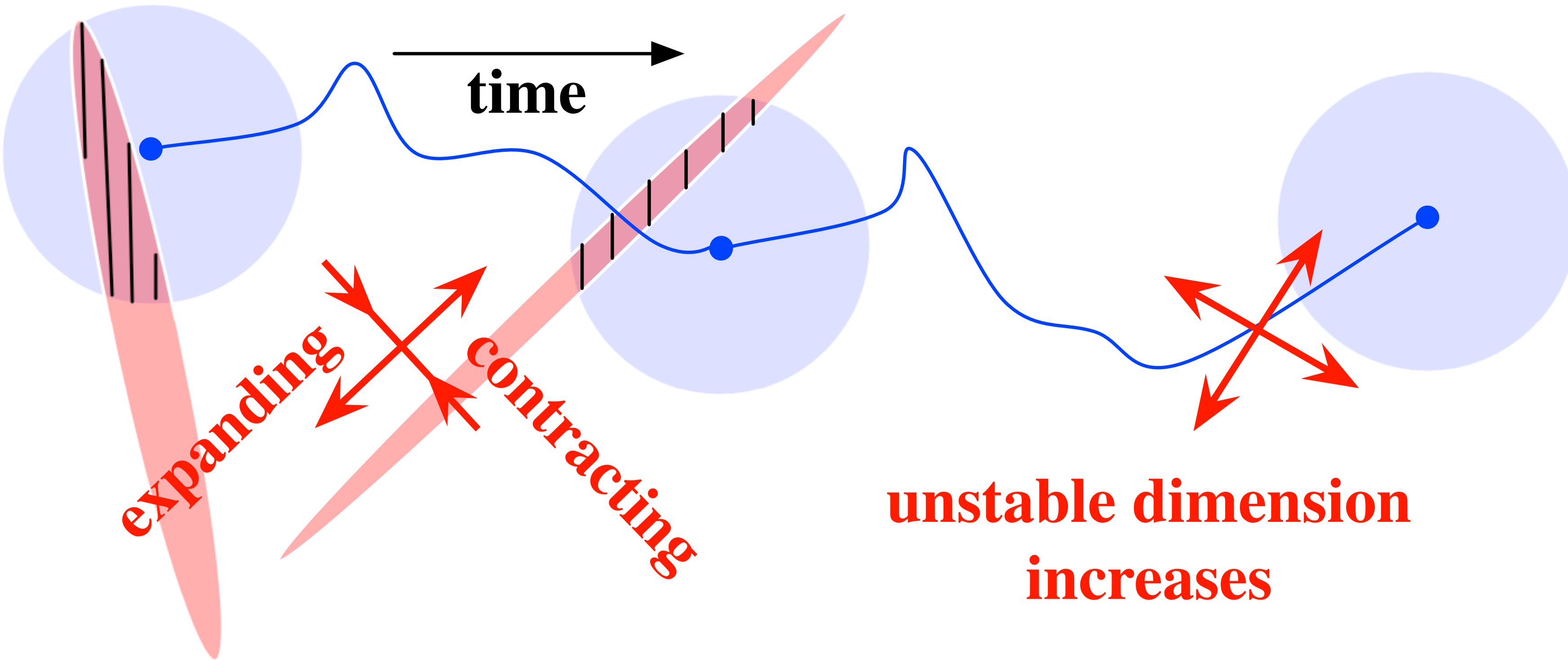


(Loading Movie) (Loading Movie)

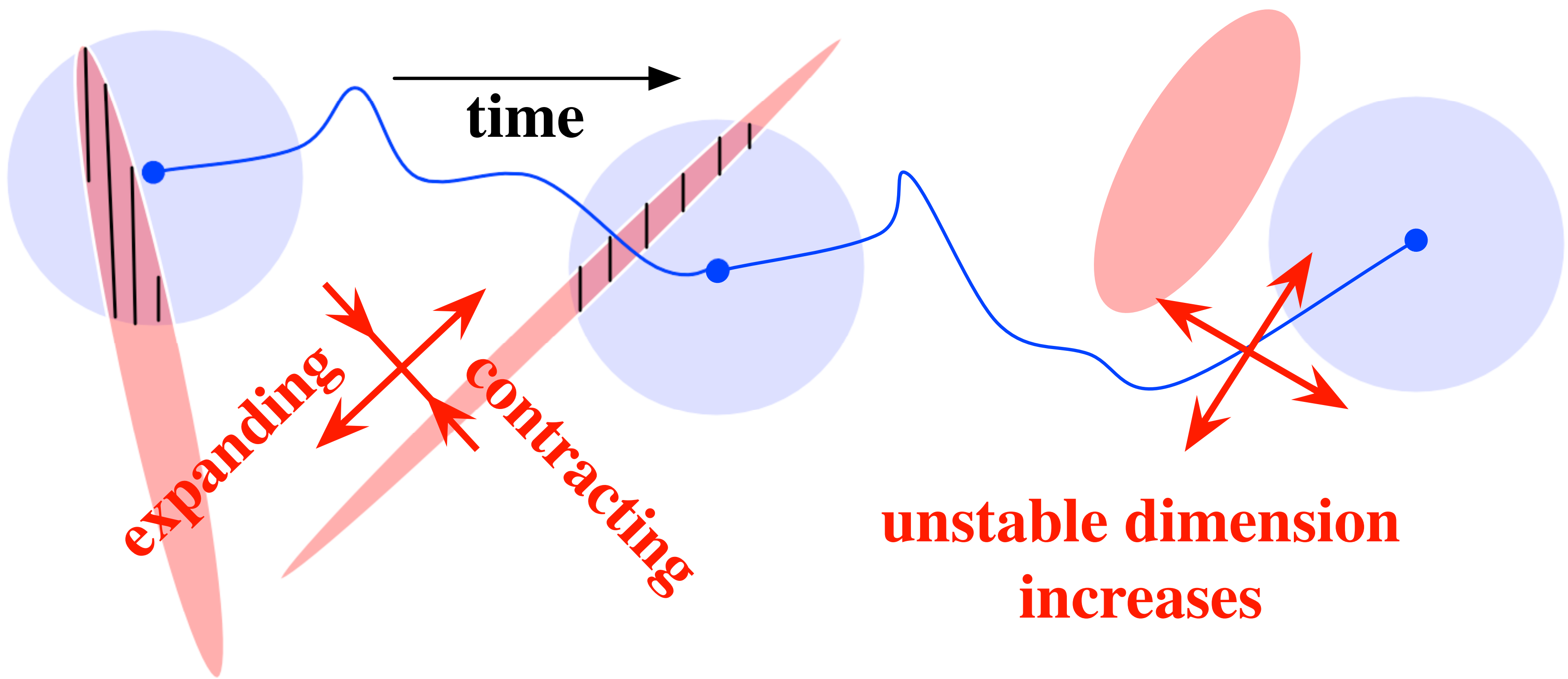
Shadowing a Non-Hyperbolic System



Shadowing a Non-Hyperbolic System

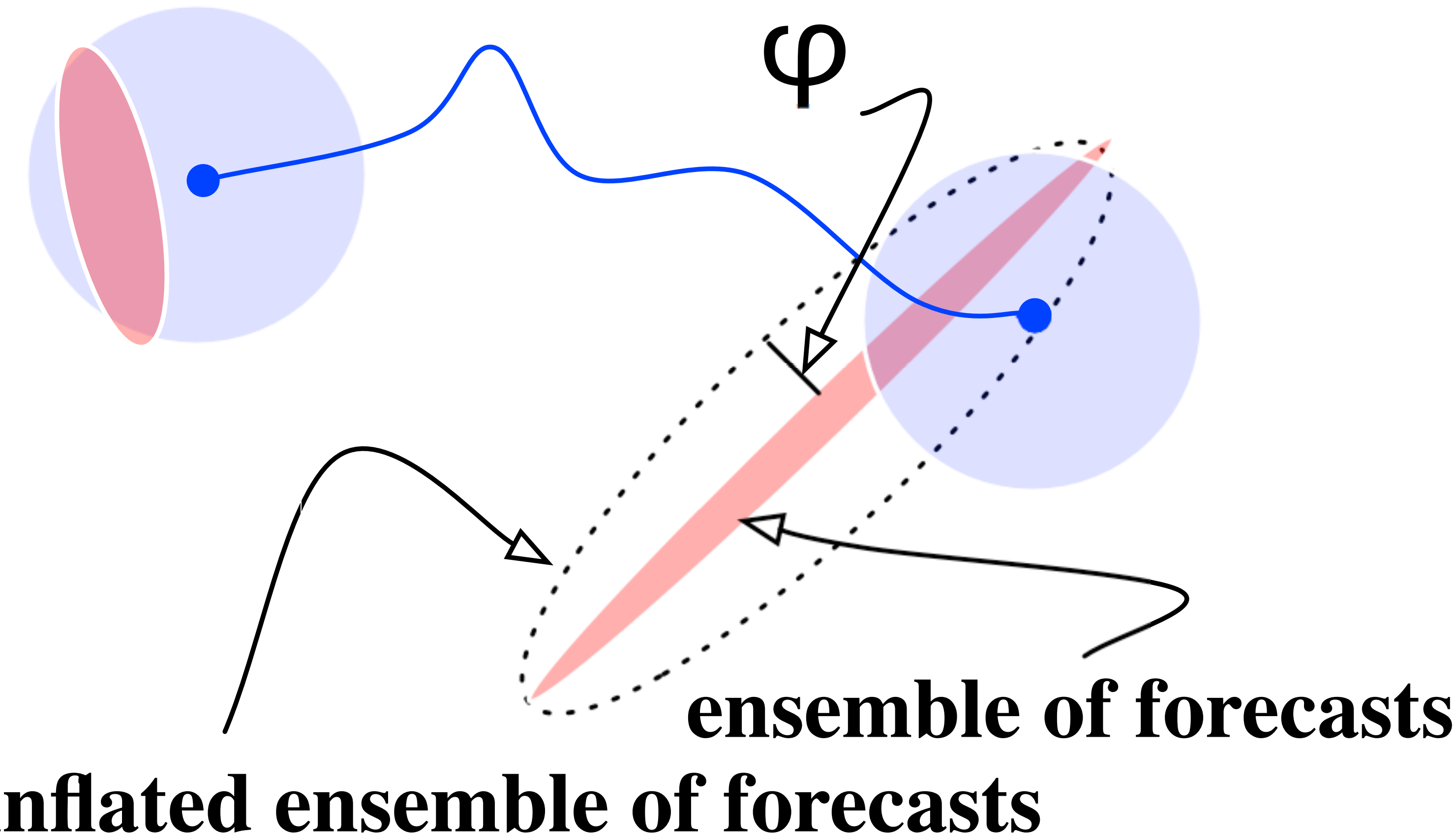


Shadowing a Non-Hyperbolic System



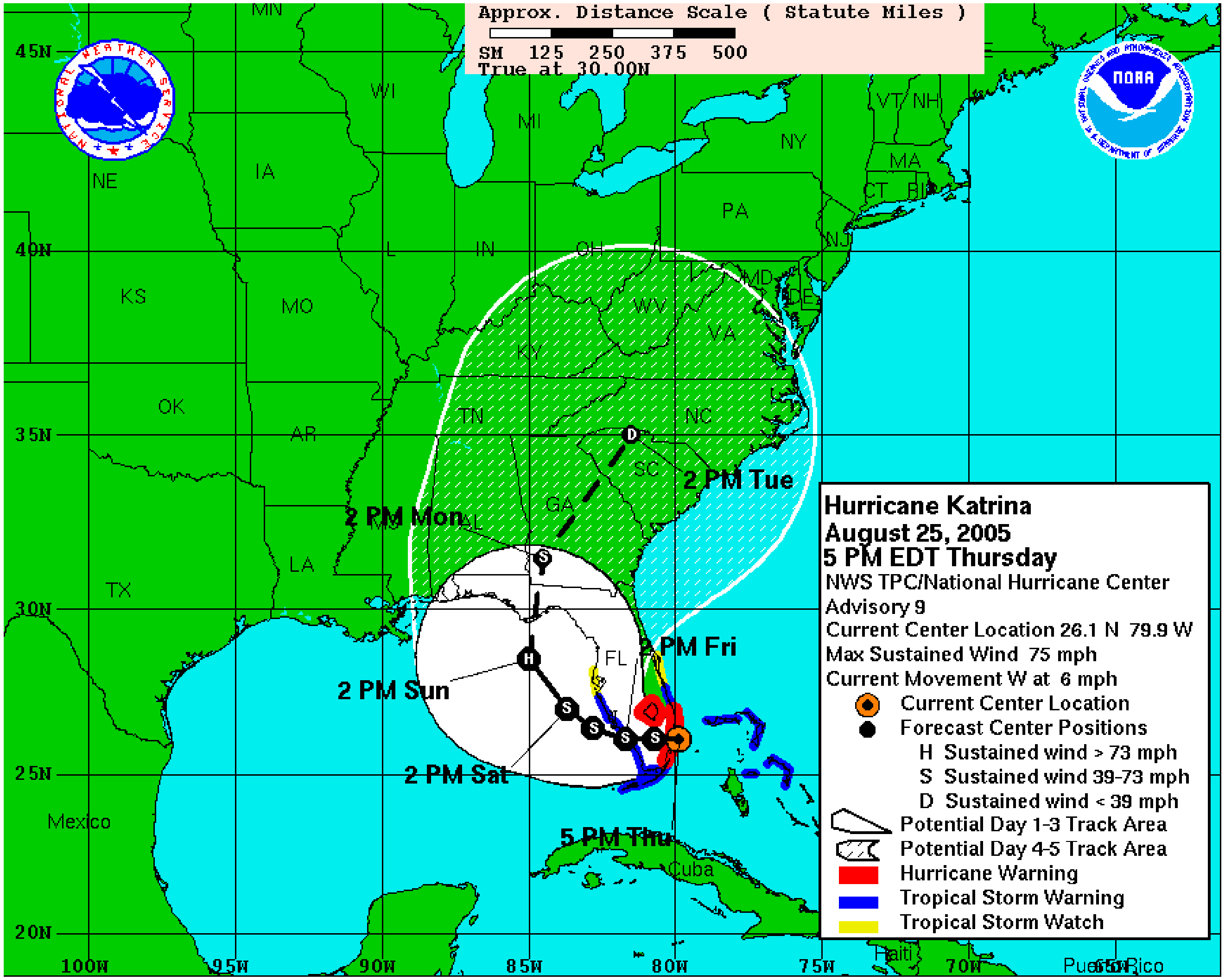
If the number of expanding directions increases, shadowing fails.

New Idea: Stalking a non-hyperbolic system H



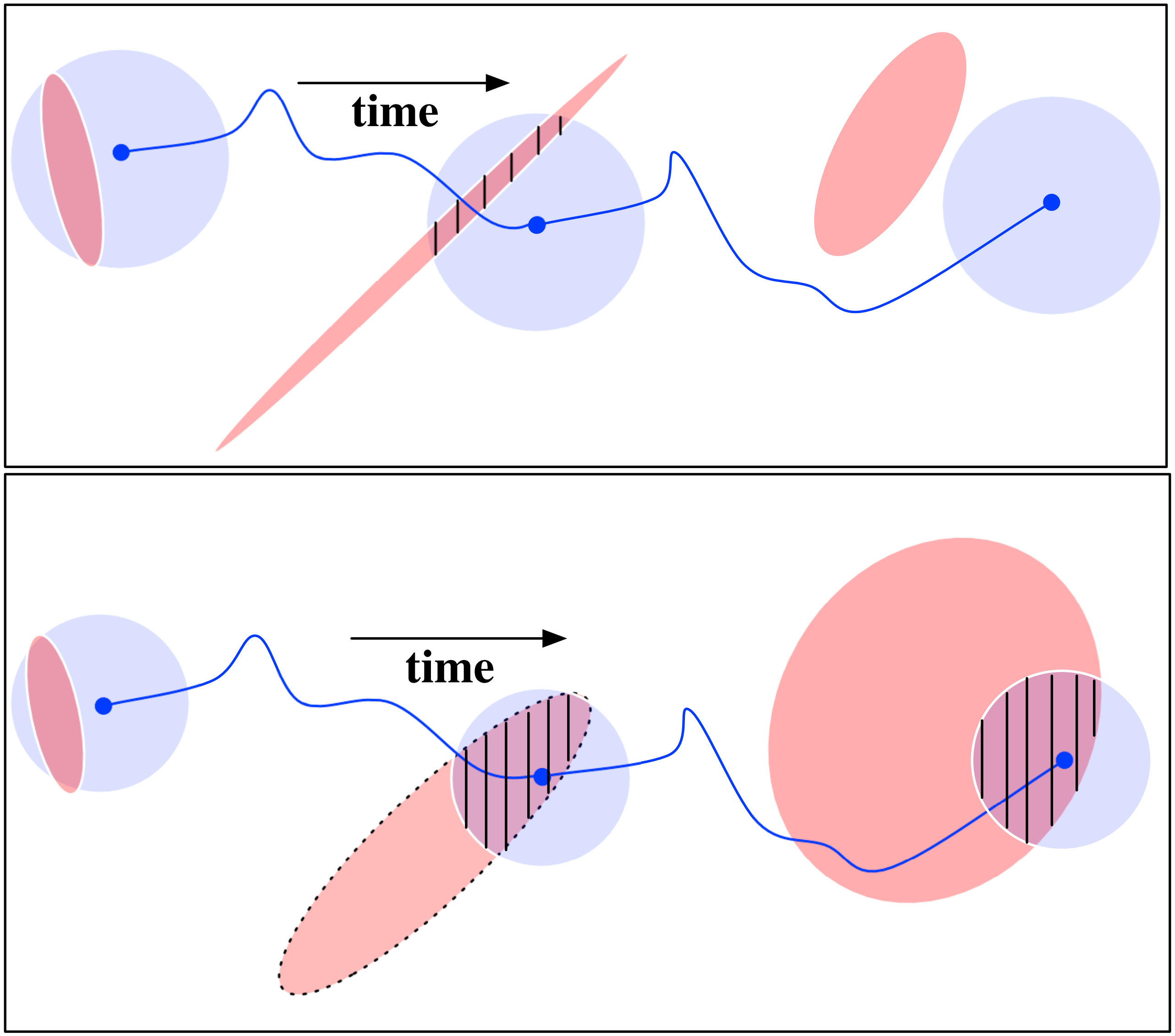
Inflate the contracting dimensions of the ensemble

Cone of Uncertainty

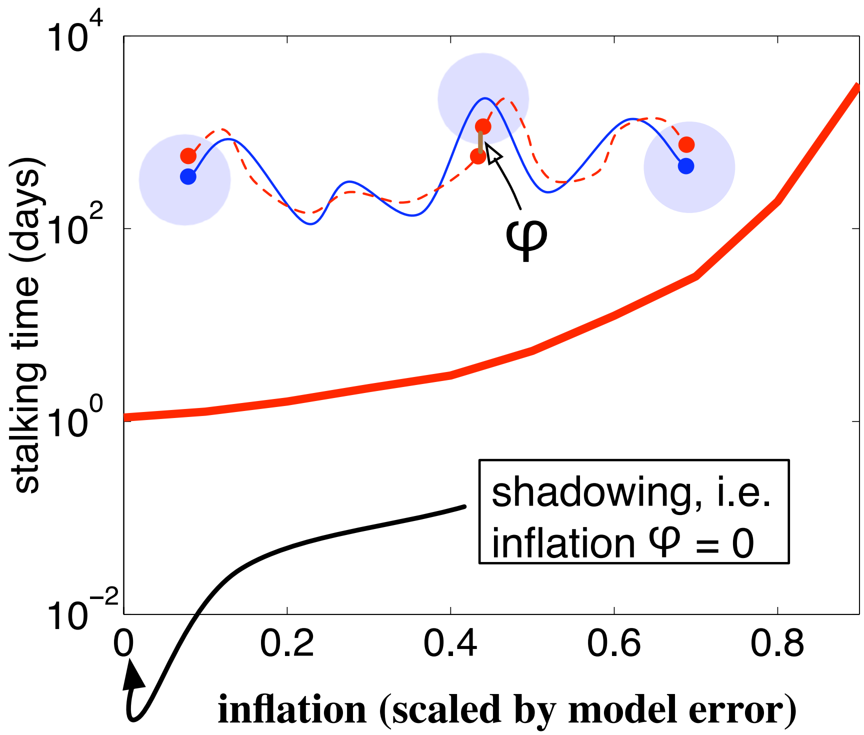


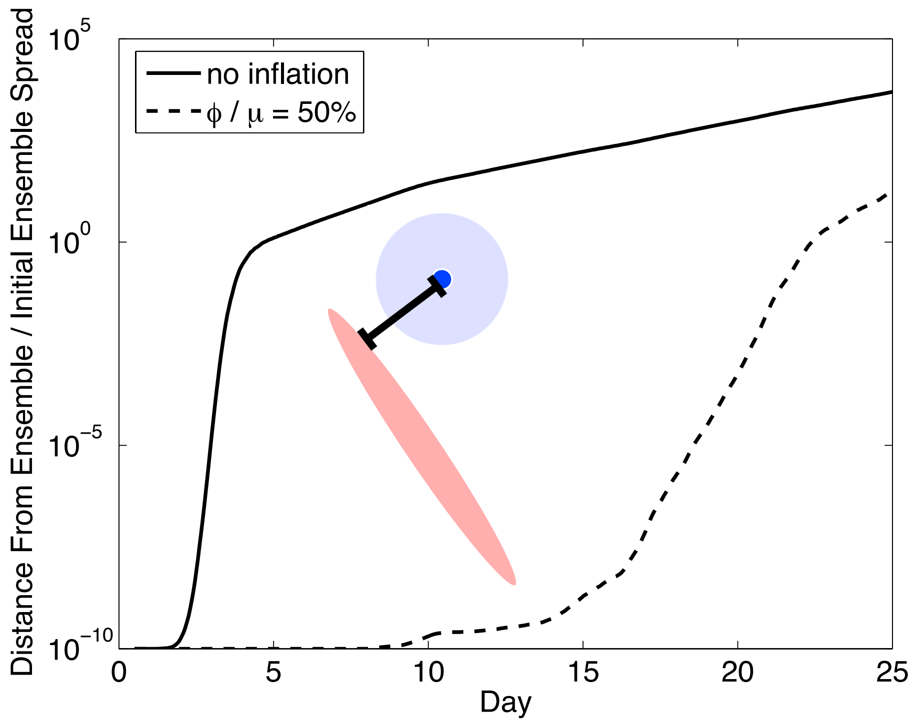
Stalking: inflate the cone, but only in dimensions whose uncertainty is currently shrinking with time.

Shadowing fails (top), stalking succeeds (bottom)



Stalking in Lorenz '96 Model





Danforth and Yorke. *Making Forecasts for Chaotic Physical Processes*. Physical Review Letters, 2006.

Three Experiments

- An experimental analog to Lorenz's 1963 model
- Stalking observations with a numerical trajectory
- *Online* empirical correction of model error

Empirical Model Error Correction Background

Leith (1978), first to formulate state-dependent correction procedure.

Given a model $\dot{\mathbf{x}} = \mathrm{M}\left(\mathbf{x}\right)$

Empirical Model Error Correction Background

Leith (1978), first to formulate state-dependent correction procedure.

Given a model $\dot{\mathbf{x}} = \mathrm{M}\left(\mathbf{x}\right)$ and a "true" solution $\mathbf{x}^{\mathrm{t}}$ (e.g. reanalysis)

Empirical Model Error Correction Background

Leith (1978), first to formulate state-dependent correction procedure.

Given a model $\dot{\mathbf{x}} = \mathrm{M}(\mathbf{x})$ and a "true" solution $\mathbf{x}^{\mathrm{t}}$ (e.g. reanalysis)

- sought an improved model of the form: $\dot{\mathbf{x}} = \mathrm{M}(\mathbf{x}) + \mathbf{L}\mathbf{x} + \mathbf{c}$

Empirical Model Error Correction Background

Leith (1978), first to formulate state-dependent correction procedure.

Given a model $\dot{\mathbf{x}} = \mathrm{M}(\mathbf{x})$ and a "true" solution $\mathbf{x}^{\mathrm{t}}$ (e.g. reanalysis)

- sought an improved model of the form: $\dot{\mathbf{x}} = \mathrm{M}(\mathbf{x}) + \mathrm{L}\mathbf{x} + \mathbf{c}$
  where $\mathbf{c}$ corrects the state-independent model error (bias)

Empirical Model Error Correction Background

Leith (1978), first to formulate state-dependent correction procedure.

Given a model $\dot{\mathbf{x}} = \mathrm{M}(\mathbf{x})$ and a "true" solution $\mathbf{x}^{\mathrm{t}}$ (e.g. reanalysis)

- sought an improved model of the form: $\dot{\mathbf{x}} = \mathrm{M}(\mathbf{x}) + \mathrm{L}\mathbf{x} + \mathbf{c}$
  where $\mathbf{c}$ corrects the state-independent model error (bias)
  and $\mathrm{L}\mathbf{x}$ corrects the state-dependent model error

Empirical Model Error Correction Background

Leith (1978), first to formulate state-dependent correction procedure.

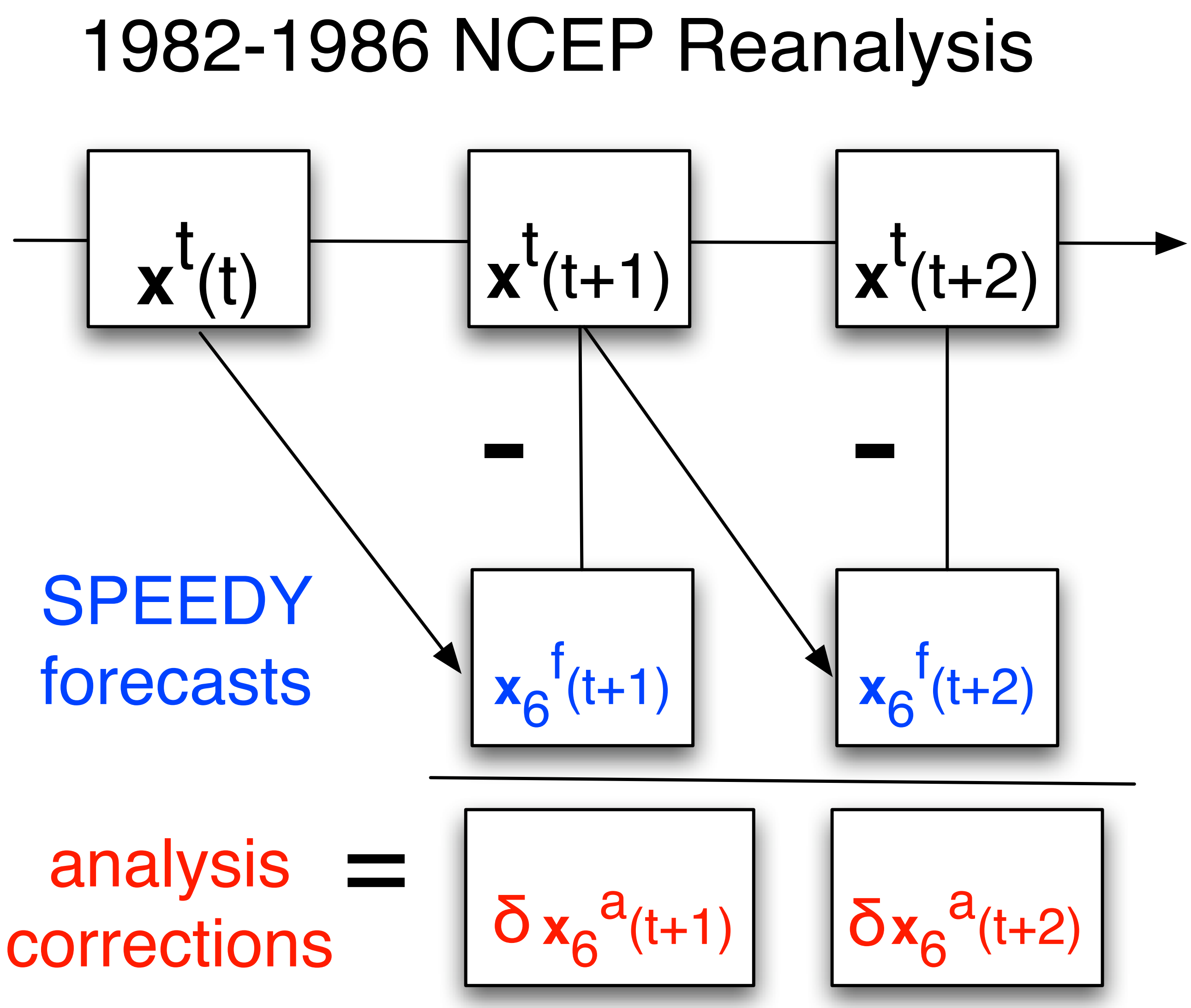
Given a model $\dot{\mathbf{x}} = \mathrm{M}\left(\mathbf{x}\right)$ and a "true" solution $\mathbf{x}^{\mathrm{t}}$ (e.g. reanalysis)

- sought an improved model of the form: $\dot{\mathbf{x}} = \mathrm{M}\left(\mathbf{x}\right) + \mathrm{L}\mathbf{x} + \mathbf{c}$
  where $\mathbf{c}$ corrects the state-independent model error (bias)
  and $\mathrm{L}\mathbf{x}$ corrects the state-dependent model error
- by minimizing the mean square tendency error of the improved model, $< \mathbf{g}^{\top}\mathbf{g} >$ where

$$\mathbf{g} = \dot{\mathbf{x}}^{\mathrm{t}} - \left(\mathrm{M}\left(\mathbf{x}^{\mathrm{t}}\right) + \mathrm{L}\mathbf{x}^{\mathrm{t}} + \mathbf{c}\right)$$

with respect to L and $\mathbf{c}$.

Generating Time Series of Forecasts and Errors



I. State-Dependent Error Estimation

Leith (1978) Empirical Correction Operator

- Forecast state covariance: $C_{x_6^f x_6^f} = < \mathbf{x}_6^{f\prime}\, \mathbf{x}_6^{f\prime\top} >$
- Correction & forecast state cross covariance: $C_{\delta x_6^a x_6^f} = < \delta\mathbf{x}_6^{a\prime}\, \mathbf{x}_6^{f\prime\top} >$

I. State-Dependent Error Estimation

Leith (1978) Empirical Correction Operator

- Forecast state covariance: $C_{x_6^f x_6^f} = < \mathbf{x}_6^{f\prime} \, \mathbf{x}_6^{f\prime\top} >$
- Correction & forecast state cross covariance: $C_{\delta x_6^a x_6^f} = < \delta\mathbf{x}_6^{a\prime} \, \mathbf{x}_6^{f\prime\top} >$

Leith's correction operator, given by $L = C_{\delta x_6^a x_6^f} \, {C_{x_6^f x_6^f}}^{-1}$, provides a state-dependent correction:

$$\dot{\mathbf{x}} = M(\mathbf{x}) + \left[L\mathbf{x}' + \mathbf{c} \right] \frac{1}{6\mathrm{hr}}$$

where $\mathbf{c} = < \delta\mathbf{x}_6^a >$

I. State-Dependent Error Estimation

Leith (1978) Empirical Correction Operator

- Forecast state covariance: $C_{x_6^f x_6^f} =< \mathbf{x}_6^{f\prime}\ \mathbf{x}_6^{f\prime\top} >$
- Correction & forecast state cross covariance: $C_{\delta x_6^a x_6^f} =< \delta\mathbf{x}_6^{a\prime}\ \mathbf{x}_6^{f\prime\top} >$

Leith's correction operator, given by $\mathrm{L} = C_{\delta x_6^a x_6^f}\ C_{x_6^f x_6^f}{}^{-1}$, provides a state-dependent correction:

$$\dot{\mathbf{x}} = \mathrm{M}(\mathbf{x}) + \left[\mathrm{L}\mathbf{x}' + \mathbf{c}\right]\frac{1}{6\mathrm{hr}}$$

where $\mathbf{c} =< \delta\mathbf{x}_6^a >$

Problem: Direct computation of $\mathrm{L}\mathbf{x}'$ requires $O(\mathrm{N}^3)$ floating point operations *every* time step!

I. State-Dependent Error Estimation

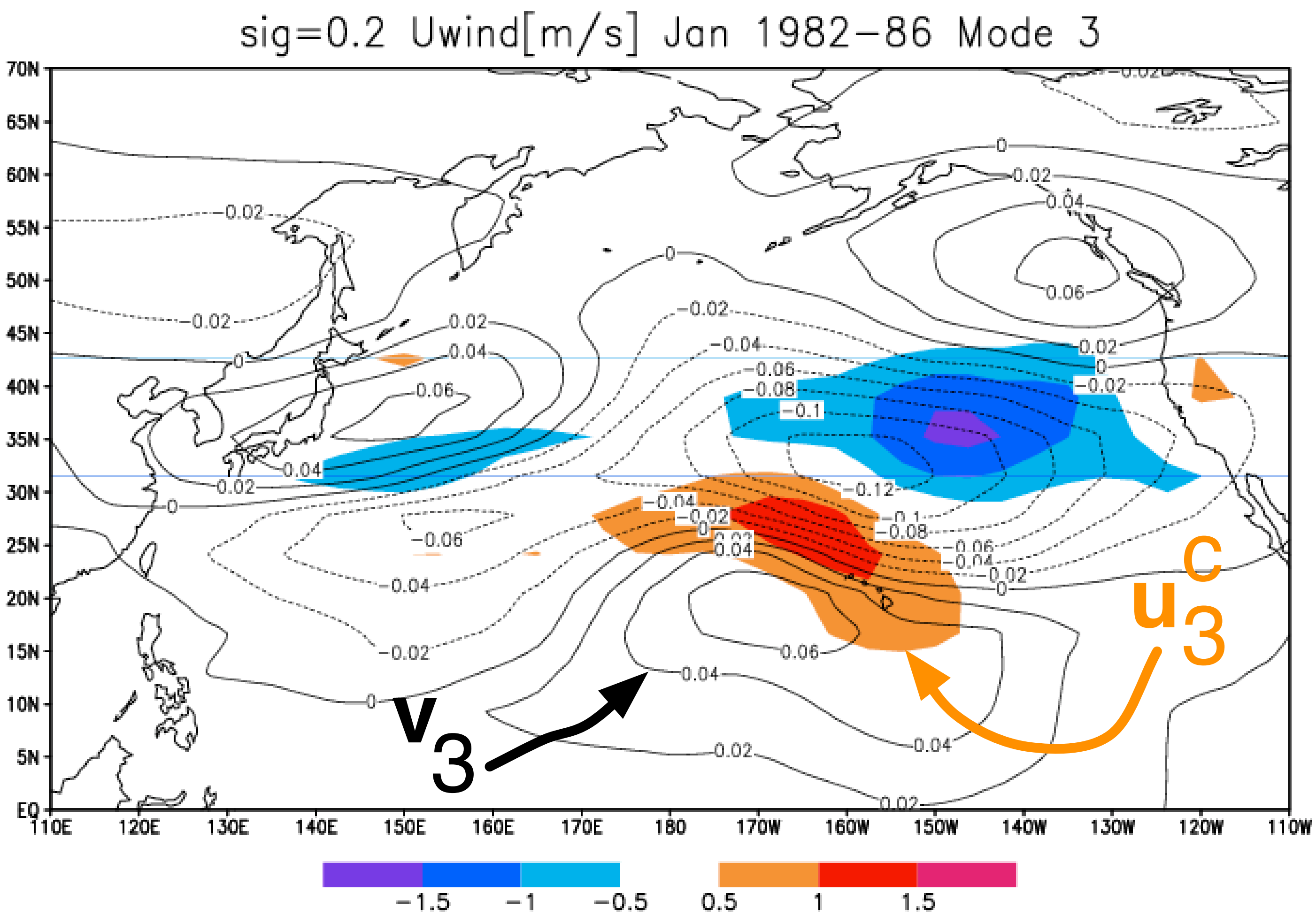
First step in our new approach:

Low-Dimensional Approximation based on regression

- Singular Value Decomposition (SVD) of the sparse analysis correction & state cross covariance: $C_{\delta \mathbf{x}_6^a \mathbf{x}_6^f} = U \Sigma V^\top$
- identifies pairs of spatial patterns or EOFs ($\mathbf{u}_k$ and $\mathbf{v}_k$) that explain as much of possible of the mean-squared temporal covariance between the analysis correction and state anomalies.

I. State-Dependent Error Estimation

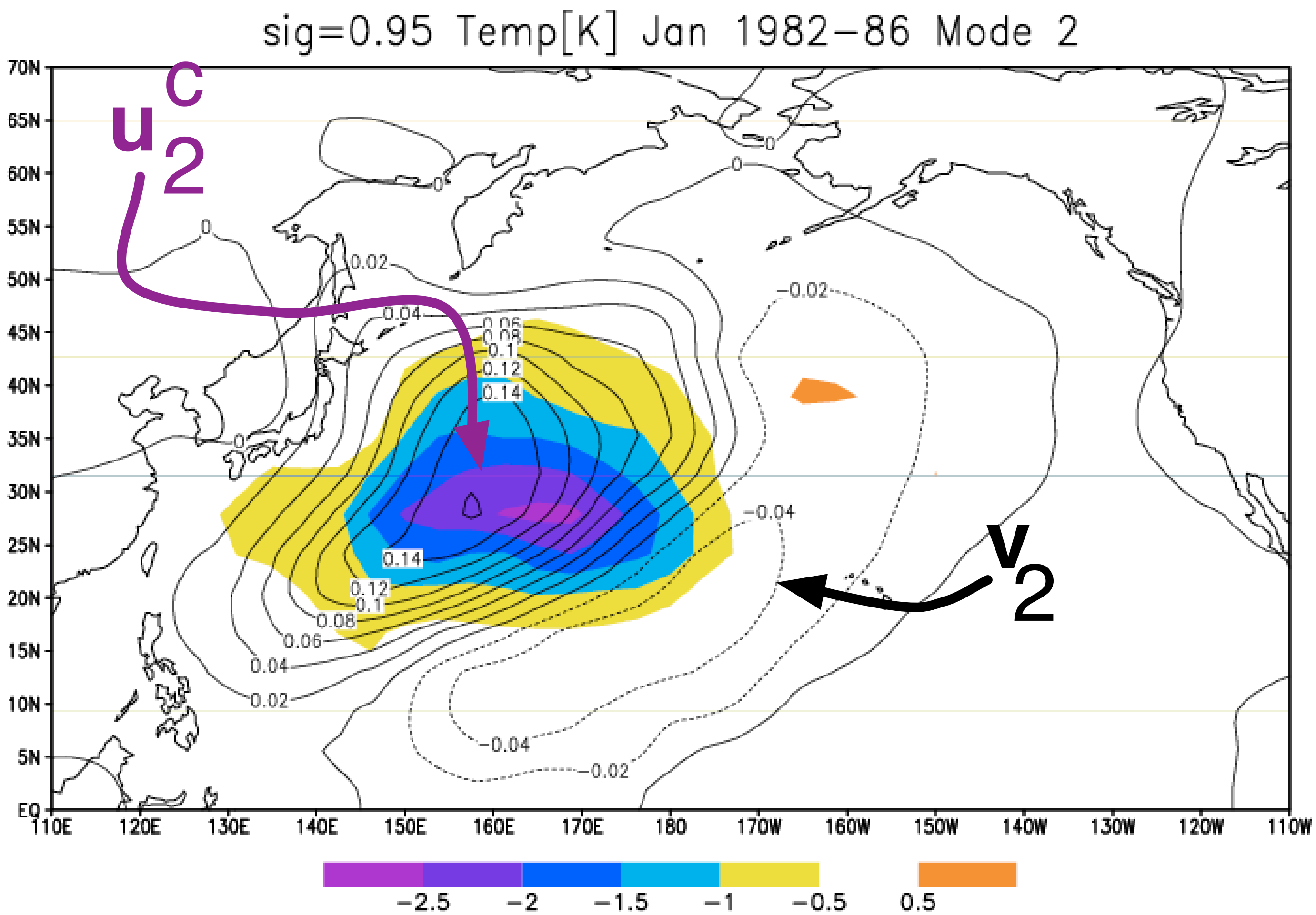
Correction (color) and state (contour) coupled signals



- $\mathbf{u}_3$ suggests shifting the anomaly $\mathbf{v}_3$ northeast (over the dependent sample)

I. State-Dependent Error Estimation

Correction (color) and state (contour) coupled signals



- $\mathbf{u}_2$ suggests damping the anomaly $\mathbf{v}_2$ (over the dependent sample)

I. State-Dependent Error Estimation

Second step in our new approach:

Leith's empirical correction involves solving $C_{x_6^f x_6^f}\mathbf{w} = \mathbf{x}'$ for $\mathbf{w}$ at each time step.

$$Lx' = C_{\delta x_6^a x_6^f}\, C_{x_6^f x_6^f}{}^{-1}\mathbf{x}'$$

I. State-Dependent Error Estimation

Second step in our new approach:

Leith's empirical correction involves solving $C_{x_6^f x_6^f}\mathbf{w} = \mathbf{x}'$ for $\mathbf{w}$ at each time step.

$$
\begin{aligned}
L\mathbf{x}' &= C_{\delta x_6^a x_6^f}\, {C_{x_6^f x_6^f}}^{-1}\mathbf{x}' \\
&= C_{\delta x_6^a x_6^f}\mathbf{w}
\end{aligned}
$$

I. State-Dependent Error Estimation

Second step in our new approach:

Leith's empirical correction involves solving $C_{x_6^f x_6^f}\mathbf{w} = \mathbf{x}'$ for $\mathbf{w}$ at each time step.

$$\begin{aligned} L\mathbf{x}' &= C_{\delta x_6^a x_6^f}\, {C_{x_6^f x_6^f}}^{-1}\mathbf{x}' \\ &= C_{\delta x_6^a x_6^f}\mathbf{w} \\ &= U\Sigma V^\top \mathbf{w} \end{aligned}$$

I. State-Dependent Error Estimation

Second step in our new approach:

Leith's empirical correction involves solving $C_{x_6^f x_6^f}\mathbf{w} = \mathbf{x}'$ for $\mathbf{w}$ at each time step.

$$\begin{aligned} L\mathbf{x}' &= C_{\delta x_6^a x_6^f}\, {C_{x_6^f x_6^f}}^{-1}\mathbf{x}' \\ &= C_{\delta x_6^a x_6^f}\mathbf{w} \\ &= U\Sigma V^{\top}\mathbf{w} \\ &\approx \sum_{k=1}^{K} \mathbf{u}_k \sigma_k \mathbf{v}_k^{\top} \cdot \mathbf{w} \end{aligned}$$

I. State-Dependent Error Estimation

Second step in our new approach:

Leith's empirical correction involves solving $C_{x_6^f x_6^f}\mathbf{w} = \mathbf{x}'$ for $\mathbf{w}$ at each time step.

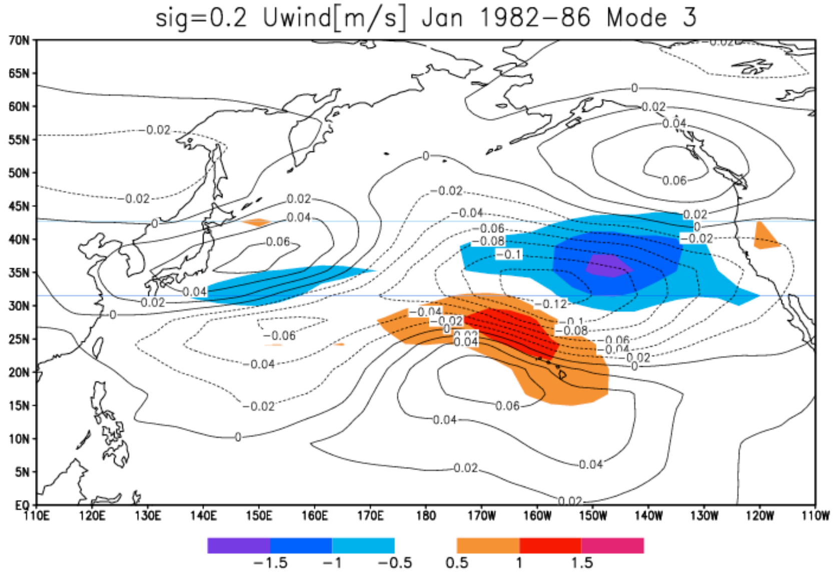
$$\begin{aligned} L\mathbf{x}' &= C_{\delta x_6^a x_6^f}\, C_{x_6^f x_6^f}{}^{-1}\mathbf{x}' \\ &= C_{\delta x_6^a x_6^f}\mathbf{w} \\ &= U\Sigma V^\top \mathbf{w} \\ &\approx \sum_{k=1}^{K} \mathbf{u}_k \sigma_k \mathbf{v}_k^\top \cdot \mathbf{w} \end{aligned}$$

However, only the component of $\mathbf{w}$ in the space spanned by the right singular vectors $\mathbf{v}_k$ can contribute to the empirical correction!!

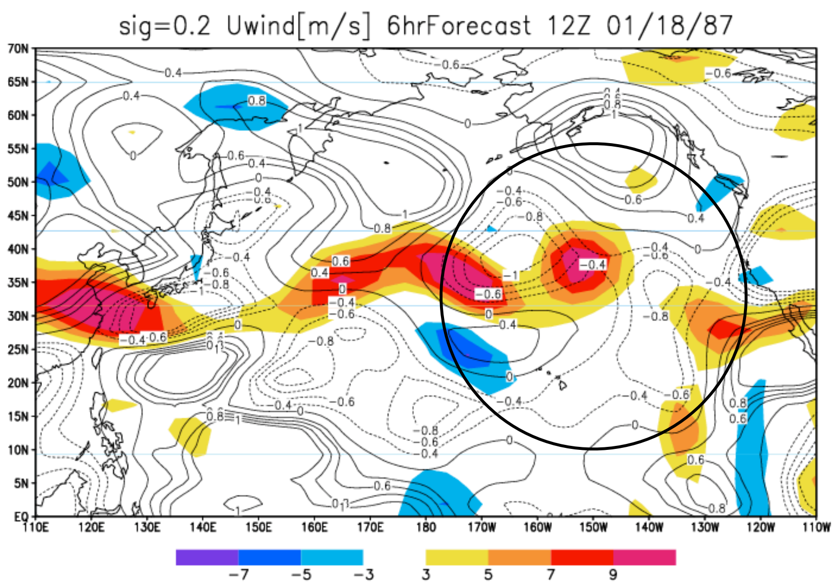
II. State-Dependent Correction

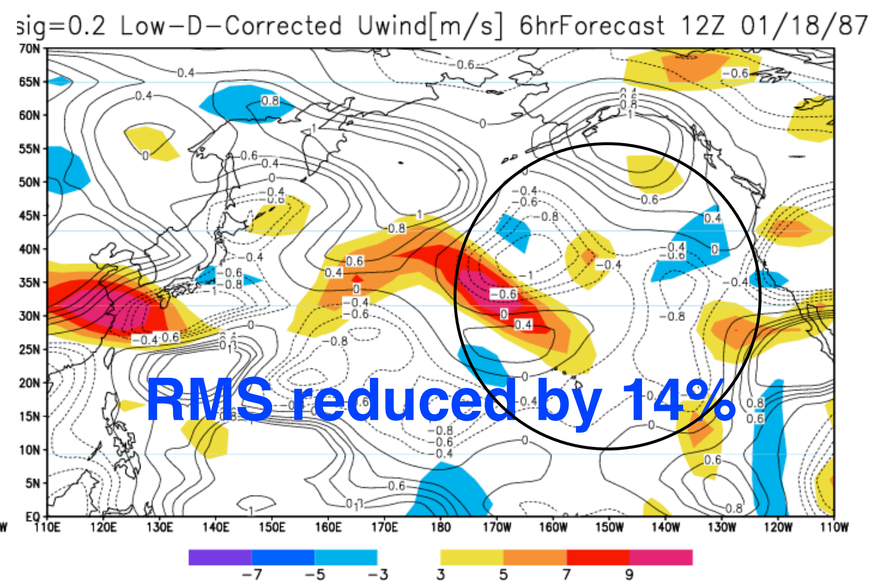
σ=0.2 U-wind
Error (shades)
and State (contour)

6-hr forecasts



debiased low-d corrected

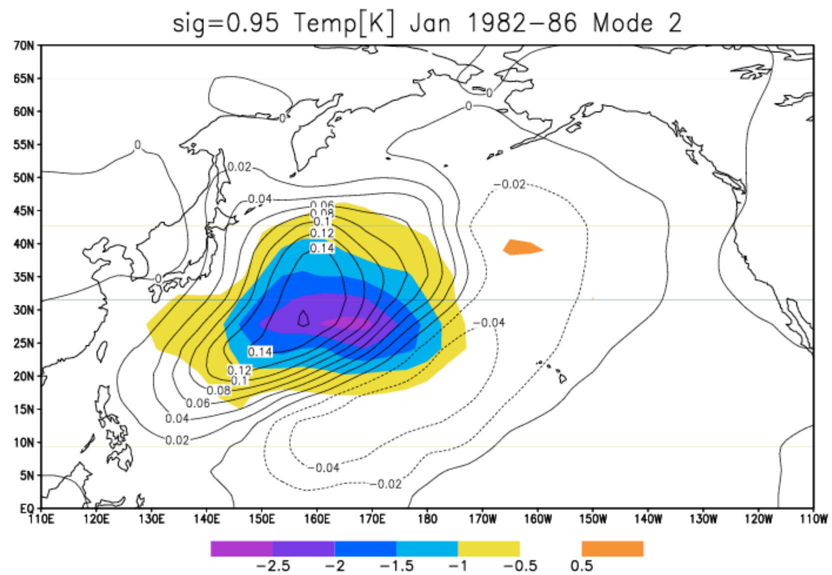




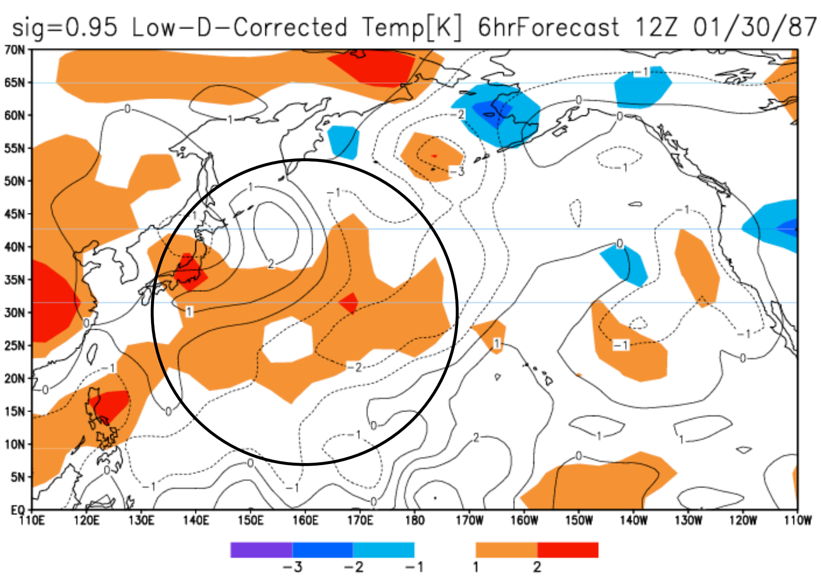
II. State-Dependent Correction

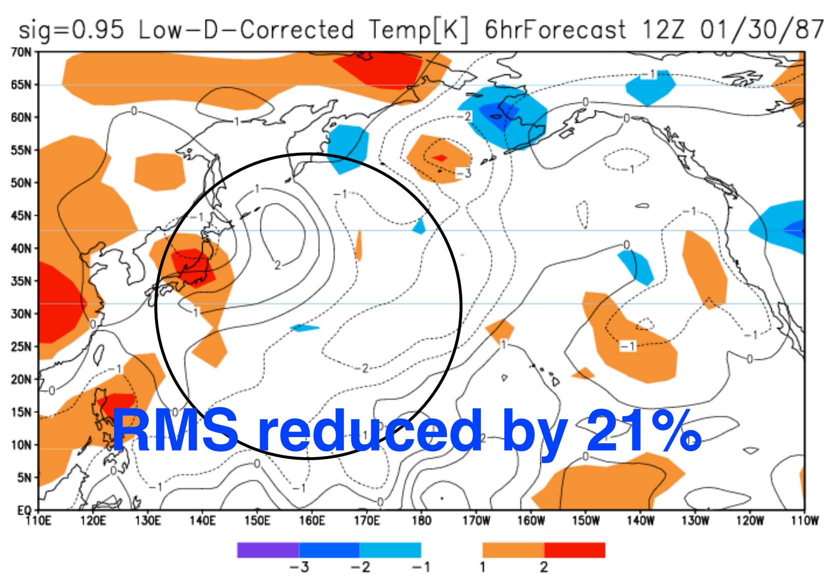
σ=0.95 Temp
Error (shades)
and State (contour)

6-hr forecasts



debiased   low-d corrected





Model Error Correction Results

- 3-day RMSE of online corrected model equal to 1-day RMSE of original (better than offline correction)
- Climate statistics of model are improved
- SVD modes *may* suggest physically meaningful errors
- Works easily with existing data assimilation and ensemble schemes (requires only the analysis increments for sampling)
- Techniques could be used to improve model predictions of *any* physical system.

Danforth, Kalnay, Miyoshi. *Estimating and Correcting Global Weather Model Error*. Monthly Weather Review, 2007.

Danforth and Kalnay. *Using Singular Value Decomposition to Parameterize State-Dependent Model Error*. Journal of the Atmospheric Sciences, 2008. (Lorenz '96 model)

Danforth and Kalnay. *The Impact of Online Empirical Model Correction on Nonlinear Error Growth*. Geophysical Research Letters, submitted.

Questions?

Thanks to ...

<u>University of Maryland</u>
Eugenia Kalnay, Atmospheric Science
James A. Yorke, Math and Physics
Robert F. Cahalan, NASA GFSC
Takemasa Miyoshi, JMO
Chaos group (Hunt, Kalnay, Kostelich, Ott, Patil, Sauer, Szunyogh, Yorke)

<u>Bates College</u>
Chip Ross, Math
Mark Semon, Physics
George Ruff, Physics

<u>University of Vermont</u>
Darren Hitt, Mechanical Engineering
Floyd Vilmont, Lab

<u>Students</u>
Kameron Harris, Undergrad
Nicholas Allgaier, Graduate
El Hassan Ridouane, Postdoc

Contact: chris.danforth@uvm.edu