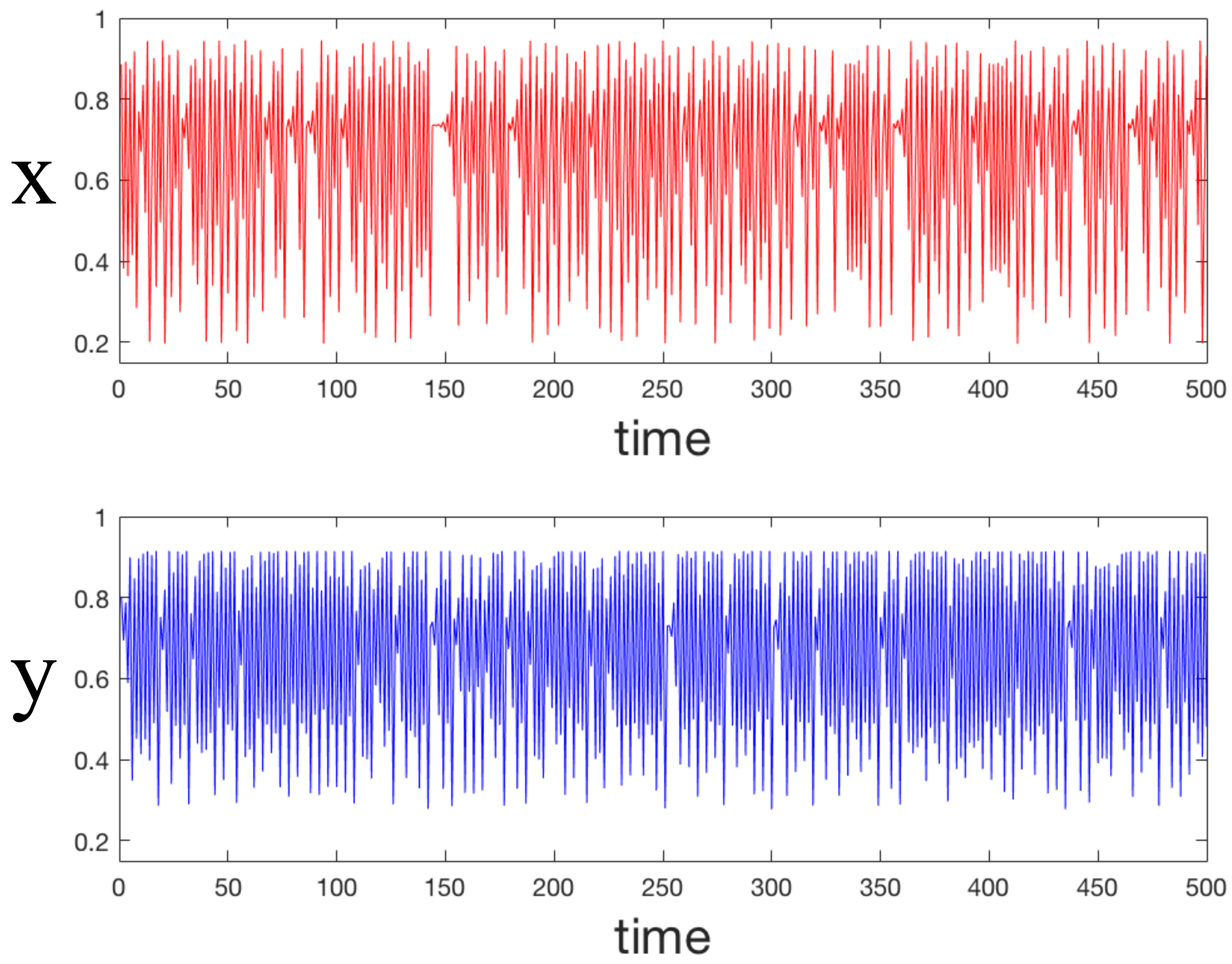
A new method for quantifying Earth system interactions from time series

David Diego
Department of Earth Science, UiB

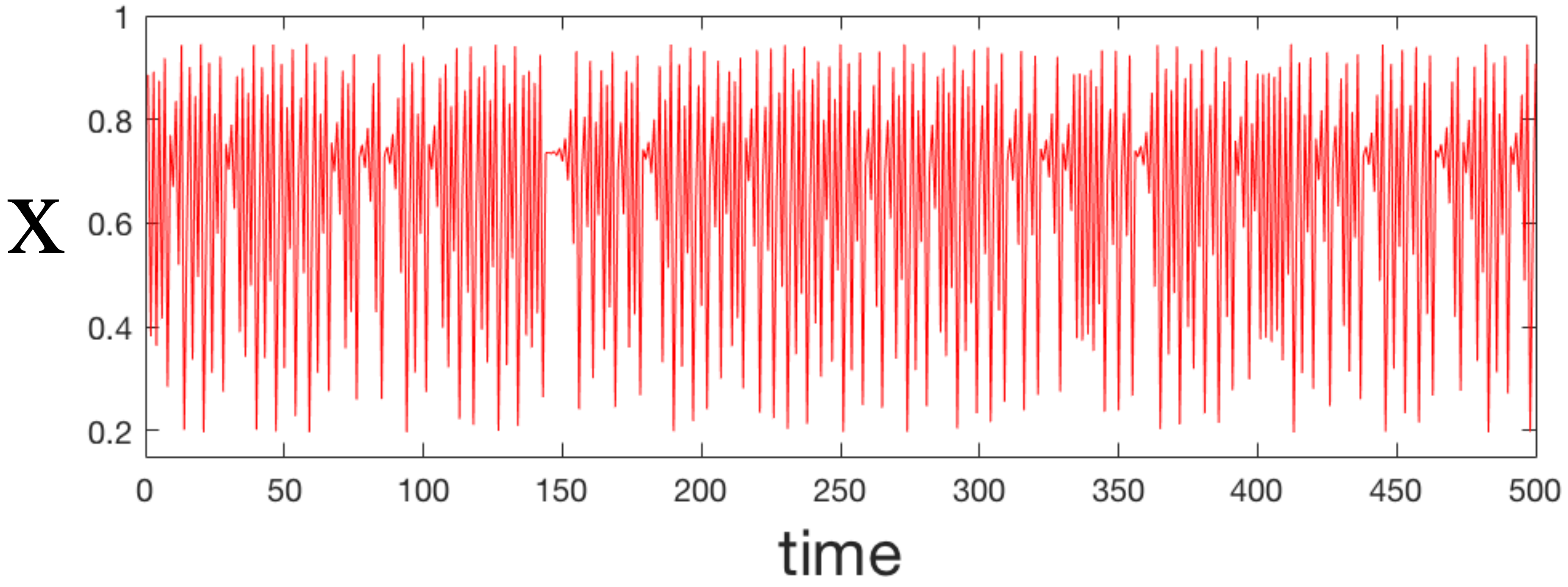
Transfer entropy computation using the Perron-Frobenius operator
D. Diego, K.A. Haaga and B. Hannisdal

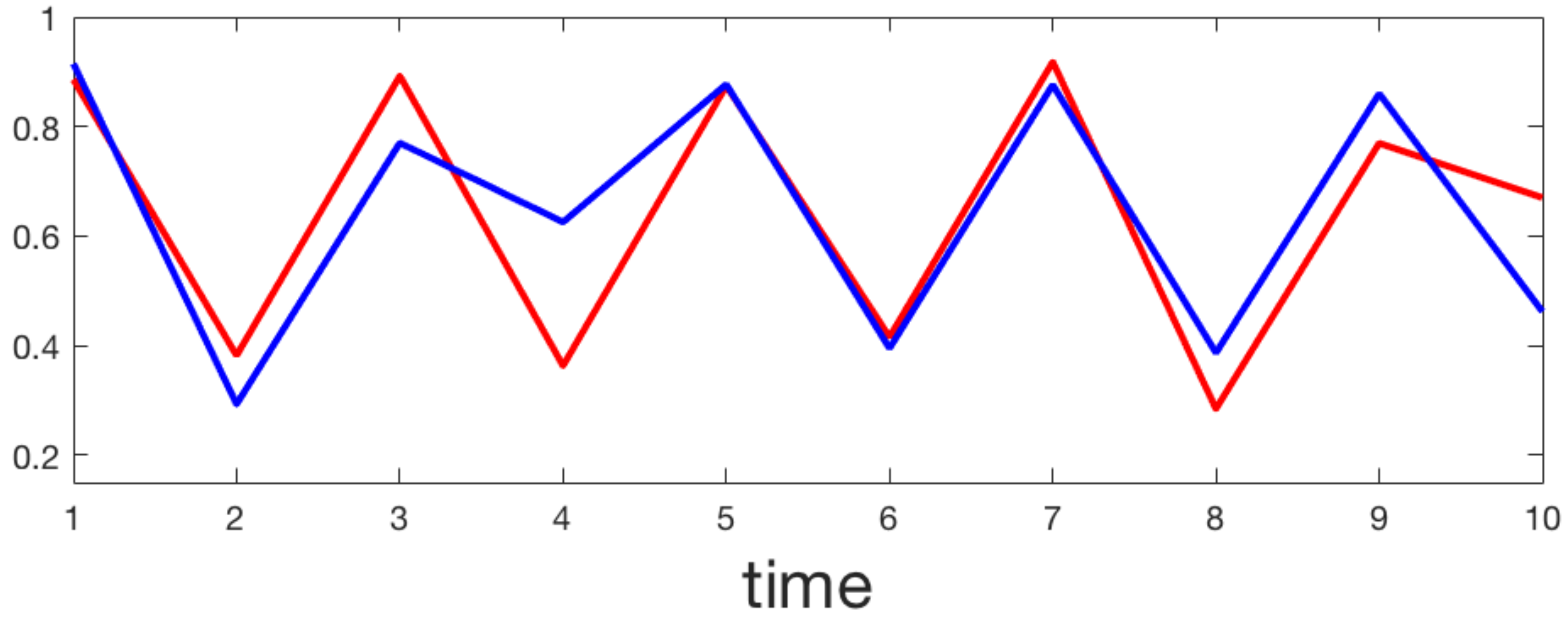
DOI: 10.1103/PhysRevE.99.042212

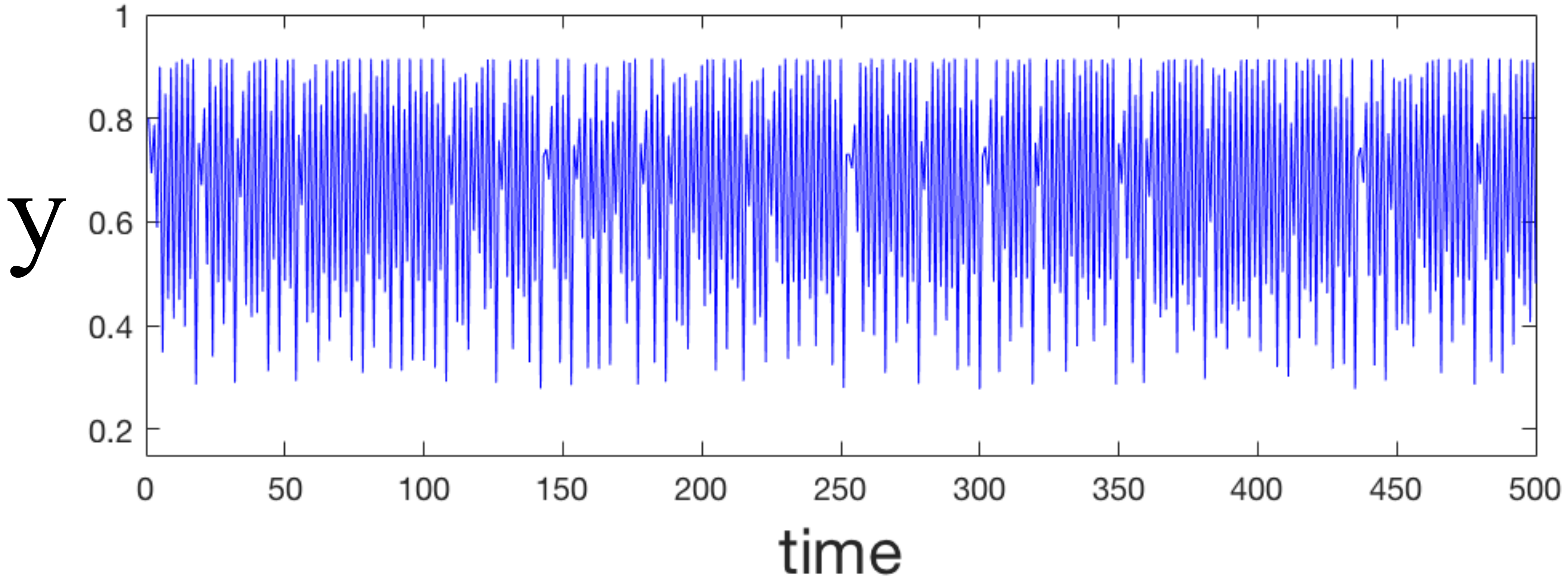
CORRELATION vs CAUSATION

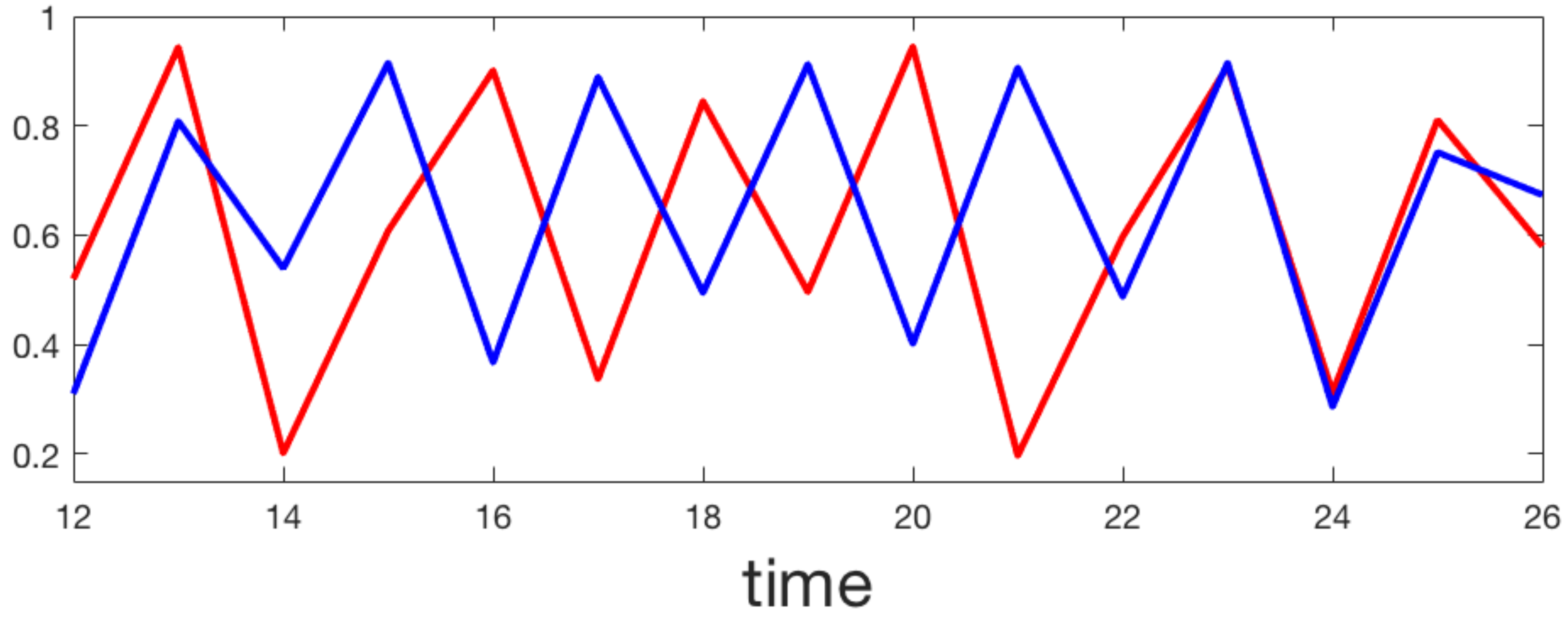


CORRELATION vs CAUSATION

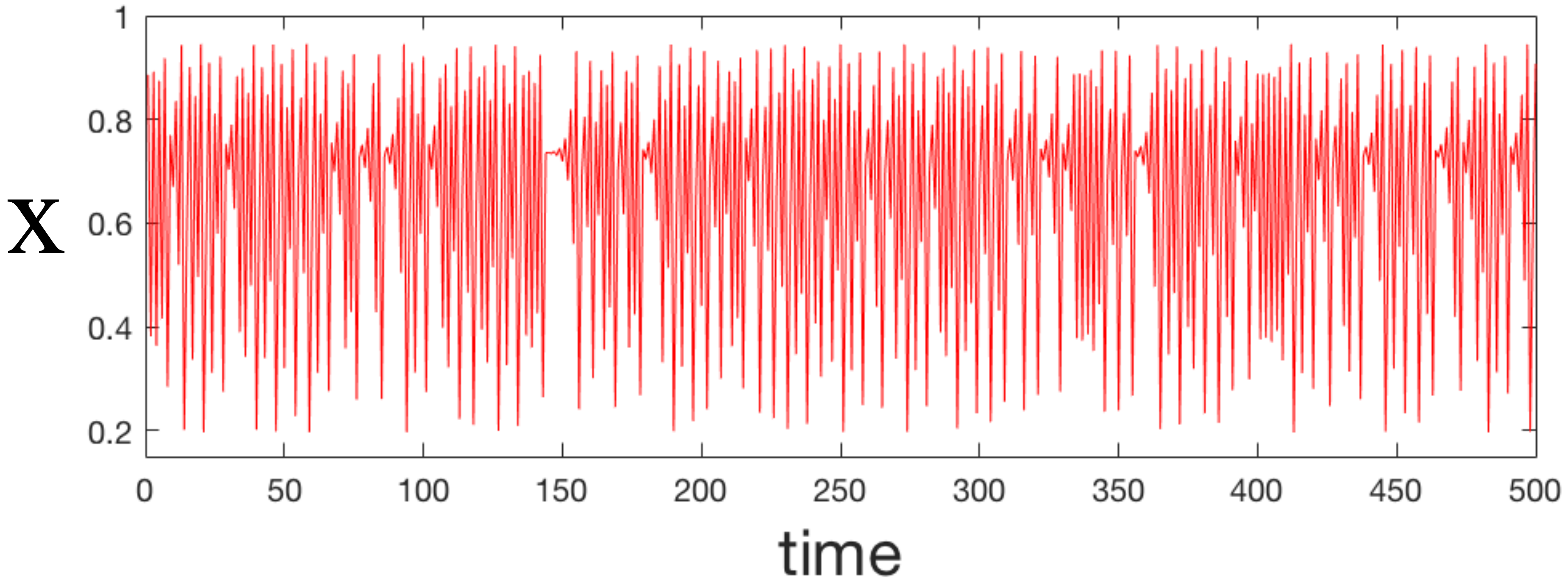


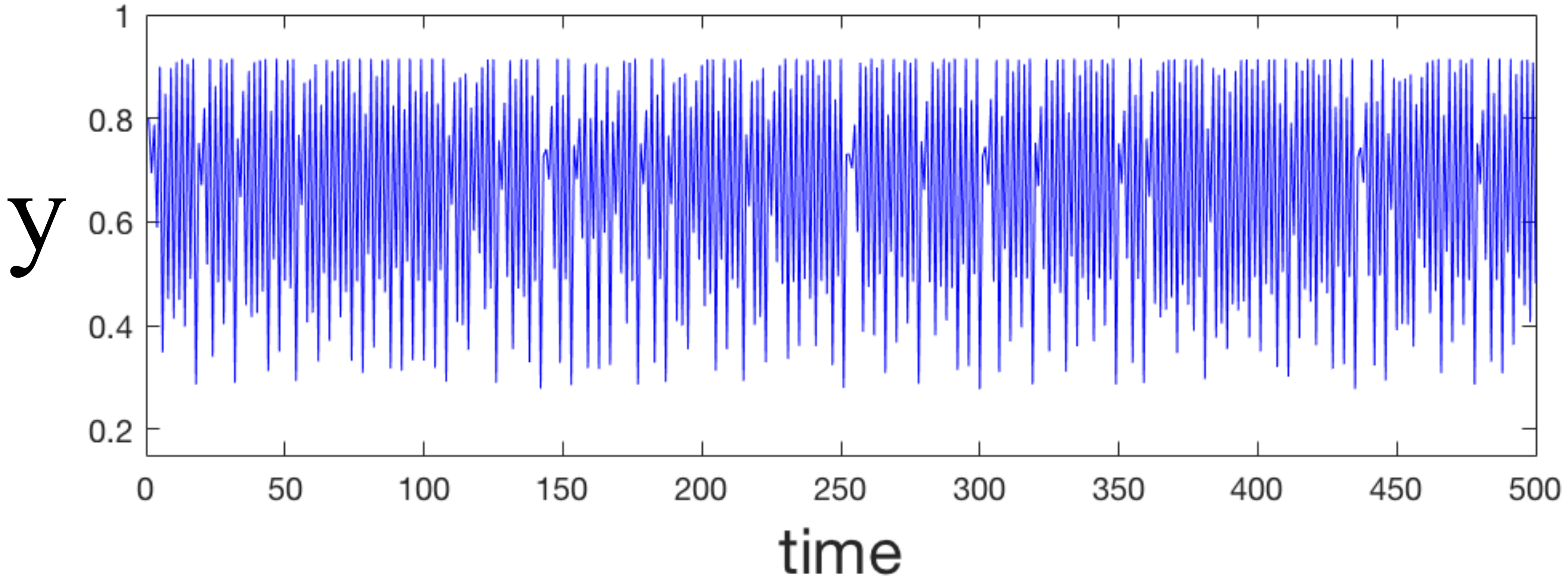


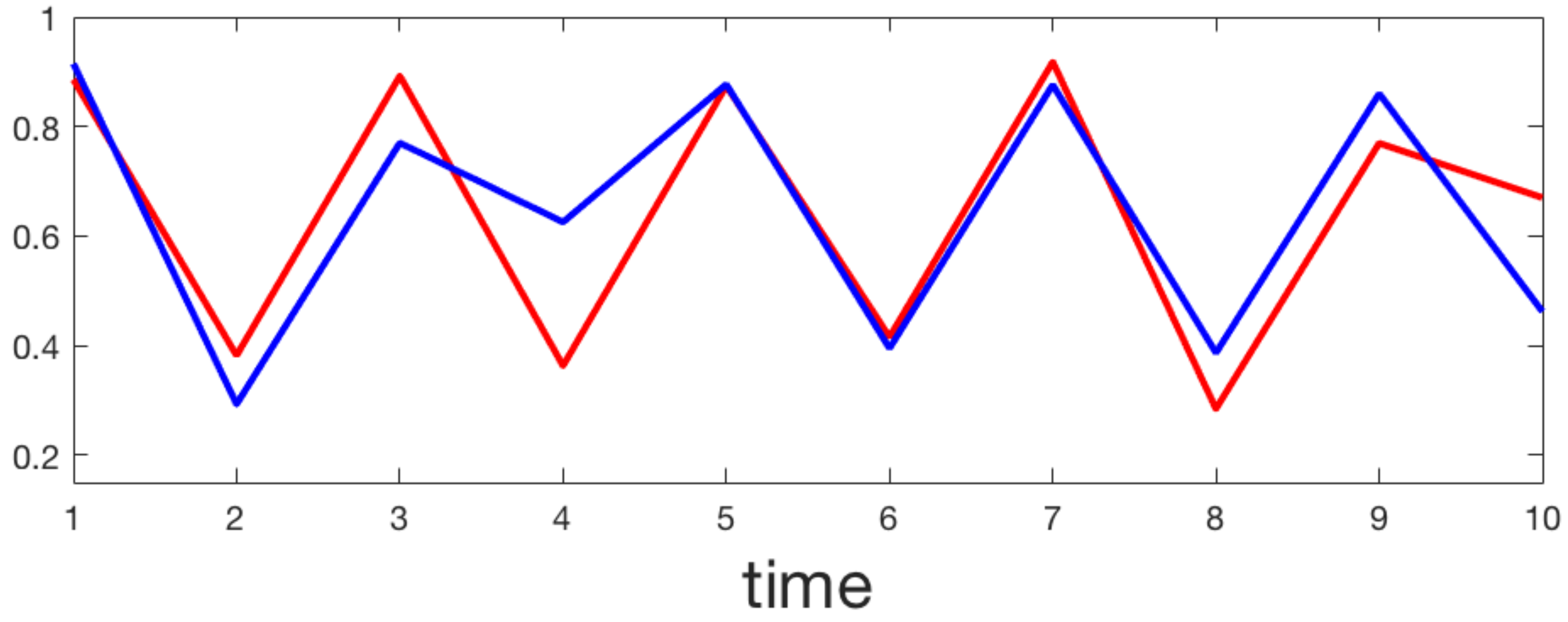


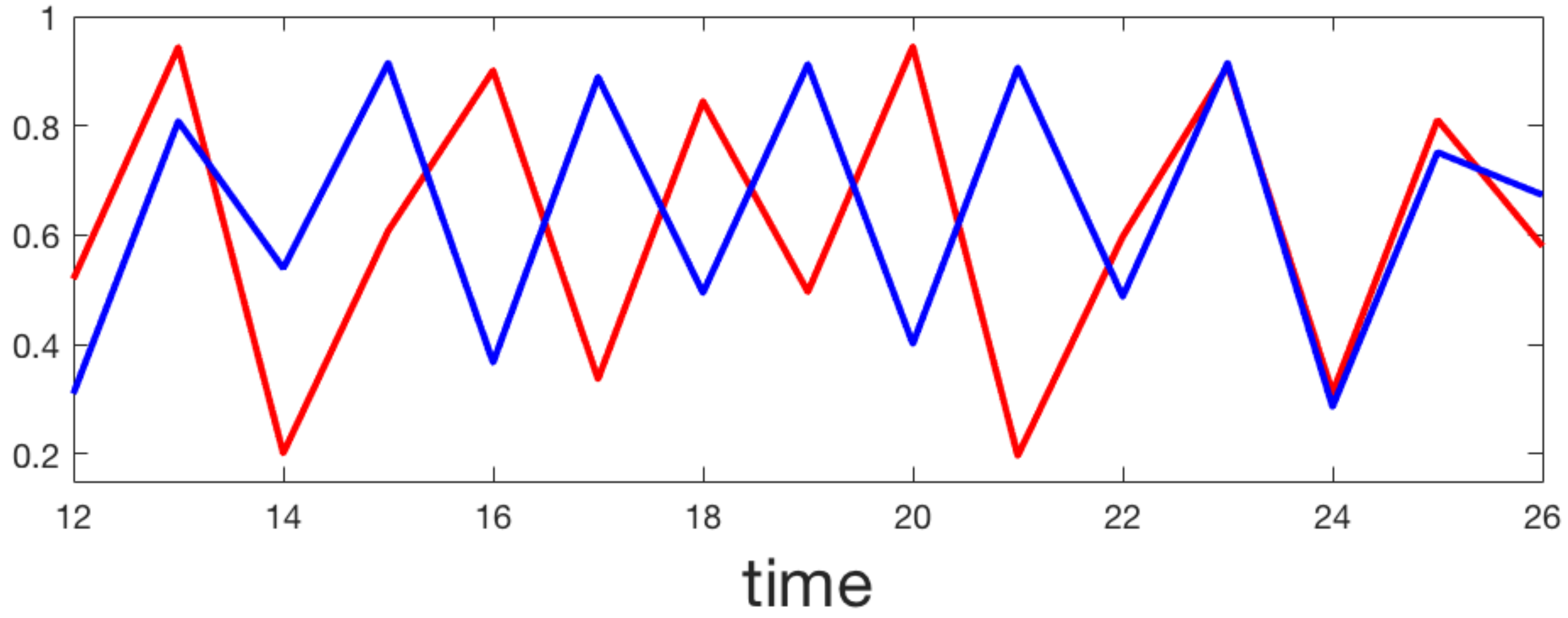


CORRELATION vs CAUSATION









$$x_{n+1} = r_1 x_n (1 - x_n)$$

$$y_{n+1} = r_2 f(x_n, y_n)\,(1 - f(x_n, y_n))$$

$$f(x, y) = \frac{y + c\,x}{1 + c}$$

INFORMATION (beyond correlation)

Given two time series of data … (almost) any data …

$$X = \{x_1, x_2, \cdots, x_n\}$$

$$Y = \{y_1, y_2, \cdots, y_n\}$$

We want to quantify (estimate) how much information about Y is contained in X and vice versa.

A standard tool is the **transfer entropy (TE)**

T. Schreiber (2000), Physical Review Letters 85, 461

Conditional probabilities and transfer entropy

Given two variables, A and B, the conditional probability for having an outcome for A given that the outcome for B is known, is defined as

$$P(a|b) = \frac{P(a,b)}{P(b)}$$

Conditional probabilities and transfer entropy

Given two variables, A and B, the conditional probability for having an outcome for A given that the outcome for B is known, is defined as

probability for obtaining a and b at the same time

$$P(a|b) = \frac{P(a,b)}{P(b)}$$

Conditional probabilities and transfer entropy

Given two variables, A and B, the conditional probability for having an outcome for A given that the outcome for B is known, is defined as

$$P(a|b) = \frac{P(a,b)}{P(b)}$$

probability for obtaining b regardless of what happens with the outcomes of A.

Conditional probabilities and transfer entropy

Given two variables, A and B, the conditional probability for having an outcome for A given that the outcome for B is known, is defined as

$$P(a|b) = \frac{P(a,b)}{P(b)}$$

$$P(b) = \sum_{a \in A} P(a,b)$$

Conditional probabilities and transfer entropy

This can be extended to three (or more) variables:

$$P(a|b,c) = \frac{P(a,b,c)}{P(b,c)}$$

Conditional probabilities and transfer entropy

Given two time series

$$X = \{x_1, x_2, \cdots, x_n\}$$

$$Y = \{y_1, y_2, \cdots, y_n\}$$

Conditional probabilities and transfer entropy

Given two time series

$$X = \{x_1, x_2, \cdots, x_n\}$$

$$Y = \{y_1, y_2, \cdots, y_n\}$$

Suppose one may compute the transition probabilities $P(x_{t+\tau}|x_t)$ and $P(x_{t+\tau}|x_t, y_t)$ for a fixed forward lag τ

Conditional probabilities and transfer entropy

Given two time series

$$X = \{x_1, x_2, \cdots, x_n\}$$

$$Y = \{y_1, y_2, \cdots, y_n\}$$

Suppose one may compute the transition probabilities $P(x_{t+\tau}|x_t)$ and $P(x_{t+\tau}|x_t, y_t)$ for a fixed forward lag τ

If $P(x_{t+\tau}|x_t, y_t) > P(x_{t+\tau}|x_t)$ then having the information about y_t helps predicting the transition $x_t \to x_{t+\tau}$

Conditional probabilities and transfer entropy

Given two time series

$$X = \{x_1, x_2, \cdots, x_n\}$$

$$Y = \{y_1, y_2, \cdots, y_n\}$$

Suppose one may compute the transition probabilities $P(x_{t+\tau}|x_t)$ and $P(x_{t+\tau}|x_t, y_t)$ for a fixed forward lag τ

If $P(x_{t+\tau}|x_t, y_t) > P(x_{t+\tau}|x_t)$ then having the information about y_t helps predicting the transition $x_t \rightarrow x_{t+\tau}$

For some other time t', the opposite could happen …

$$P(x_{t'+\tau}|x_{t'}, y_{t'}) < P(x_{t'+\tau}|x_{t'})$$

Conditional probabilities and transfer entropy

A way of checking whether there is any improvement in the average is the **Transfer Entropy**

$$\boxed{TE(Y \rightarrow X) = \sum_t P(x_{t+\tau}, x_t, y_t) \, \log_2 \frac{P(x_{t+\tau}|x_t, y_t)}{P(x_{t+\tau}|x_t)}}$$

$TE(Y \rightarrow X) \geq 0$

$TE(Y \rightarrow X) = 0$ if and only if $P(x_{t+\tau}|x_t, y_t) = P(x_{t+\tau}|x_t)$ for all t !

Transfer entropy comes from entropy

$$\boxed{TE(Y \to X) = \sum_t P(x_{t+\tau}, x_t, y_t) \log_2 \frac{P(x_{t+\tau}|x_t, y_t)}{P(x_{t+\tau}|x_t)}}$$

The entropy quantifies how much information is required to predict the outcome of a random process with probability distribution:
$P = (p_1, \cdots, p_n)$

$$H(P) = -\sum_{i=1}^{n} p_i \log_2(p_i)$$

$$0 \leq H(P) \leq \log_2(n)$$

Introduced by Shannon in 1948

why the log?

$$\log_2(a \cdot b) = \log_2 a + \log_2 b$$

$$H(P_1 \cdot P_2) = H(P_1) + H(P_2)$$

Transfer entropy comes from entropy

$$TE(Y \to X) = \sum_t P(x_{t+\tau}, x_t, y_t) \, \log_2 \frac{P(x_{t+\tau}|x_t, y_t)}{P(x_{t+\tau}|x_t)}$$

The entropy quantifies how much information is required to predict the outcome of a random process with probability distribution:
$P = (p_1, \cdots, p_n)$

$$H(P) = -\sum_{i=1}^{n} p_i \log_2(p_i)$$

$$0 \leq H(P) \leq \log_2(n)$$

Introduced by Shannon in 1948

why the log?

$$\log_2(a \cdot b) = \log_2 a + \log_2 b$$

$$H(P_1 \cdot P_2) = H(P_1) + H(P_2)$$

(silly?) examples

$$H(P) = -p_h \log_2(p_h) - p_t \log_2(p_t)$$

- Fair coin: $p_h = 0.5$, $p_t = 0.5$ — 1 bit
- Unfair coin: $p_h = 0.8$, $p_t = 0.2$ — 0.72 bits
- Cheating coin: $p_h = 1$, $p_t = 0$ — 0 bits

Transfer entropy (TE)

$TE(Y \to X)$ measures the relevance of the information about Y to predict the evolution of X

$TE(Y \to X)$ is taken as the flow of information from Y to X

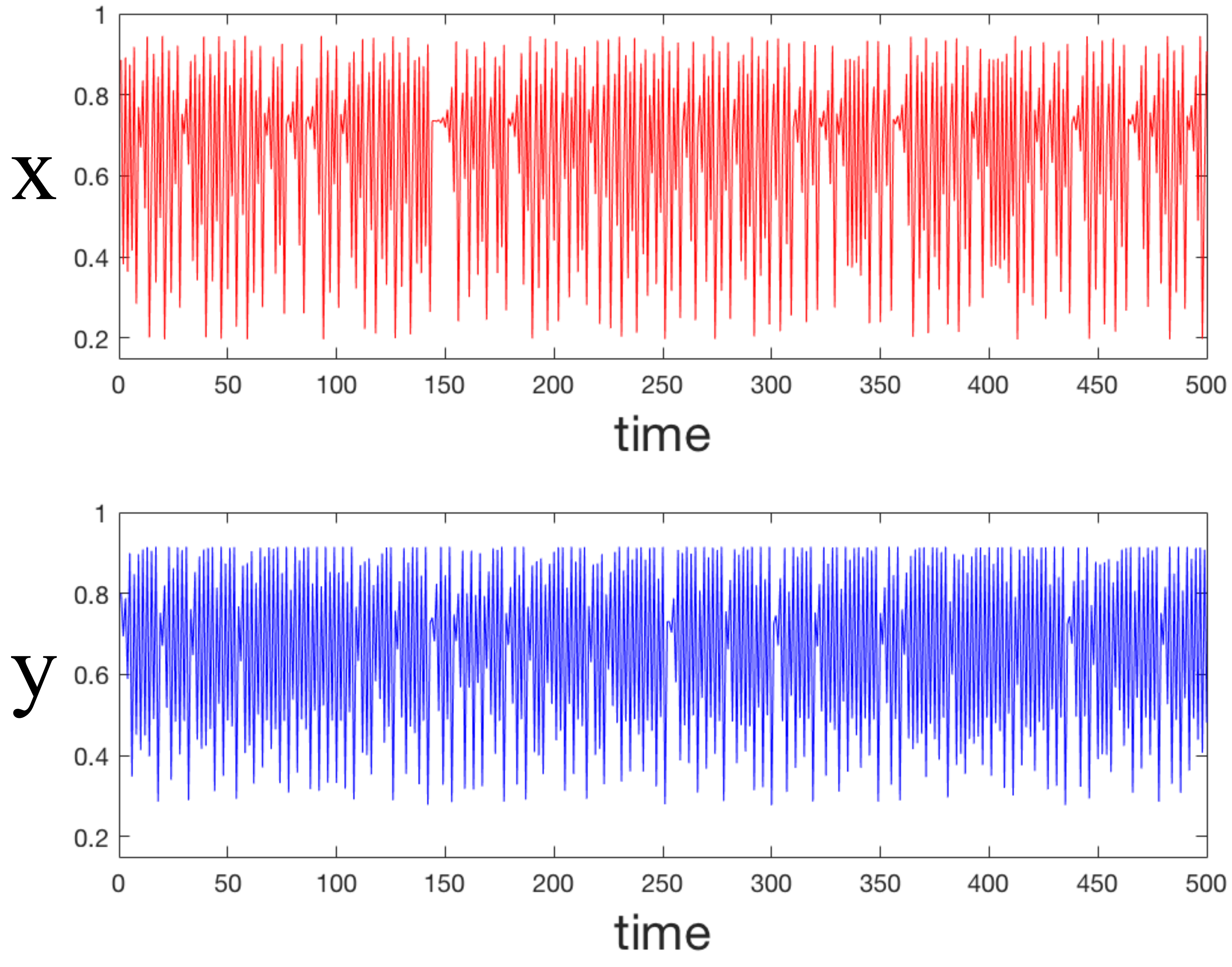
To estimate TE from Y to X we first need to get the probability distribution

$$P(x_{t+\tau}, x_t, y_t)$$

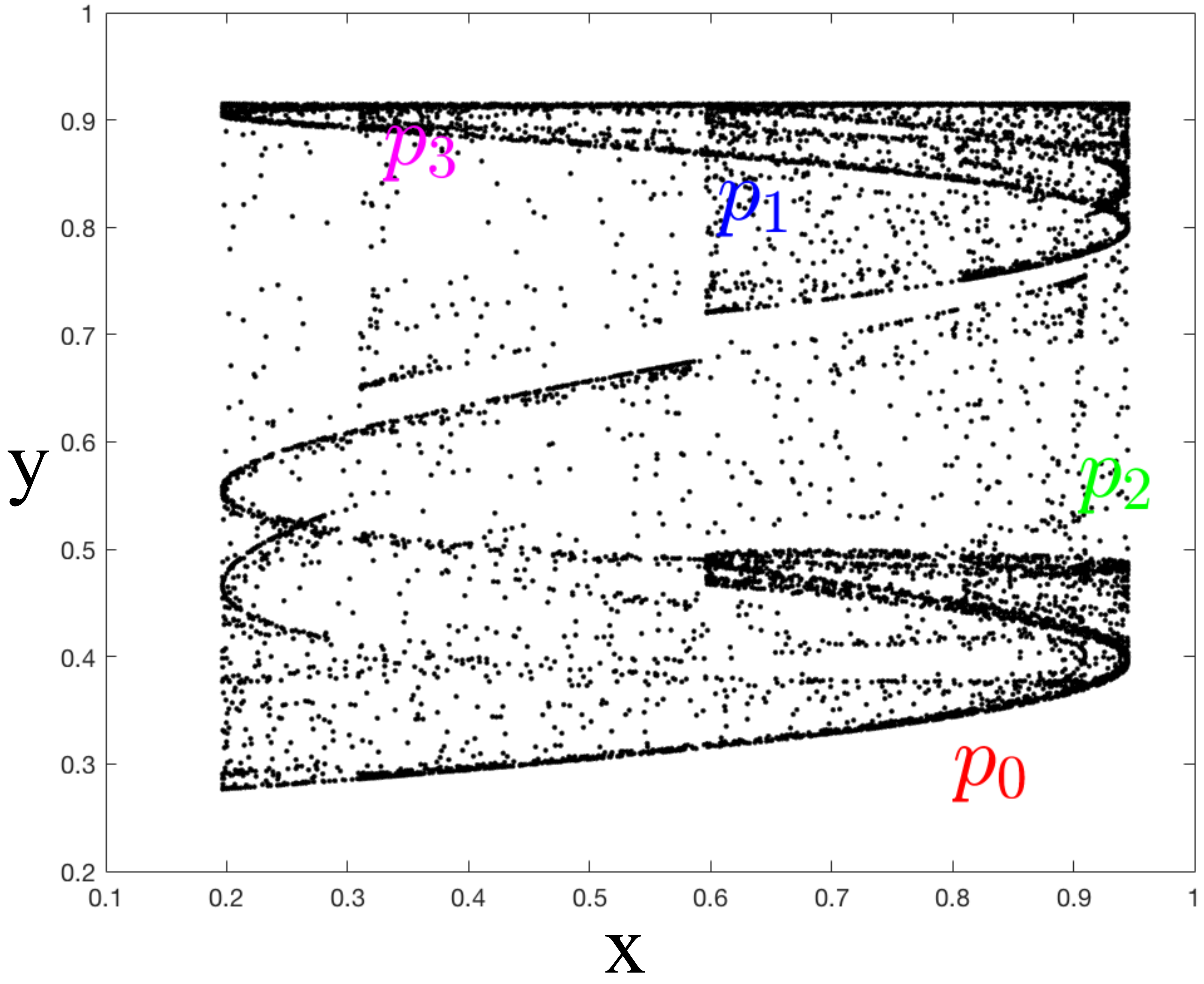
and from it, the transition probabilities $P(x_{t+\tau}|x_t)$ **and** $P(x_{t+\tau}|x_t, y_t)$

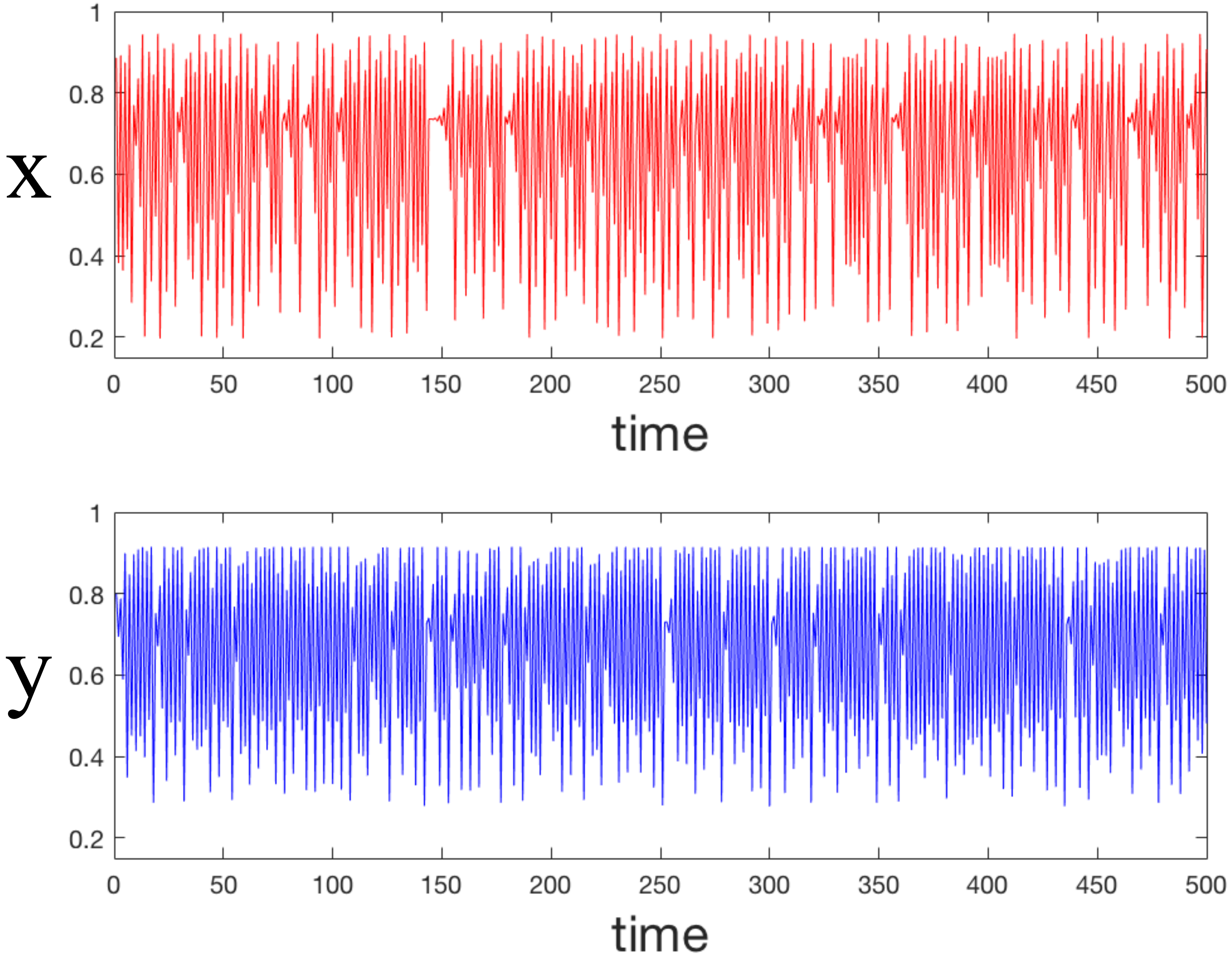
We use the invariant density associated to the dynamics underlying the time series X and Y (call the invariant density ρ)

WHAT IS AN INVARIANT DENSITY?

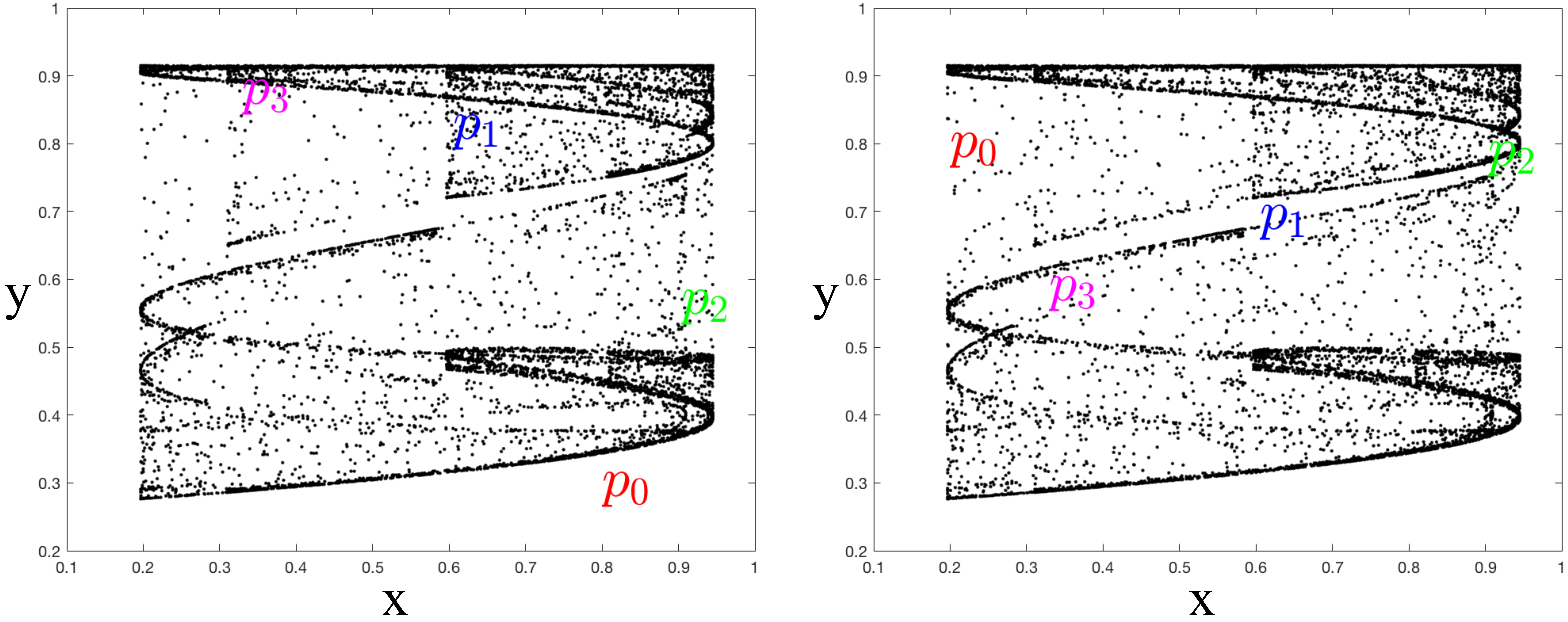


WHAT IS AN INVARIANT DENSITY?





WHAT IS AN INVARIANT DENSITY?

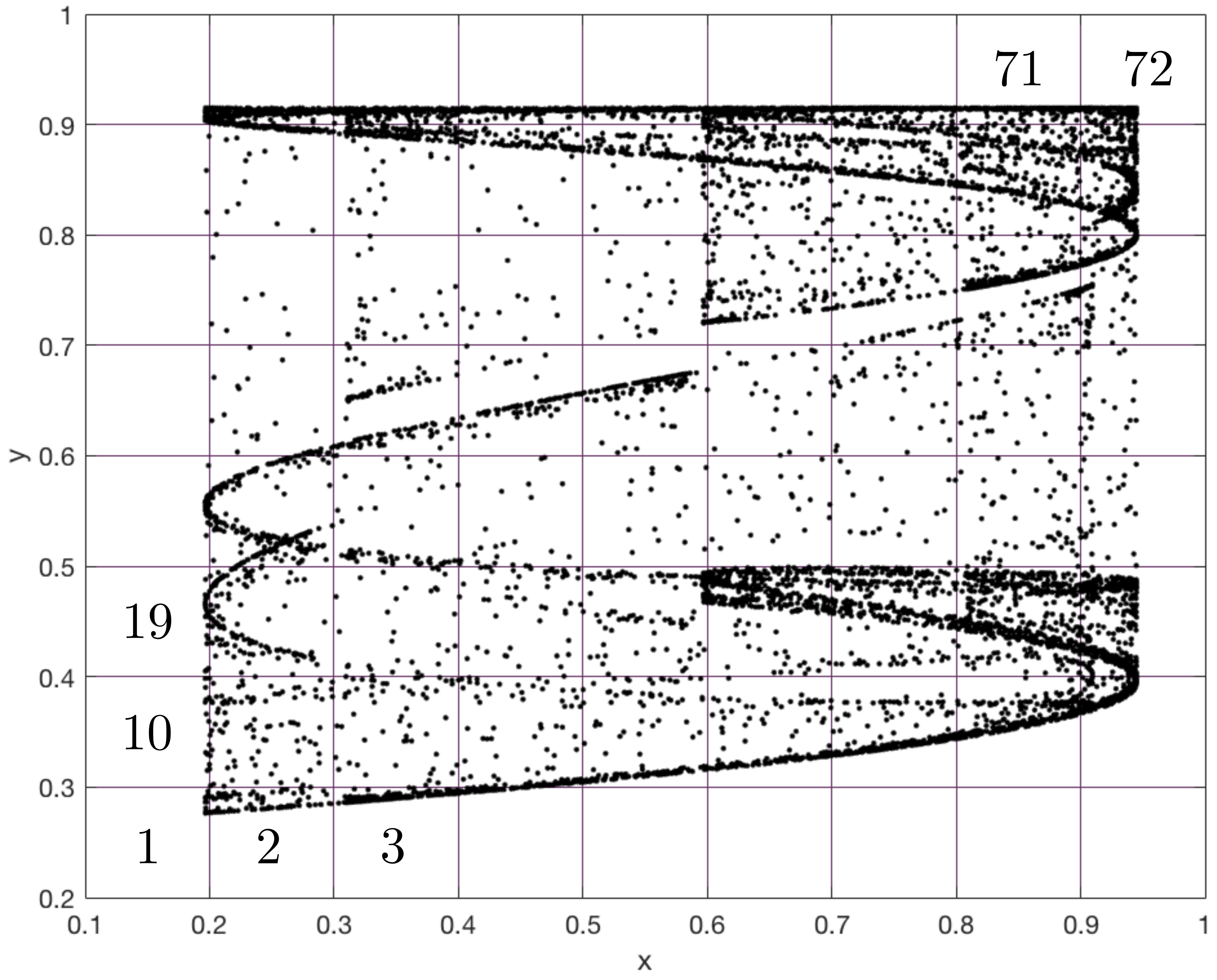


Ways to estimate the invariant density:

- Visitation frequency

Label the bins in this grid as

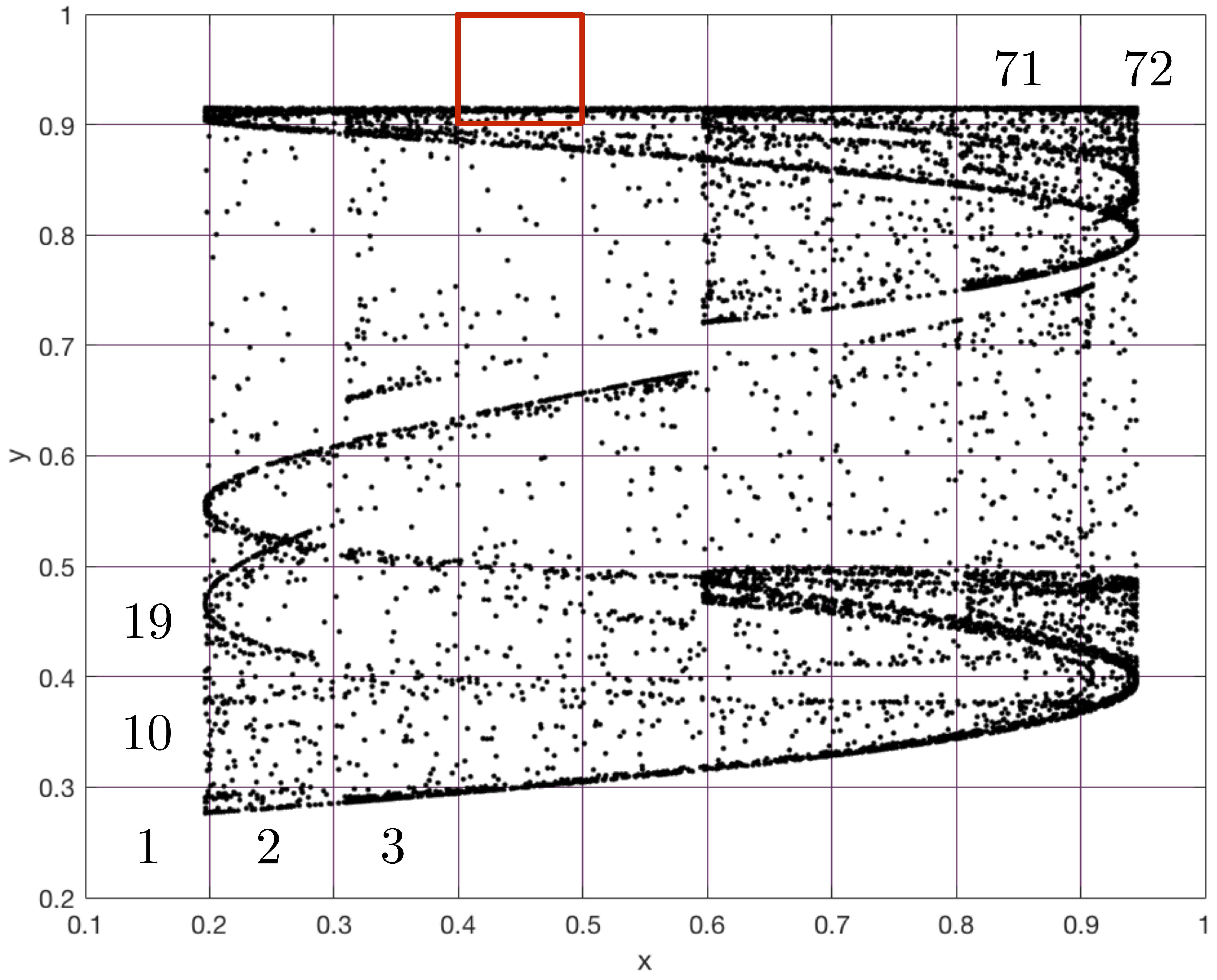
$$1, 2, \cdots, 72$$



Total number of points in the orbit: 10^4

Ways to estimate the invariant density:

- Visitation frequency



Label the bins in this grid as

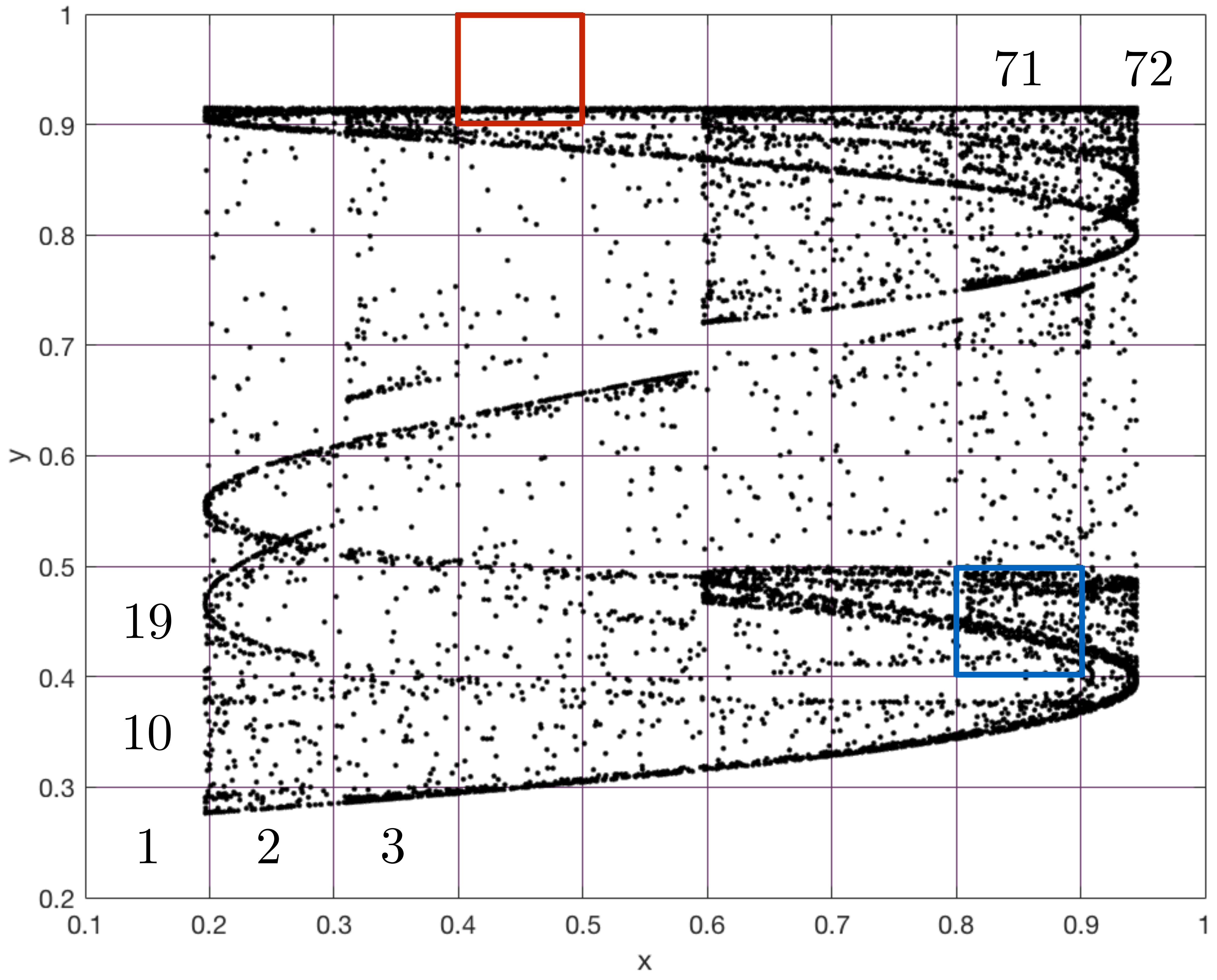
$$1, 2, \cdots, 72$$

$$\rho_{67} \sim \frac{n}{N} = \frac{189}{10^4} \simeq 0.019$$

Total number of points in the orbit: 10^4

Ways to estimate the invariant density:

- Visitation frequency



Label the bins in this grid as

$$1, 2, \cdots, 72$$

$$\rho_{67} \sim \frac{n}{N} = \frac{189}{10^4} \simeq 0.019$$

$$\rho_{26} \sim \frac{n}{N} = \frac{518}{10^4} \simeq 0.052$$

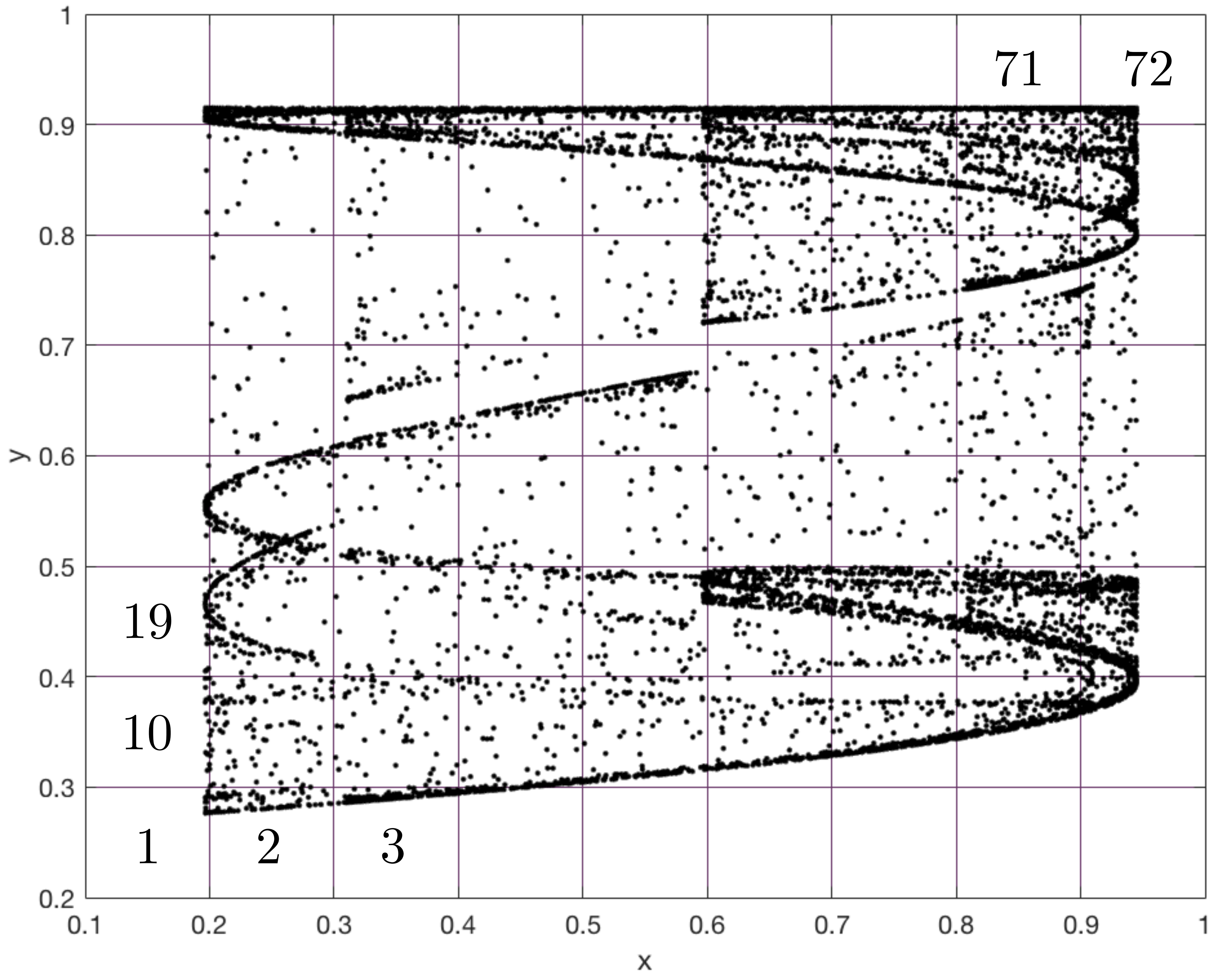
Total number of points in the orbit: 10^4

Ways to estimate the invariant density:

- Via the transfer matrix (Ulam's approximation)

Approximate the unknown function generating the time series by a matrix of transition probabilities (transfer matrix):

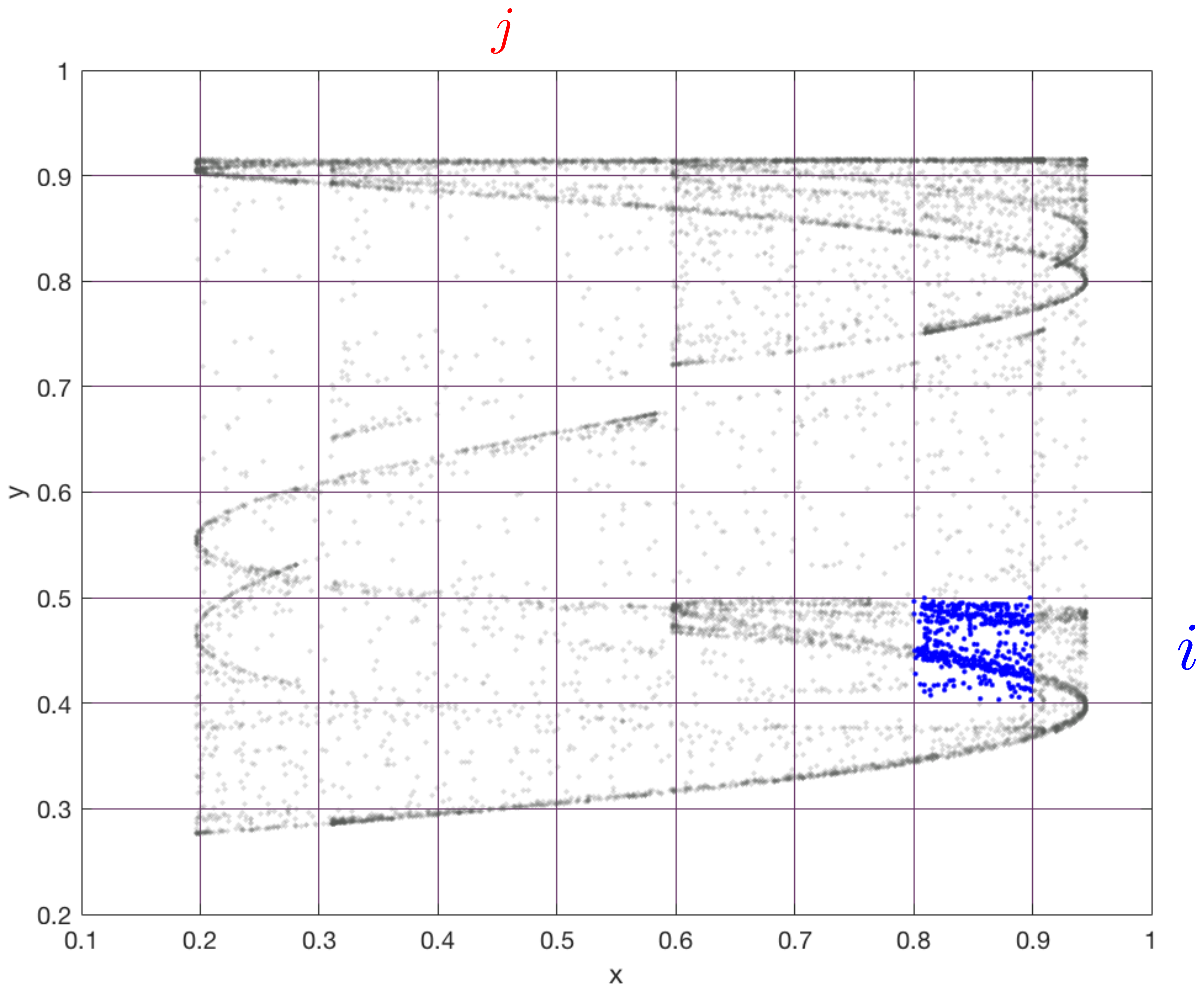
$$T_{ij}\ , \qquad i,j = 1,\cdots,72$$



Ways to estimate the invariant density:

- Via the transfer matrix (Ulam's approximation)

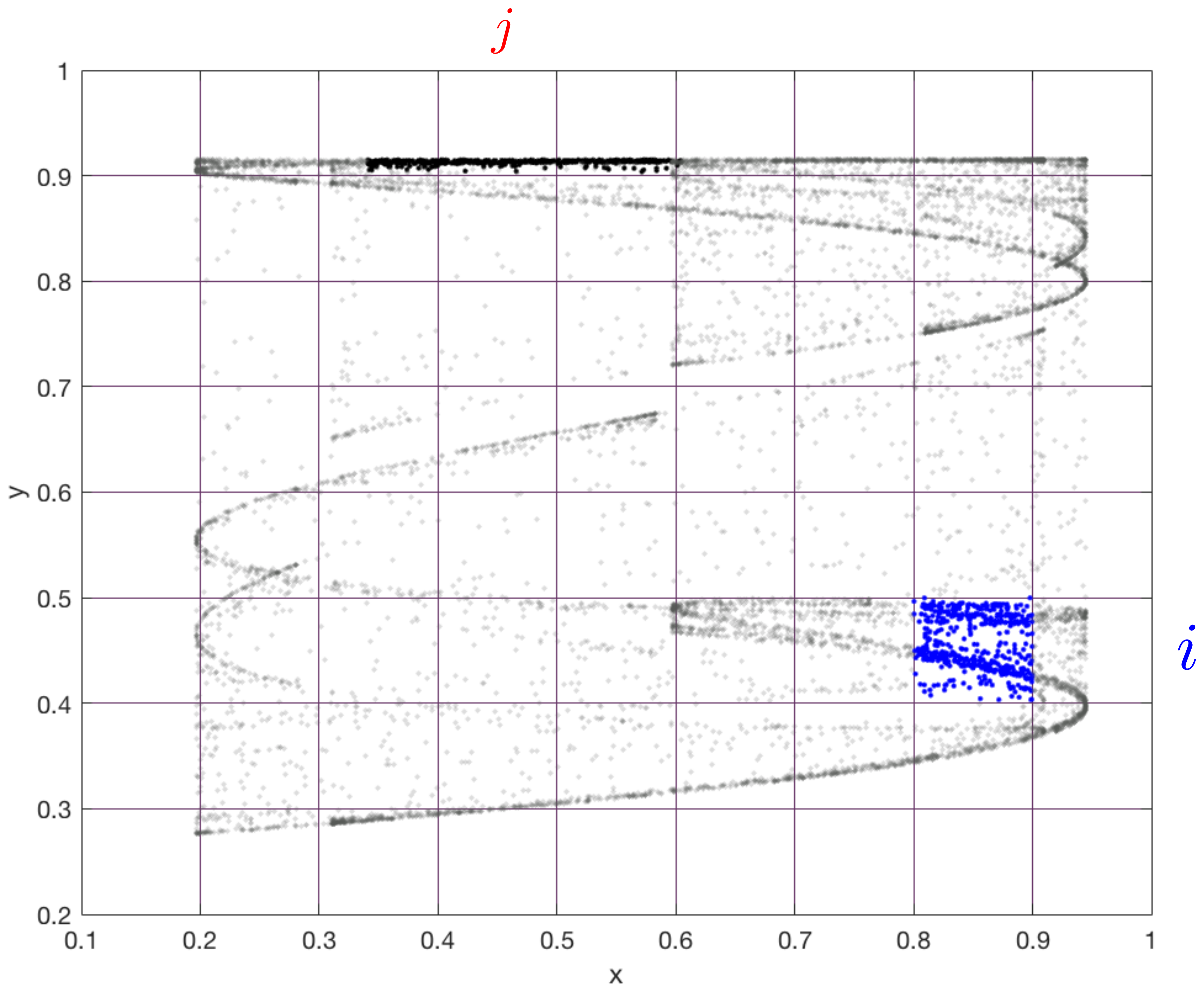
$$T_{ij}\,?$$



Suppose $p_1, p_2, \cdots, p_{N_i}$ are the points in the orbit falling in the i-th bin …

Ways to estimate the invariant density:

- Via the transfer matrix (Ulam's approximation)



$$T_{ij}\ ?$$

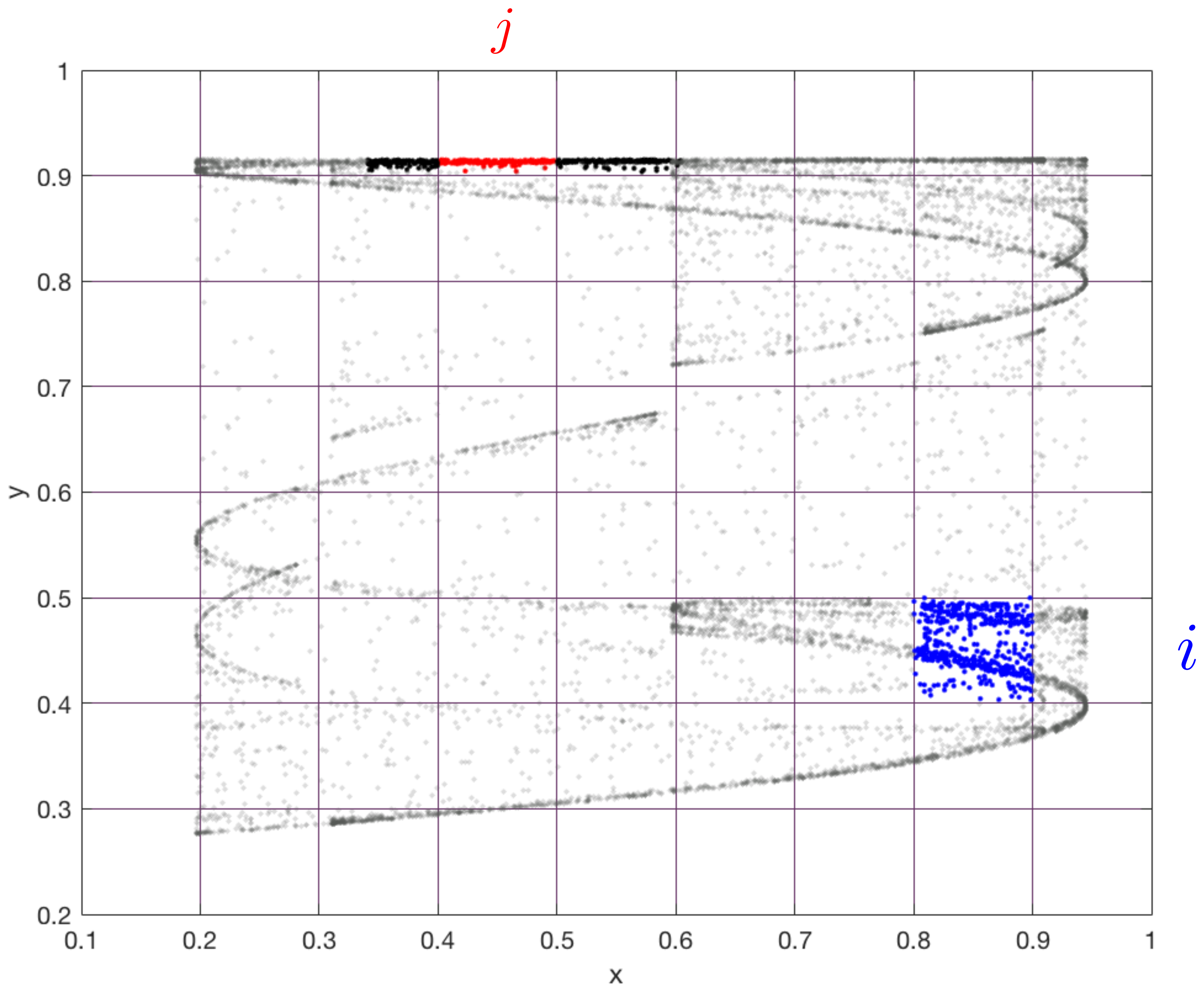
Suppose $p_1, p_2, \cdots, p_{N_i}$ are the points in the orbit falling in the i-th bin …

where will those points be in the next time step?

Ways to estimate the invariant density:

- Via the transfer matrix (Ulam's approximation)

$$T_{ij}\,?$$



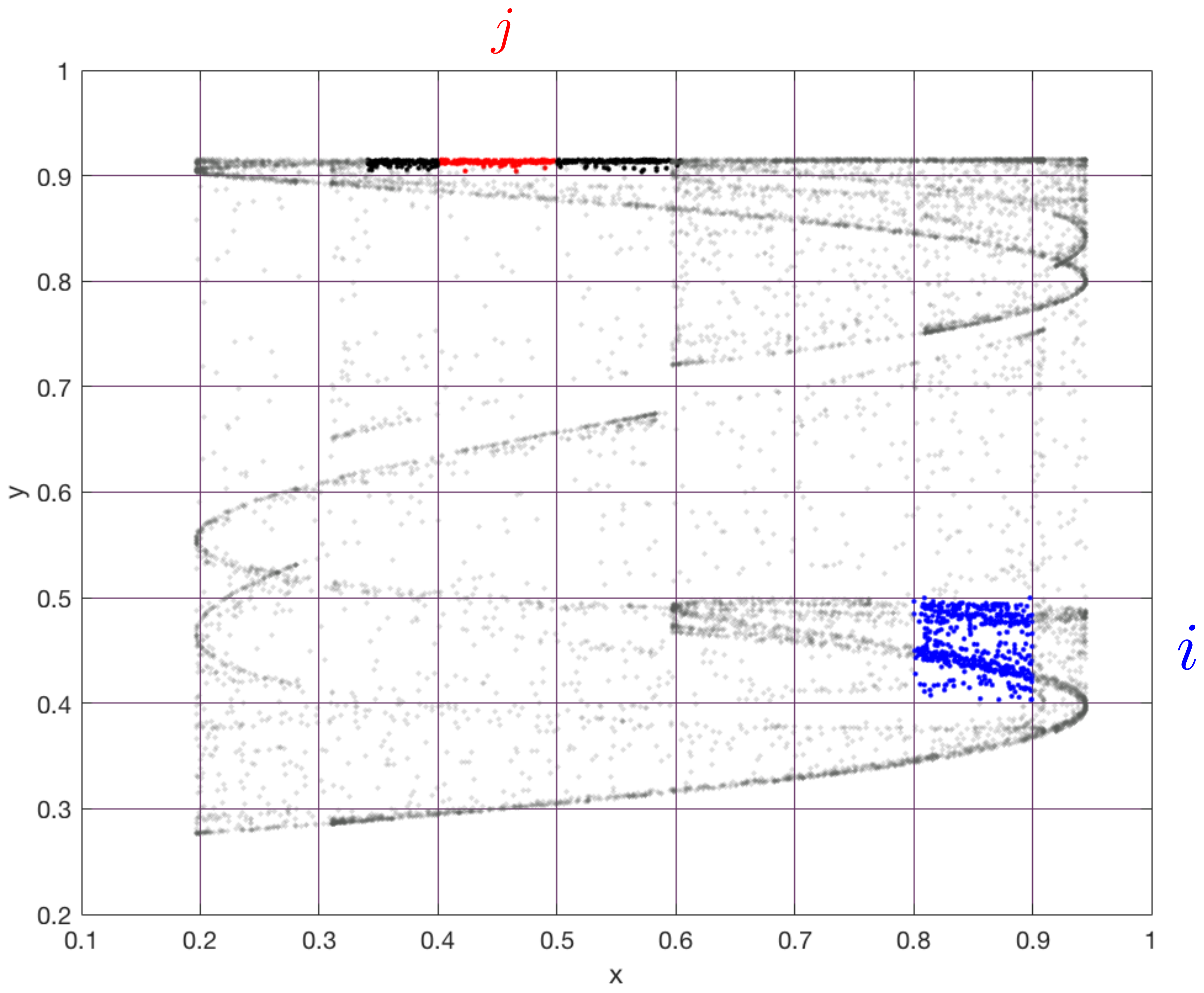
Suppose $p_1, p_2, \cdots, p_{N_i}$ are the points in the orbit falling in the i-th bin …

where will those points be in the next time step?

suppose that n_j fall into the j-th bin …

Ways to estimate the invariant density:

- Via the transfer matrix (Ulam's approximation)



$$T_{ij}\,?$$

Suppose $p_1, p_2, \cdots, p_{N_i}$ are the points in the orbit falling in the i-th bin …

where will those points be in the next time step?

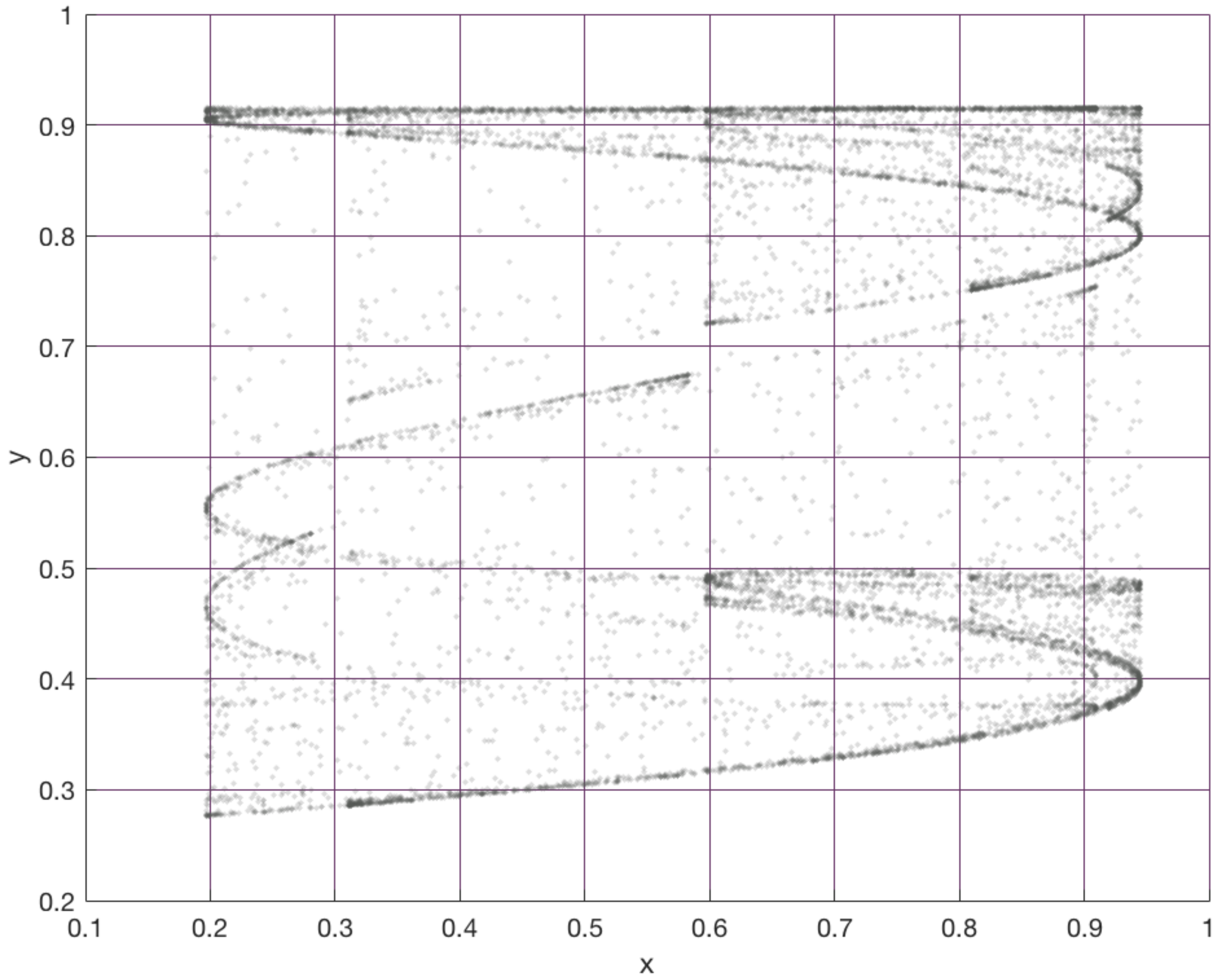
suppose that n_j fall into the j-th bin …

Then

$$T_{ij} = \frac{n_j}{N_i}$$

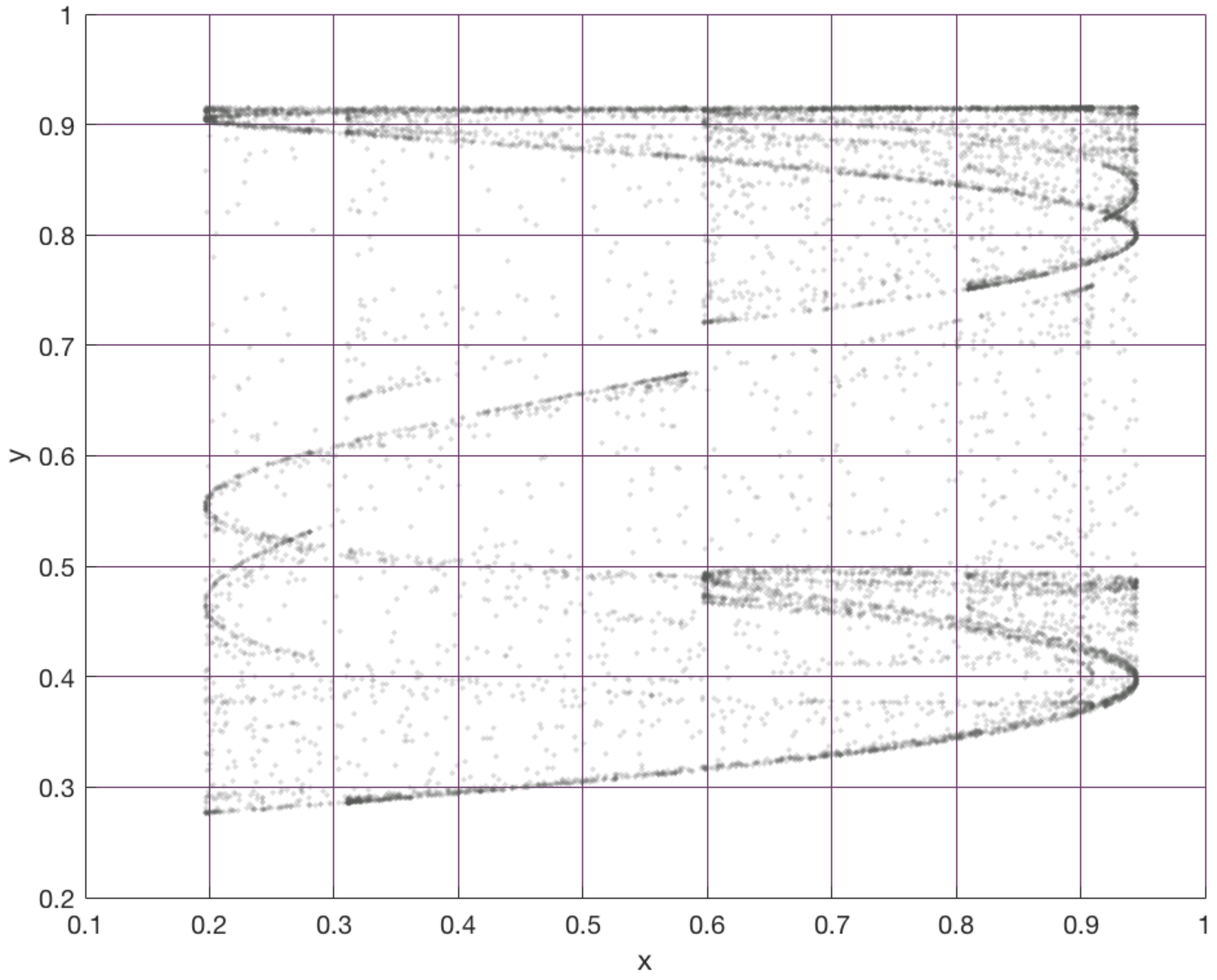
Probability of going from bin i to bin j

How to compute the invariant density from T?



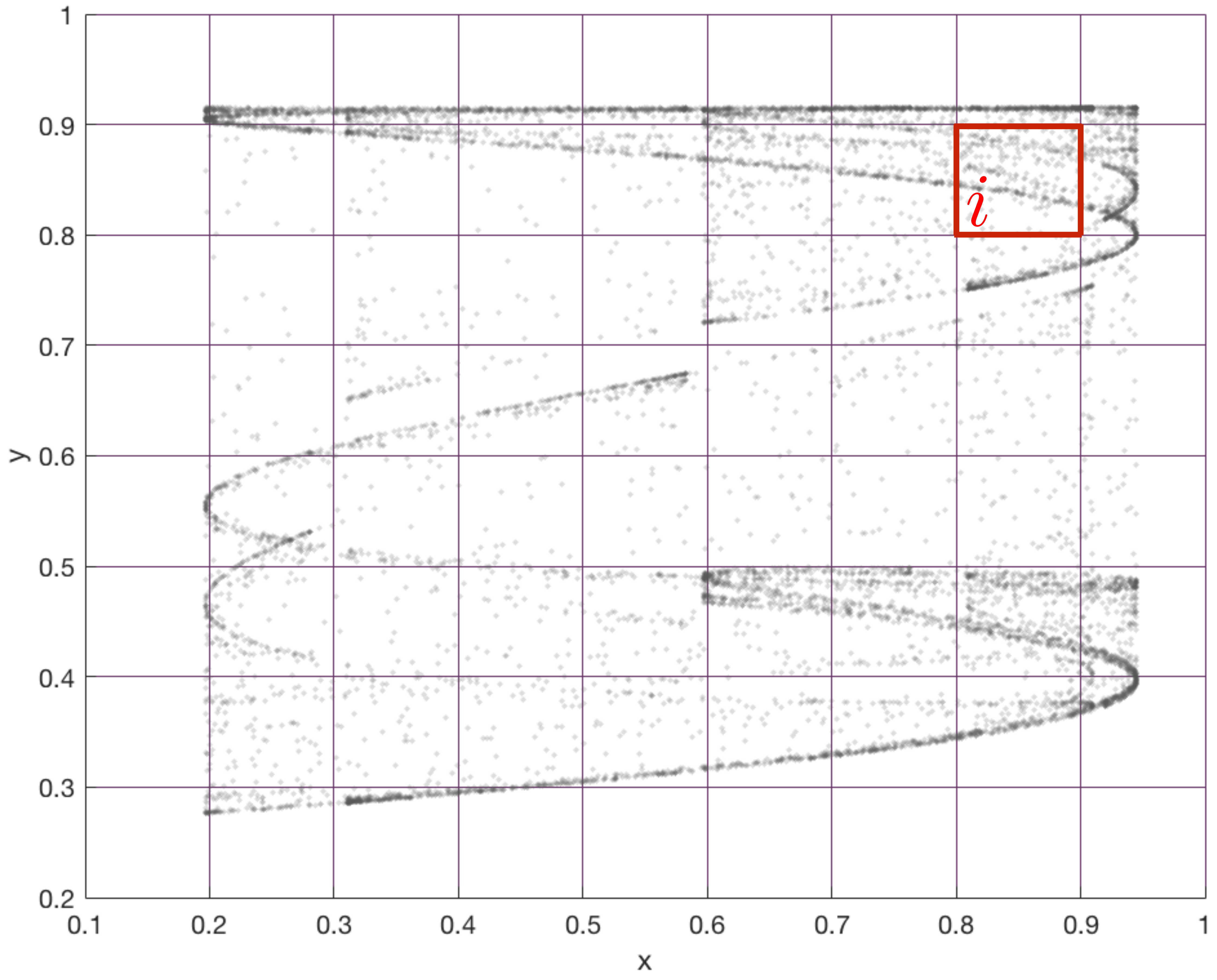
How to compute the invariant density from T ?

Let $P = (P_1, \cdots, P_{72})$ be a probability distribution over the bins.



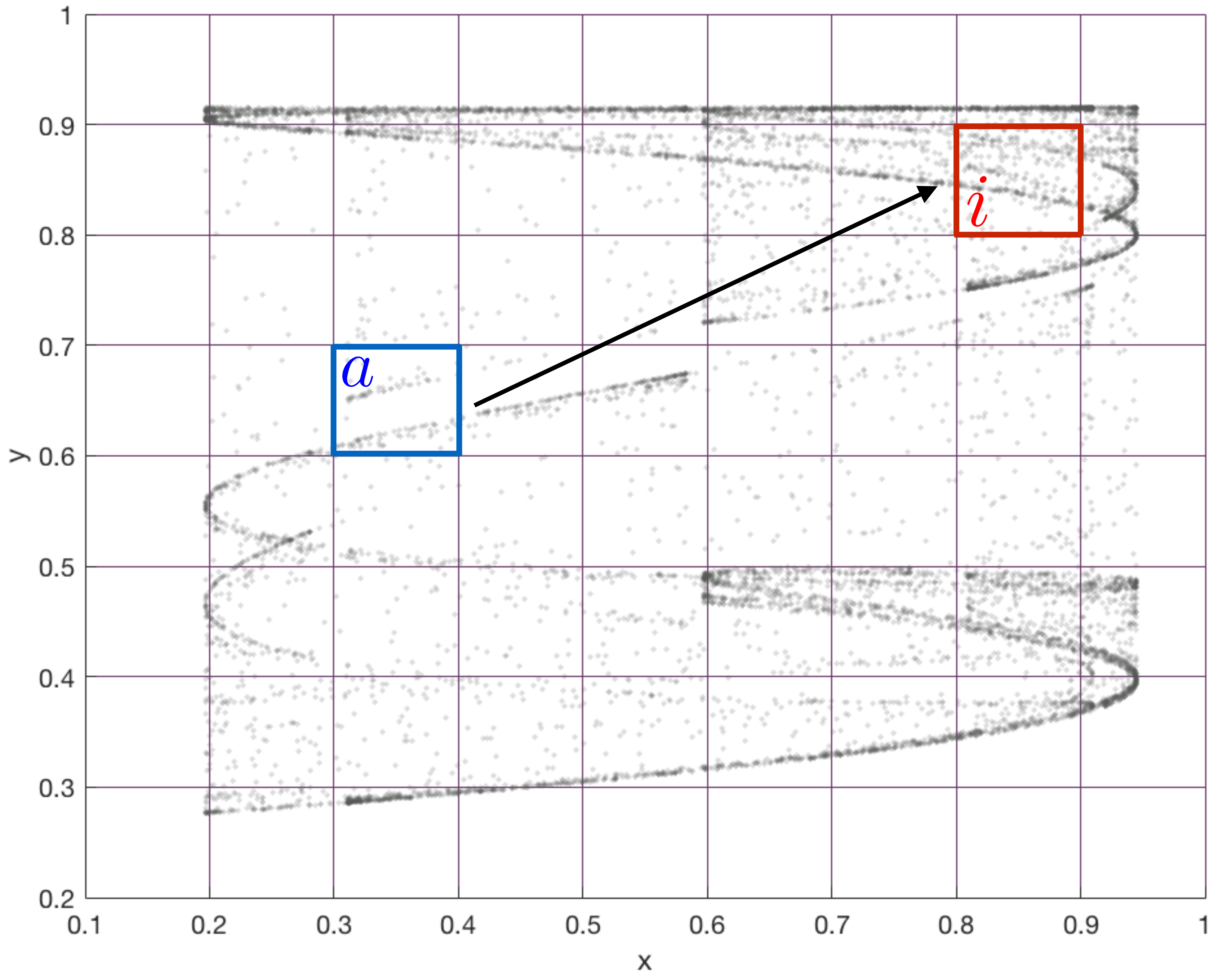
How to compute the invariant density from T ?

Let $P = (P_1, \cdots, P_{72})$ be a probability distribution over the bins.



How to compute the invariant density from T?

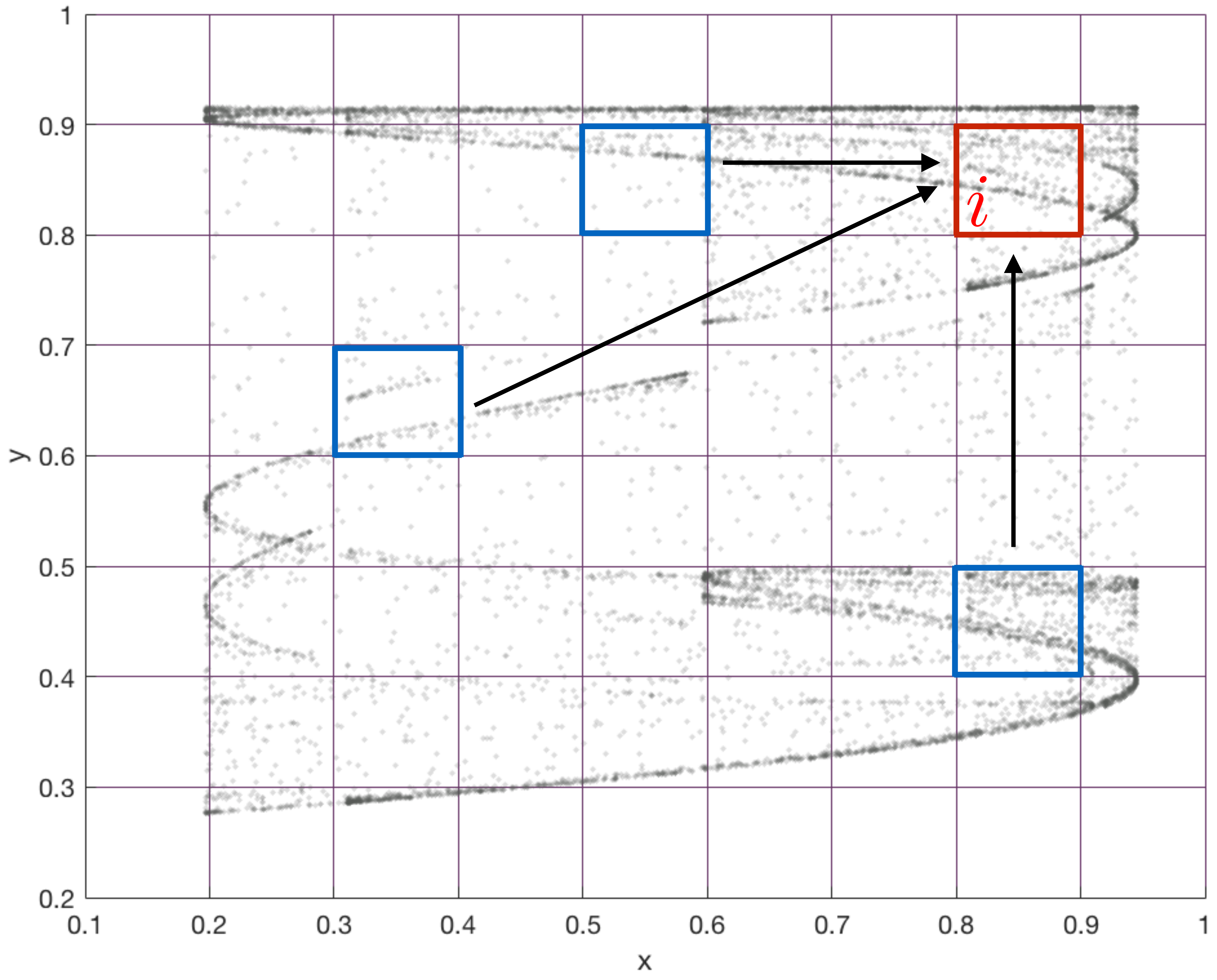
Let $P = (P_1, \cdots, P_{72})$ be a probability distribution over the bins.



$P_{a}\, T_{ai}$

How to compute the invariant density from T?

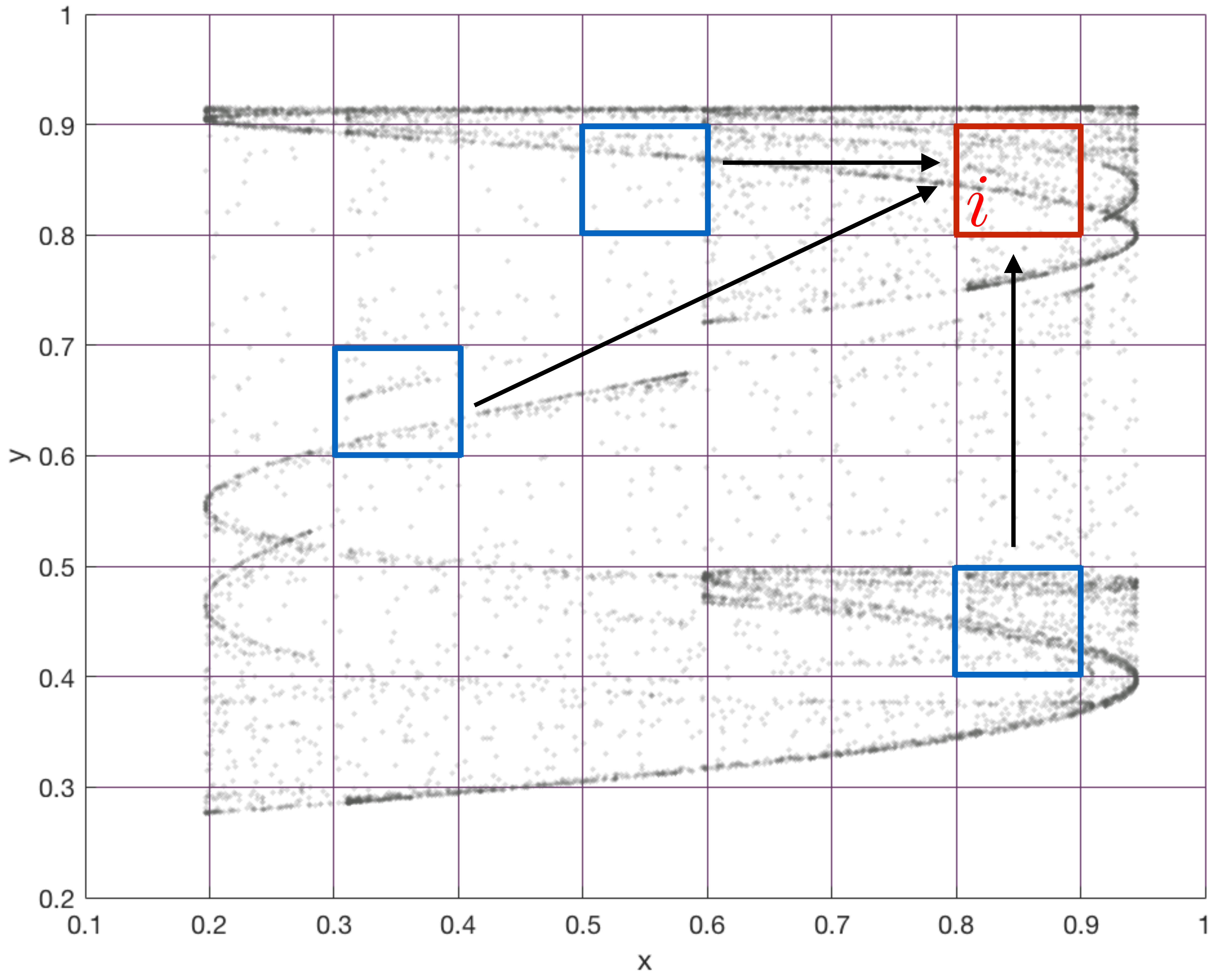
Let $P = (P_1, \cdots, P_{72})$ be a probability distribution over the bins.



$$P'_{i} = \sum_{a=1}^{72} P_a T_{ai} \neq P_i$$

How to compute the invariant density from T ?

Let $P = (P_1, \cdots, P_{72})$ be a probability distribution over the bins.

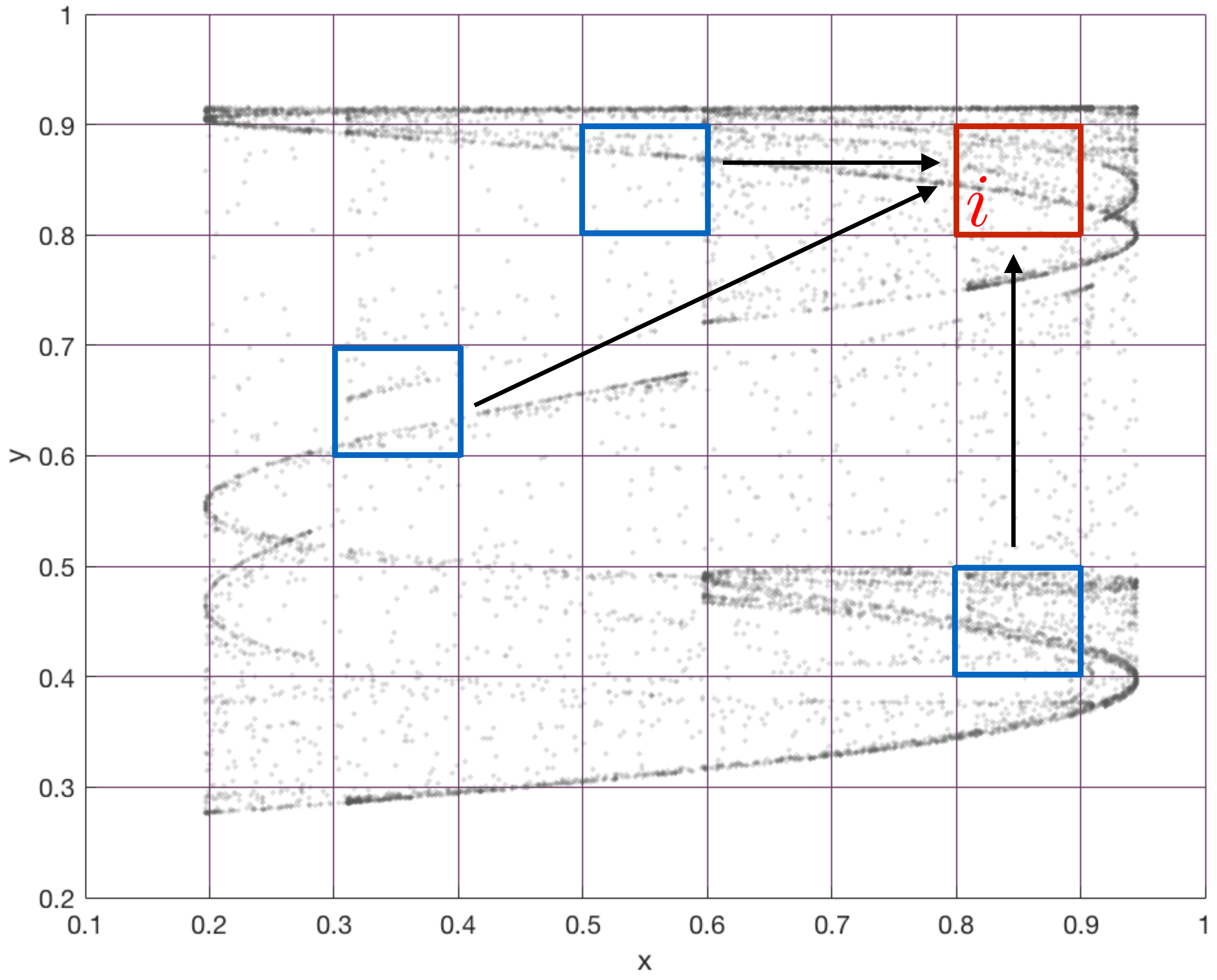


$$P_i' = \sum_{a=1}^{72} P_a T_{ai} \neq P_i$$

$$\begin{aligned} P(1) &= P \cdot T \\ P(2) &= P(1) \cdot T \\ P(3) &= P(2) \cdot T \\ &\vdots \end{aligned} \implies \boxed{\lim_{n \to \infty} P(n) = \rho}$$

How to compute the invariant density from T?

Let $P = (P_1, \cdots, P_{72})$ be a probability distribution over the bins.



$$P'_i = \sum_{a=1}^{72} P_a T_{ai} \neq P_i$$

$$P(1) = P \cdot T$$
$$P(2) = P(1) \cdot T$$
$$P(3) = P(2) \cdot T$$
$$\vdots$$

$$\Longrightarrow \boxed{\lim_{n\to\infty} P(n) = \rho}$$

$$\sum_{a=1}^{72} \rho_a T_{ai} = \rho_i$$

ρ_i is the probability for the system to occupy the i-th bin

ρ is invariant under the action of T

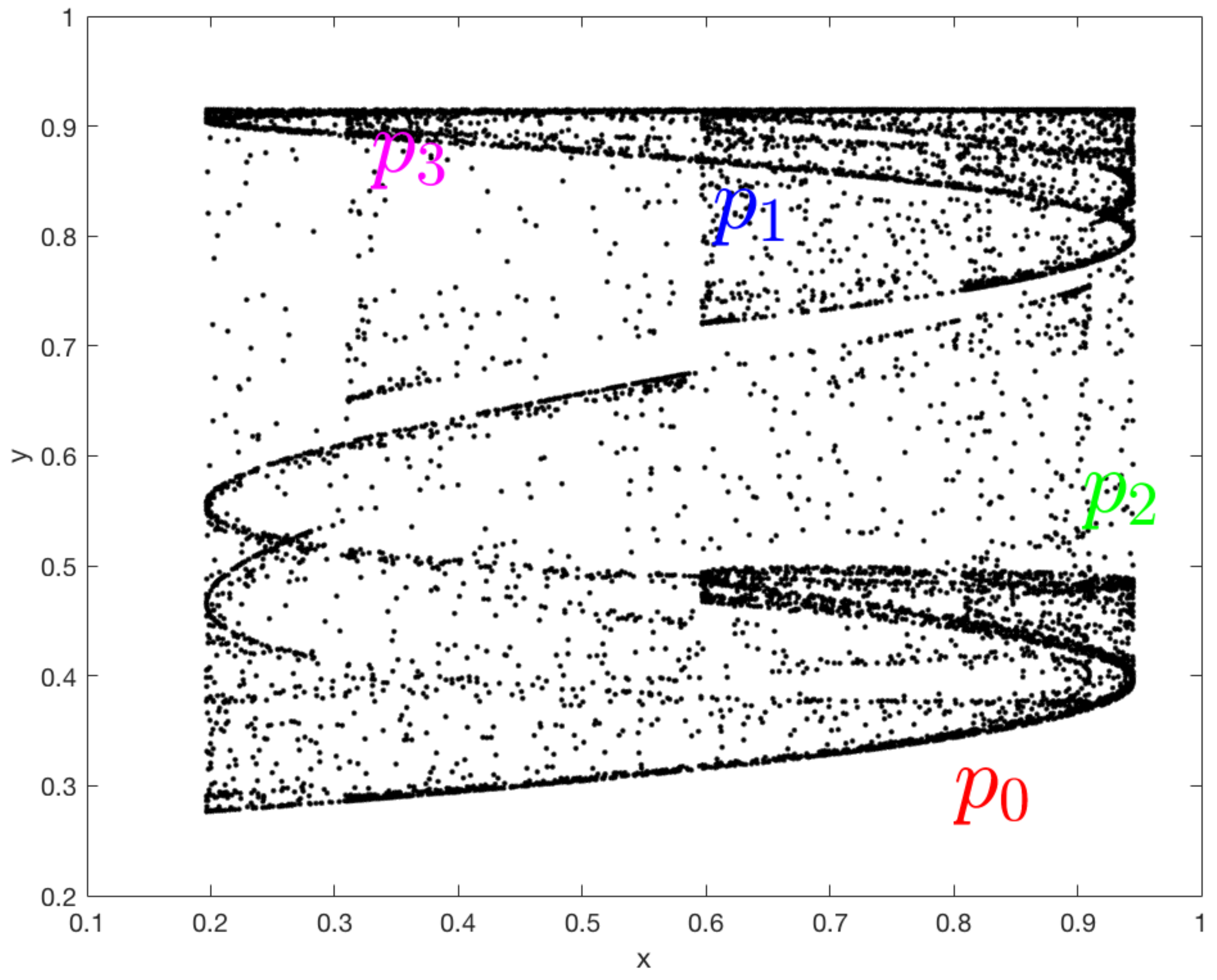
Starting from (almost) any probability distribution

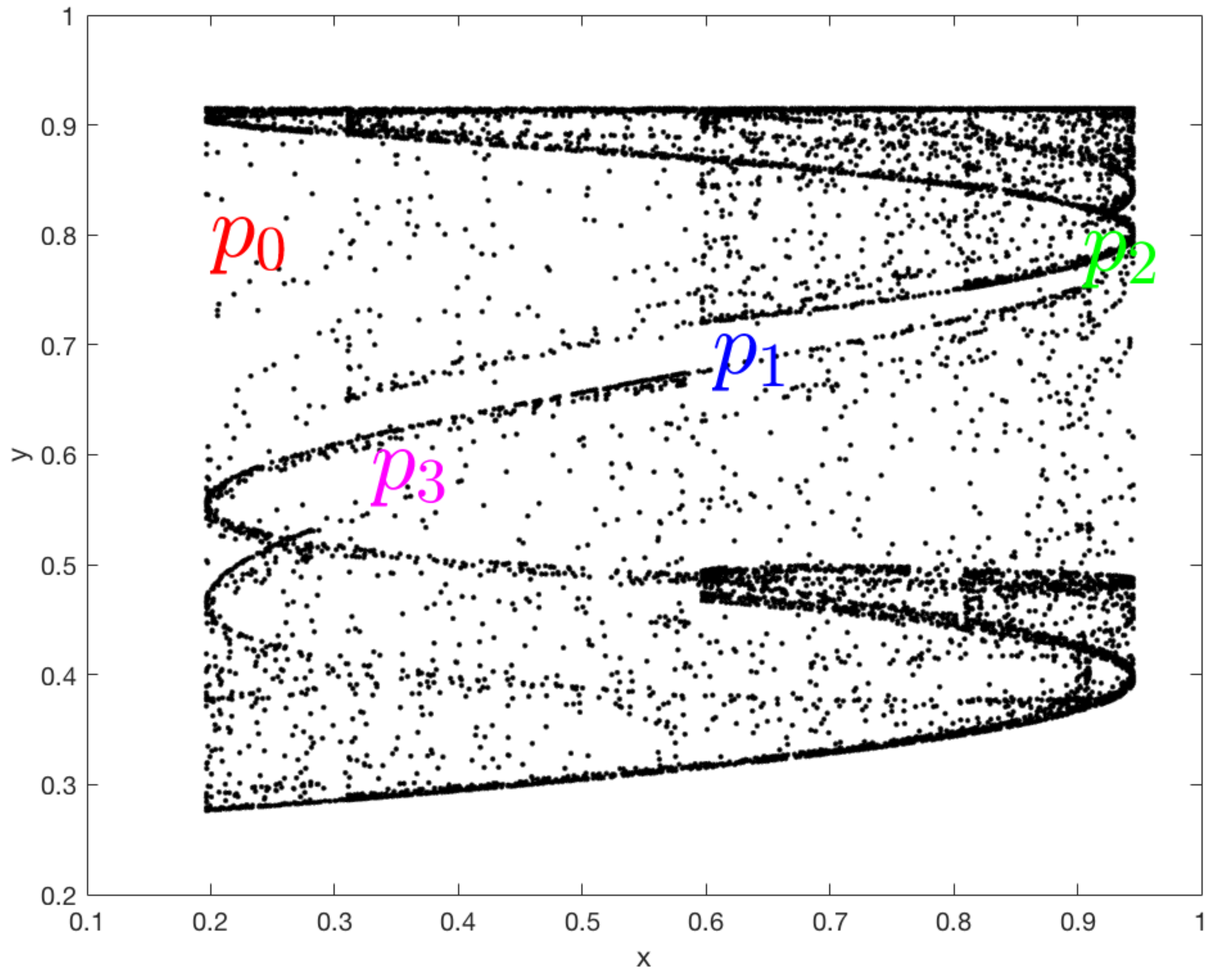
$P(0)$, the sequence …

$$P(1) = P(0) \cdot T$$
$$P(2) = P(1) \cdot T$$
$$P(3) = P(2) \cdot T$$
$$\vdots$$

has the same limit

$$\boxed{\lim_{n \to \infty} P(n) = \rho}$$



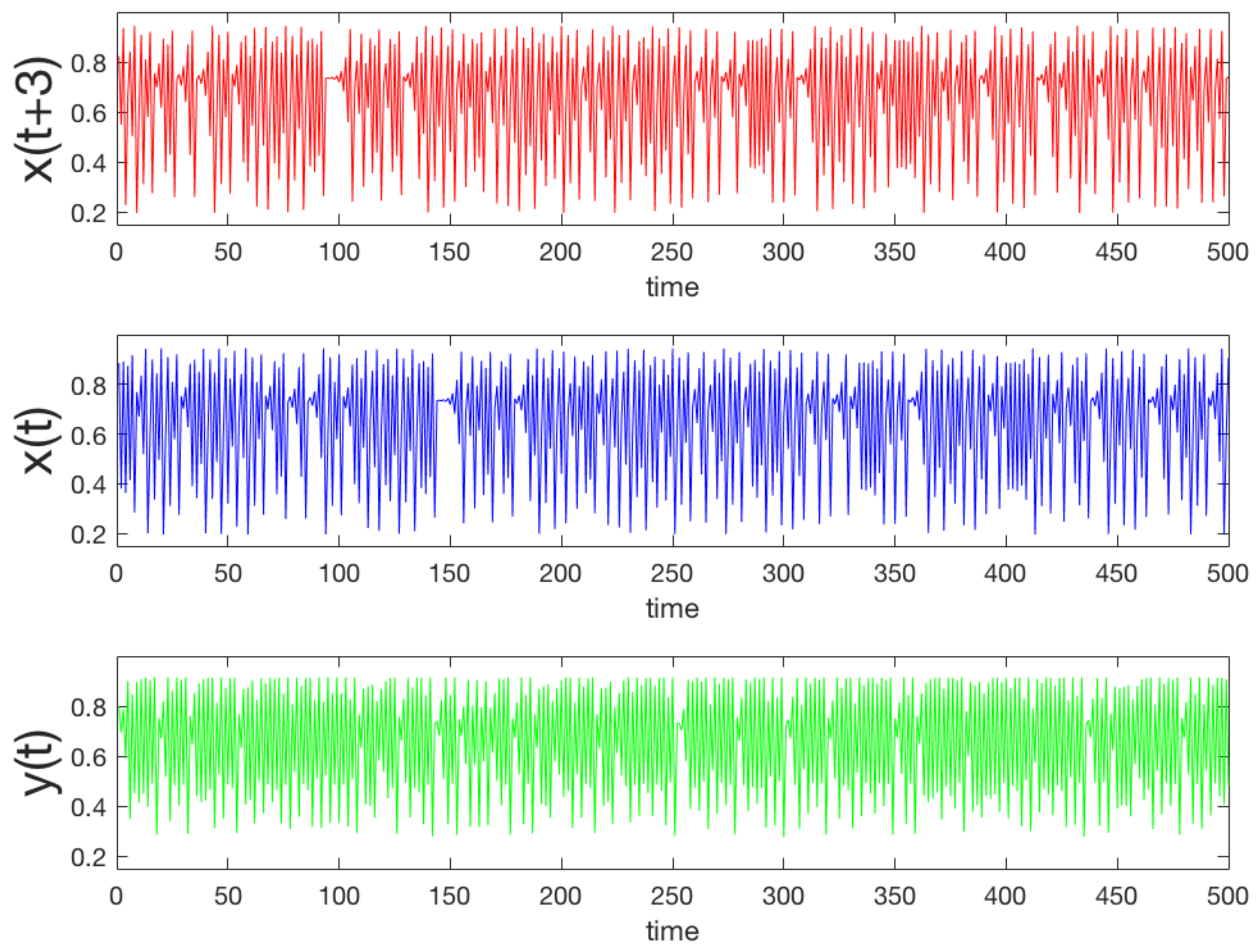


How then do we compute $P(x_{t+\tau}, x_t, y_t)$?

We start off by computing the invariant density …

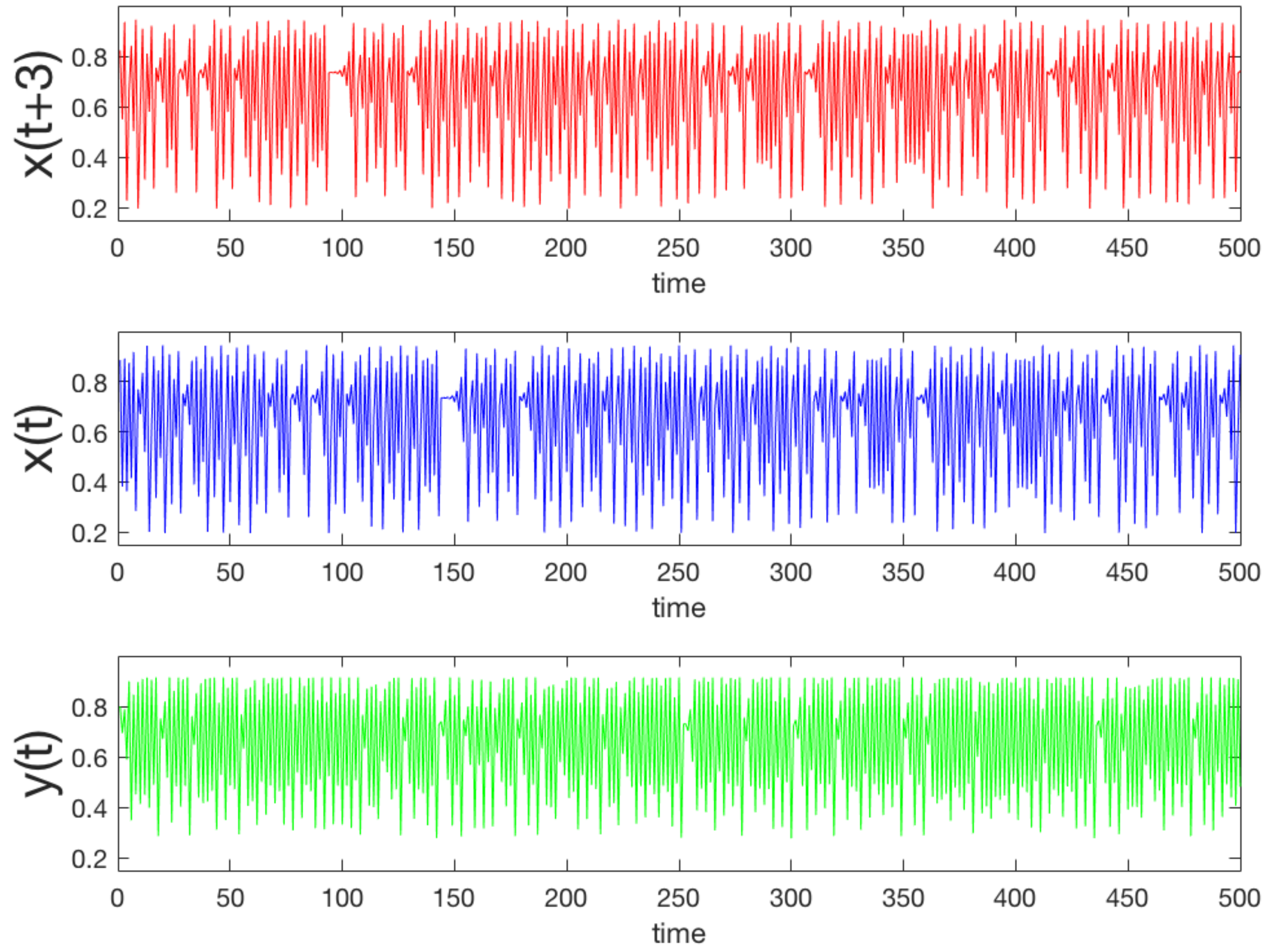
How then do we compute $P(x_{t+\tau}, x_t, y_t)$?

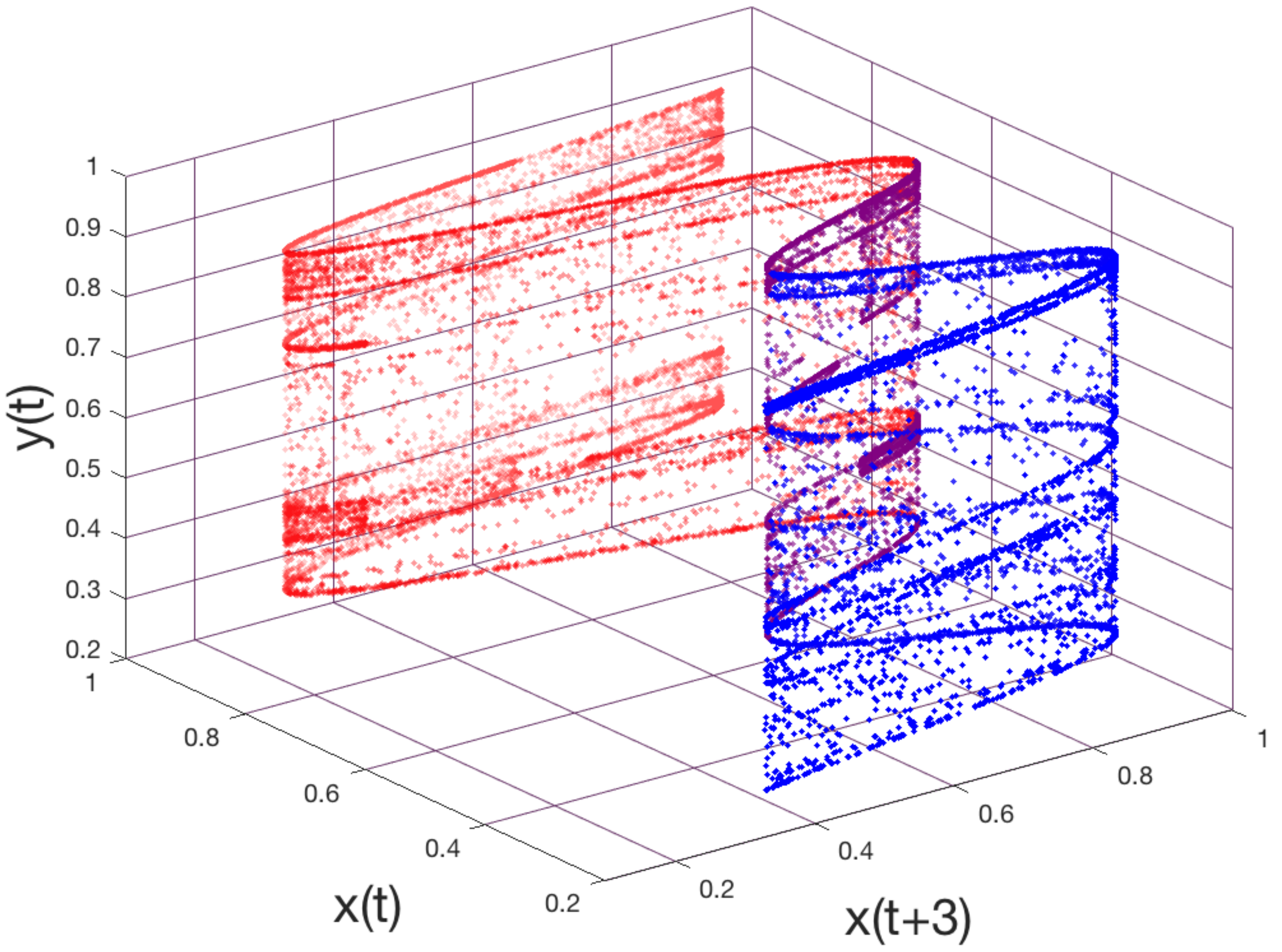
We start off by computing the invariant density …



How then do we compute $P(x_{t+\tau}, x_t, y_t)$?

We start off by computing the invariant density …

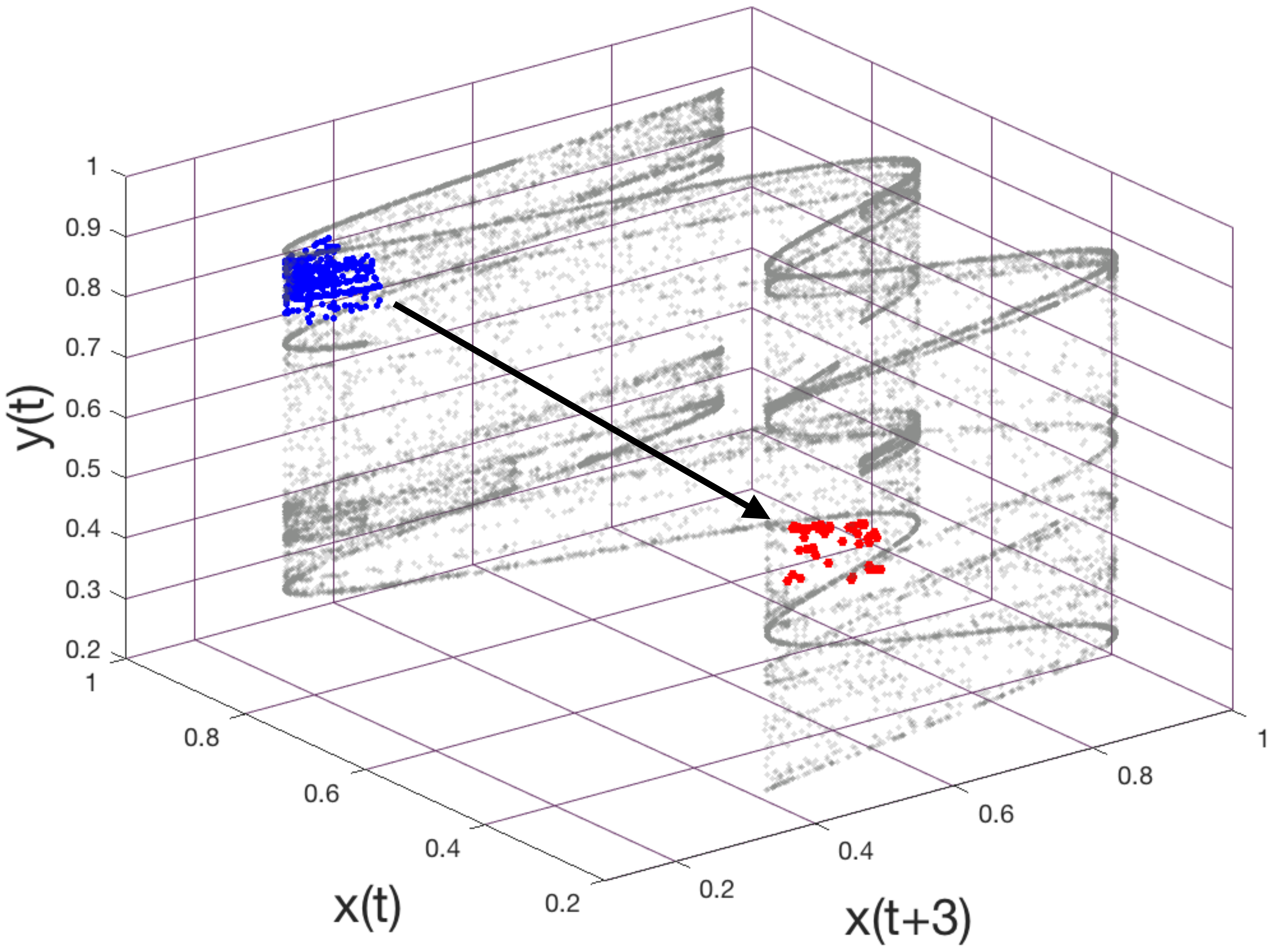




How then do we compute $P(x_{t+\tau}, x_t, y_t)$?

We start off by computing the invariant density …

$$T_{ij} = \frac{n_j}{N_i}$$

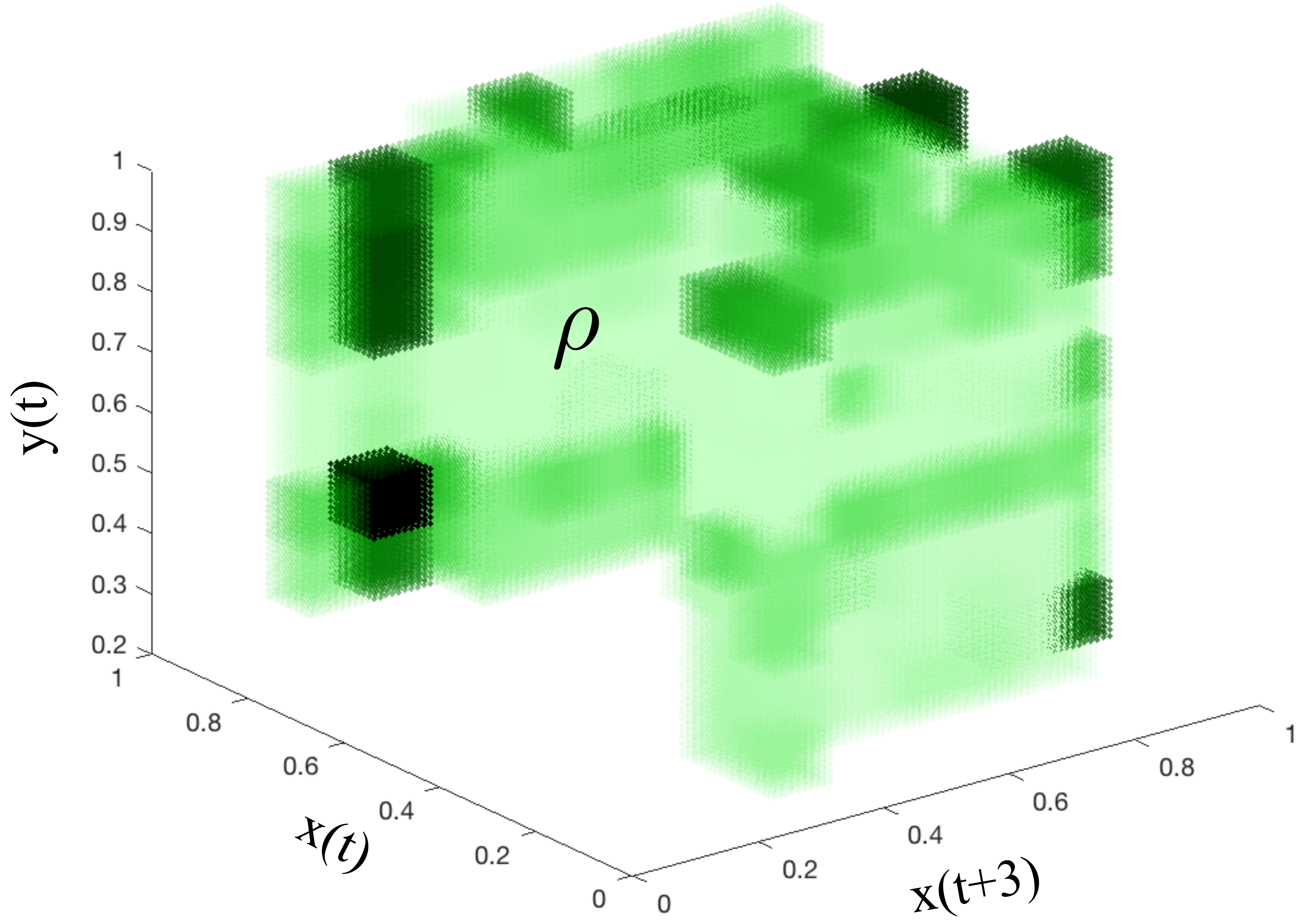


$$P(1) = P(0) \cdot T$$

$$\vdots$$

$$\rho = \lim_{n \to \infty} P(n)$$

$$P(x_{t+3}, x_t, y_t)$$



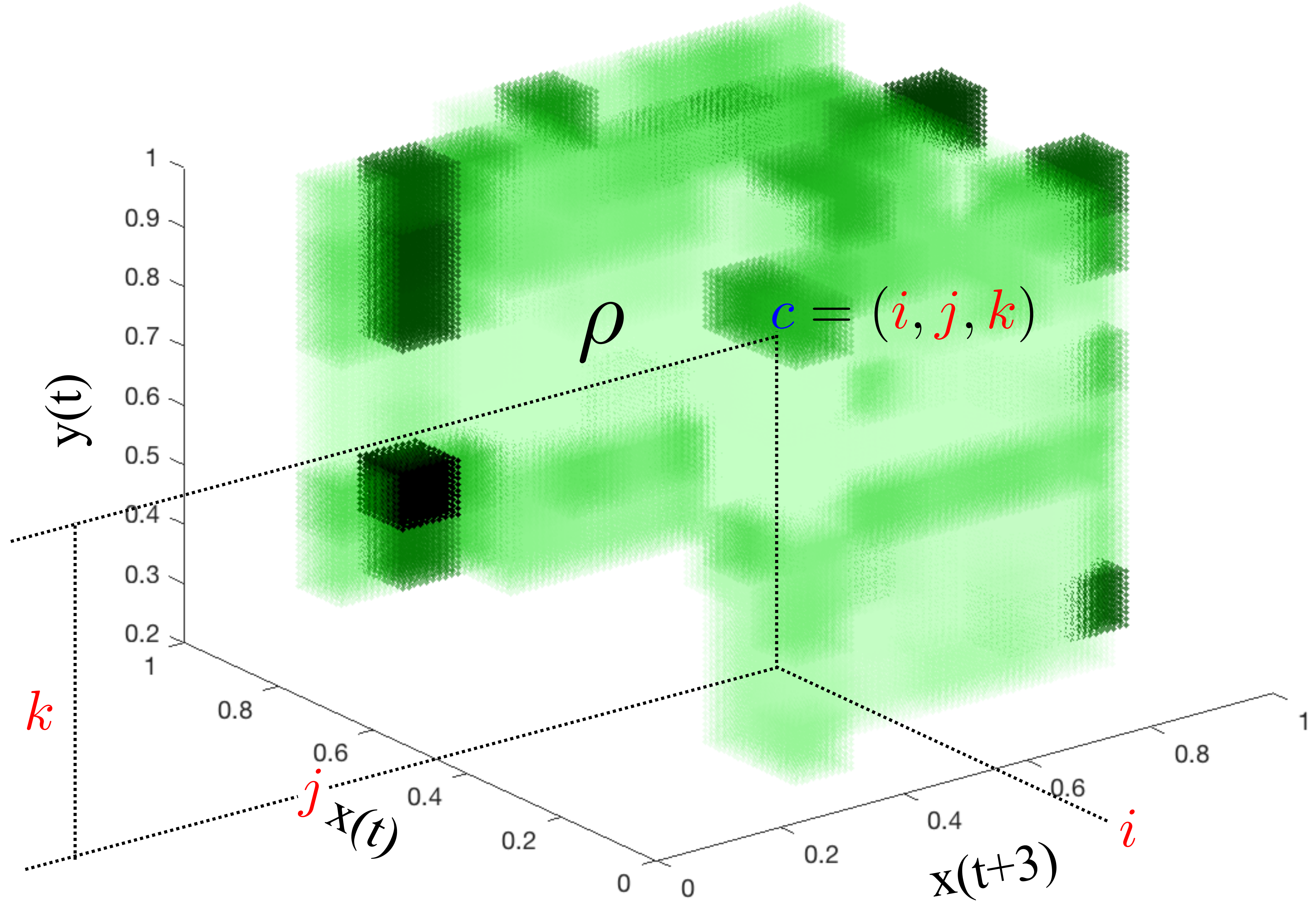
$$P(1) = P(0) \cdot T$$

$$\vdots$$

$$\rho = \lim_{n \to \infty} P(n)$$

$$P(i, j, k) = \rho(c) \longleftarrow P(x_{t+3}, x_t, y_t)$$

Probability for the system
to occupy the bin with center ***c***

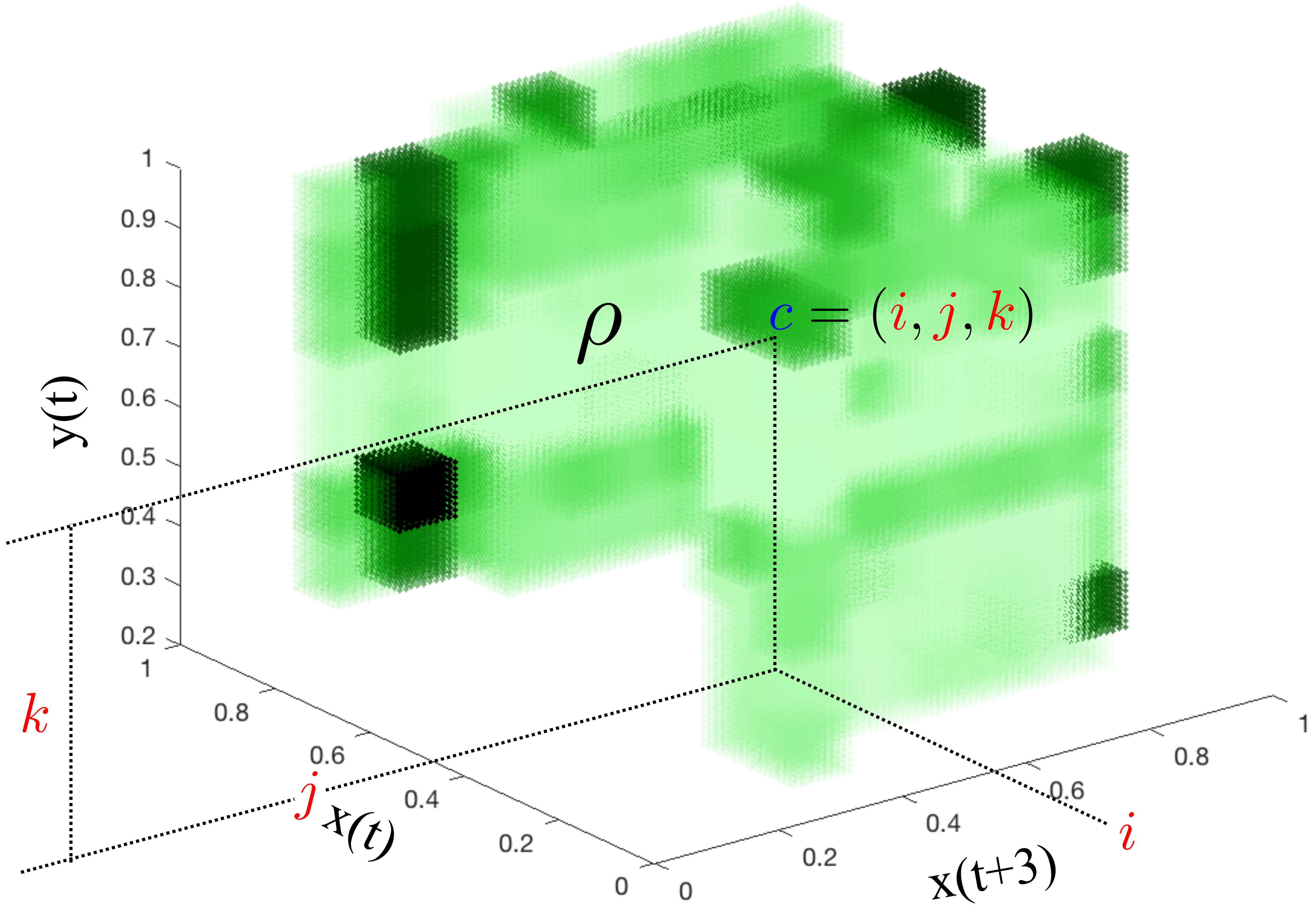


$$P(1) = P(0) \cdot T$$
$$\vdots$$
$$\rho = \lim_{n \to \infty} P(n)$$

$$P(i, j, k) = \rho(c) \longleftarrow P(x_{t+3}, x_t, y_t)$$

$$P(i|j, k) = \frac{P(i, j, k)}{\sum_m P(m, j, k)} \longleftarrow P(x_{t+3}|x_t, y_t)$$

$$P(i|j) = \frac{\sum_m P(i, j, m)}{\sum_r \sum_s P(r, j, s)} \longleftarrow P(x_{t+3}|x_t)$$

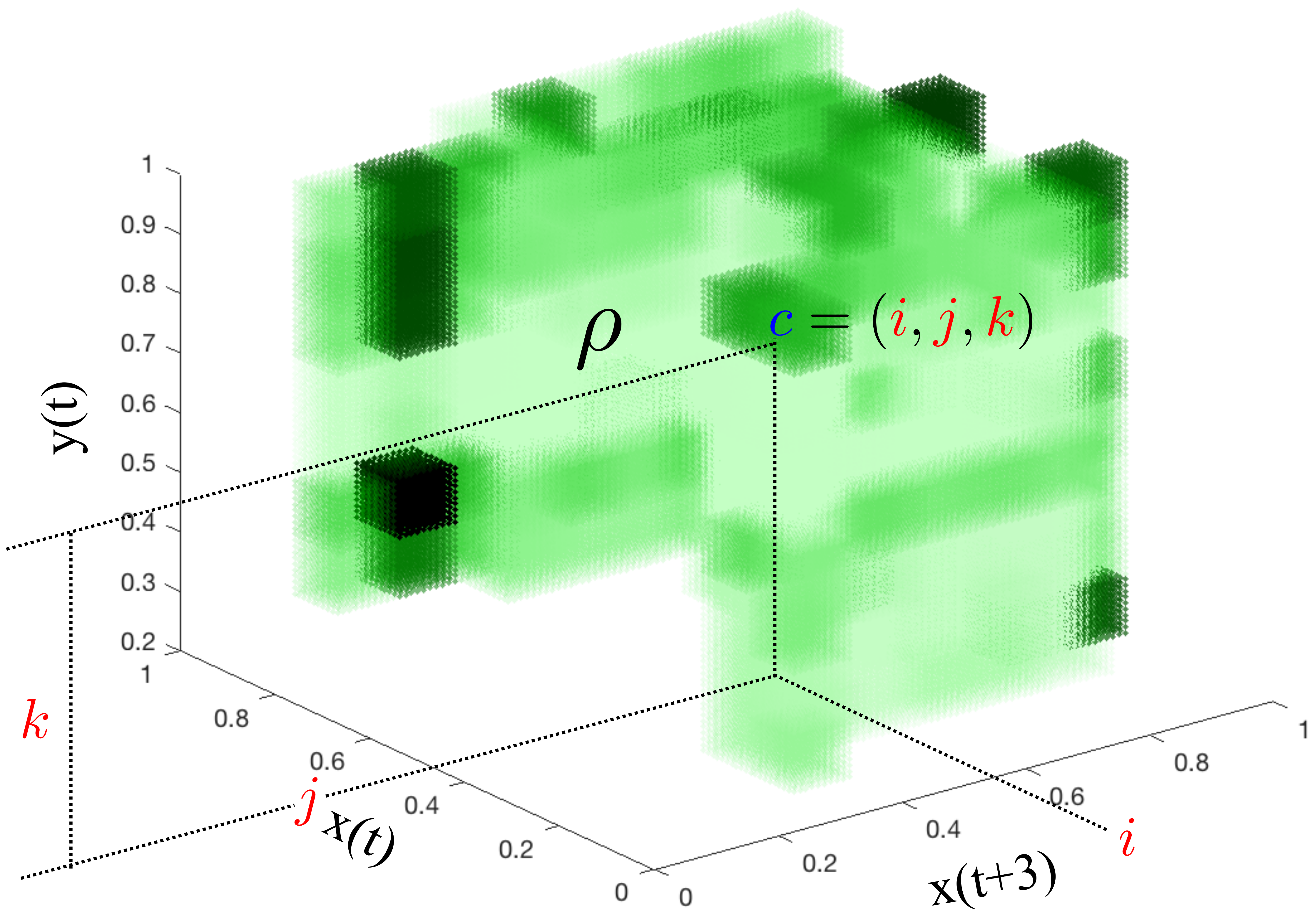


$$P(1) = P(0) \cdot T$$

$$\vdots$$

$$\rho = \lim_{n \to \infty} P(n)$$

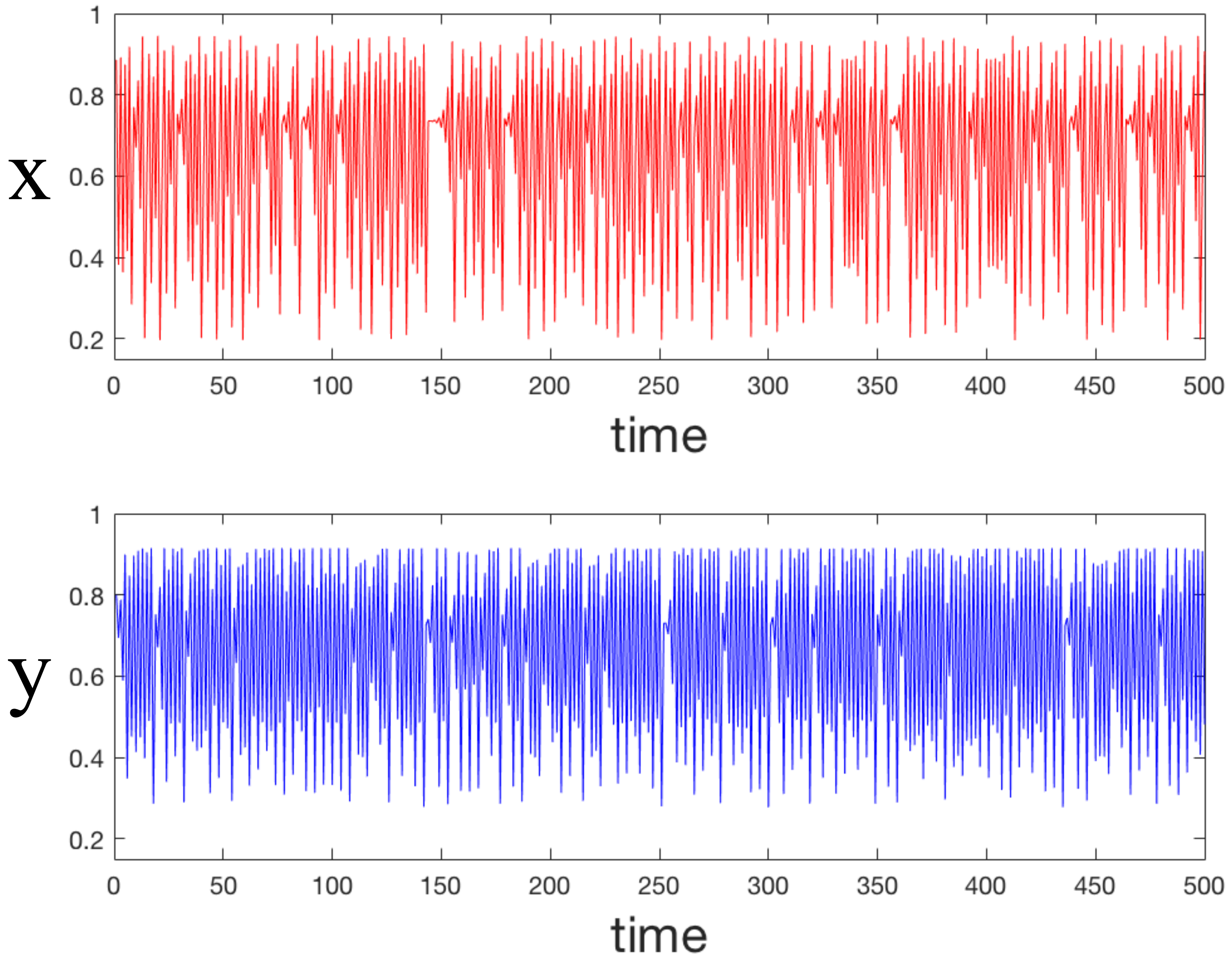
$$P(i, j, k) = \rho(c) \longleftarrow P(x_{t+3}, x_t, y_t)$$

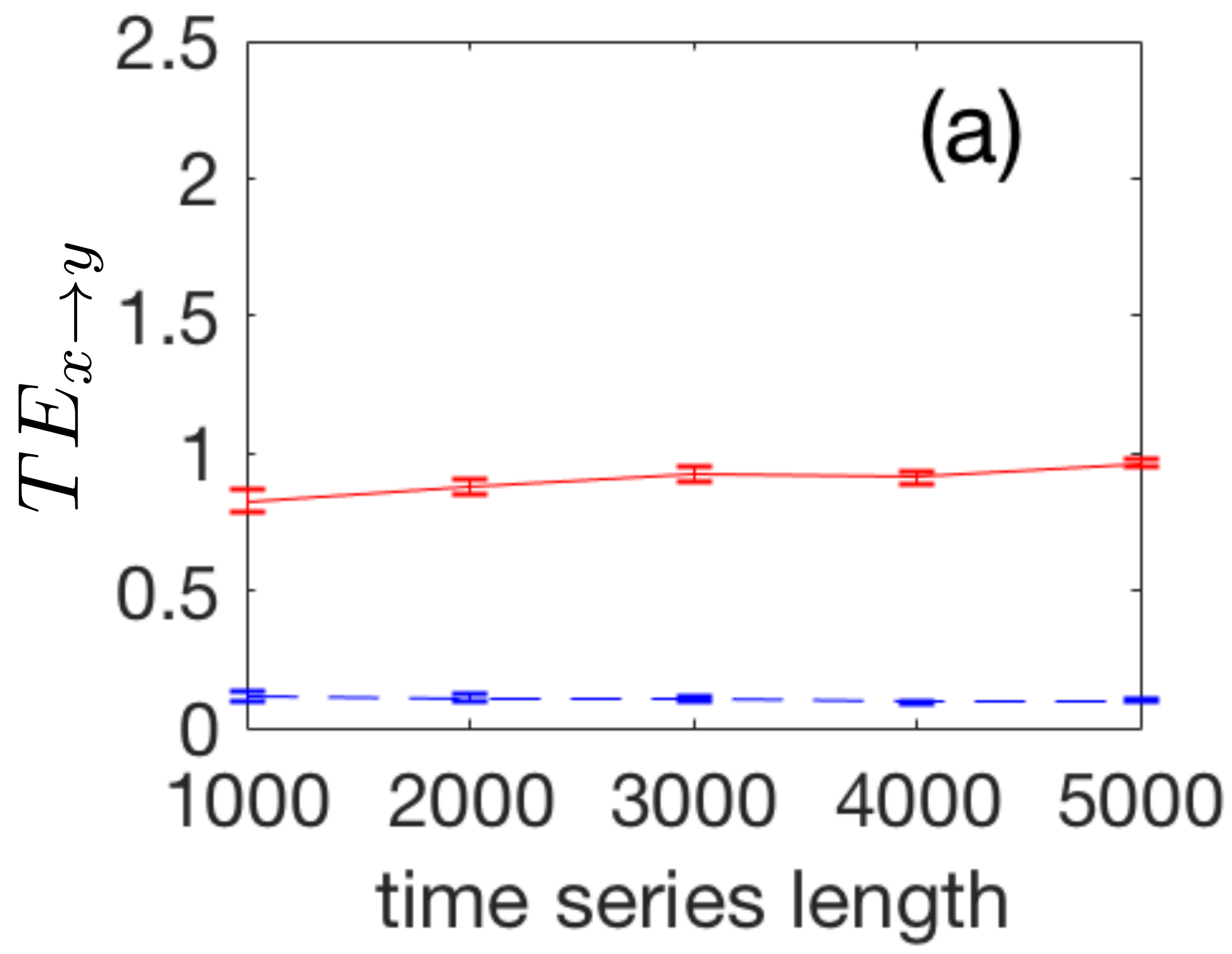


$$P(i|j, k) = \frac{P(i, j, k)}{\sum_m P(m, j, k)}$$

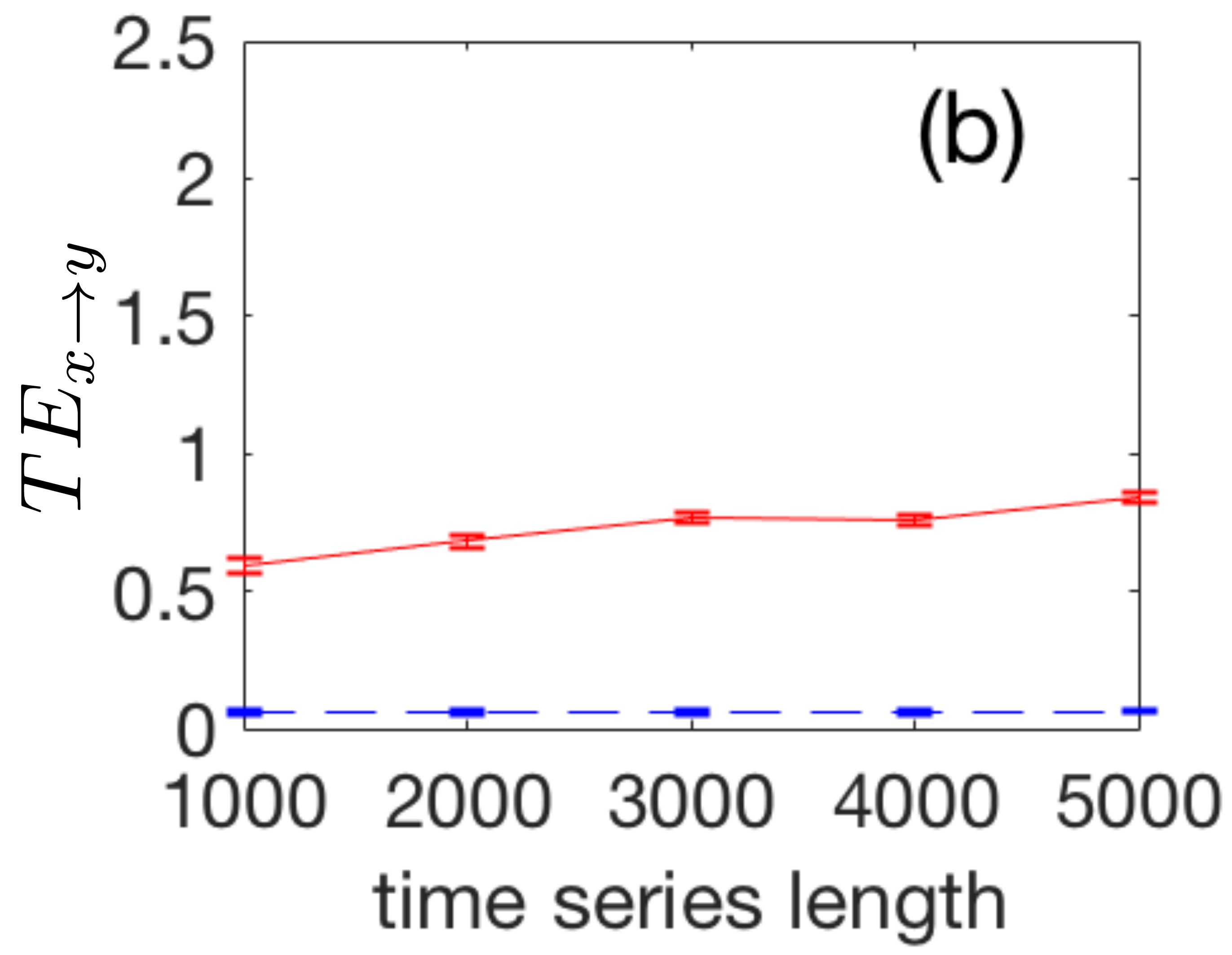
$$P(i|j) = \frac{\sum_m P(i, j, m)}{\sum_r \sum_s P(r, j, s)}$$

$$TE(Y \to X) = \sum_{i,j,k} P(i, j, k) \log_2 \frac{P(i|j, k)}{P(i|j)}$$





(a) From the transfer matrix



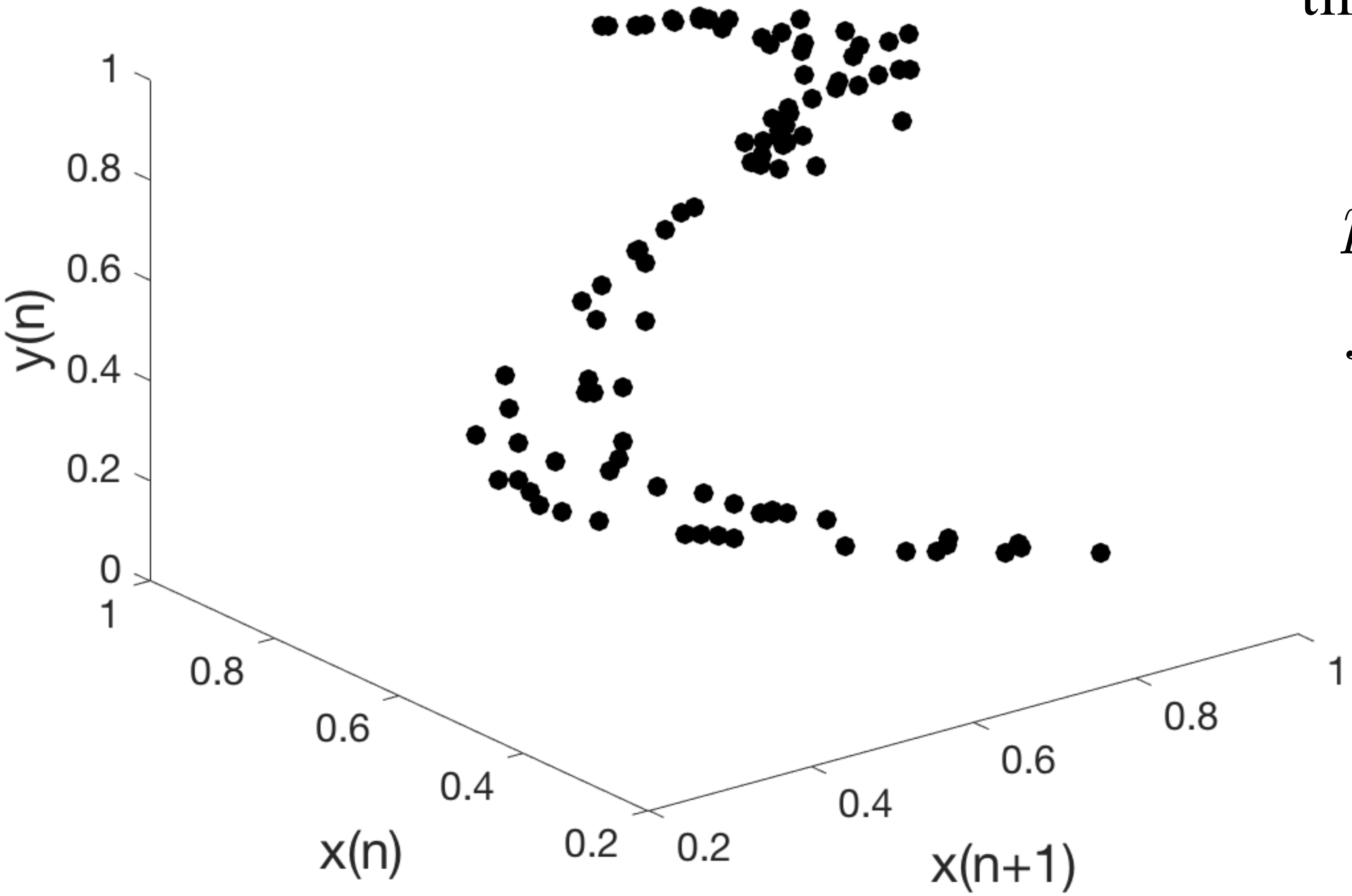
(b) From kernel density estimation

Steuer et al. (2002), Bioinformatics 18, S231

We know that the coupling is unidirectional

$$x \to y$$

Alternative approach for sparse time series: **Triangulation approach**

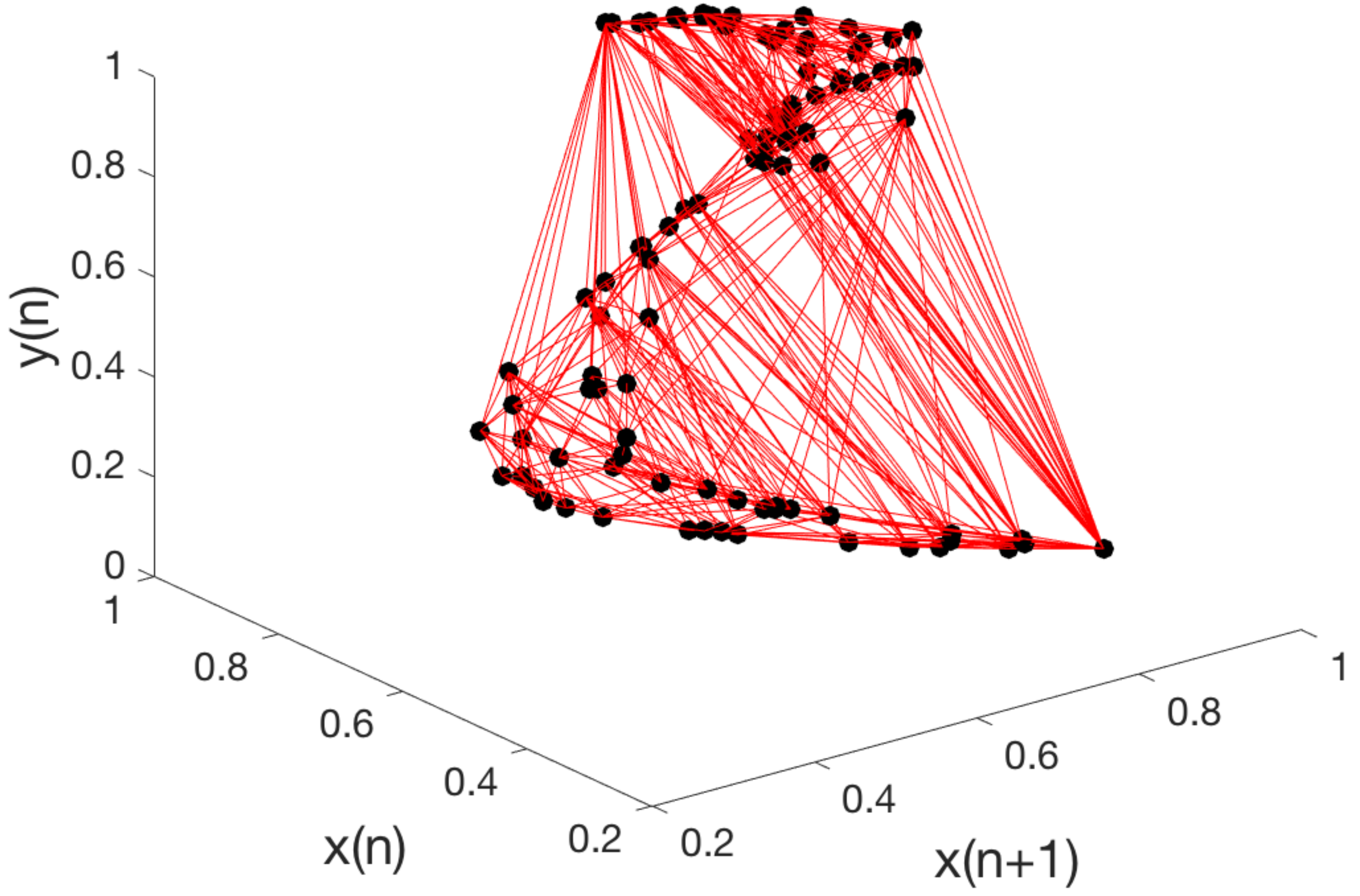


time series length: 100 observations

$$p_k = (x(k+1), x(k), y(k))$$
$$f(p_k) = p_{k+1}$$

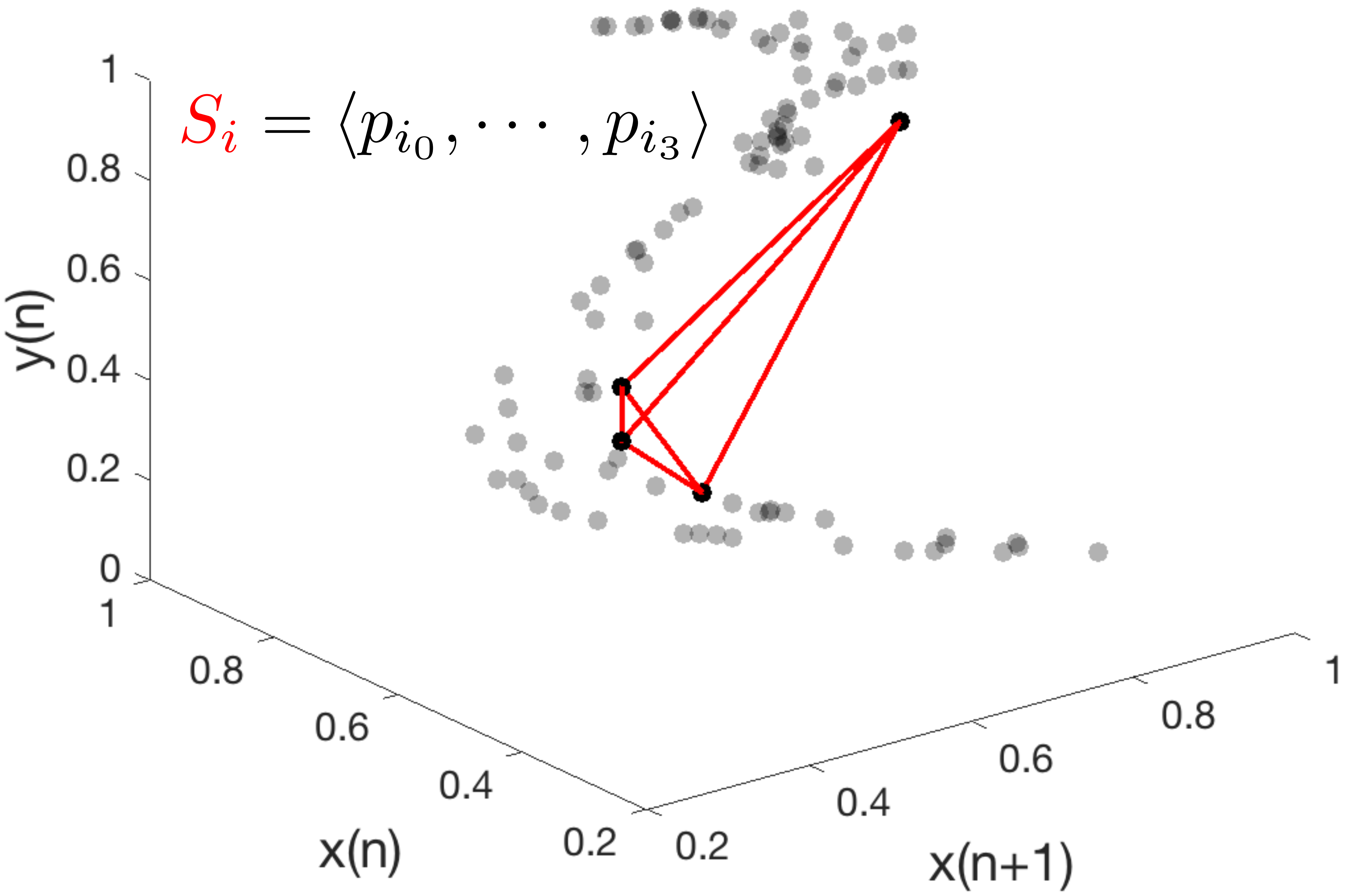
Triangulation approach

Each simplex is formed by 4 embedded points



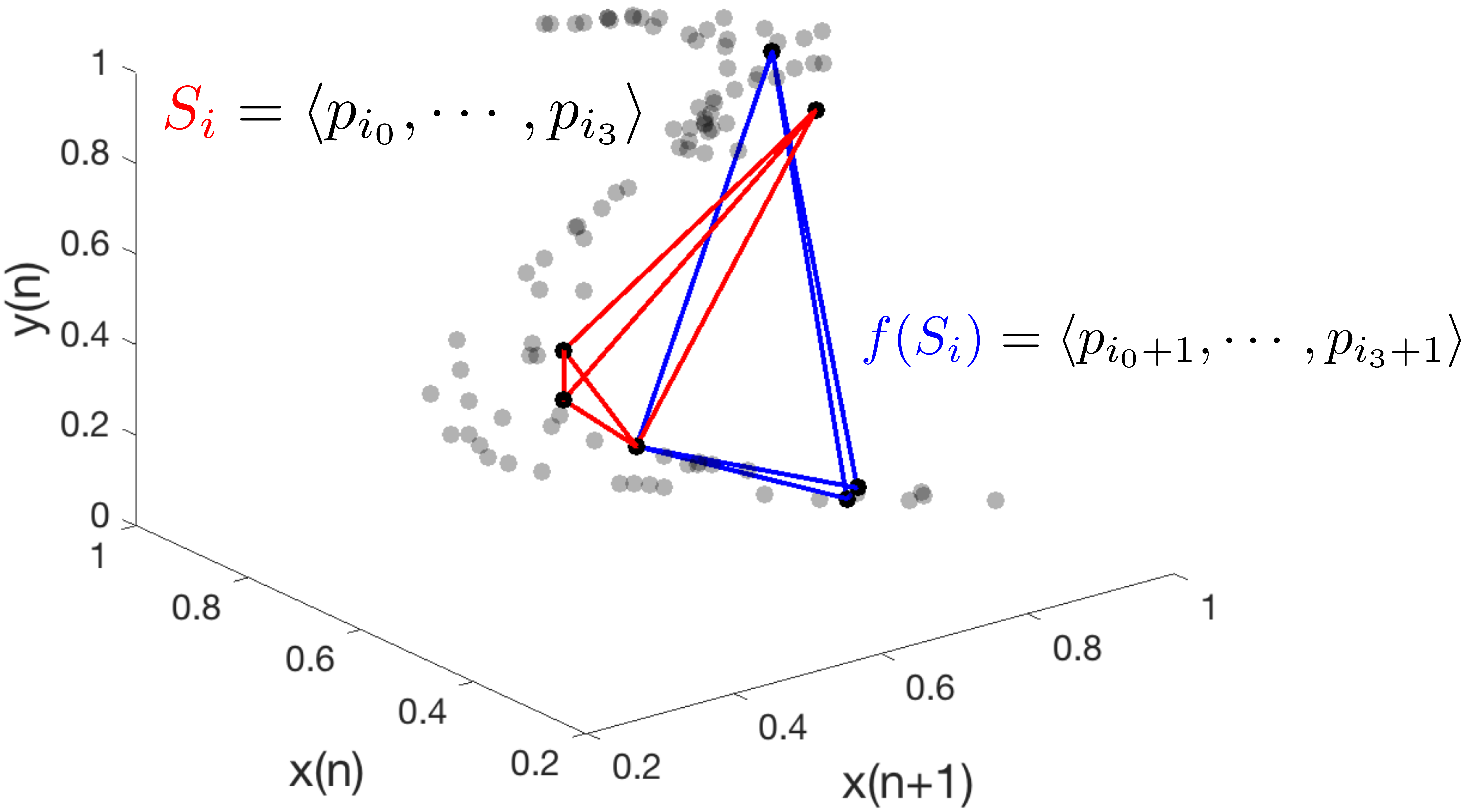
Triangulation approach

Each simplex is formed by 4 embedded points

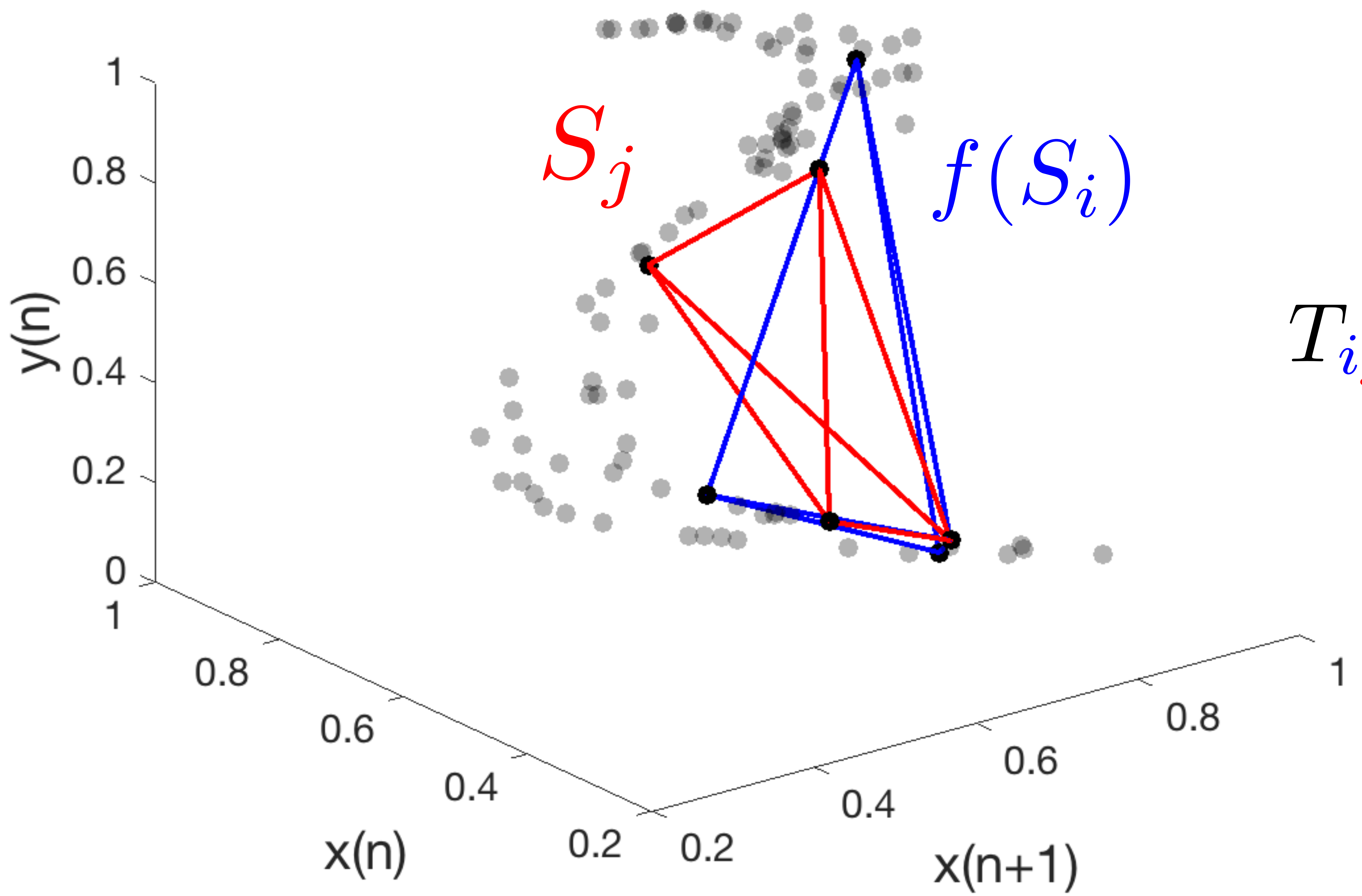


Triangulation approach

Each simplex is formed by 4 embedded points

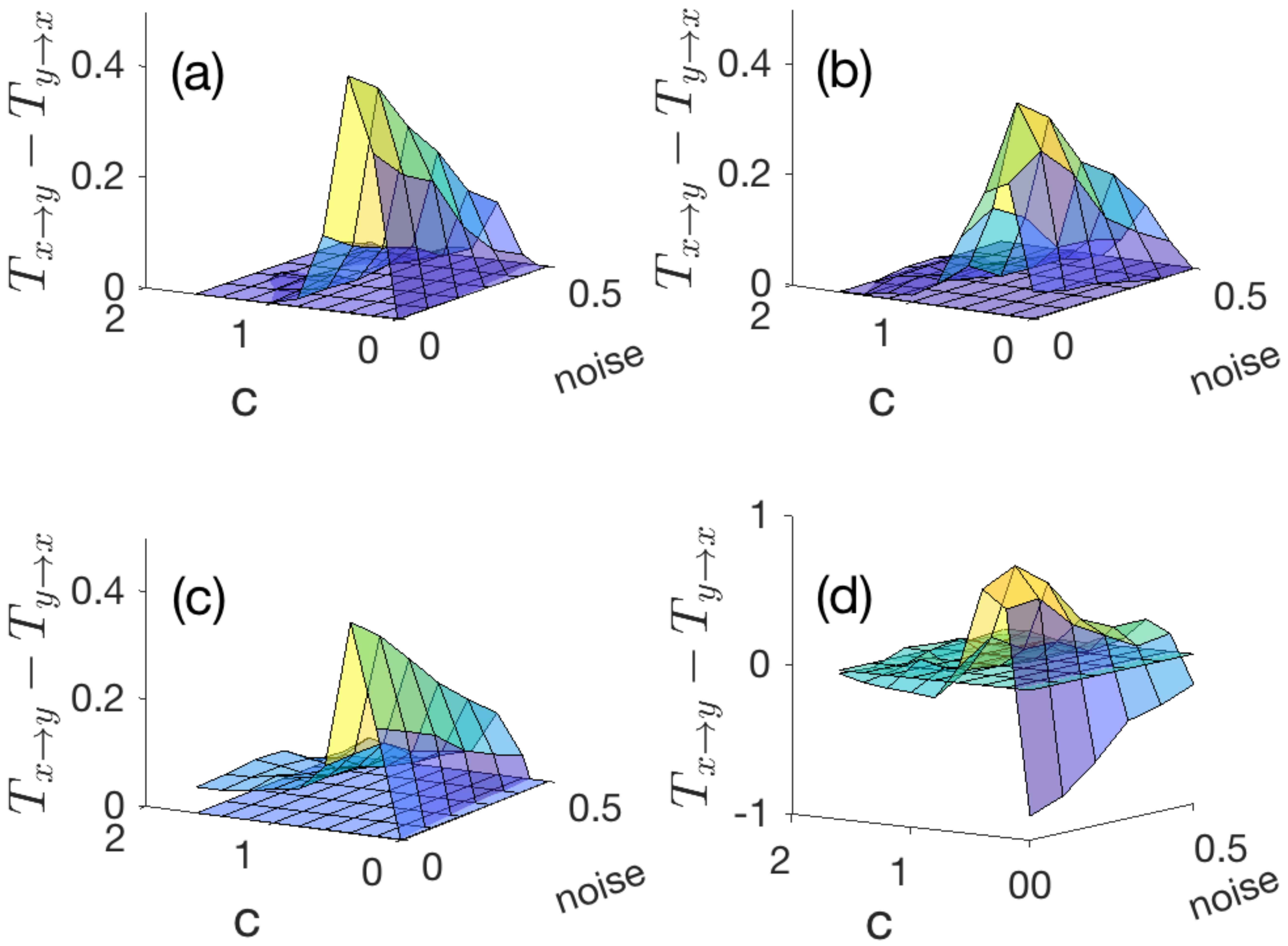


Triangulation approach



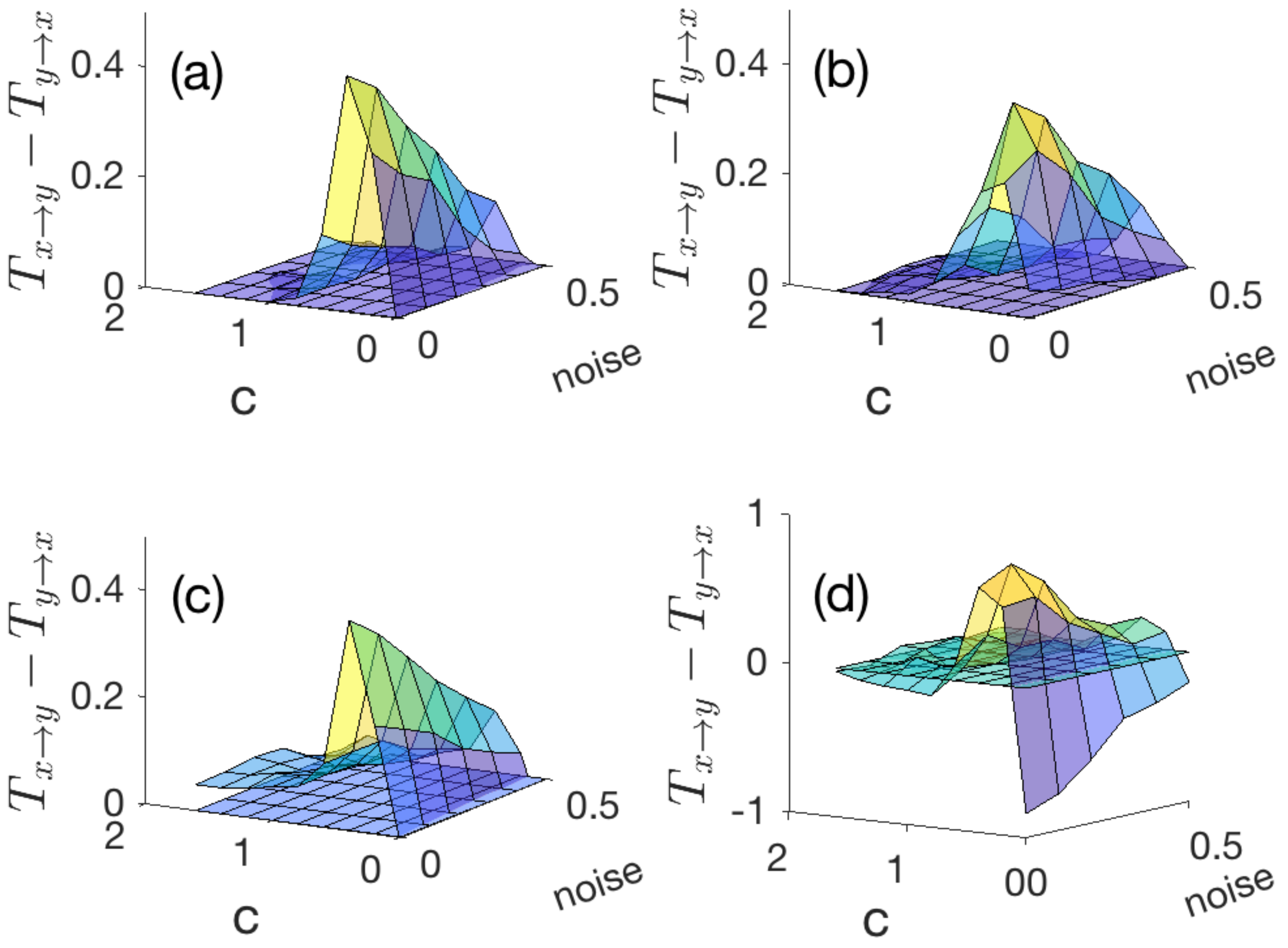
$$T_{ij} = \frac{vol\left(f\left(S_i\right) \cap S_j\right)}{vol\left(f\left(S_i\right)\right)}$$

The transfer matrix approach is robust for sparse and noisy time series



time series length: 100 observations

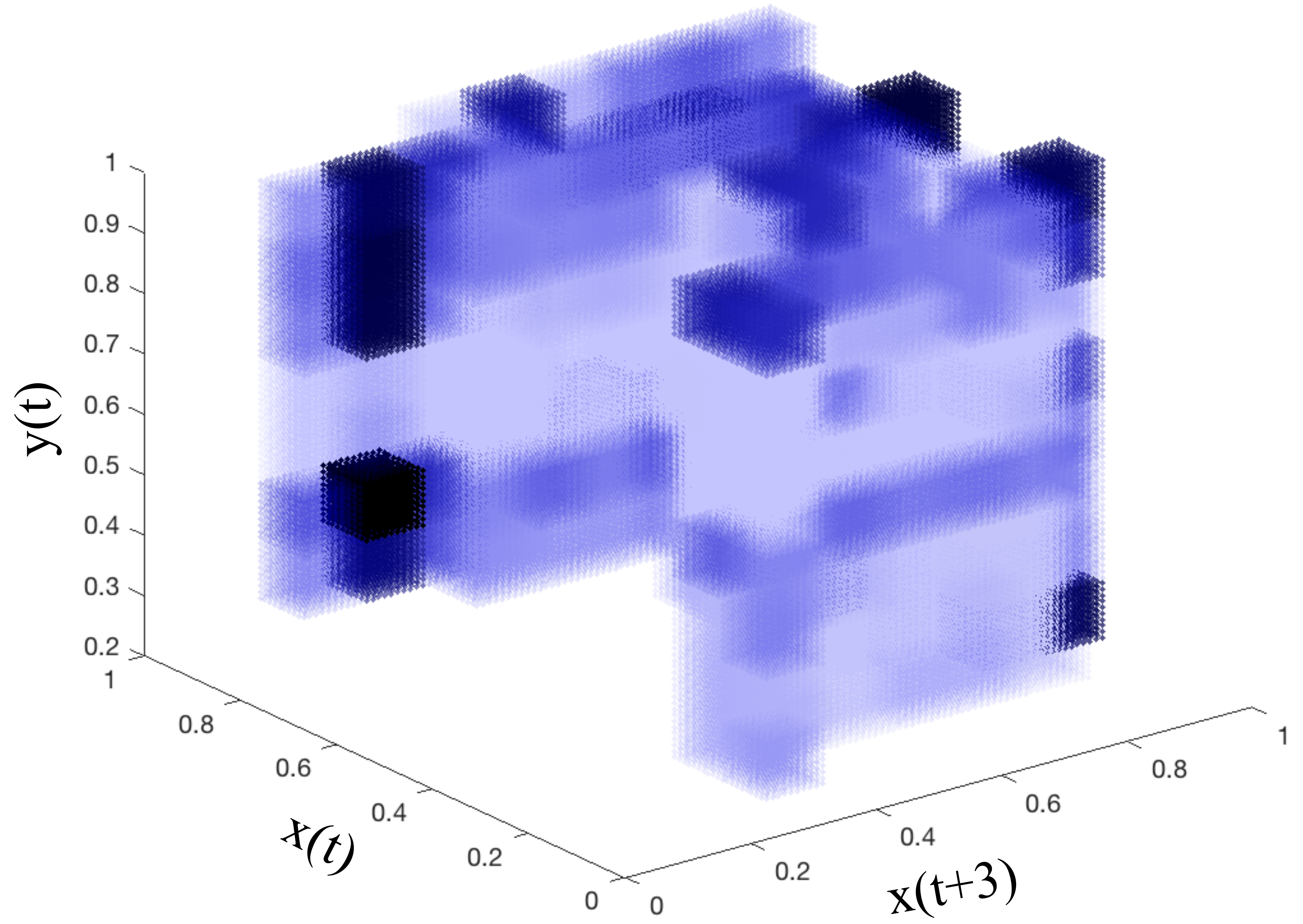
The transfer matrix approach is robust for sparse and noisy time series



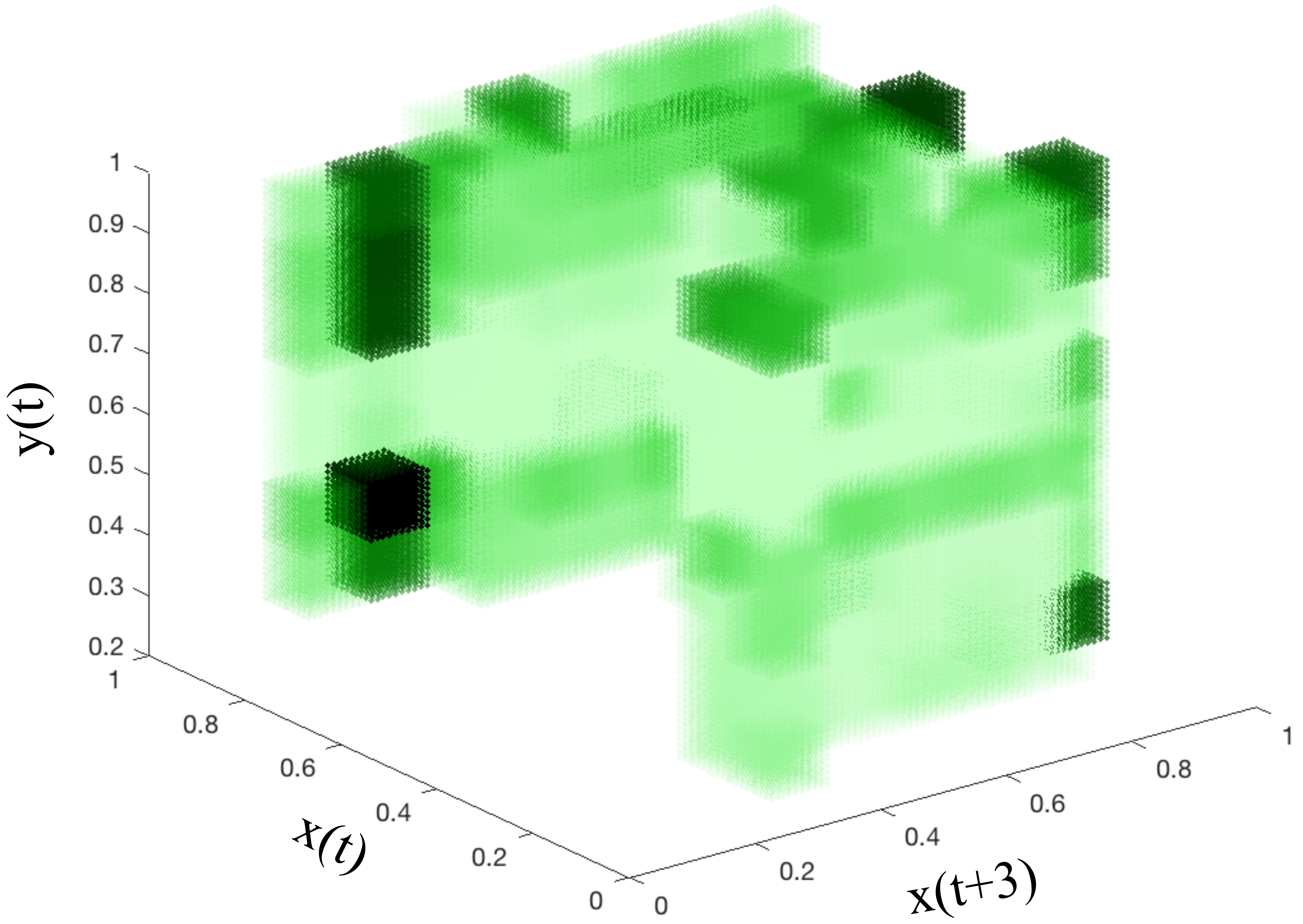
(a) trans. matrix
(b) trans. matrix (triangulation approach)
(c) kernel density estimation
(d) nearest neighbors counting
Kraskov et al. (2004), Phys Rev E 69

time series length: 100 observations

ρ from the visitation freq.



ρ from the transfer matrix



Why then bother with the transfer matrix?

The transfer matrix approximates the dynamical behavior

It enables a number of possibilities:

- Extension and/or interpolation of time series.

- Generation of null hypotheses for causality testing.

- Revealing important aspects of long term behavior:

  Attracting regions on the phase space

  Estimating fractal dimensions of the attractor

In summary

More info/details in:

https://github.com/kahaaga/CausalityTools.jl

and

DOI: 10.1103/PhysRevE.99.042212

THANKS