Real Time Series analysis and modelling

Aid machine learning with dynamical insight

Hailiang Du
Department of Mathematical Sciences, Durham University
hailiang.du@durham.ac.uk
Data Science Institute, London School of Economics and Political Science
h.l.du@lse.ac.uk

All theorems are true, All models are wrong. All data are inaccurate. What are we to do?

The aim of this course is to teach you how to deal with real data, to increase your scepticism regarding reliable modelling in practice, and to expand the tool box you carry to include nonlinear techniques, both deterministic and stochastic with the aid of dynamical insight.

In short: to get you to think before you compute (and perhaps afterwards too.)

Hailiang Du
Department of Mathematical Sciences, Durham University
hailiang.du@durham.ac.uk
Data Science Institute, London School of Economics and Political Science
h.l.du@lse.ac.uk

About me…

My Research Interests

- Chaos (theory of nonlinear dynamics)

  ◇ Data Assimilation

  ◇ Parameter Estimation

  ◇ Predictability

  ◇ Model error correction

My Research Interests

- Forecast Interpretation and Evaluation

  ◇ (Multi-model) Ensemble Forecasts

  ◇ From Ensemble to PDF

  ◇ Skill Scores

  ◇ Imprecise Probability/Nonprobability odds

My Research Interests

- Uncertainty Quantification

  ◇ Emulation

  ◇ Bayesian Linear Analysis

  ◇ History Matching (Model Calibration)

  ◇ (Energy system) optimization

My Research Interests

- Machine Learning

  ◇ Tree-based methods

  ◇ Radio basis model

  ◇ Deep Learning/Neural Networks

Questions are always welcome!

Feel free to correct my English and Chinese.

Lecture 1

Overview, Motivation and Convention

Stochastic vs Deterministic, Linear vs Nonlinear

Overview of Course

- Developing a (mathematical) language to talk about physical processes (stock markets, weather, pendulums) – *dynamical systems theory*

- Develop a statistical toolbox with which to analyse data from physical processes –*time series analysis*

- Introduction to a selection of (machine learning) modelling techniques – *time series modelling*

- Overview of forecast interpretation and *evaluation* methods

- Overview of *uncertainty quantification* and approaches to address *model discrepancy*

we will discuss and criticise aspects across the board.

What do I want you to take away from this course?

- A more general, geometric, view of the analysis of time series.

- An expanded toolkit for analysing data from nonlinear systems, nonlinear models of actual systems, actual observations and forecasts.

- A focus on information content and model inadequacy.

- A better understanding of
  - deterministic nonlinear dynamical system, its descriptive statistics, and relevance to real-world modelling.
  - the implications nonlinearity holds for statistical analysis, model building and forecasting of actual systems.
  - the interpretation and use of probabilistic forecasts
  - the different beliefs as to what a “model” is and should be.

be more vigilant!

The Modeller’s Mantra 口头禅

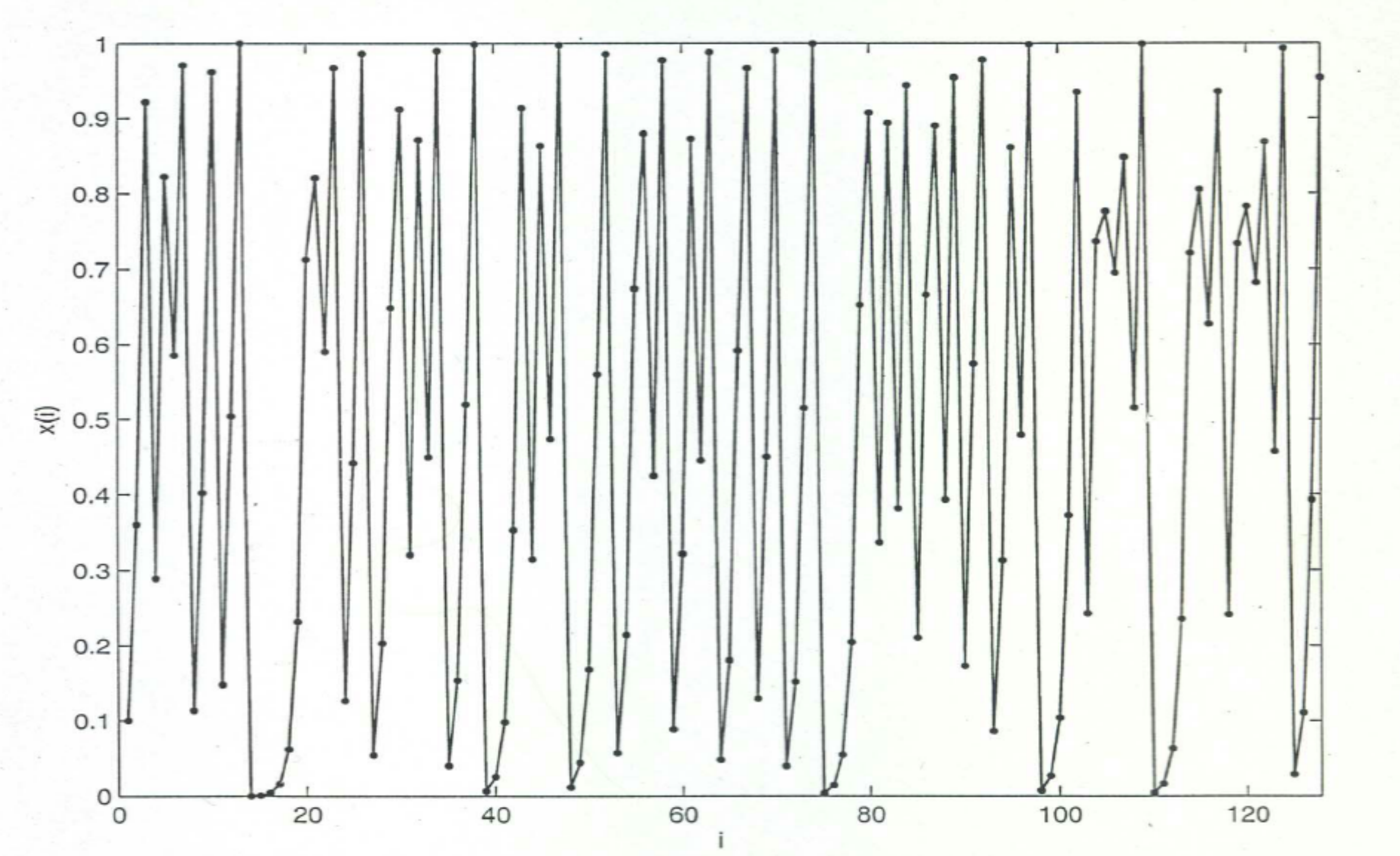
This is the best available information, so it must be of value.
Everyone knows the limitations. Everyone understands the implications of these assumptions.
This is better than nothing.
No one has proven this is wrong.
There is no systematic error, on average. The systematic errors don't matter.
The systematic errors are accounted for in the post processing.
Normality is always a good first approximation. In the limit, it has to be normally distributed, at least approximately.
Everyone assumes it is normally distributed to start with.
Everyone makes approximations like that.
Everyone makes this approximation.
The users demand this. The users will not listen to us unless we give them the level of detail they ask for.
If we do not do this, the user will try and do it themselves.
There is a commercial need for this information, and it is better supplied by us than some cowboy.
Refusing to answer a question is answering the question.
Refusing to use a model is still using a model.
No model is perfect.
No model is useless if interpreted correctly. It is easy to criticise.
Sure this model is not perfect, but it is not useless.
This model is based on fundamental physics.
The probabilities follow from the latest developments in Bayesian statistics.
Think of the damage a decision maker might do without these numbers.
Any rational user will agree.
Things will get better with time, we are making real progress.
You have to start somewhere. What else can we do? It might work, can you deny that?
What damage will it do?
We are following the science.

So, what is a:

Real Time Series analysis and modelling

Aid machine learning with dynamical insight

X(t)



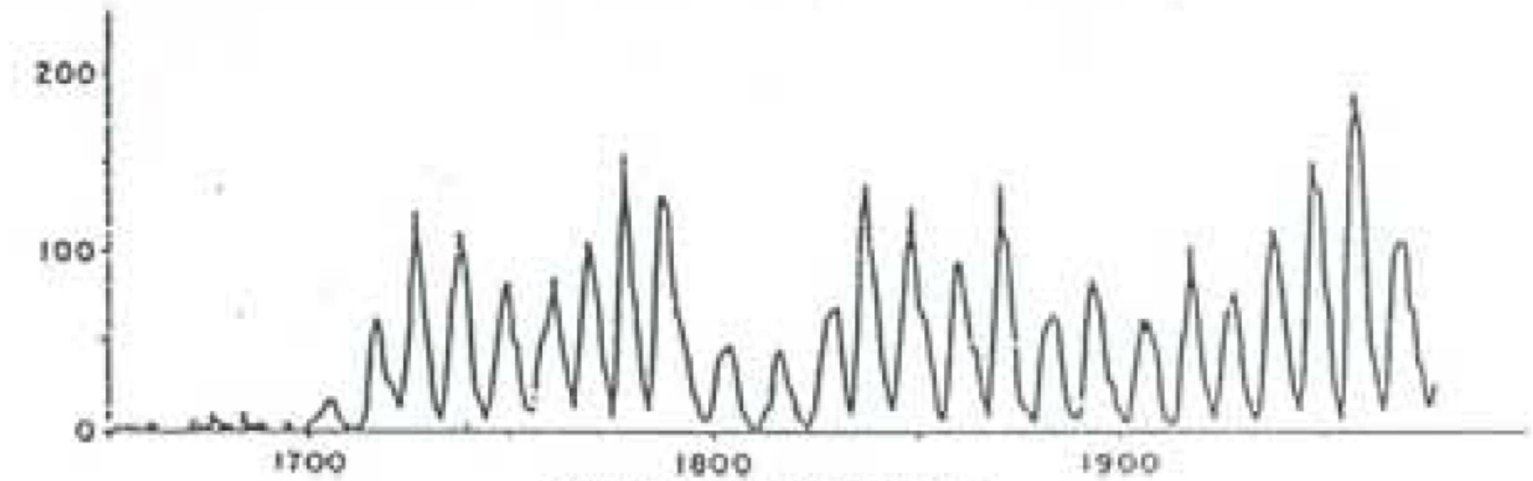
time: could be “t” or “i”

a series of data points indexed in time order

So, what is a:

Real Time Series analysis and modelling

Aid machine learning with dynamical insight

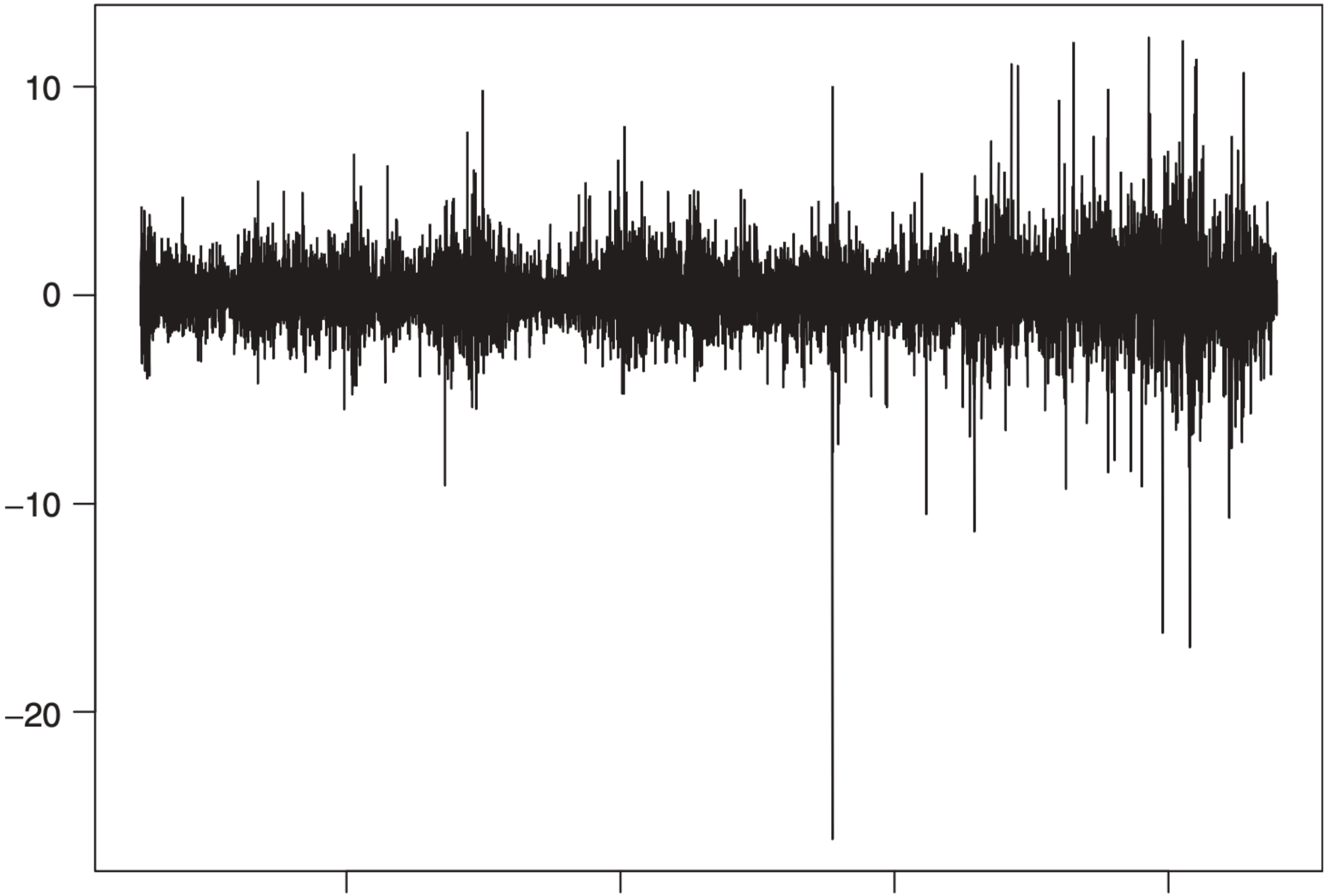


How many of you recognise this time series?

So, what is a:

Real Time Series analysis and modelling

Aid machine learning with dynamical insight



How many of you recognise this time series?

More about

Real Time Series analysis and modelling

Aid machine learning with dynamical insight

When we make an observation, the measurement is never exact in a mathematical sense, so there is always some uncertainty as to the “true” value.

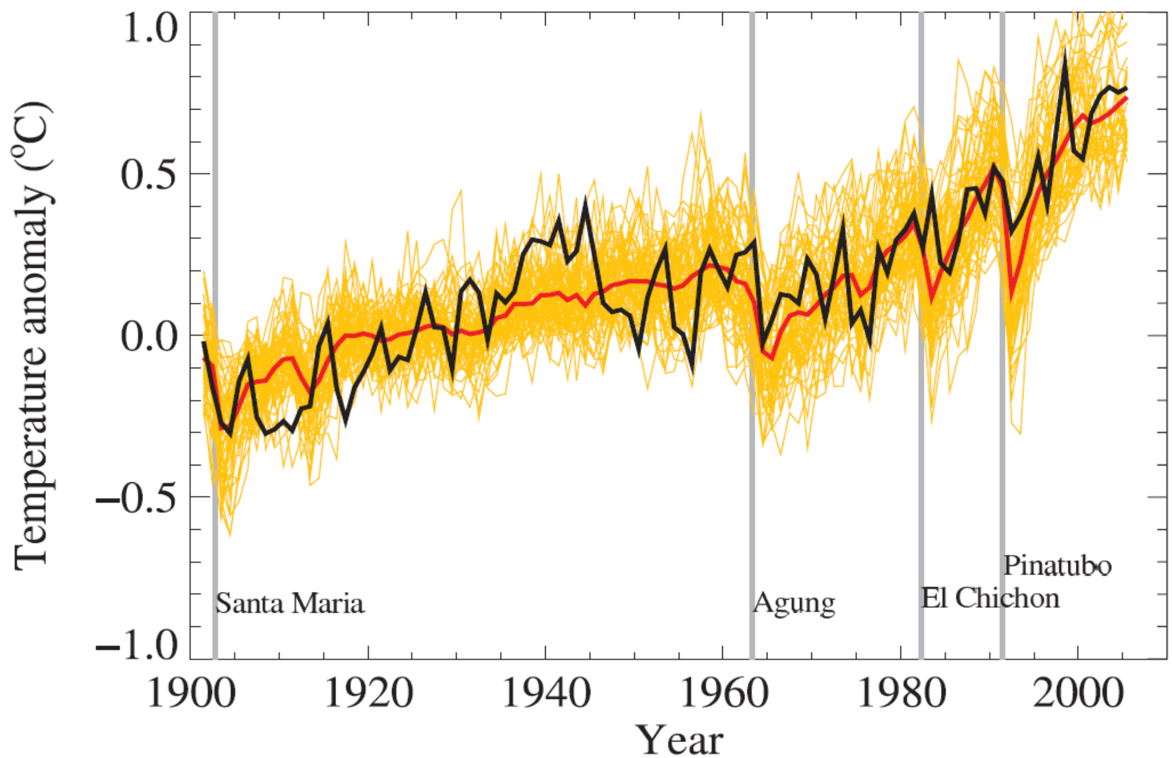
It is important not to confuse the system (that generates the observation), the time series itself (a collection of numbers which are the observation), the model we employ, and potential time-series generated by the model.

So, what is a:

Real Time Series analysis and modelling

Aid machine learning with dynamical insight

FAQ 8.1, Figure 1. *Global mean near-surface temperatures over the 20th century from observations (black) and as obtained from 58 simulations produced by 14 different climate models driven by both natural and human-caused factors that influence climate (yellow). The mean of all these runs is also shown (thick red line). Temperature anomalies are shown relative to the 1901 to 1950 mean. Vertical grey lines indicate the timing of major volcanic eruptions. (Figure adapted from Chapter 9, Figure 9.5. Refer to corresponding caption for further details.)*



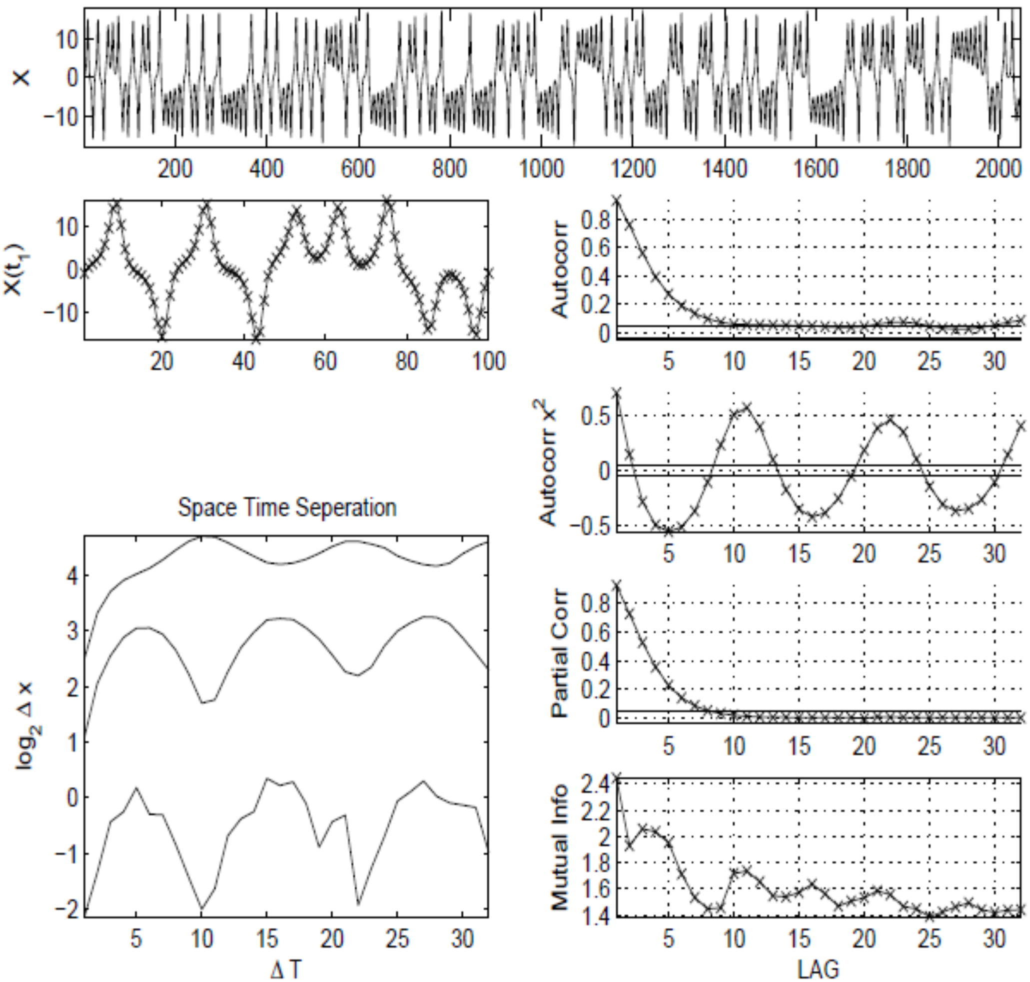
Time series are ubiquitous (无处不在).

We can have time series of observations, or model simulations, or both, and even both for the same “time”.

So, what is:

Real Time Series analysis and modelling

Aid machine learning with dynamical insight



The analysis of time series can aim to understand, to model, to forecast, and/or to provide decision support. These different tasks require different tools.

Questions of analysing the past are rather different than those of predicting the future.

to learn the underlying system

So, what is:

Real Time Series analysis and modelling

Aid machine learning with dynamical insight

A fancy name of statistical learning…..

Thanks to the development of computer science

So, what is:

Real Time Series analysis and modelling

Aid machine learning with dynamical insight

what is dynamics?

The “rules” that specify how the dynamical system evolves; how the system changes from one state to the next

Stock Market

Laws of Physics

Mathematical functions

An example of

Real Time Series analysis and modelling

Aid machine learning with dynamical insight

Vulcan dining series

Breakfast → Supper → Lunch → Supper → Lunch → Supper → Breakfast →

→ Supper → Breakfast → Supper → Lunch → Supper → Breakfast → Supper →

→ Lunch → Supper → Breakfast → ????

What's the unconditional distribution?

So, what is the underlying system of a:

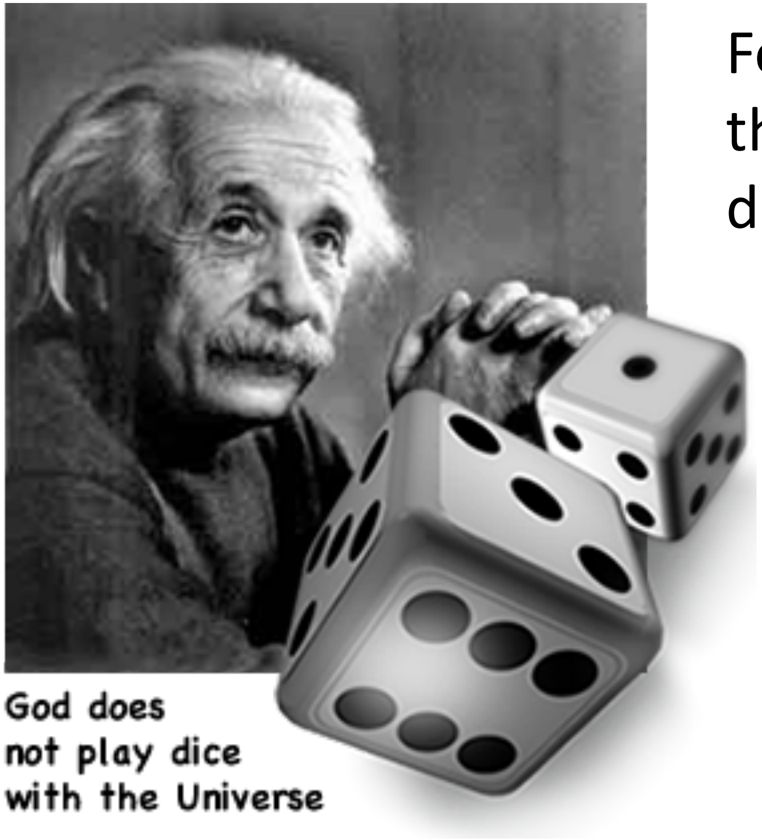
Real Time Series analysis and modelling

Aid machine learning with dynamical insight

Stochastic dynamical system evolves randomly

For deterministic dynamical system, the dynamics and initial condition define the future state unambiguously.

Deterministic vs Stochastic



For deterministic dynamical system, the dynamics and initial condition define the future state unambiguously.

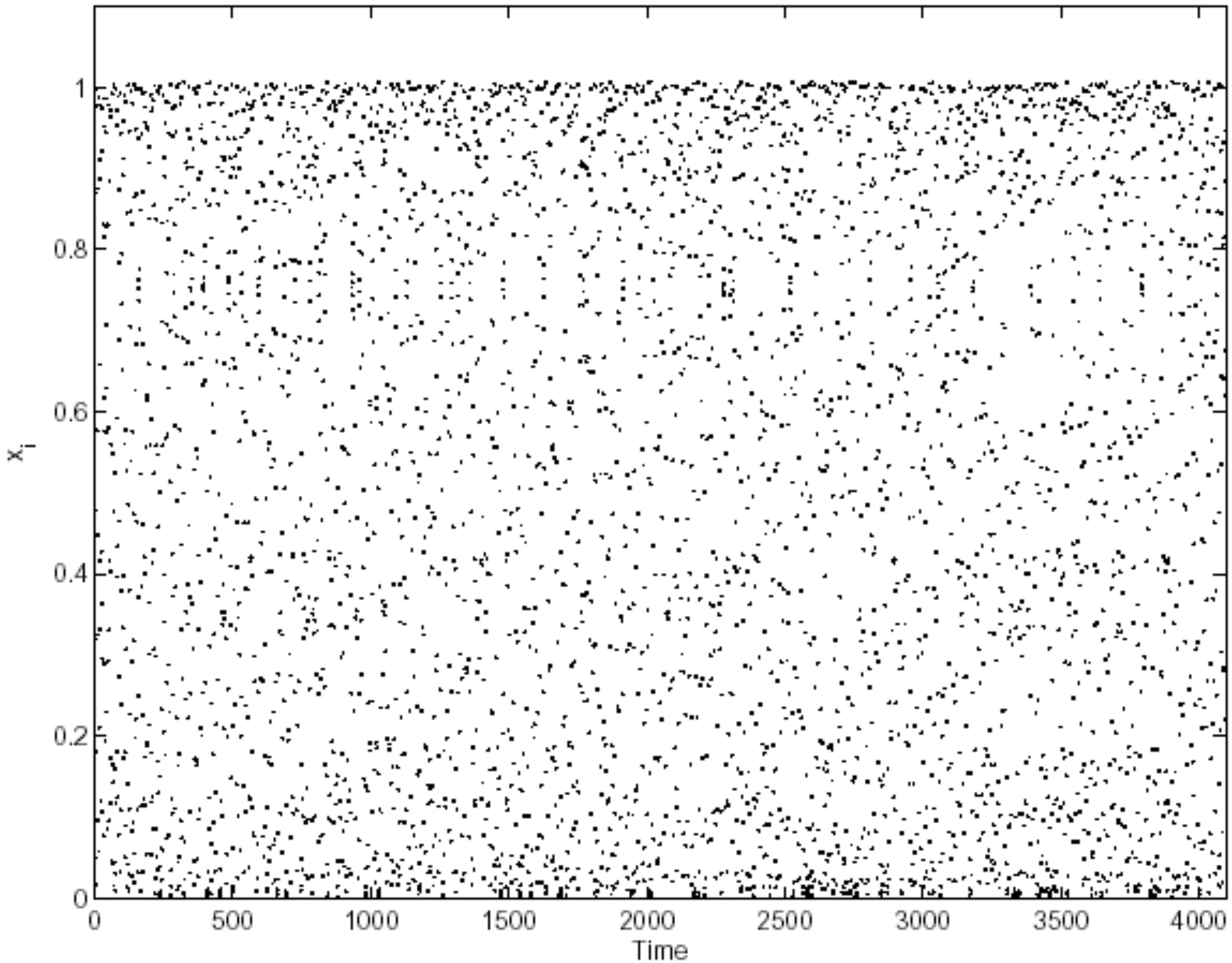
How would I play dice?

Deterministic vs Stochastic

"Random"?

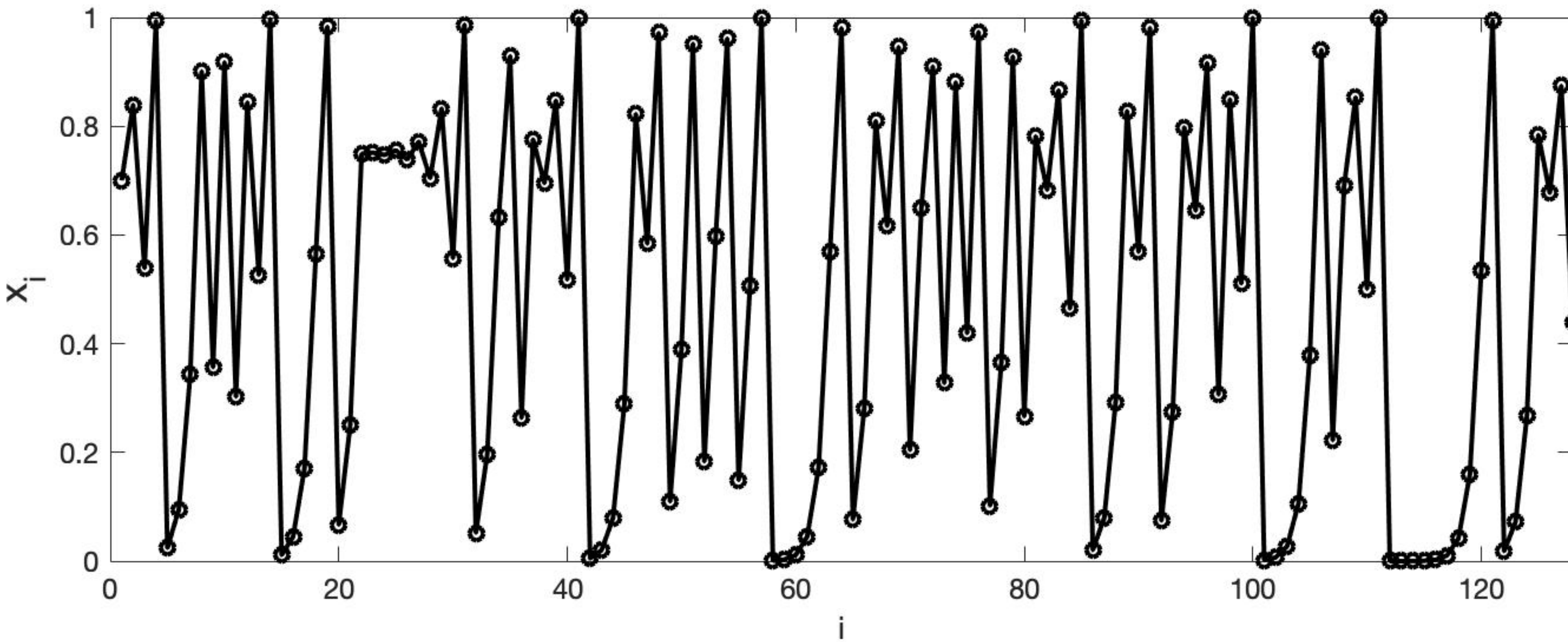
"Unpredictable"?

Deterministic vs Stochastic



Random?

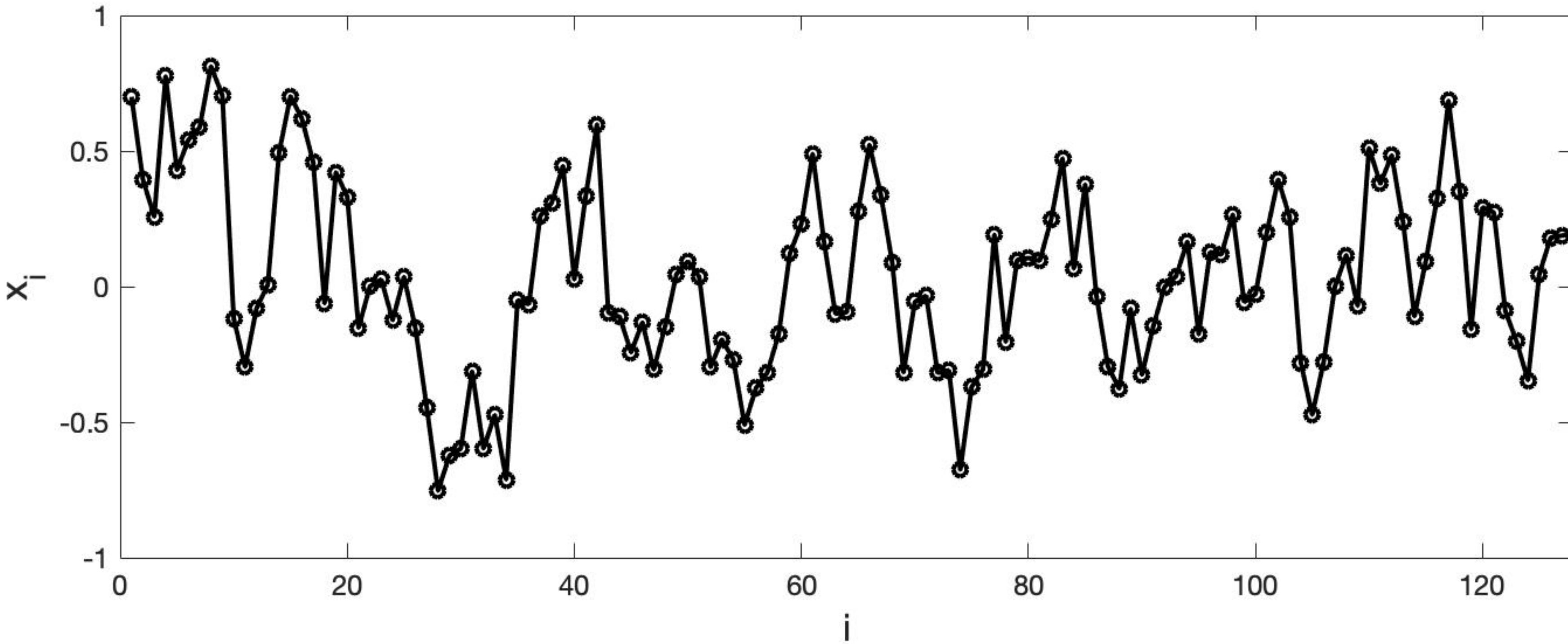
Deterministic vs Stochastic



Population model: $x_{i+1} = 4\, x_i\, (1 - x_i)$

NOT random!

Deterministic vs Stochastic



AR(1) model: $x_{i+1} = 0.7x_i + \epsilon_i$, $\epsilon \sim N(0,0.25^2)$

Stochastic

Dynamical Systems Jargon

State: the configuration of the dynamical system at any time, where it is and what it is doing.

$$\mathbf{x} = (x_1, x_2, ..., x_n)$$

State space: the collection of all possible states $\mathbf{x} \in \mathbb{S}$

Dynamics: the set of rules that determine the evolution of the state

$$\mathbf{x}_t = F^t(\mathbf{x}_0)$$

$\mathbf{x}_0$ is the starting state called the initial condition

Map: discrete dynamical system takes place at regular time intervals, for example $t \in \mathbb{Z}$

Flow: continuous dynamical system whose evolution takes place continuously

$$\frac{d\mathbf{x}}{dt} = F(\mathbf{x}) \qquad t \in \mathbb{R}$$

Stochastic dynamical system evolves randomly, for example AR(1)

For deterministic dynamical system, the dynamics and initial condition define the future state unambiguously.

More Dynamical Systems Jargon

State: the configuration of the dynamical system at any time, where it is and what it is doing.

$$\mathbf{x} = (x_1, x_2, \ldots, x_n)$$

State space: the collection of all possible states $\mathbf{x} \in \mathbb{S}$

Dynamics: the set of rules that determine the evolution of the state

$$\mathbf{x}_t = F^t(\mathbf{x}_0)$$

$\mathbf{x}_0$ is the starting state called the initial condition

Observation: some function of the state

for example: $\mathbf{s}_t = \mathbf{x}_t + \epsilon_t$ Not true in reality

Observational noise only effects observations

Dynamical noise effects both observations and states $\mathbf{x}_t = F^t(\mathbf{x}_0, \delta_t)$

$$x_{t+1} = x_t + \delta_t$$

So, what is:

Linear vs Nonlinear

I am going to cheat a bit on this one, but honestly, and define nonlinear as anything that is not linear.

A linear dynamics is one for which the evolution depends only linearly on the variables. Its complement defines nonlinear dynamics.

That is fair in a way, but imagine trying to develop the biology of “non-elephants”

Why statisticians loves linear system so much?

under linear assumption, Gaussian uncertainty remains Gaussian

Linear Statistical tools can be misinterpreted

Correlation “score”: sample correlation coefficient:

Take values between -1 and 1

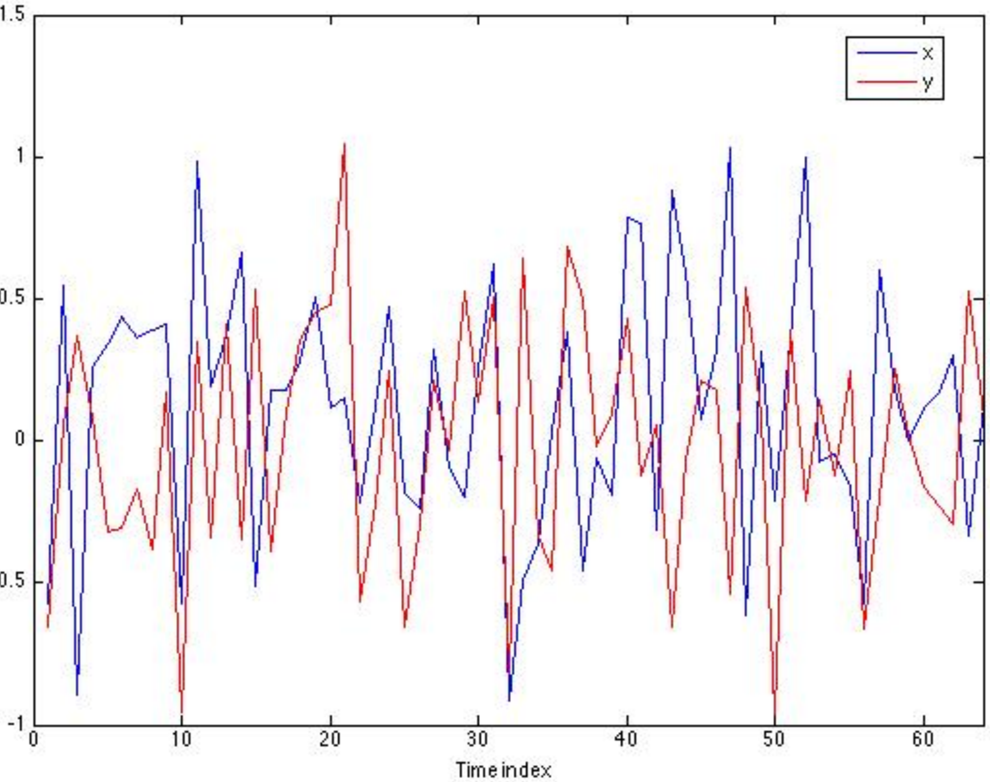
$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}},$$

There continues to be very high confidence1 that models reproduce observed large-scale mean surface temperature patterns (pattern correlation of ~0.99), though systematic errors of several degrees are found in some regions, particularly over high topography, near the ice edge in the North Atlantic, and over regions of ocean upwelling near the equator. (IPCC P.743)

“The correlation skill in XXX region remains moderate, varying from 0.3 to 0.5”

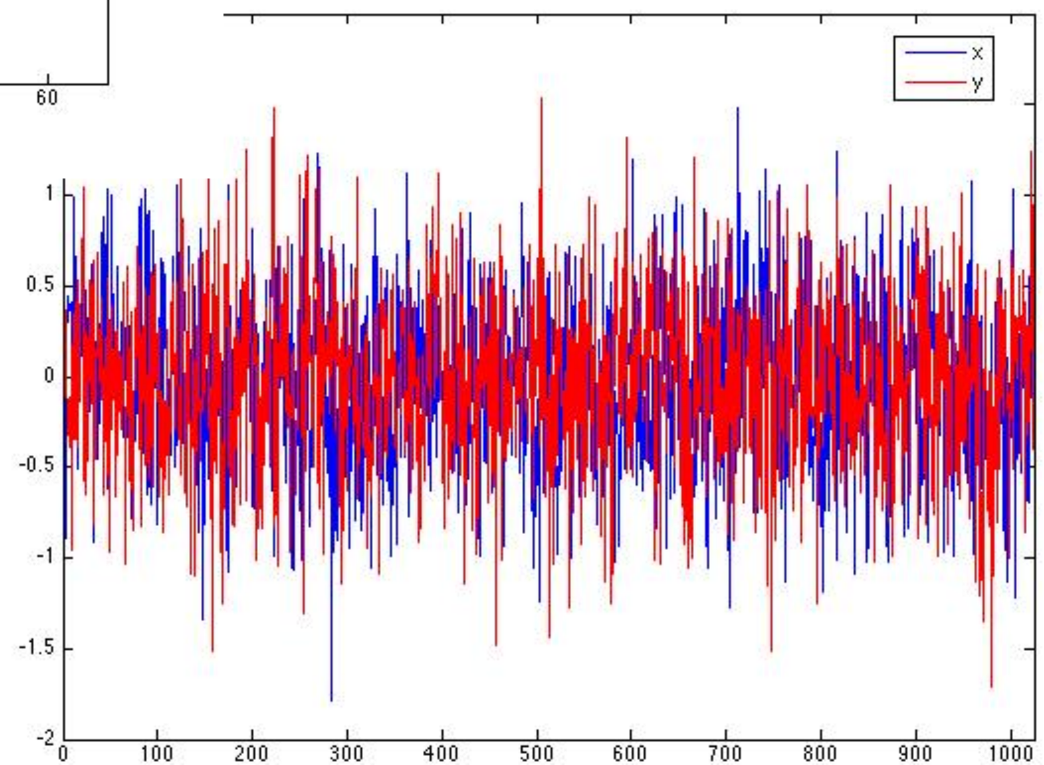
“a forecast with an anomaly correlation that exceeds 0.6 in xxxx places xxxx roughly in the correct position and thus provides a useful guidance on xxxxx. “

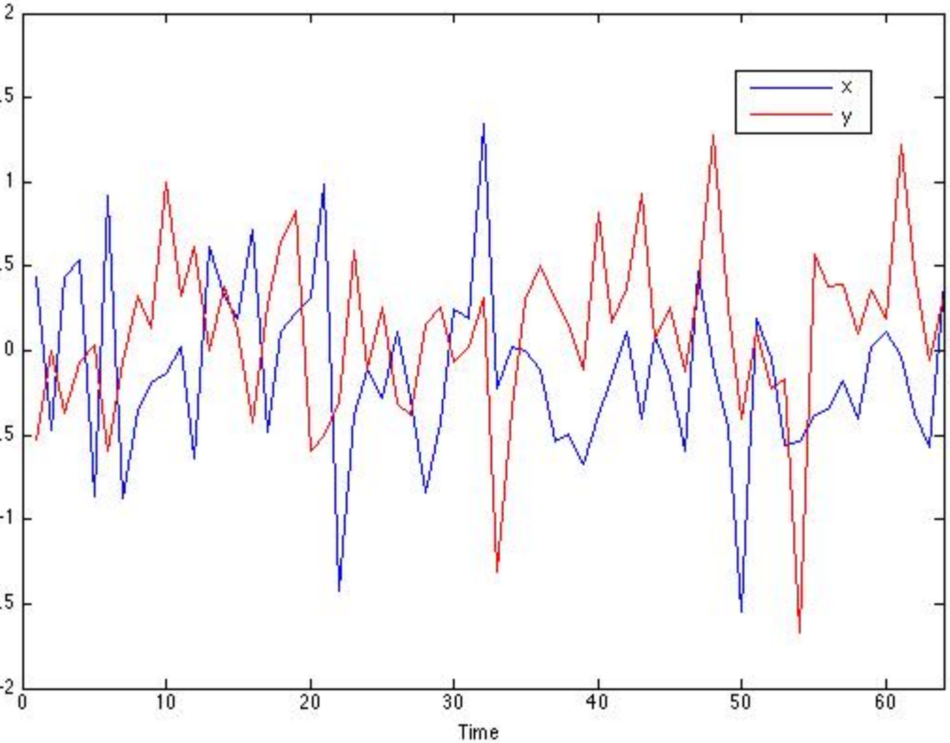
“many models carry correlations in excess of 0.9 that shows the great improvements ….. “



Example A

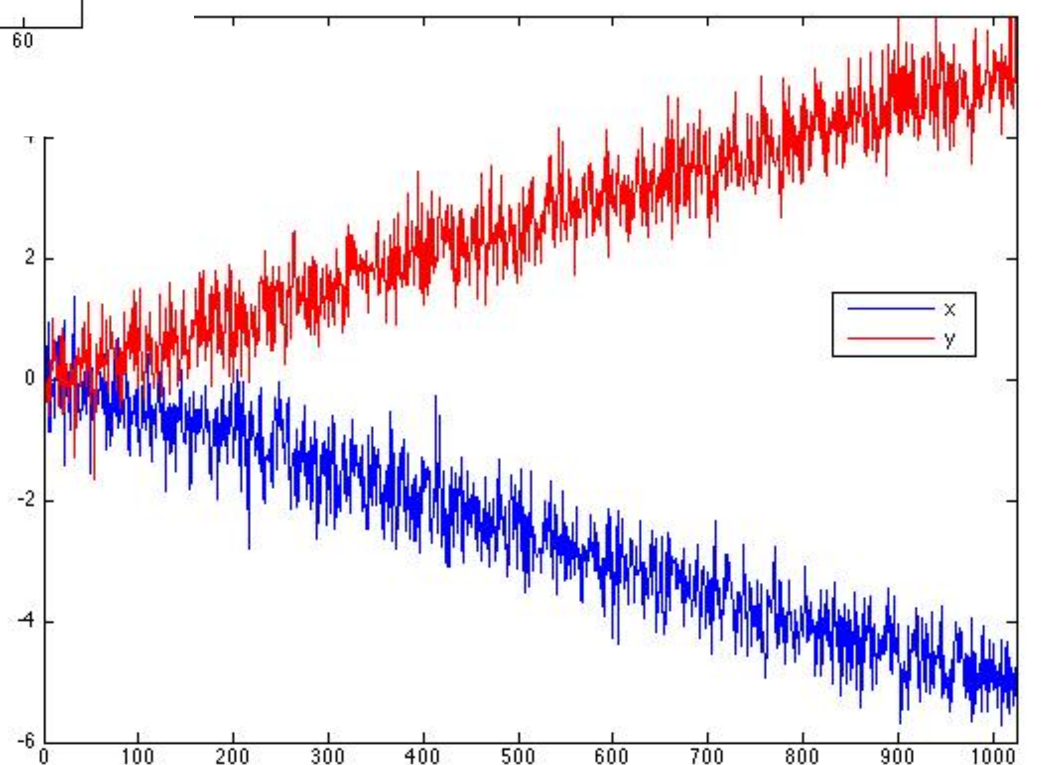
What is the absolute value of the sample correlation between x and y ?

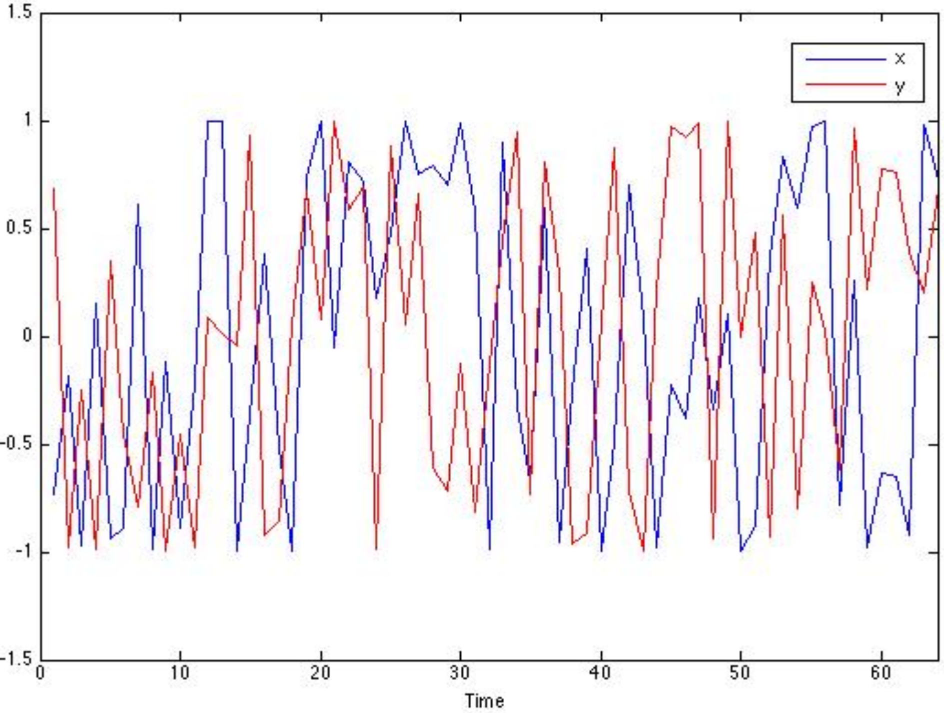




Example B

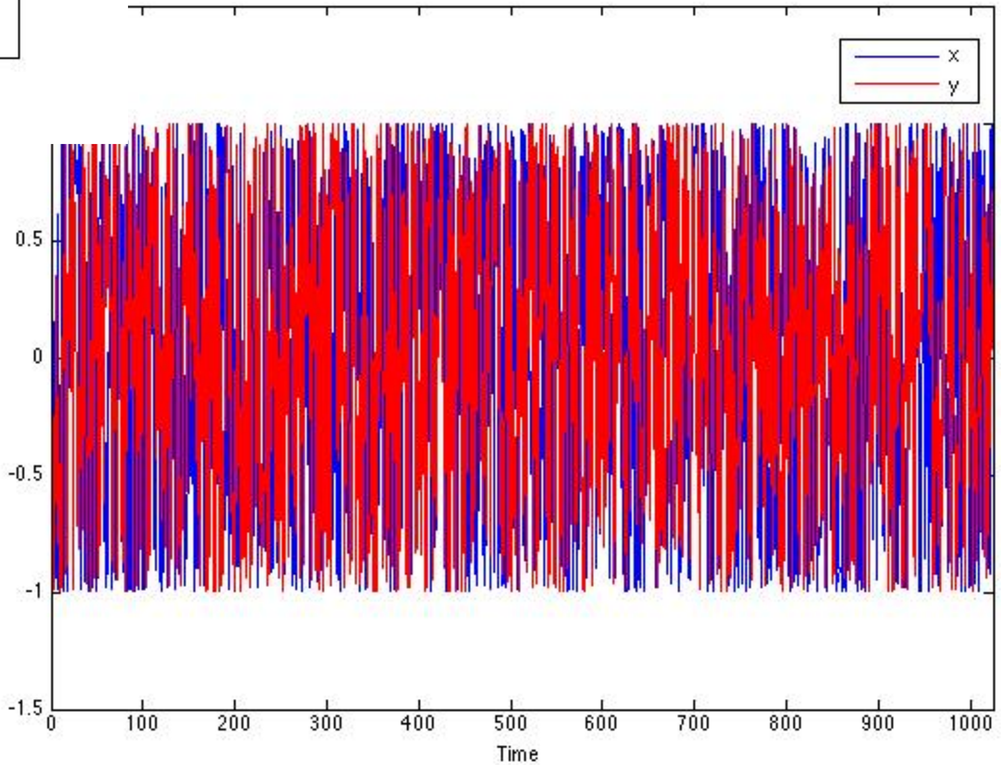
What is the absolute value of the sample correlation between x and y ?

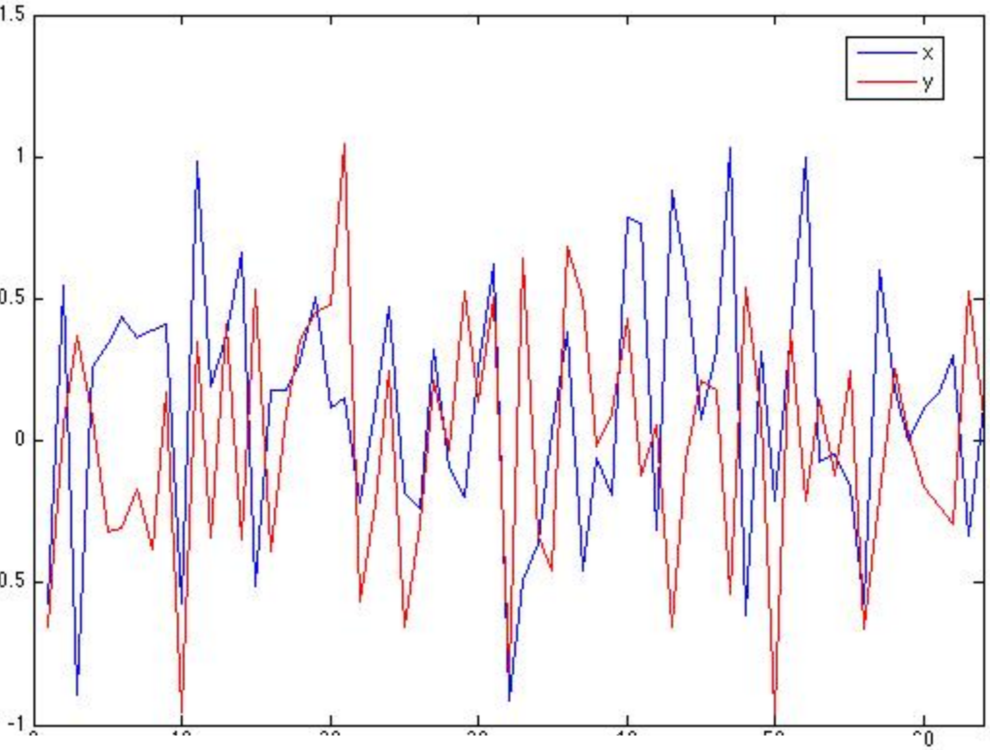




Example C

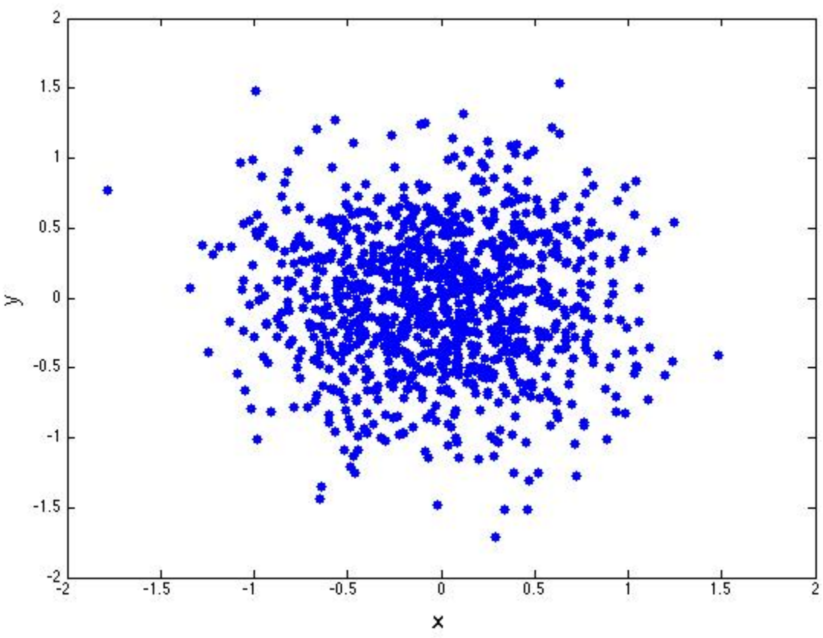
What is the absolute value of the sample correlation between x and y ?

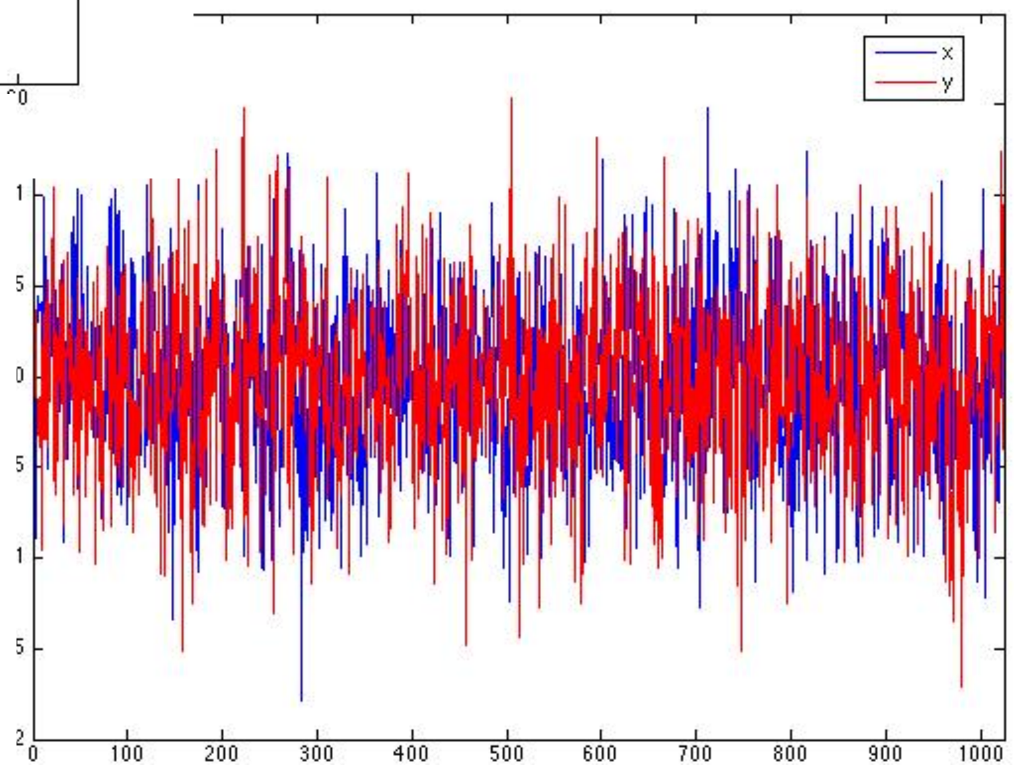


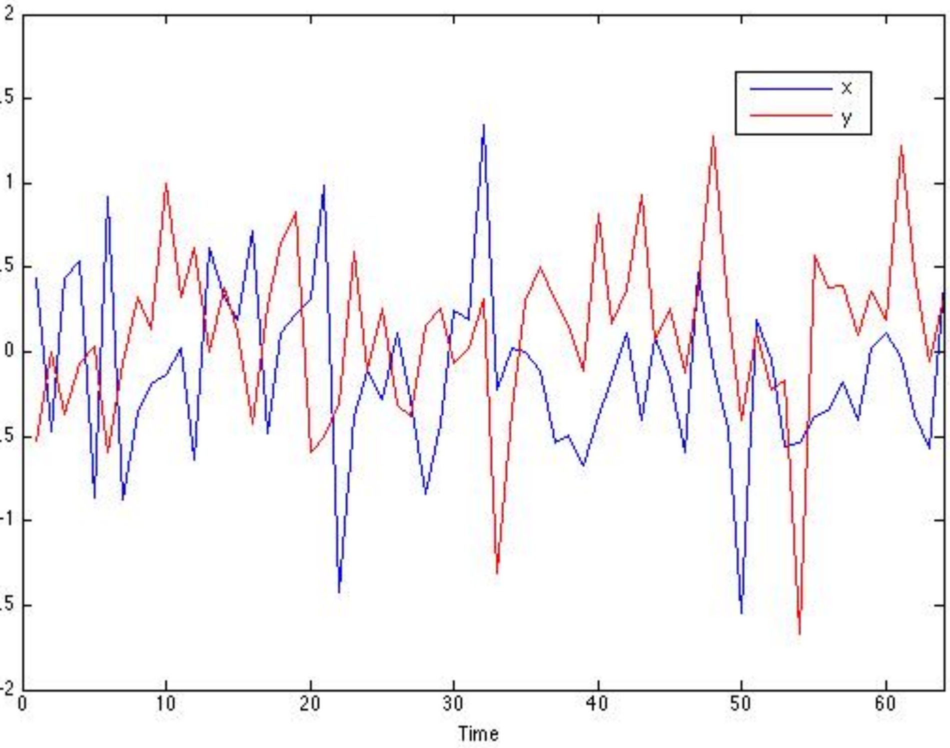


Example A

What is the absolute value of the sample correlation between x and y ?

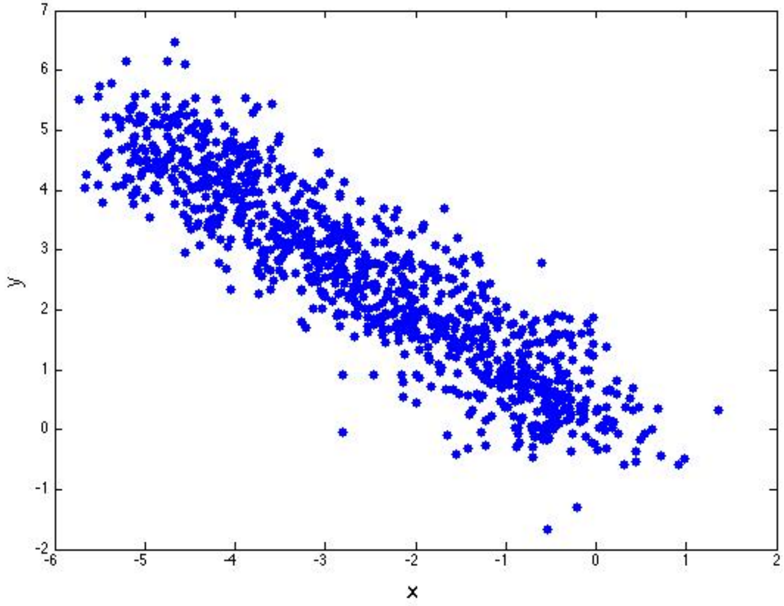


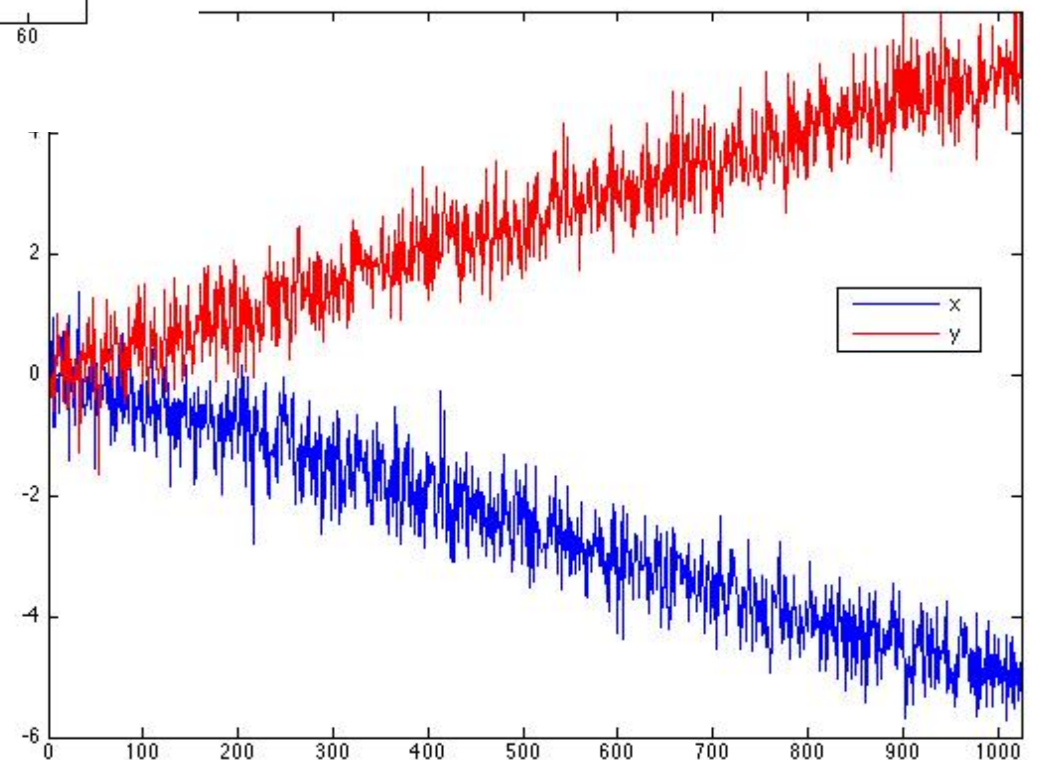


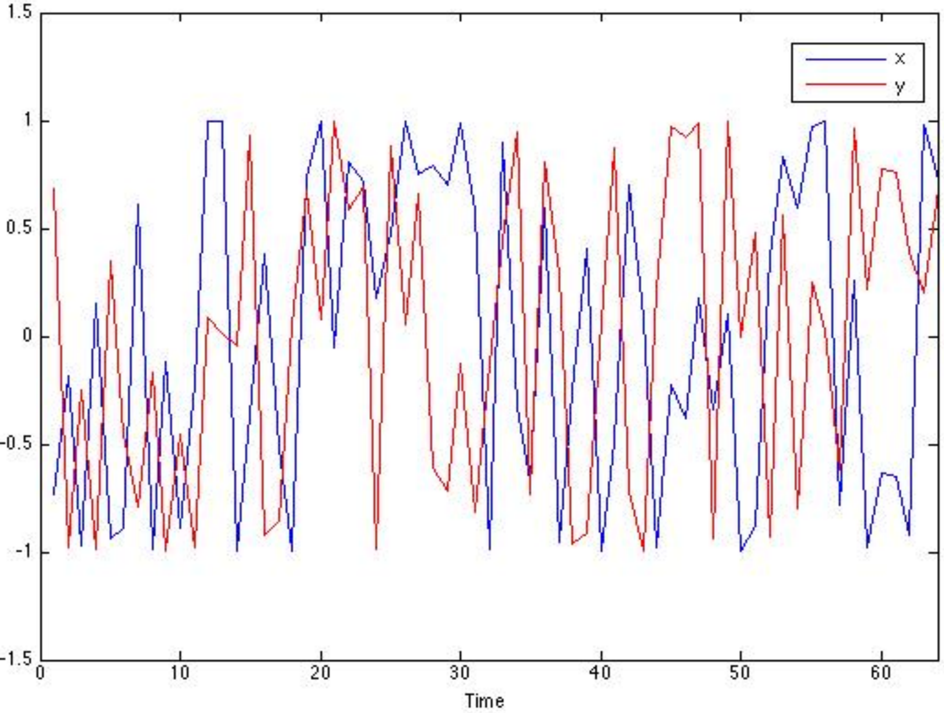


Example B

What is the absolute value of the sample correlation between x and y?

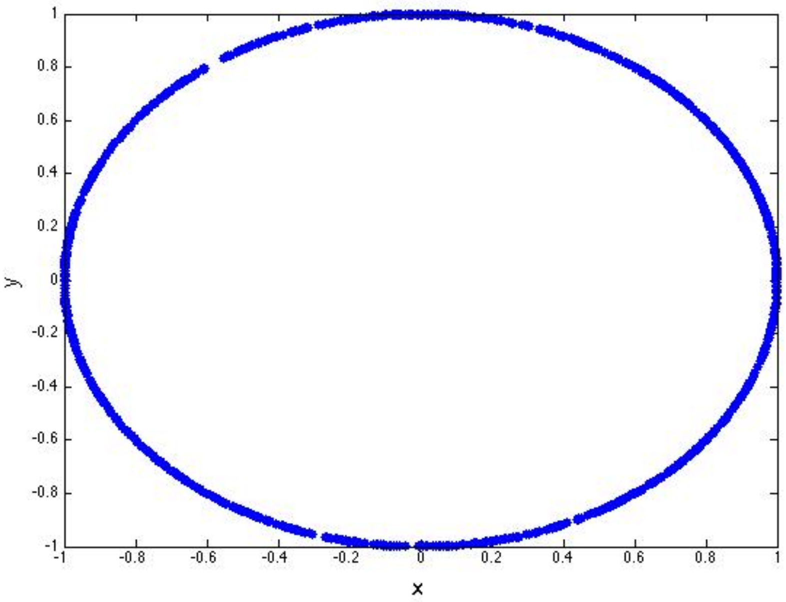


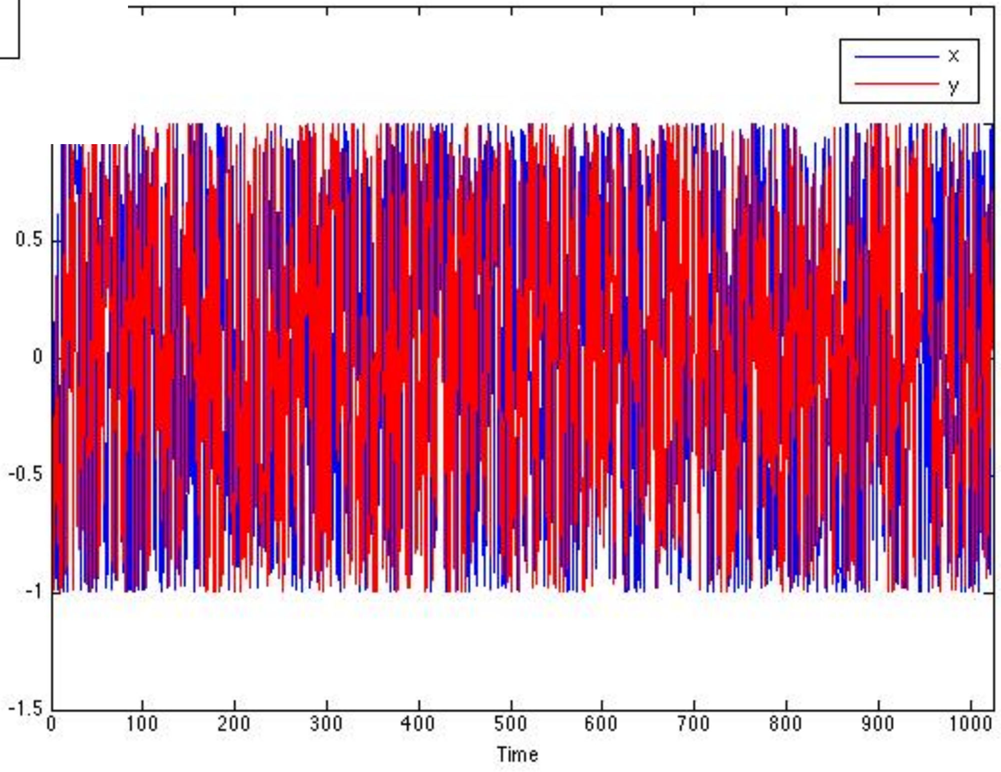


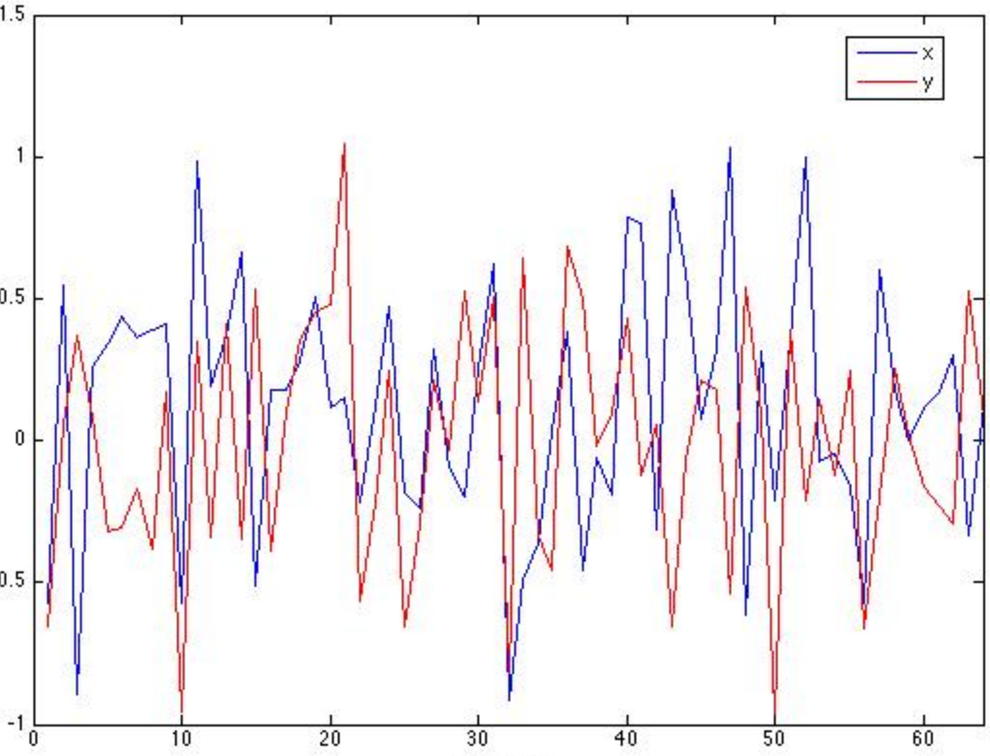


Example C

What is the absolute value of the sample correlation between x and y?



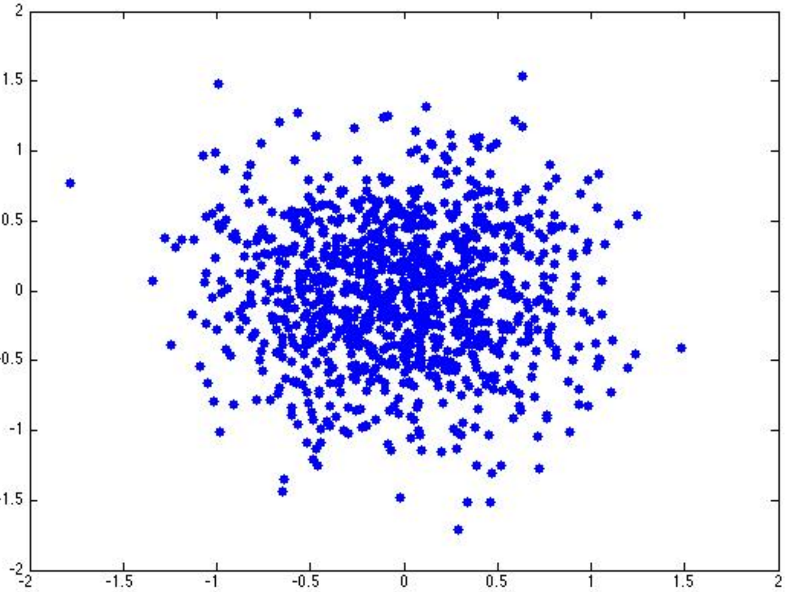


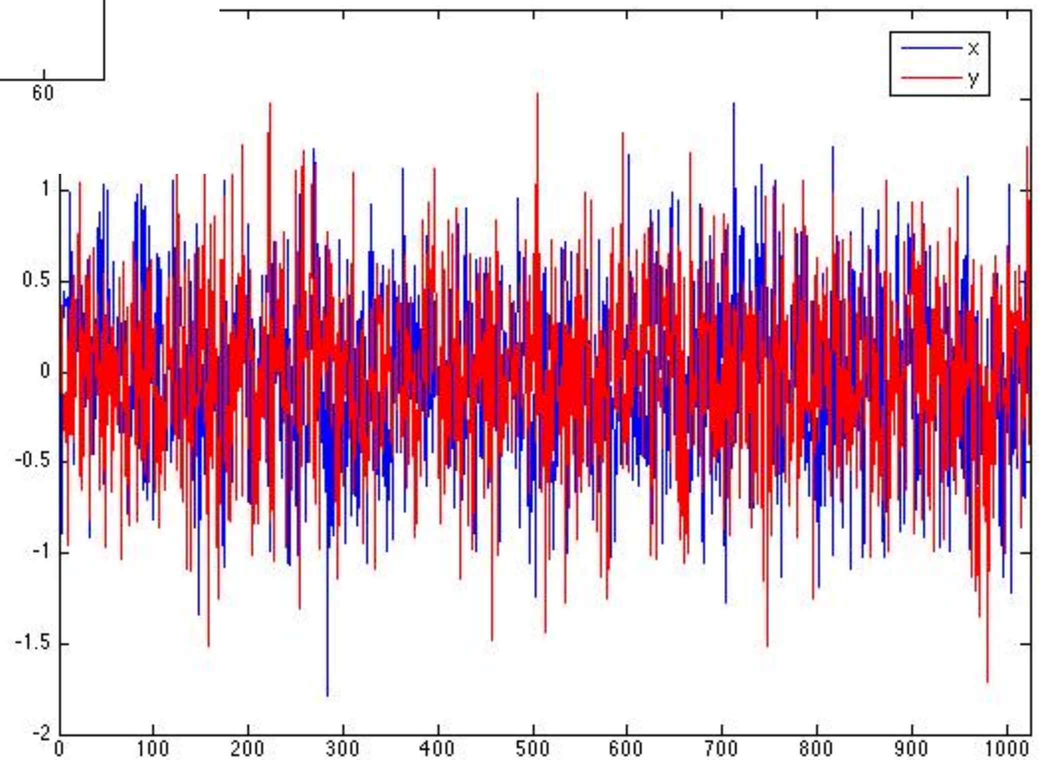


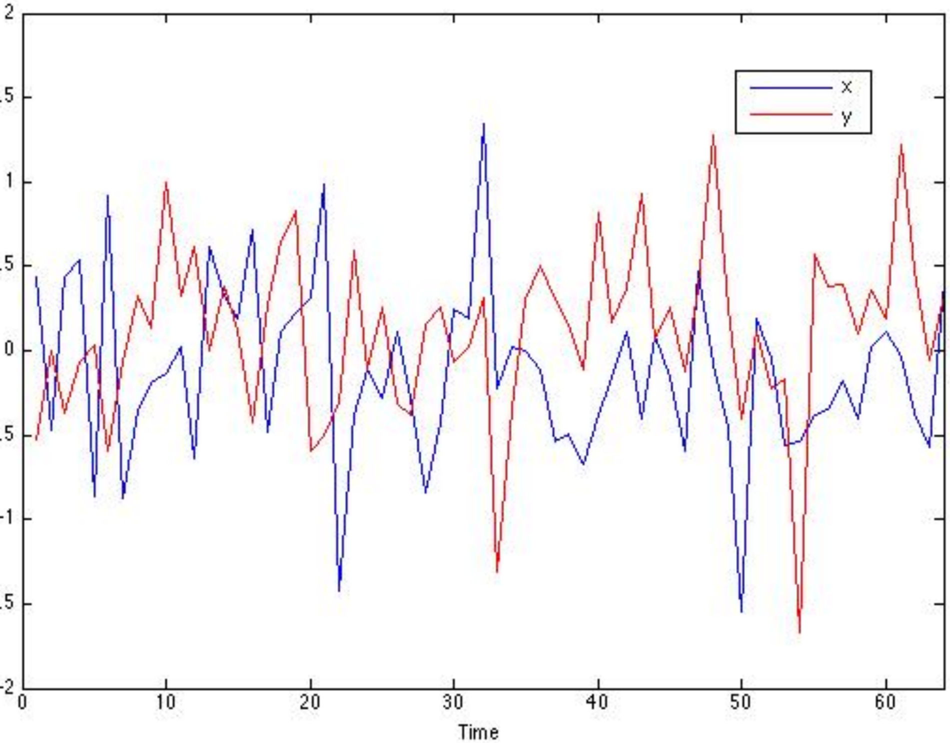
Example A

The absolute value of the sample correlation between x and y is

0.02



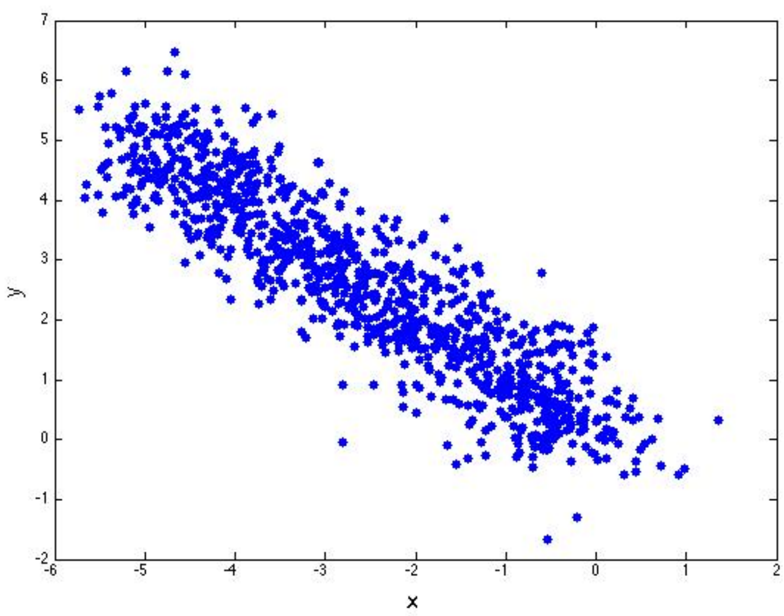


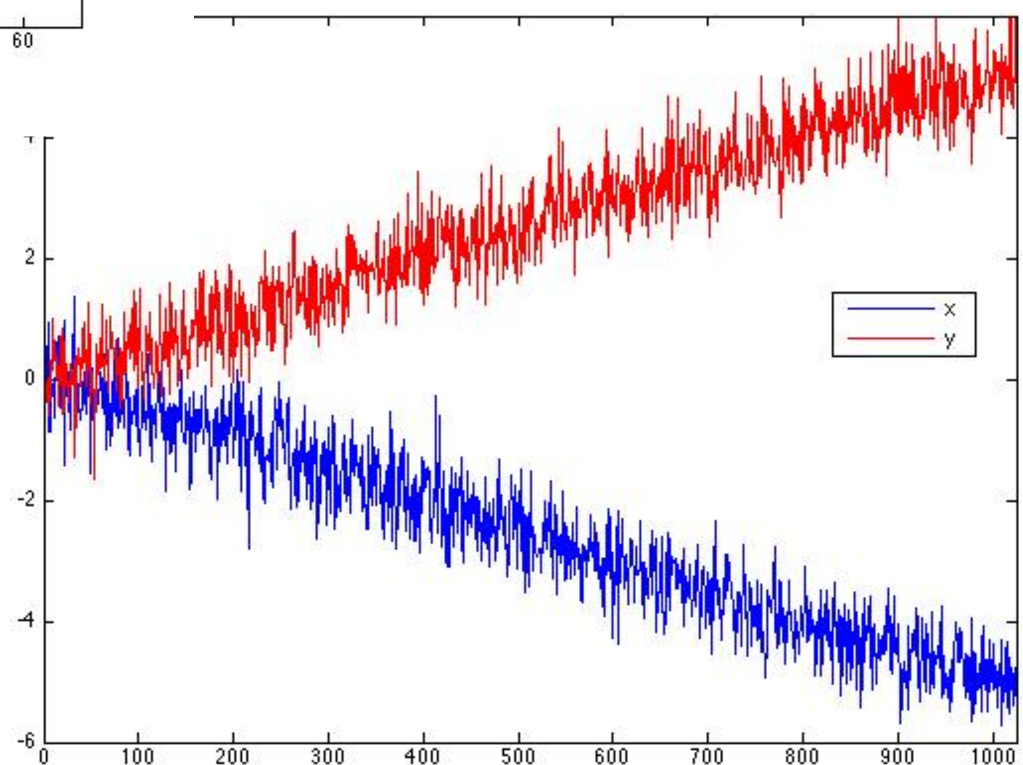


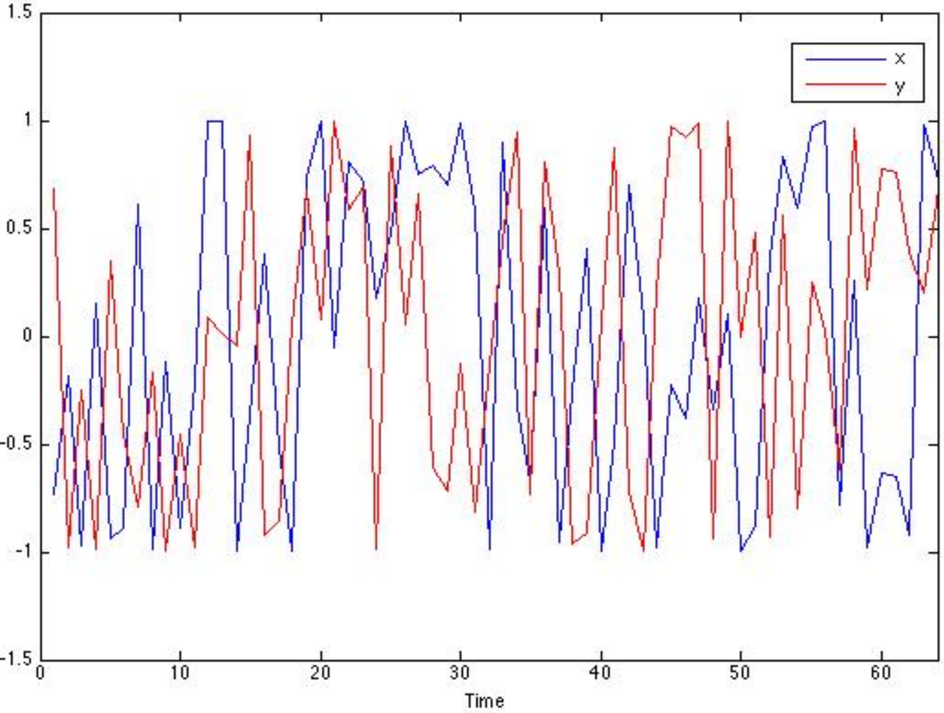
Example B

The absolute value of the sample correlation between x and y is

0.91



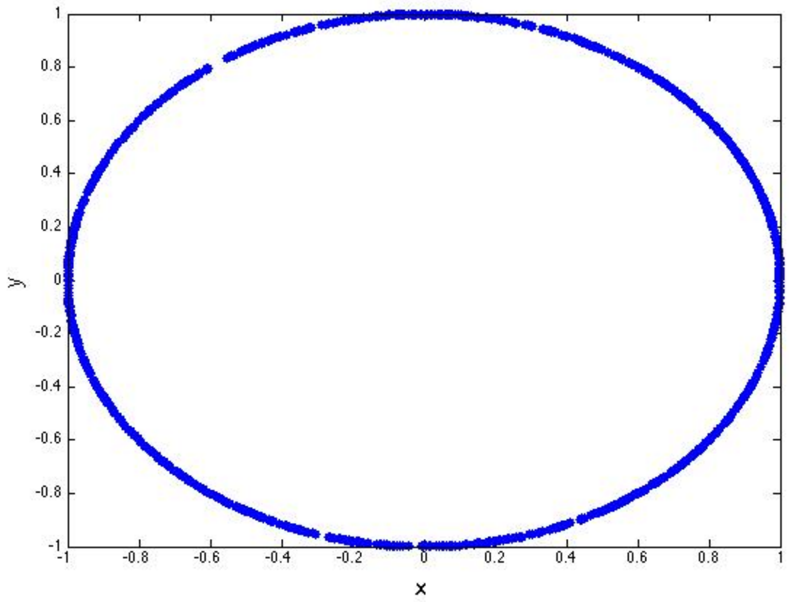


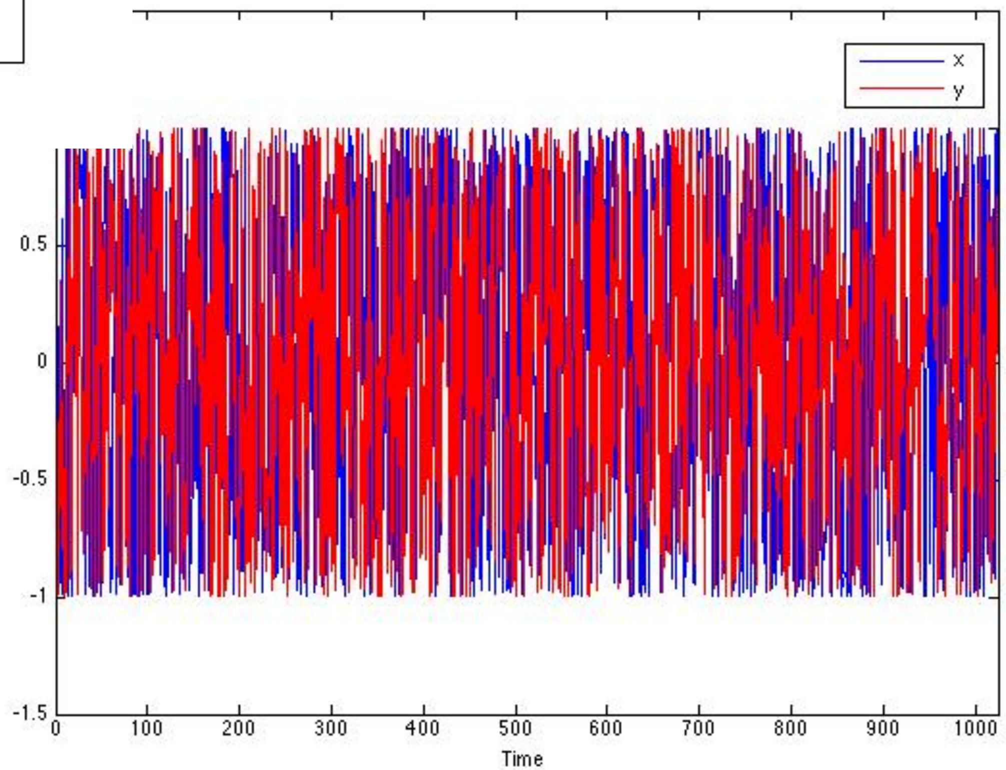


Example C

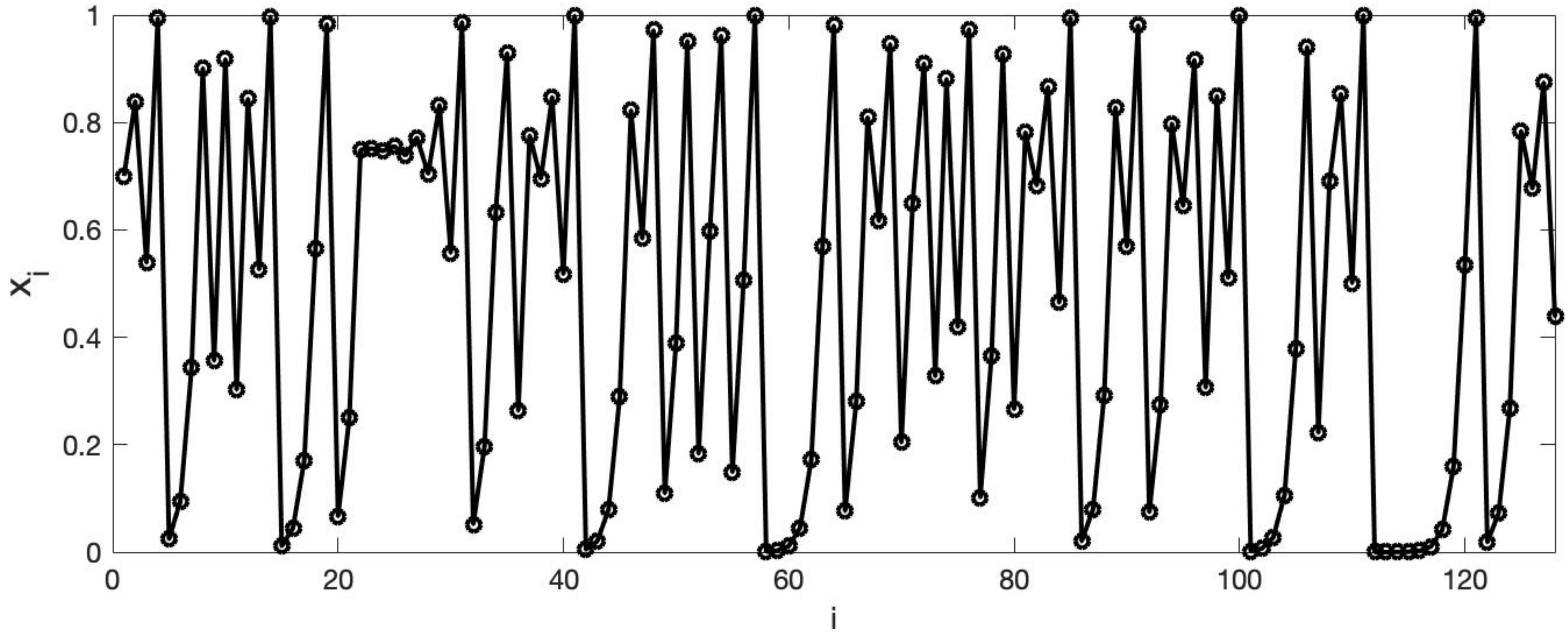
The absolute value of the sample correlation between x and y is

0.02

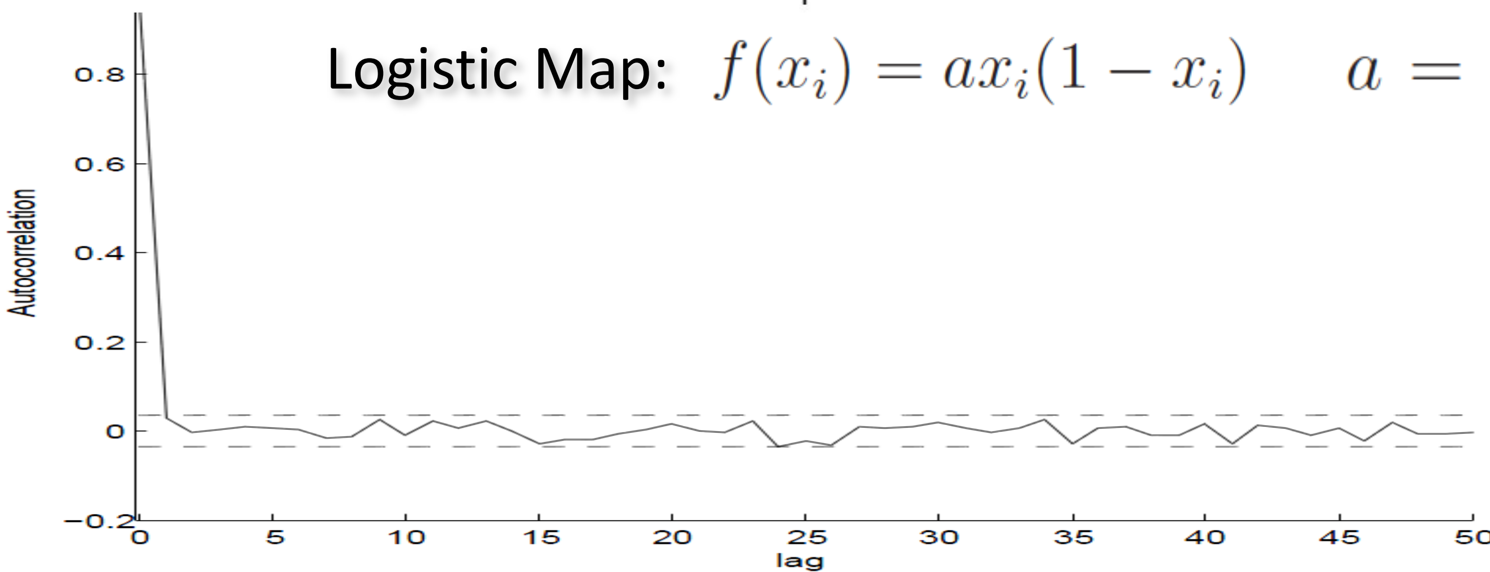




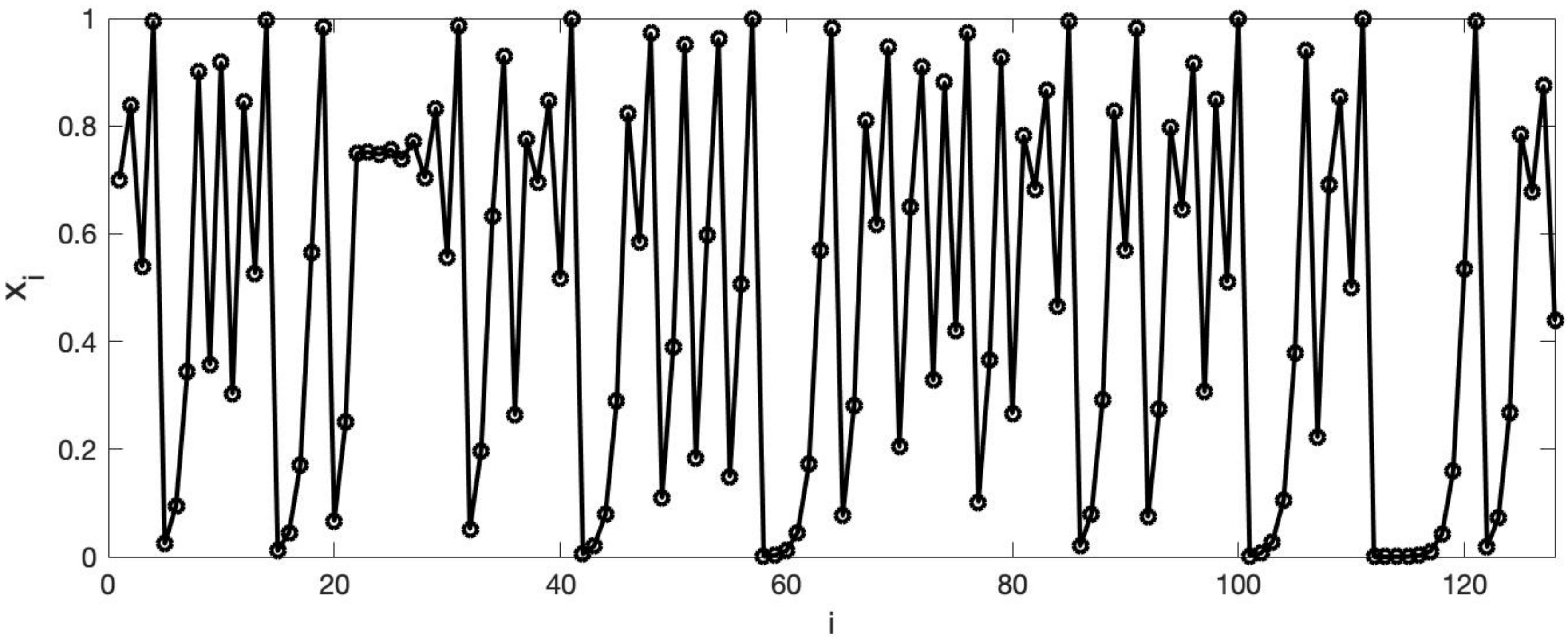
Apply Linear approach to Time Series of Nonlinear system



Logistic Map: $f(x_i) = ax_i(1 - x_i) \qquad a = 4$

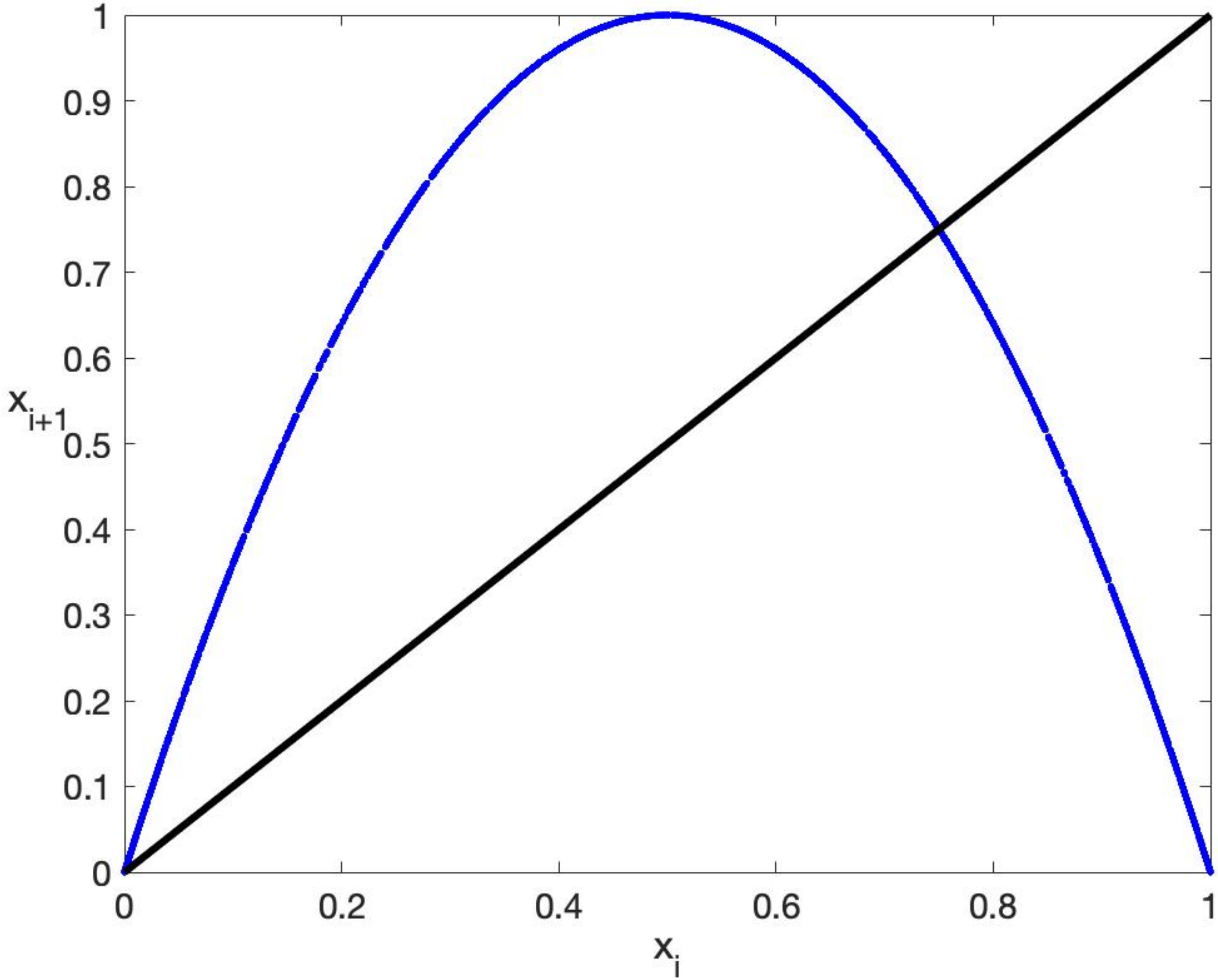


What does the previous value tell us about the current value?
(What would this look like in "delay space"?)



Logistic Map: $f(x_i) = ax_i(1 - x_i) \quad a = 4$

A two dimensional delay reconstruction

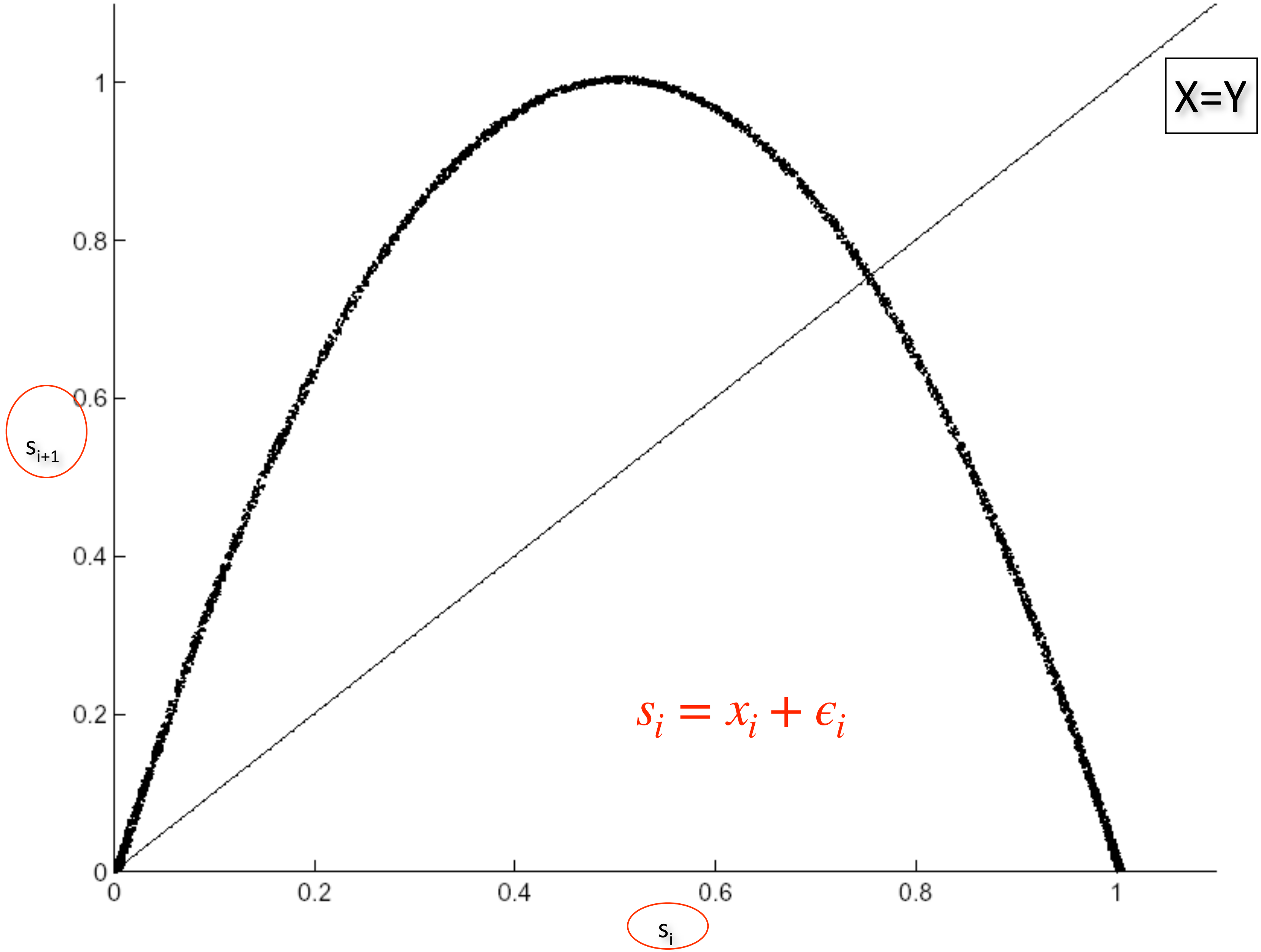


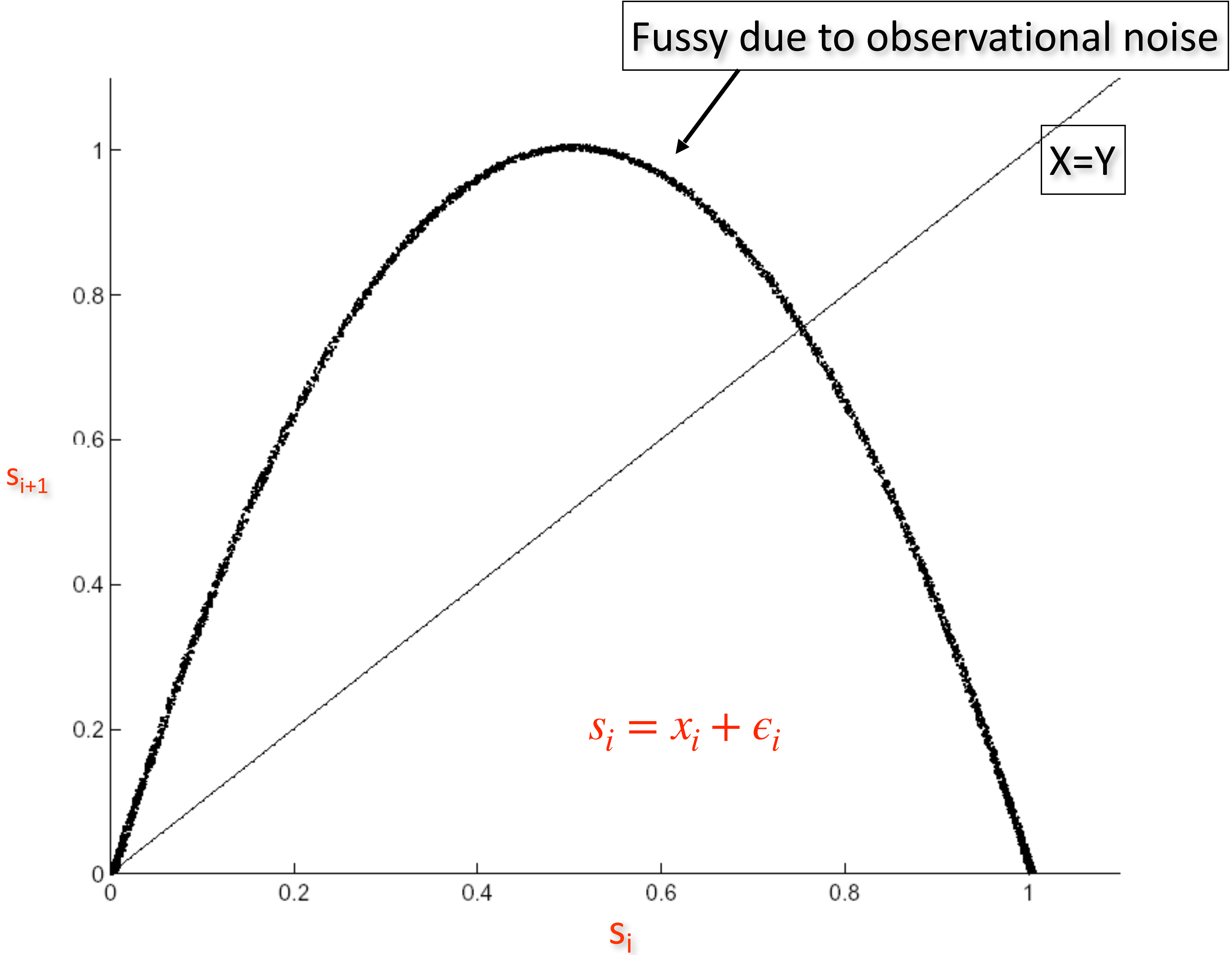
X=Y

Nonlinear relationship!

Logistic Map: $f(x_i) = ax_i(1 - x_i) \qquad a = 4$

A two dimensional delay reconstruction

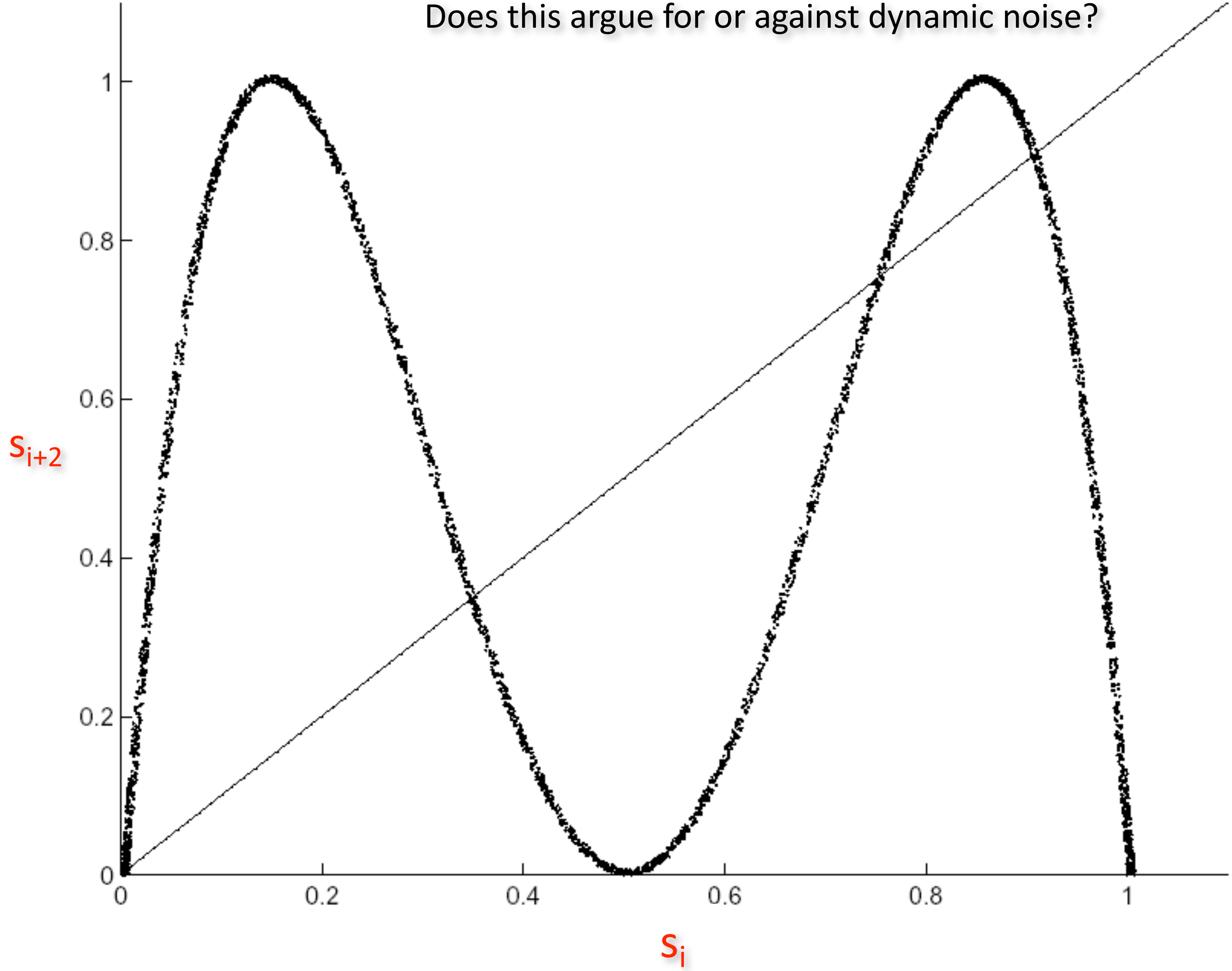




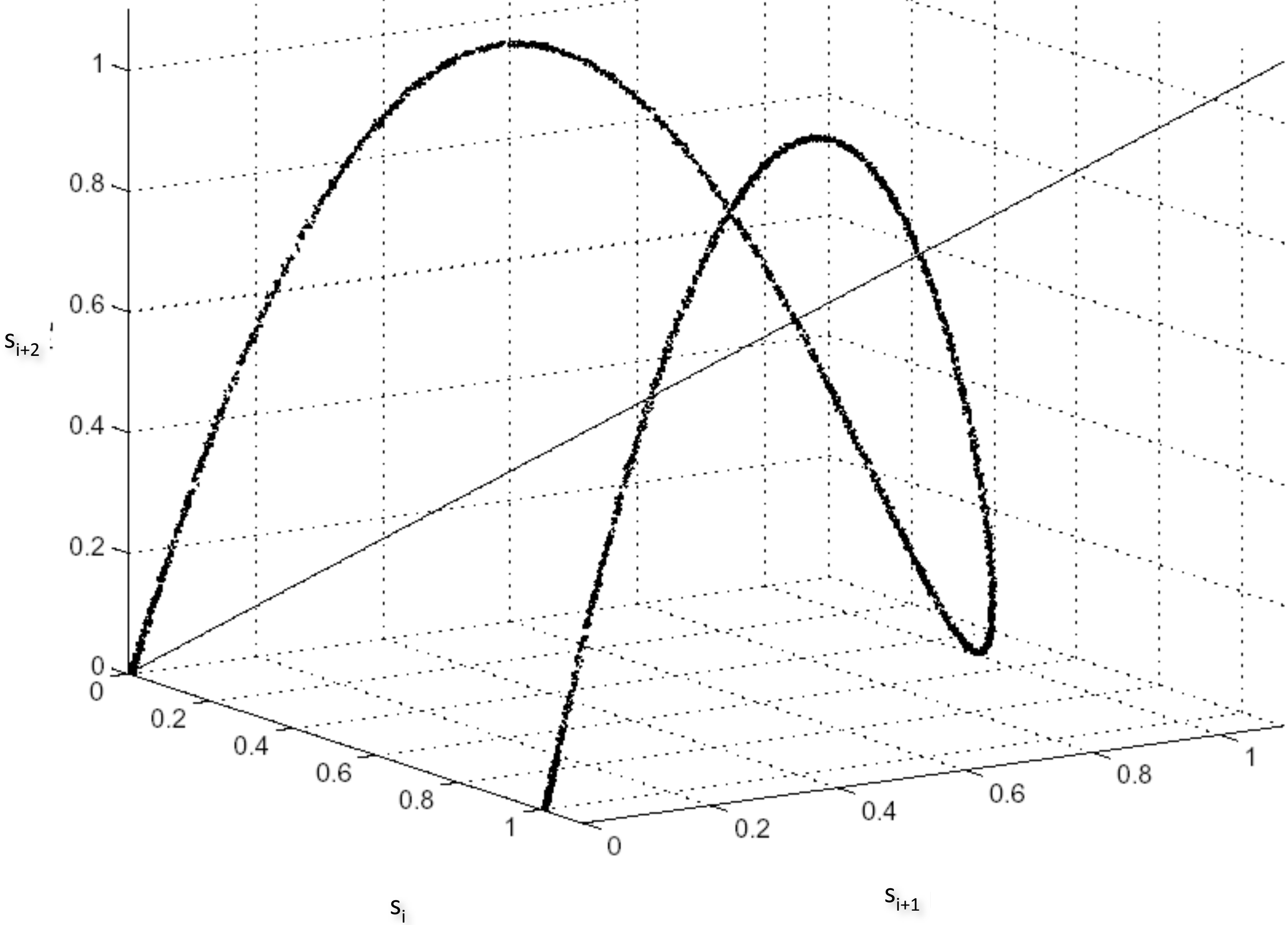
Fussy due to observational noise
X=Y
1
0.8
0.6
0.4
0.2
0
0.2
0.4
0.6
0.8
1
s_{i+1}
s_i
$s_i = x_i + \epsilon_i$

A two dimensional delay reconstruction

Does this argue for or against dynamic noise?



This is a three dimensional delay plot.

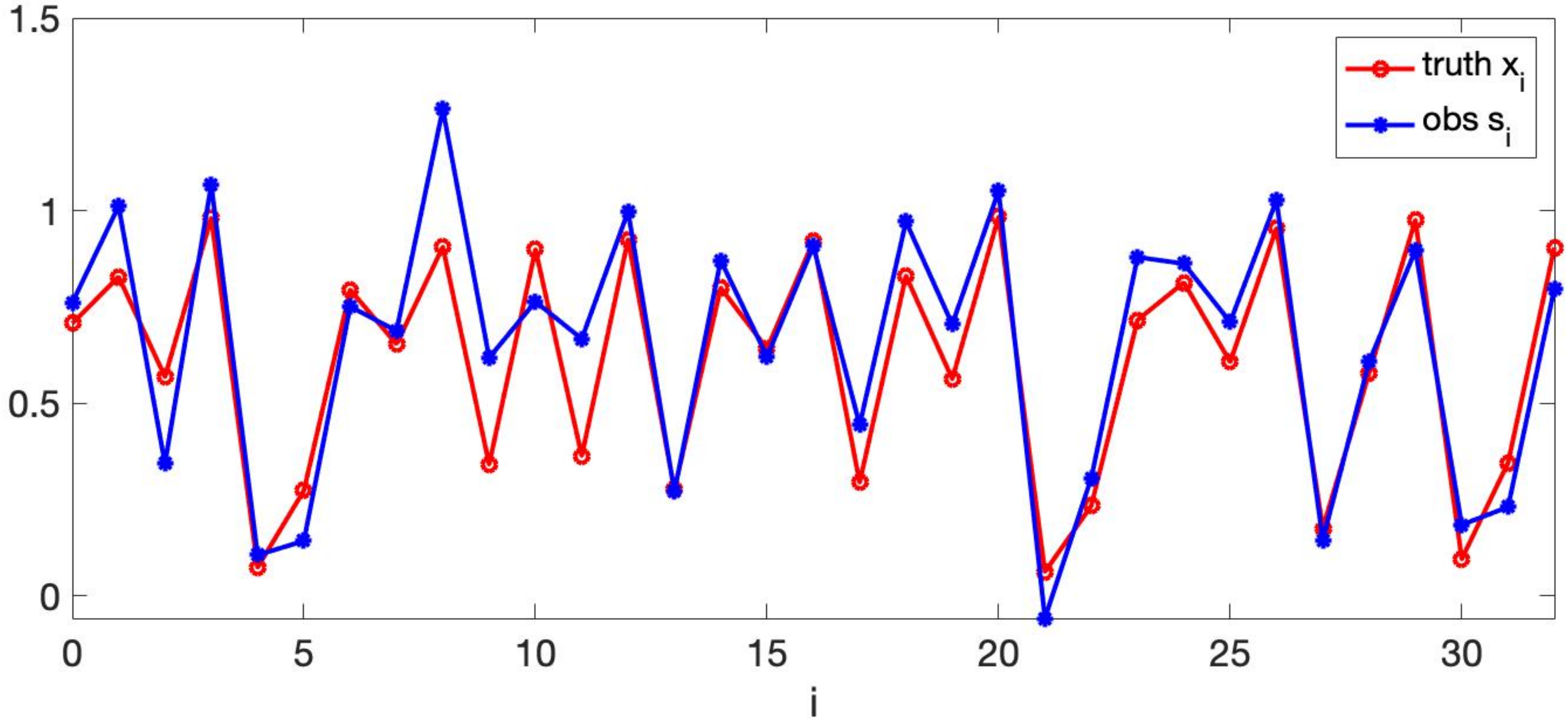


Given s_i, does knowing s_{i-1} help?

Strange behavior of likelihood functions

Logistic Map: $f(x_i) = ax_i(1 - x_i)$ $a = 4$

Assuming additive observational noise δ_i yields observations, $s_i = \tilde{x}_i + \delta_i$ where $\tilde{x}$ is the true system state (Truth) and the observational noise, δ_i, is Independent Normally Distributed (IND, $\delta_i \sim N(0, \sigma^2)$).



Given a sequence of observations and *f*, how to identify the true states?

Strange behavior of likelihood functions

Logistic Map: $f(x_i) = ax_i(1 - x_i)$ $a = 4$

Assuming additive observational noise δ_i yields observations, $s_i = \tilde{x}_i + \delta_i$ where $\tilde{x}$ is the true system state (Truth) and the observational noise, δ_i, is Independent Normally Distributed (IND, $\delta_i \sim N(0, \sigma^2)$). Under this normality assumption, the log-likelihood (LLik) function

$$LLik(x_0) = -\sum_{i=0}^{n-1} (s_i - f^i(x_0))^2 / 2\sigma^2$$

Strange behavior of likelihood functions

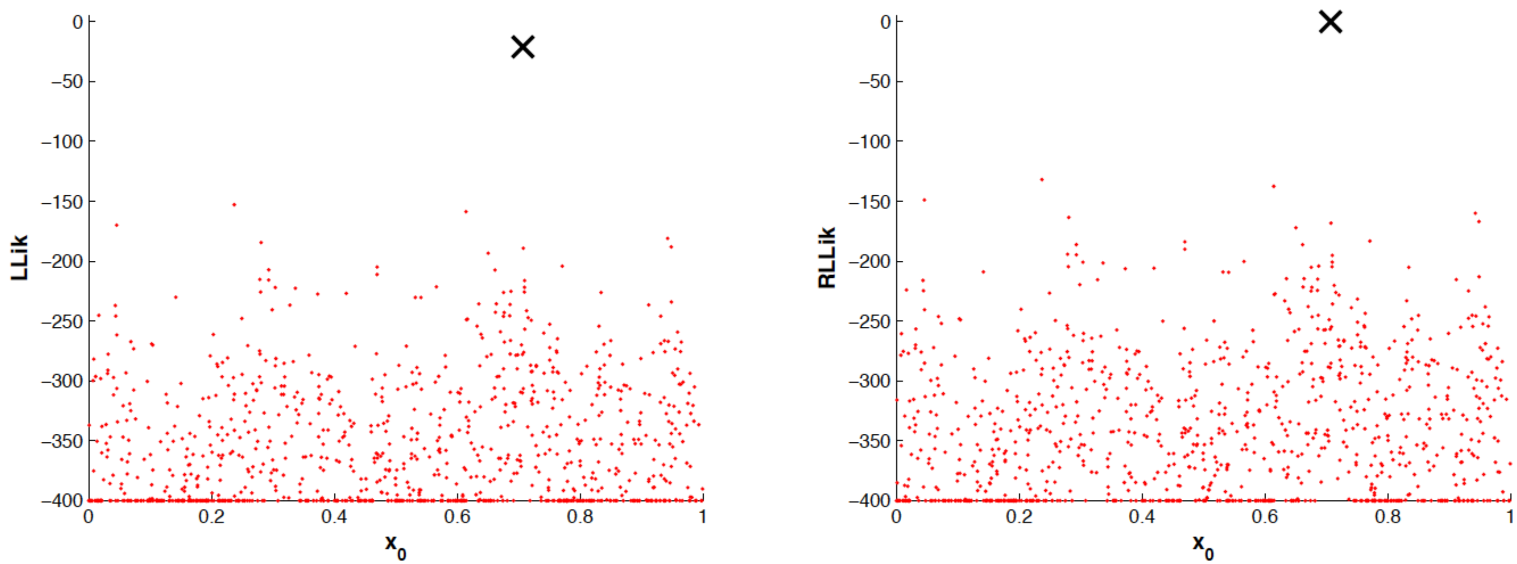


Figure 1: Typical log-likelihood of 1024 states (uniformly distributed on $[0, 1]$) for the Logistic Map. The true initial condition $\tilde{x}_0 = \sqrt{2}/2$, $\sigma = 0.1$ and $n = 32$. a) Log-likelihood function, b) Relative log-likelihood to $\tilde{x}_0$ (denoted by '$\times$'), states which have LLik/RLLik less than -400 are plotted on the -400 horizontal line. All logarithms are using natural base.

Strange behavior of likelihood functions

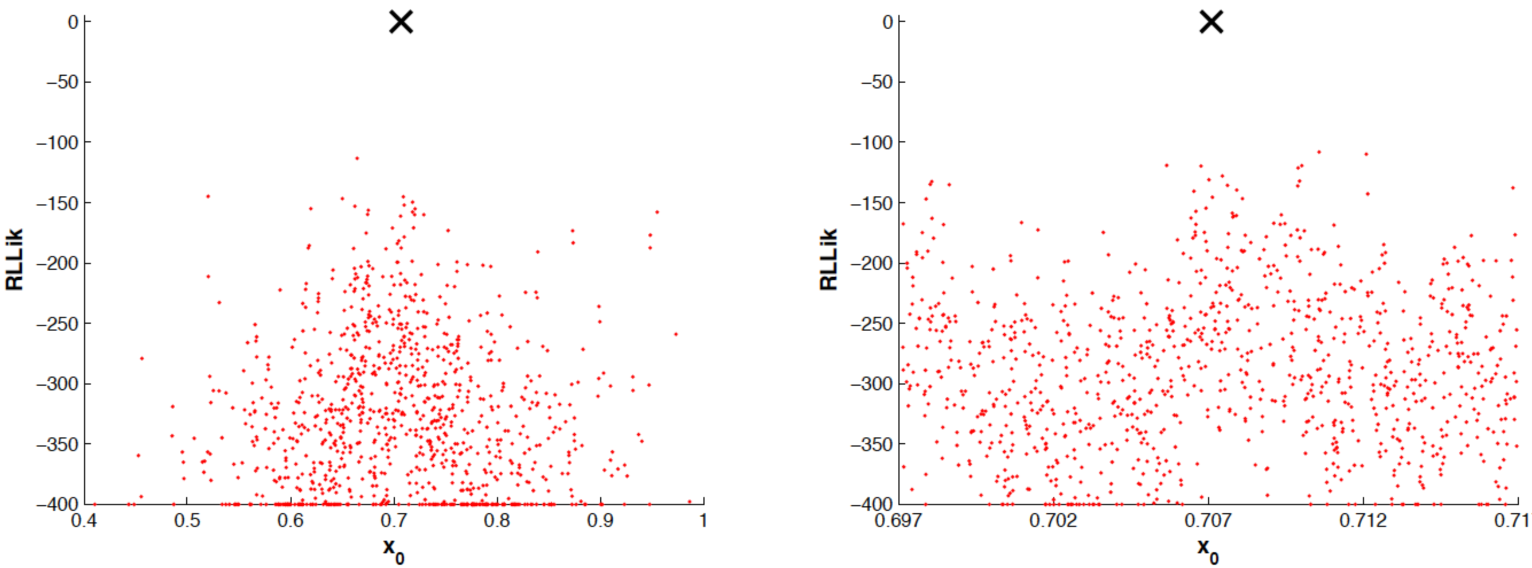


Figure 2: Log-likelihood of 1024 states for the Logistic Map, the true initial condition $\tilde{x}_0 = \sqrt{2}/2$, $\sigma = 0.1$ and $n = 32$. a) sampled from inverse observational noise, b) uniformly sampled from $[\tilde{x}_0 - \frac{\sigma}{10}, \tilde{x}_0 + \frac{\sigma}{10}]$.

Strange behavior of likelihood functions

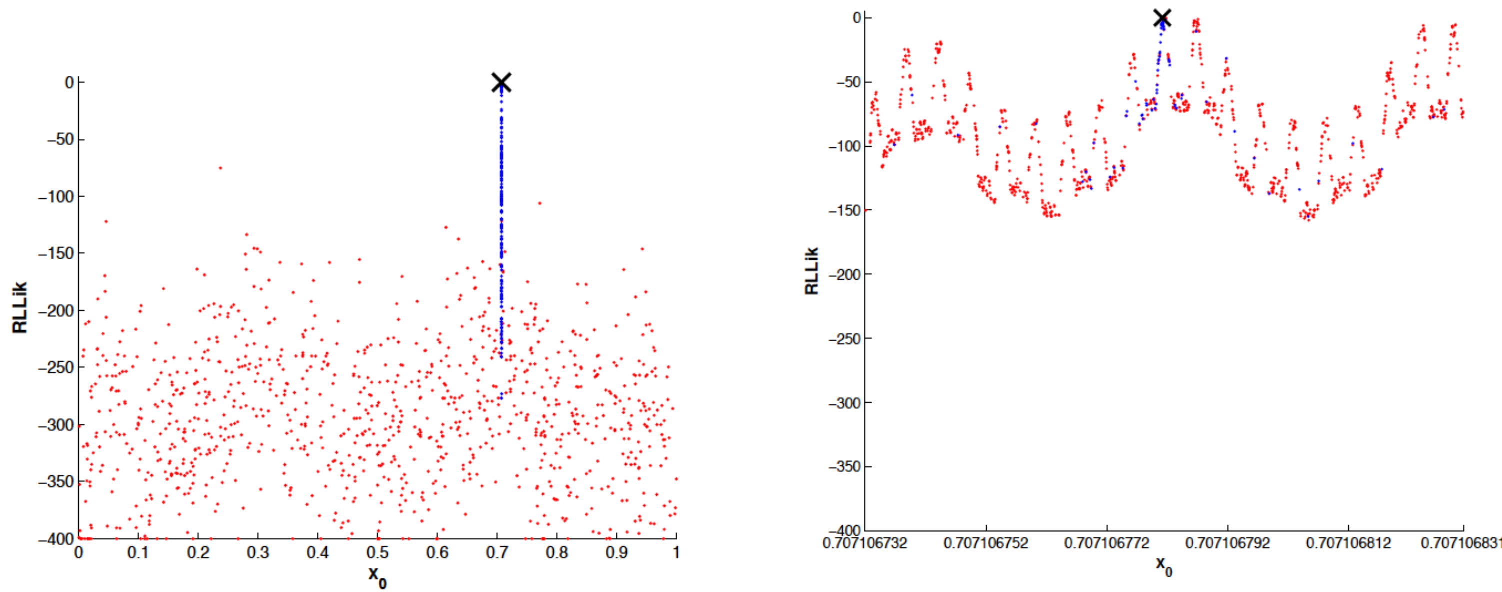
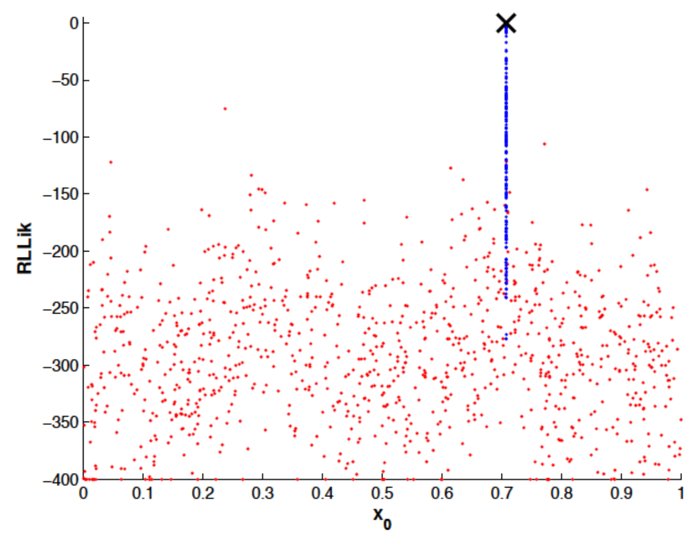
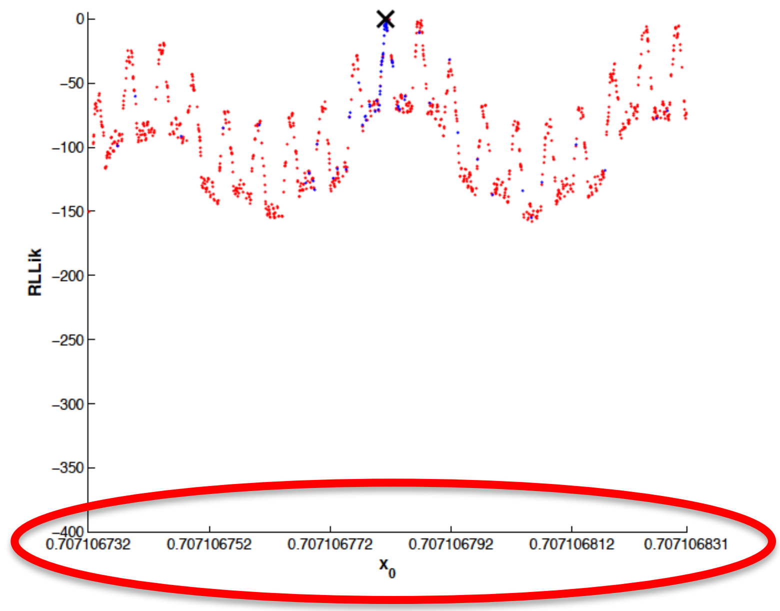


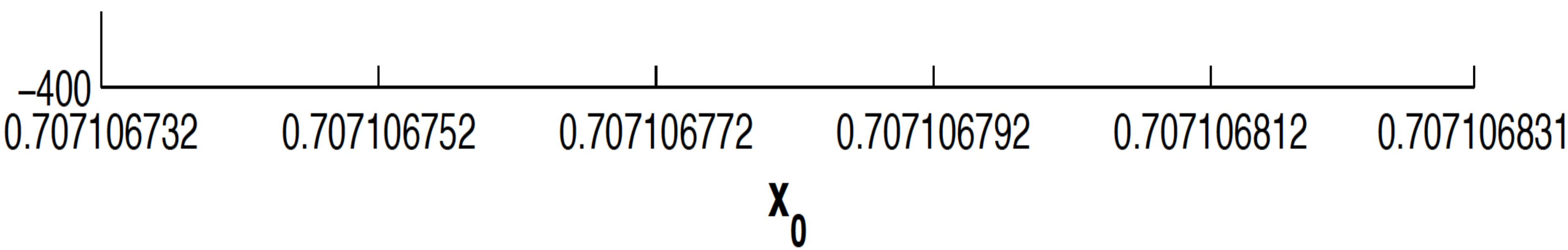
Figure 3: a) Following Figure 1b, add relative log-likelihood of 1024 states (blue), which are extremely close to $\tilde{x}_0$, generated by spiral sampling around the $\tilde{x}_0$; b) zoom in of a).

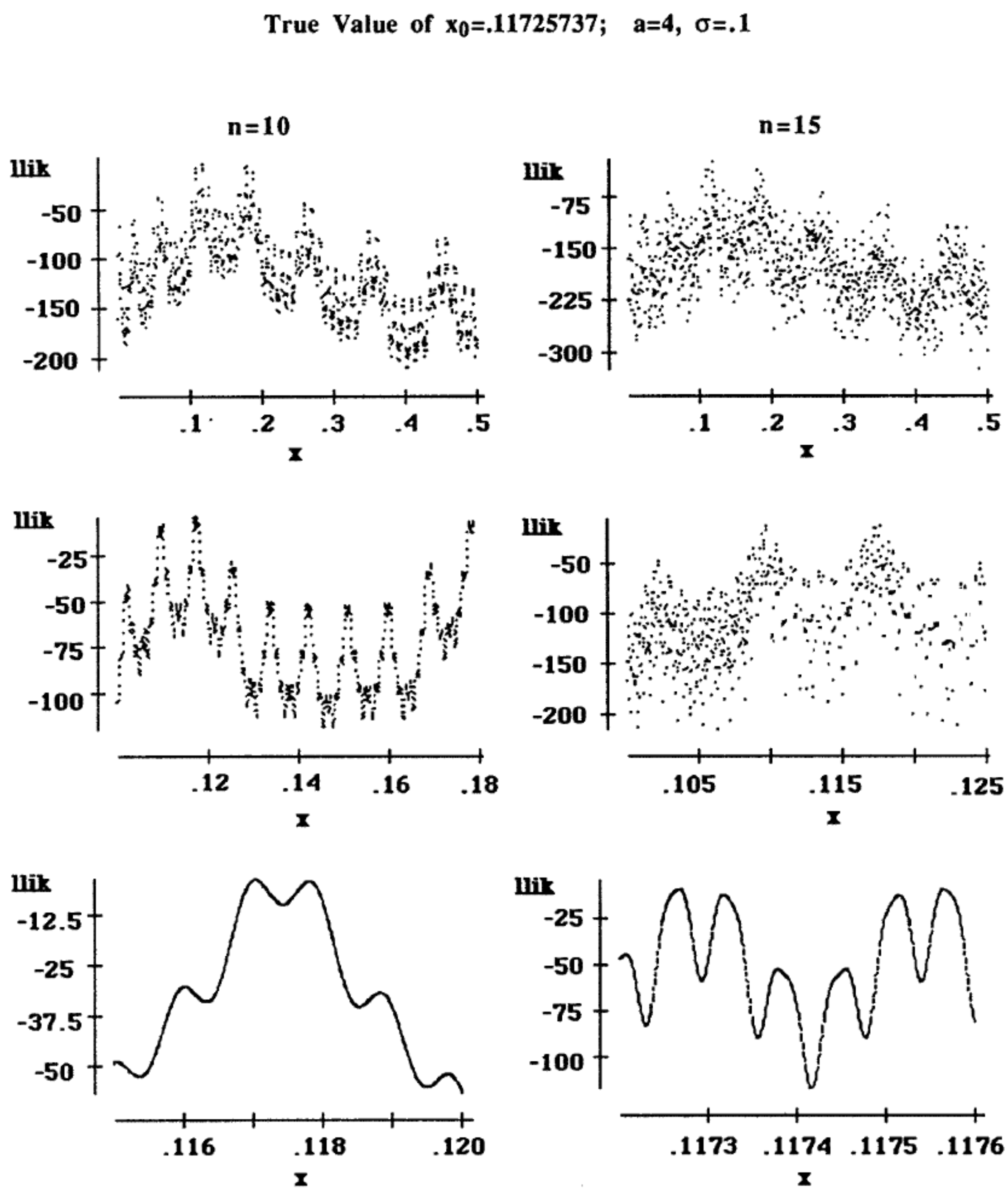
1024 points are generated by $\tilde{x}_0 + 2^{-(10+\frac{60i}{1024})}\epsilon_i$, $i = 1, 2, ..., 1024$ where ϵ_i is random drawn from $U(0, 1)$

Strange behavior of likelihood functions





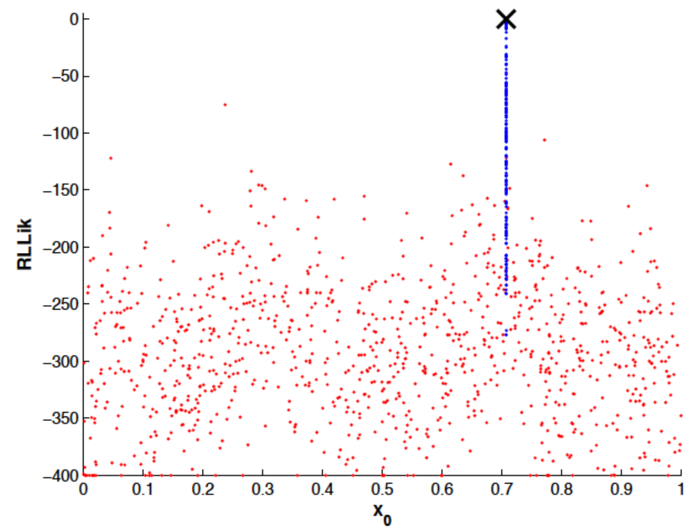


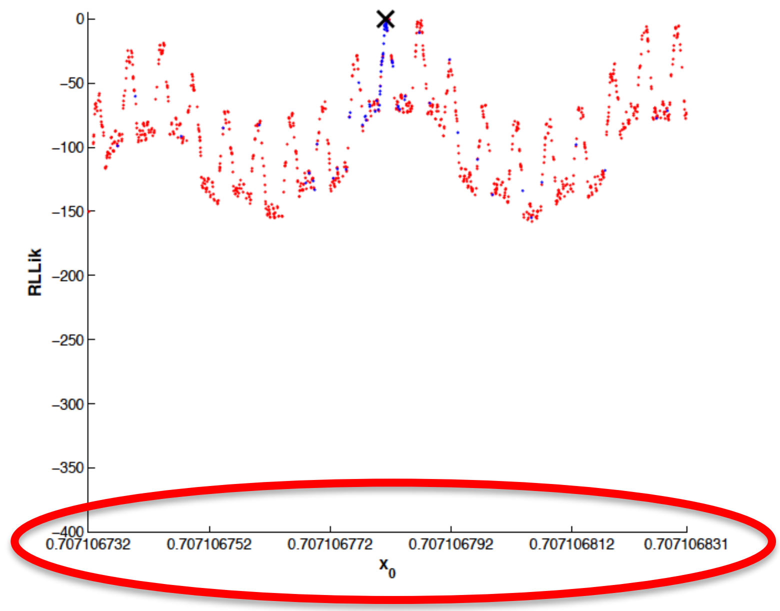


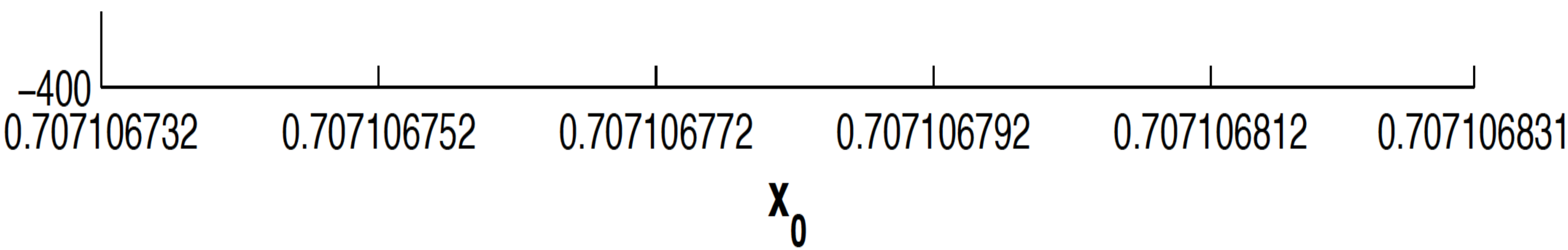
M.L. Berliner, Likelihood and bayesian prediction for chaotic systems. J. Am. Stat. Assoc. 86:938, 1991

Strange behavior of likelihood functions

Dynamical information!!





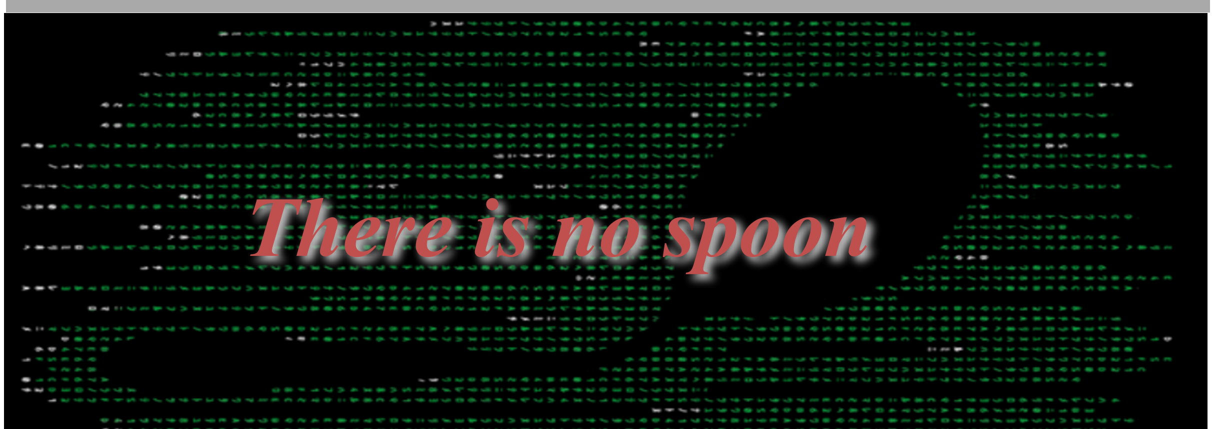


Du, H., Smith, L.A. (2017) 'Rising Above Chaotic Likelihoods', SIAM/ASA Journal on Uncertainty Quantification, (5) 246-258.

This course will give you a feel for predictability and nonlinear dynamical system using precise maths, real science, and relevant real-world statistical tools.

The story divides into a conflict between two houses, which are distinguished by their slogans:

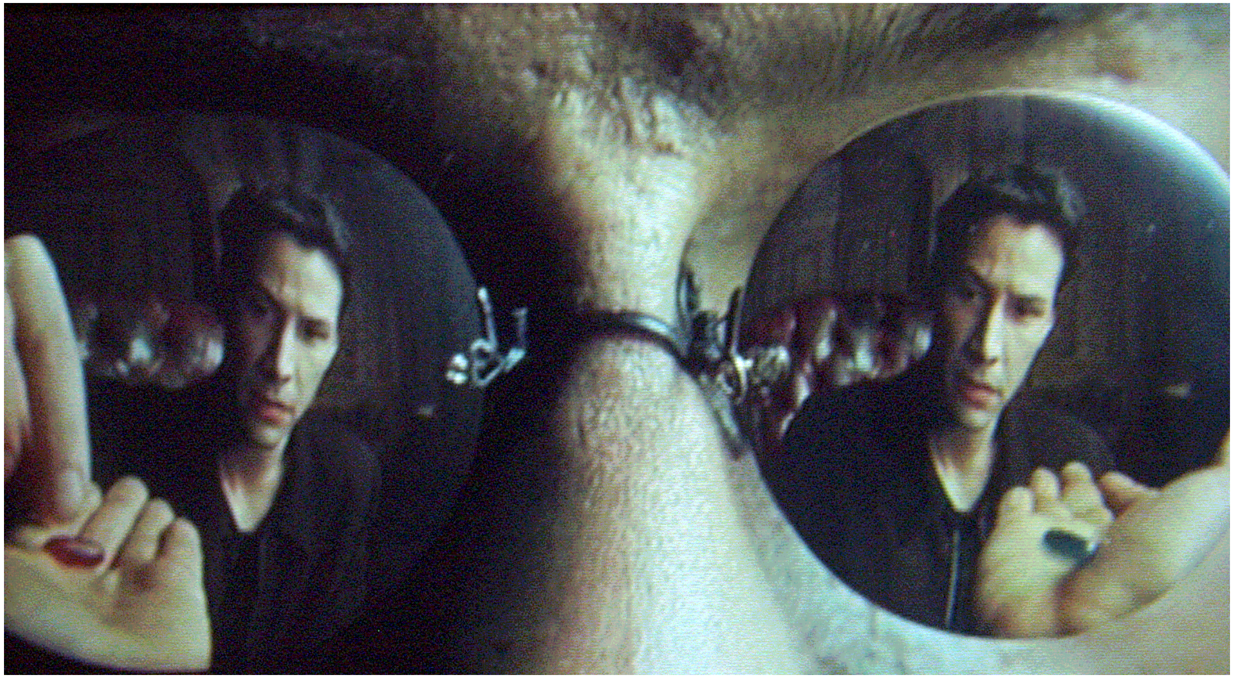
(Maths)



(Physics)

Statisticians have the freedom to move between maths and the physics!

Now you have to make a choice. You take the blue pill and the lecture ends, you wake–up in your bed and happily do mathematics…



You take the red pill, and try to give coherent statistical advice knowing all models are wrong.

"Remember that all I am offering is the truth. Nothing more"

Morpheus