Causality analysis

András Telcs

PATTERN / WIGNER MTA
telcs.szit.bme@gmail.com

HUNGARIAN BRAIN RESEARCH PROGRAM
NEMZETI AGYKUTATÁSI PROGRAM

NAP-B PATTERN (2015-2017)
Population Activity Research Unit
MTA WIGNER RESEARCH CENTRE FOR PHYSICS

joint work with

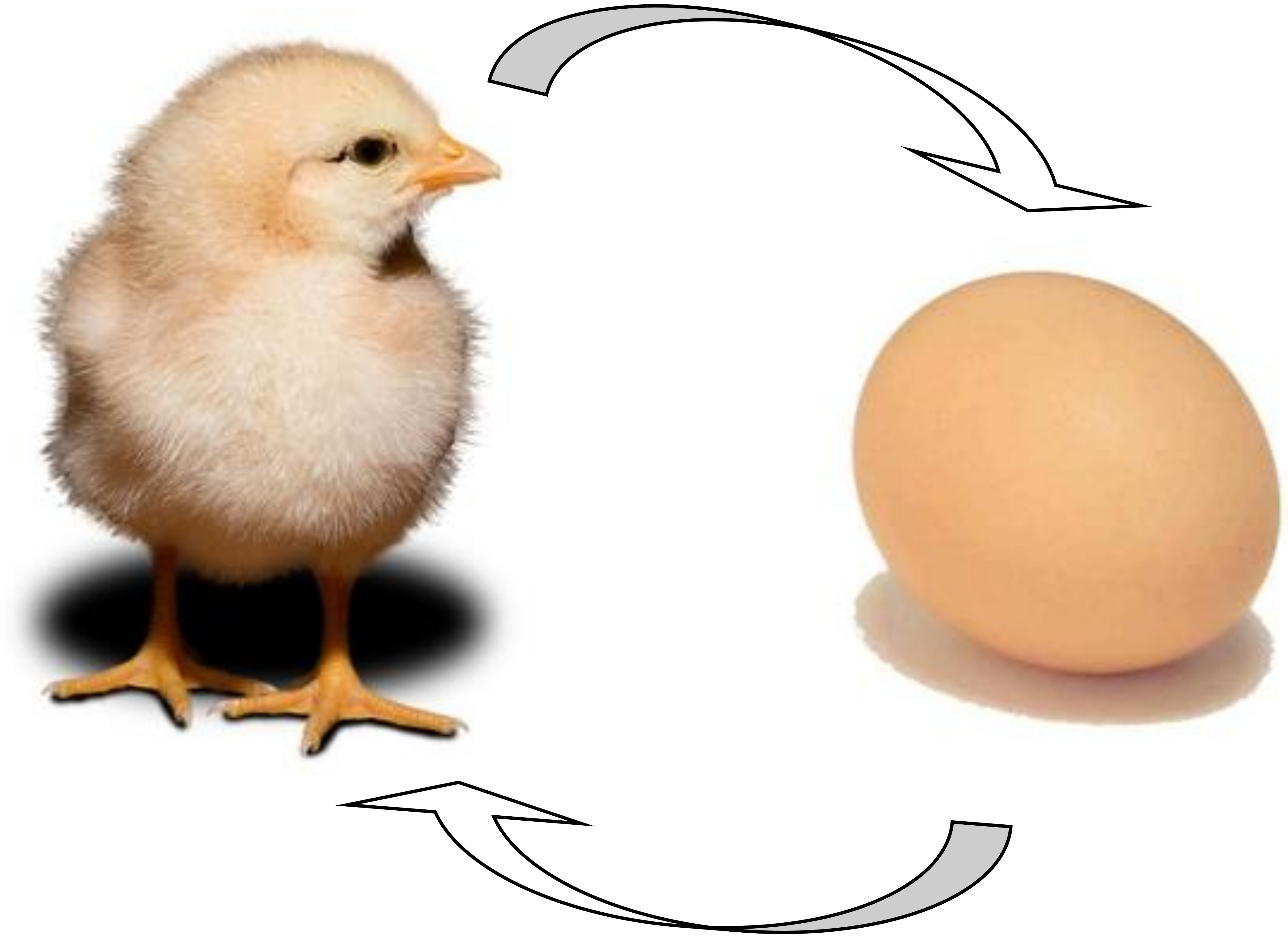
Zsigmond Benkő,

Ádám Zlatniczki,

Dániel Fabó,

Zoltán Somogyvári

Which was first?



Puzzle for you:

What on earth to do with GPU about causality?

Find the right spot in the lecture!

Wiener-Granger causality

1. Axiom – cause precedes caused

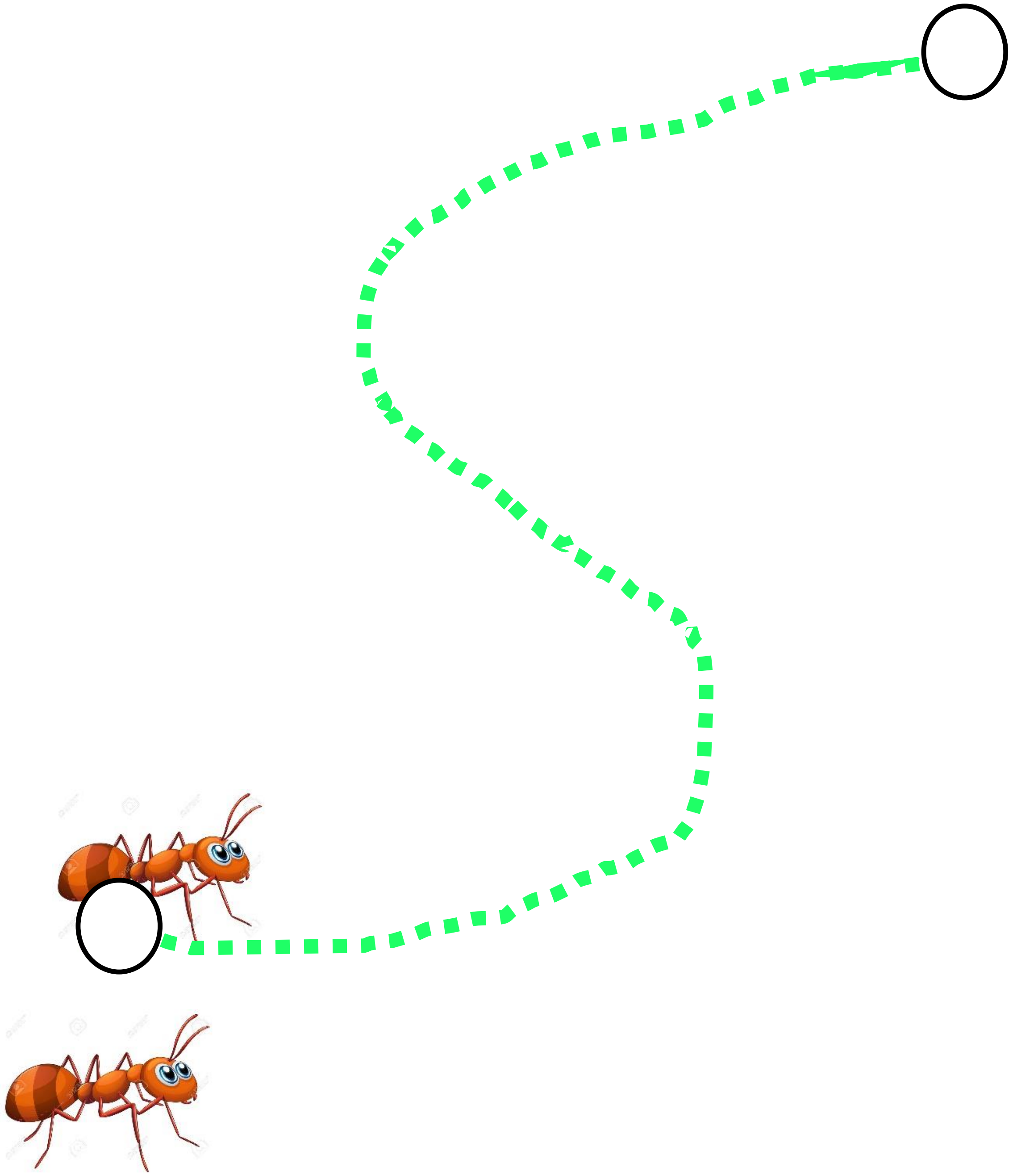
2. Axiom – Using the past of the cause improves the forecast of the caused based solely on its own past.

Granger causality

Life of the ants

Granger causality

Life of the ants



Takens' Theorem

Let $f_t: M \rightarrow M$

$a_{t+1} = f_t(a_t)$

the map for a discrete time dynamical system with a strange attractor $\mathcal{X}$ with box counting dimension $d_{\mathcal{X}}$.

$x_t = g(a_t)$ *observation*

g must be twice-differentiable observation function,

$m > 2d_{\mathcal{X}}$ then, the delay embedding $\mathcal{X}$

$X_t = (x_t, x_{t-1}, \ldots x_{t-m+1})$ *reconstruct (up to ...) the state space of* a

embeds $\mathcal{X}$ into R^m and left $d_{\mathcal{X}}$ invariant.

Time delay embedding

Embedding of single variable

(x_t, x_{t-1})

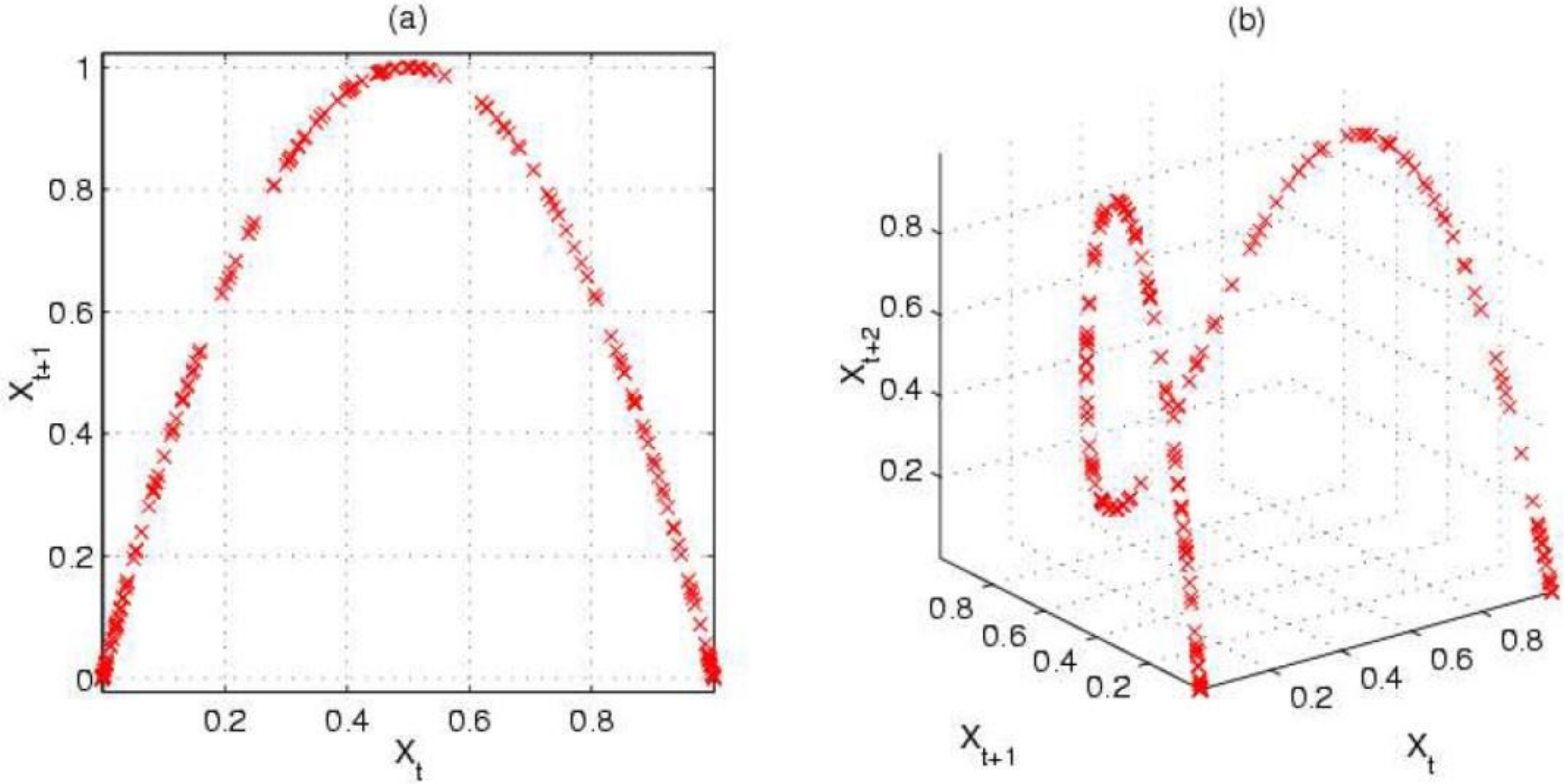
$m=2$

(x_t, x_{t-1}, x_{t-2})

$m=3$

Example: logistic map

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



Embedded in D=2,3, the manifold is still one dimensional.

Time delay embedding

Embedding of single variables

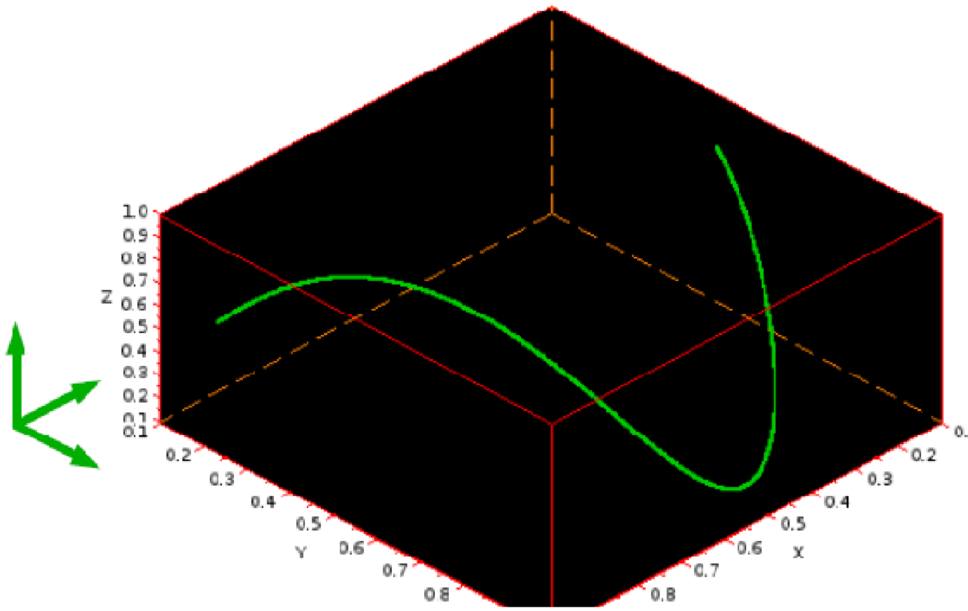
(x_t, x_{t-1}, x_{t-2}) (y_t, y_{t-1}, y_{t-2})

Joint embedding

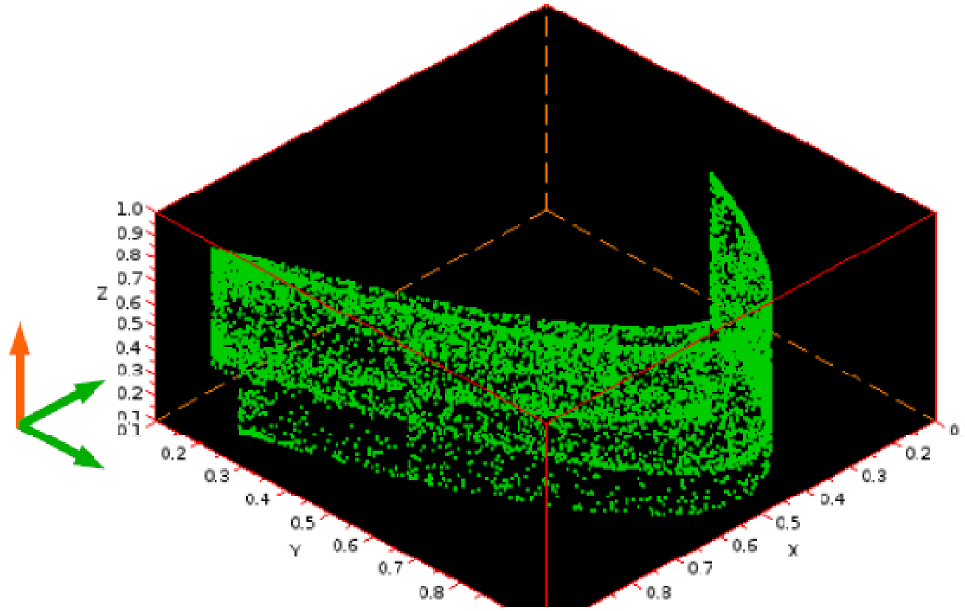
(x_t, x_{t-1}, y_t)

Time delay embedding

$$(y_{t-2}, y_{t-1}, y_t)$$



$$(y_{t-2}, y_{t-1}, x_t)$$

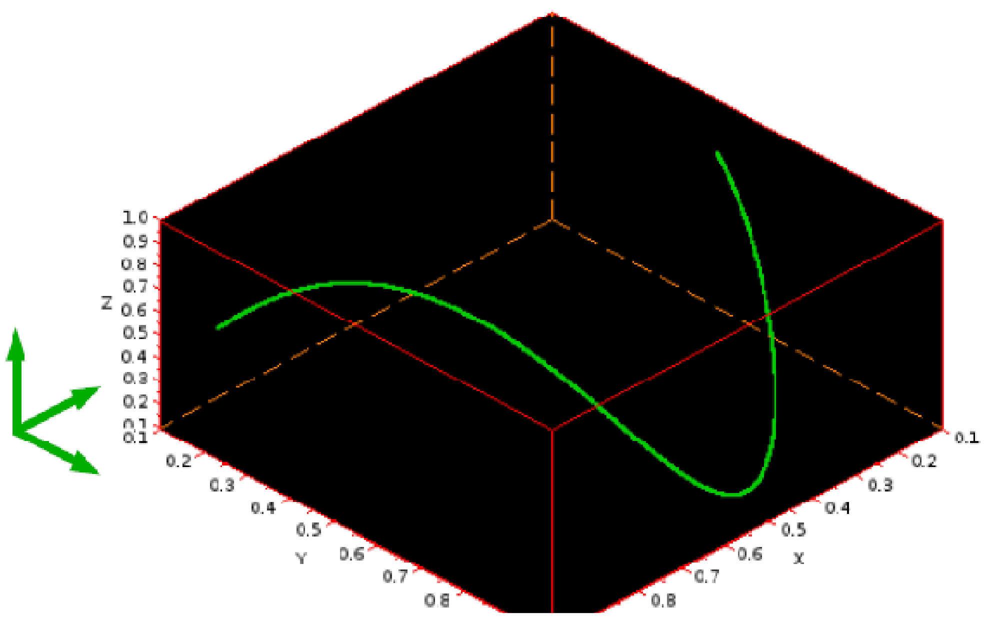


Example 1.

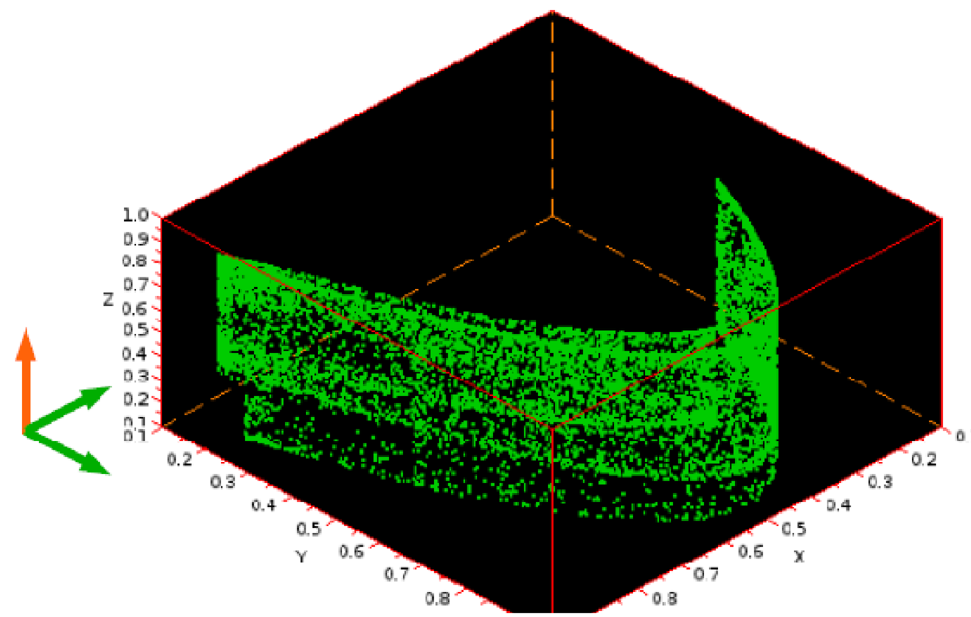
Time delay embedding

$$(y_{t-2}, y_{t-1}, y_t)$$

$$(y_{t-2}, y_{t-1}, x_t)$$



y d=1 (in D=3)



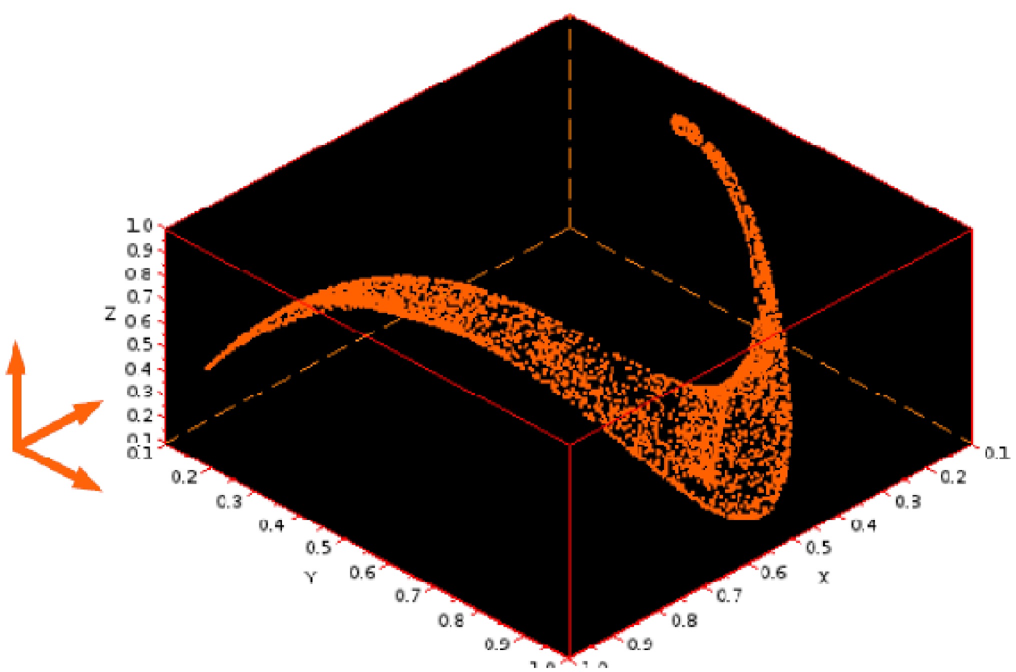
Joint d=2 (in D=3)

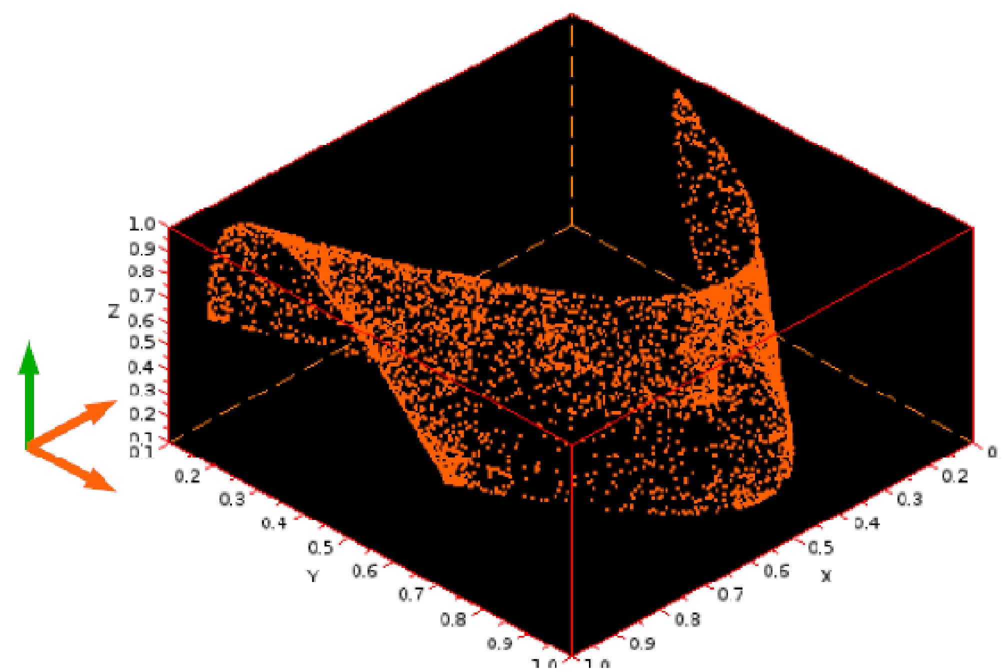
Dimension increase indicates independence

Time delay embedding

$$(x_{t-2}, x_{t-1}, x_t)$$

$$(x_{t-2}, x_{t-1}, y_t)$$



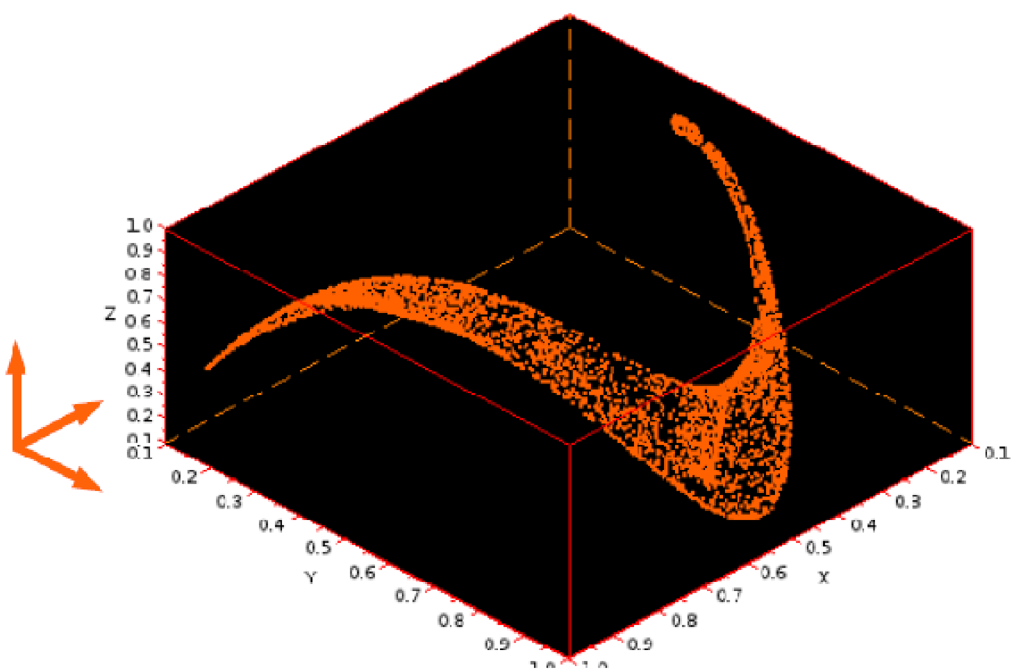


Example 2.

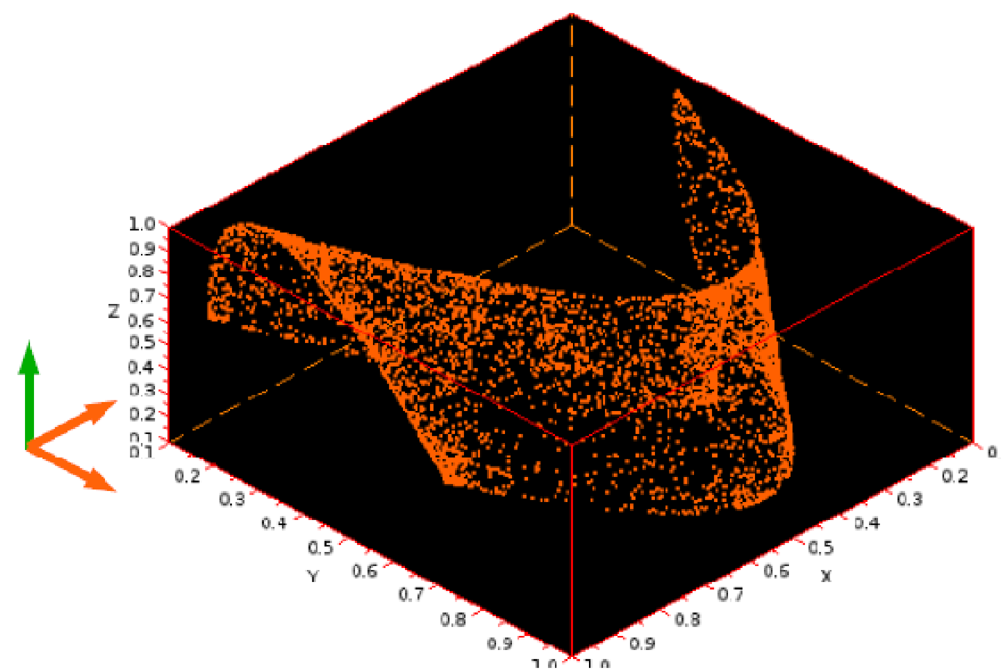
Time delay embedding

(x_{t-2}, x_{t-1}, x_t)

(x_{t-2}, x_{t-1}, y_t)



2d in 3D



joint embedding is still 2d

Lack of dimension increase indicates causality, y causes x

Embedding - a new look

X_t , stationary time series $t=1,\ldots.n$

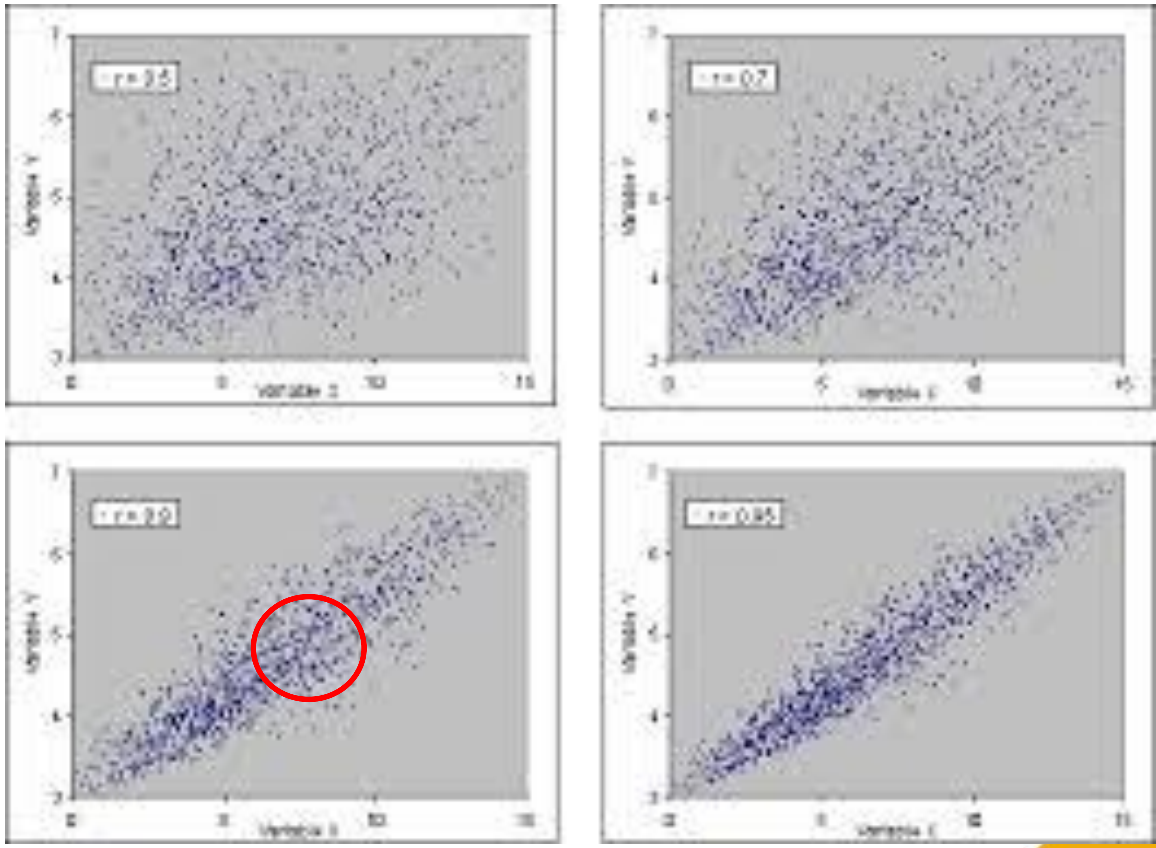
$X_t = (x_t, x_{t-1}, \ldots x_{t-m+1})$ *embedded in* R^m

$N(X_t,r) = \# \{ s : |X_s - X_t| < r \}$

$N(X_t,r) \approx r^{d(X)}$ *local dimensions*

d_X *average of local dimensions*

d_X is the intrinsic dimension of the X-manifold



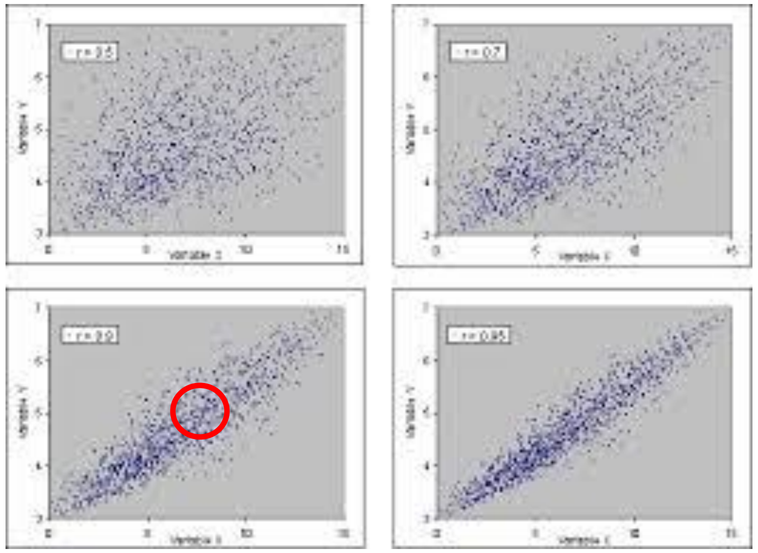
Intrinsic dimension

intrinsic dimension = information dimension

Joint embedding

X_t , Y_t time series

$X_t = (x_t, x_{t-1}, \ldots x_{t-m+1})$, $Y_t = (y_t, y_{t-1}, \ldots y_{t-m+1})$ *embedded in* R^m



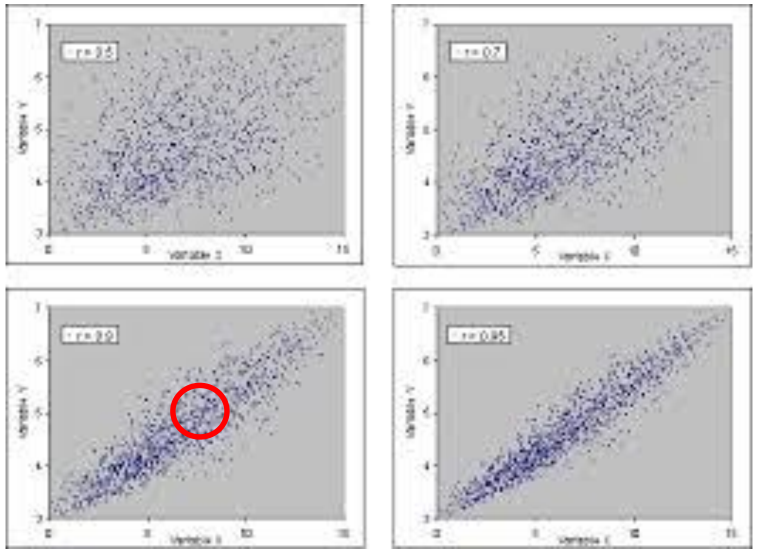
Joint embedding

X_t , Y_t time series

$X_t = (x_t, x_{t-1}, \dots x_{t-m+1})$, $Y_t = (y_t, y_{t-1}, \dots y_{t-m+1})$ *embedded in* R^m

And the joint:

$J_t = (X_t, Y_t) = (x_t, x_{t-1}, \dots x_{t-m+1}, y_t, y_{t-1}, \dots y_{t-m+1})$



Joint embedding

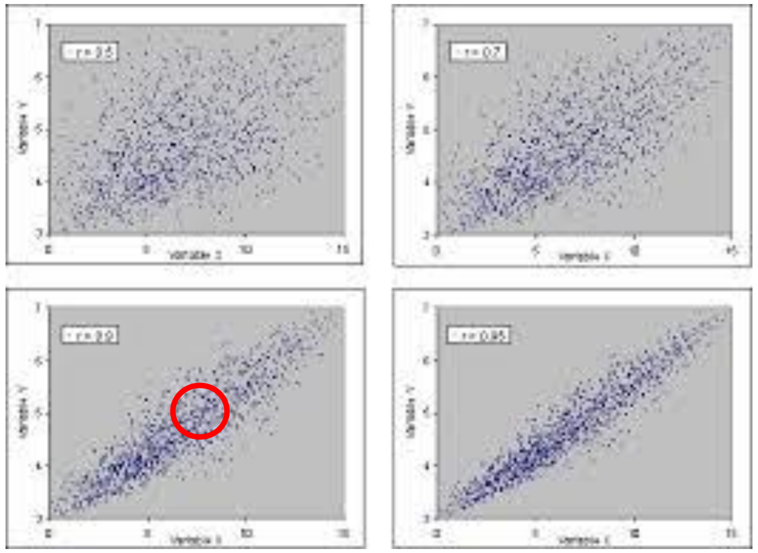
X_t , Y_t time series

$X_t = (x_t, x_{t-1}, \ldots x_{t-m+1})$, $Y_t = (y_t, y_{t-1}, \ldots y_{t-m+1})$ *embedded in* R^m

And the joint:

$J_t = (X_t, Y_t) = (x_t, x_{t-1}, \ldots x_{t-m+1}, y_t, y_{t-1}, \ldots y_{t-m+1})$

d_X, d_Y, d_J is the intrinsic dimension of the manifold, X,Y and the joint variable



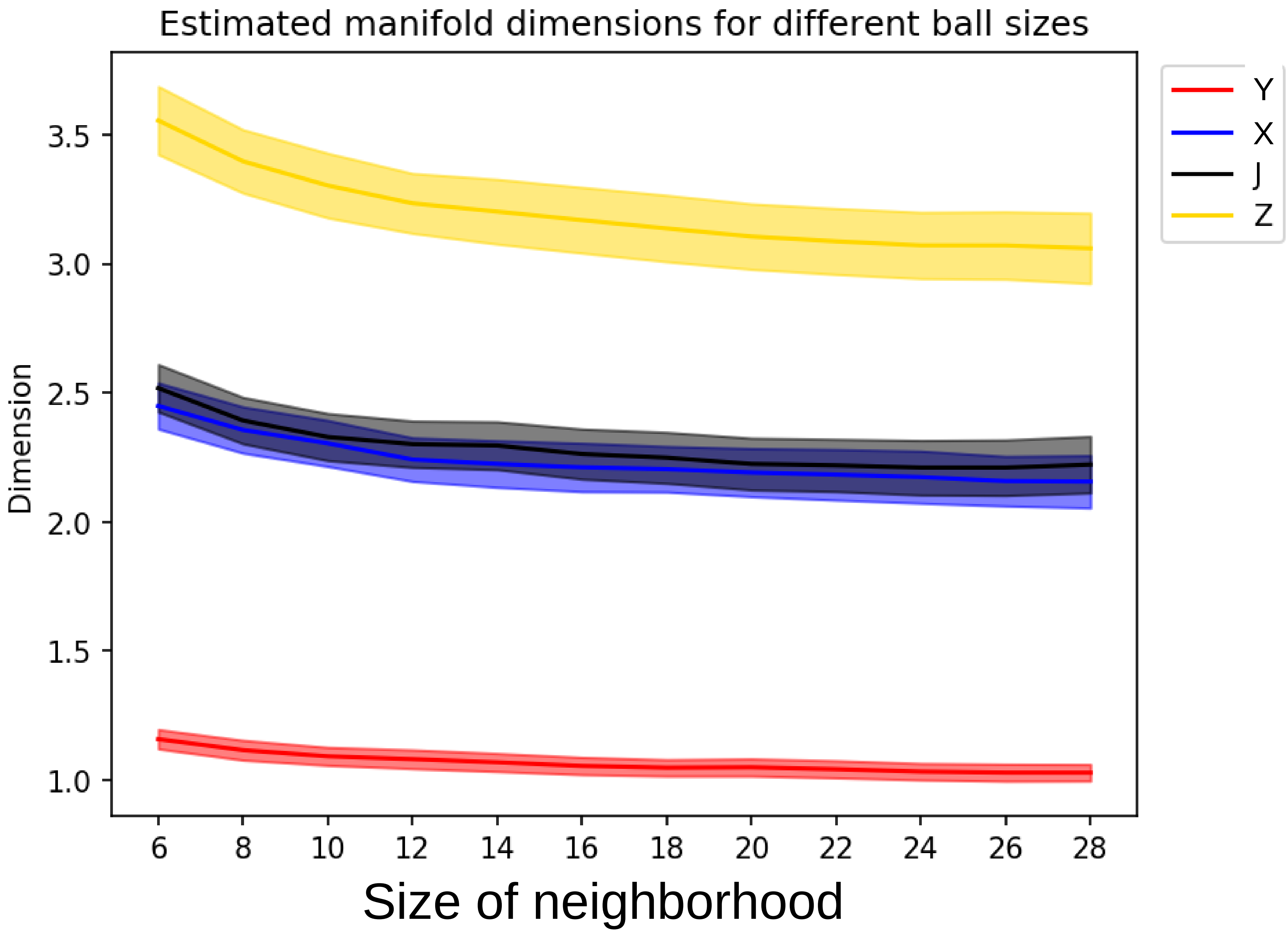
Dimensions

In general

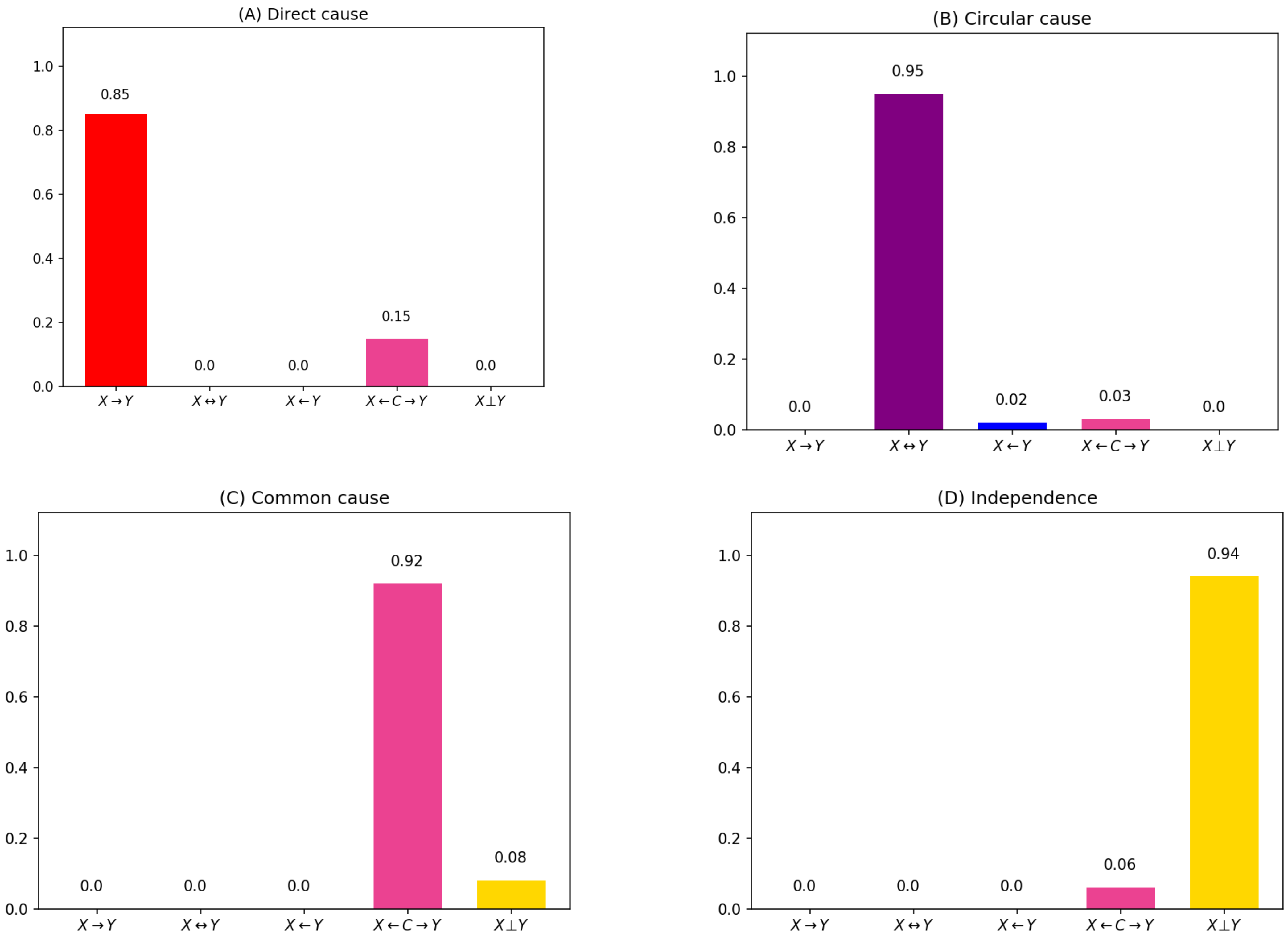
$$\max\left\{d_X, d_Y\right\} \leq d_J \leq d_X + d_Y$$

Dimensions		Causal relation
$d_X < d_Y = d_J$	$\Longleftrightarrow$	X drives Y
$d_Y < d_X = d_J$	$\Longleftrightarrow$	Y drives X
$d_X = d_Y = d_J$	$\Longleftrightarrow$	X circular Y
$\max\{d_X, d_Y\} < d_J = d_X + d_Y$	$\Longleftrightarrow$	X and Y are independent
$\max\{d_X, d_Y\} < d_J < d_X + d_Y$	$\Longleftrightarrow$	X,Y have a common cause

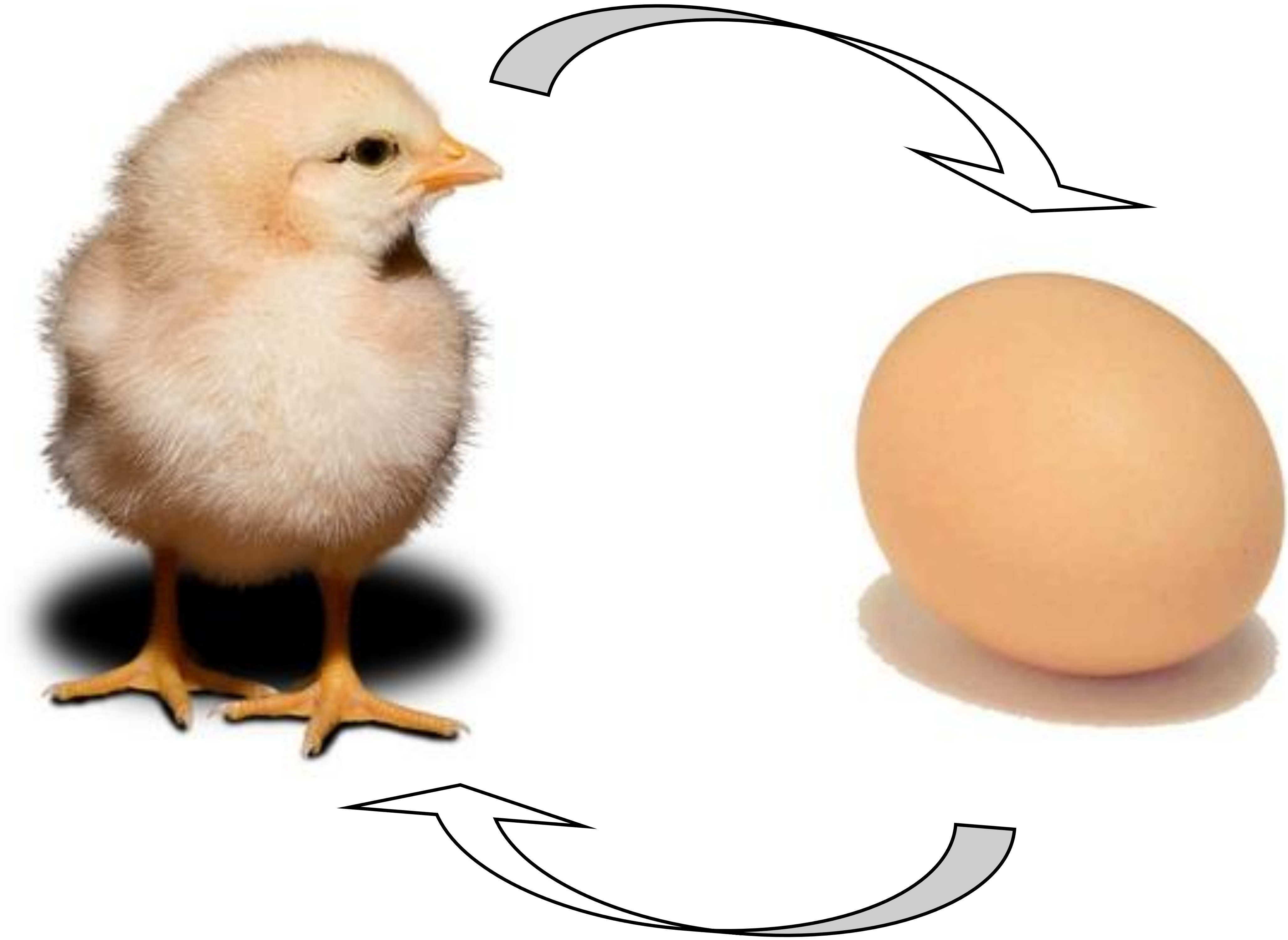
Example and test the logistic map



Logistic map



Which one came first?



1930-1983 egg production and chicken population

Chickens, Eggs, and Causality, or Which Came First?

Walter N. Thurman and Mark E. Fisher *

Time-series evidence from the United States indicates unidirectional causality from eggs to chickens.

Key words: causality, chickens, eggs.

Granger's seminal paper entitled "Investigating Causal Relations" has spawned a vast and influential literature. In macroeconomics, for example, the causal relationship between money and income has been investigated time (Sims) and again (Barth and Bennett; Williams, Goodhart, and Gowland; Ciccolo; Feige and Pearce; Hsiao). Some authors have taken exception to Granger's definition of causality *qua* causality (Zellner; Jacobs, Leamer, and Ward; Conway et al.), and even Granger has suggested "a better term might be temporally related" (Granger and Newbold, p. 225). We find ourselves in agreement with the temporal ordering interpretation of Granger causality. In fact, we believe that the most natural application of tests for Granger causality (temporal ordering) has until now been overlooked. We refer, of course, to: "Which came first, the chicken or the egg?" Our purpose in this study is to provide an empirical answer to this venerable question, which theory alone has not resolved.

This measure excludes chickens raised only for meat. Eggs are measured in millions of dozens and include all eggs produced annually in the United States. All are potentially fertilizable.

The notion of Granger causality is simple: If lagged values of *X* help predict current values of *Y* in a forecast formed from lagged values of both *X* and *Y*, then *X* is said to Granger cause *Y*. We implement this notion by regressing eggs on lagged eggs and lagged chickens; if the coefficients on lagged chickens are significant as a group, then chickens cause eggs. A symmetric regression tests the reverse causality.[1] We perform the Granger causality tests using one to four lags. The number of lags in each equation is the same for eggs and chickens.

To conclude that one of the two "came first," we must find unidirectional causality from one to the other. In other words, we must reject the noncausality of the one to the other and at the same time fail to reject the noncausality of the other to the one. If either both cause each other or neither causes the other,

**Mark E Fisher ≠ Ronald Fisher father of modern statistics*

, American Journal of Agricultural Economics *(1988)*

Which one came first?

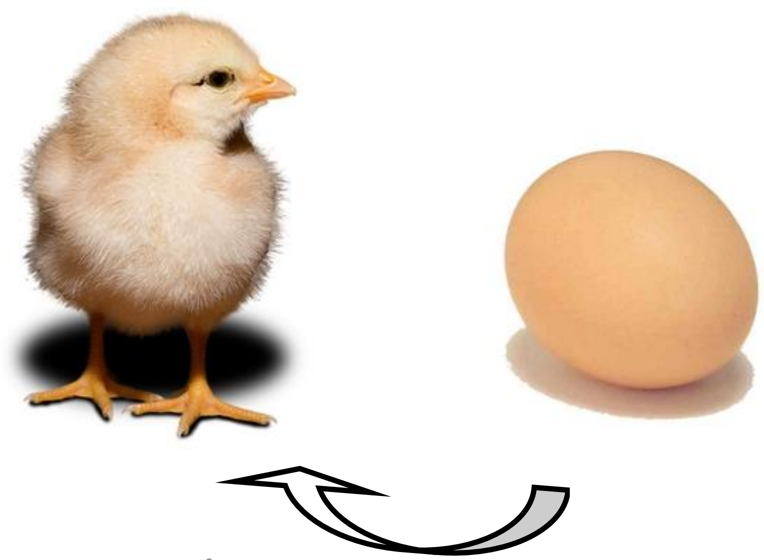


Table 1. Granger Causality Tests

Part 1: Did the Chicken Come First?

The following equation was estimated by OLS:

$$Eggs_t = \mu = \sum_{i=1}^{L} \alpha_i \, Eggs_{t-i} = \sum_{i=1}^{L} \beta_i \, Chickens_{t-i} + \epsilon_t;$$

$H_0 : \beta_1 = \ldots \beta_L = 0$ (chickens do not Granger cause eggs).

L = no. of lags	F-statistic	P-value	R^2 of the regression
1	.04	.85	.96
2	1.71	.19	.97
3	1.10	.36	.97
4	.79	.54	.97

Part 2: Did the Egg Come First?

The following equation was estimated by OLS:

$$Chickens_t = \mu + \sum_{i=1}^{L} \alpha_i \, Chickens_{t-i} + \sum_{i=1}^{L} \beta_i \, Eggs_{t-i};$$

$H_0 : \beta_1 = \ldots = \beta_L = 0$ (eggs do not Granger cause chichens).

L = no. of lags	F-statistic	P-value	R^2 of the regression
1	1.23	.27	.73
2	10.36	.0002	.81
3	5.85	.0019	.81
4	4.71	.0032	.82

Data source: U.S. Department of Agriculture, 1983 and others.
Note: The data are annual, 1930–83.

We perform the Granger causality tests using one to four lags. The number of lags in each equation is the same for eggs and chickens.

To conclude that one of the two "came first," we must find unidirectional causality from one to the other. In other words, we must reject the noncausality of the one to the other and at the same time fail to reject the noncausality of the other to the one. If either both cause each other or neither causes the other, the question will remain unanswered. The test results are presented in table 1. They indicate a clear rejection of the hypothesis that eggs do not Granger cause chickens. They provide no such rejection of the hypothesis that chickens do not Granger cause eggs. Therefore, we conclude that the egg came first.[2]

Which one came first?

terpretation of Granger causality.
believe that the most natural appli-
sts for Granger causality (temporal
as until now been overlooked. We
ourse, to: "Which came first, the
the egg?" Our purpose in this study
le an empirical answer to this ven-
stion, which theory alone has not

Results

e annual U.S. time series from 1930
gg production and chicken popula-
ount as chickens the 1 December
of a
broi
f the
all chickens that lay or fertilize
ll chickens capable of causing eggs.

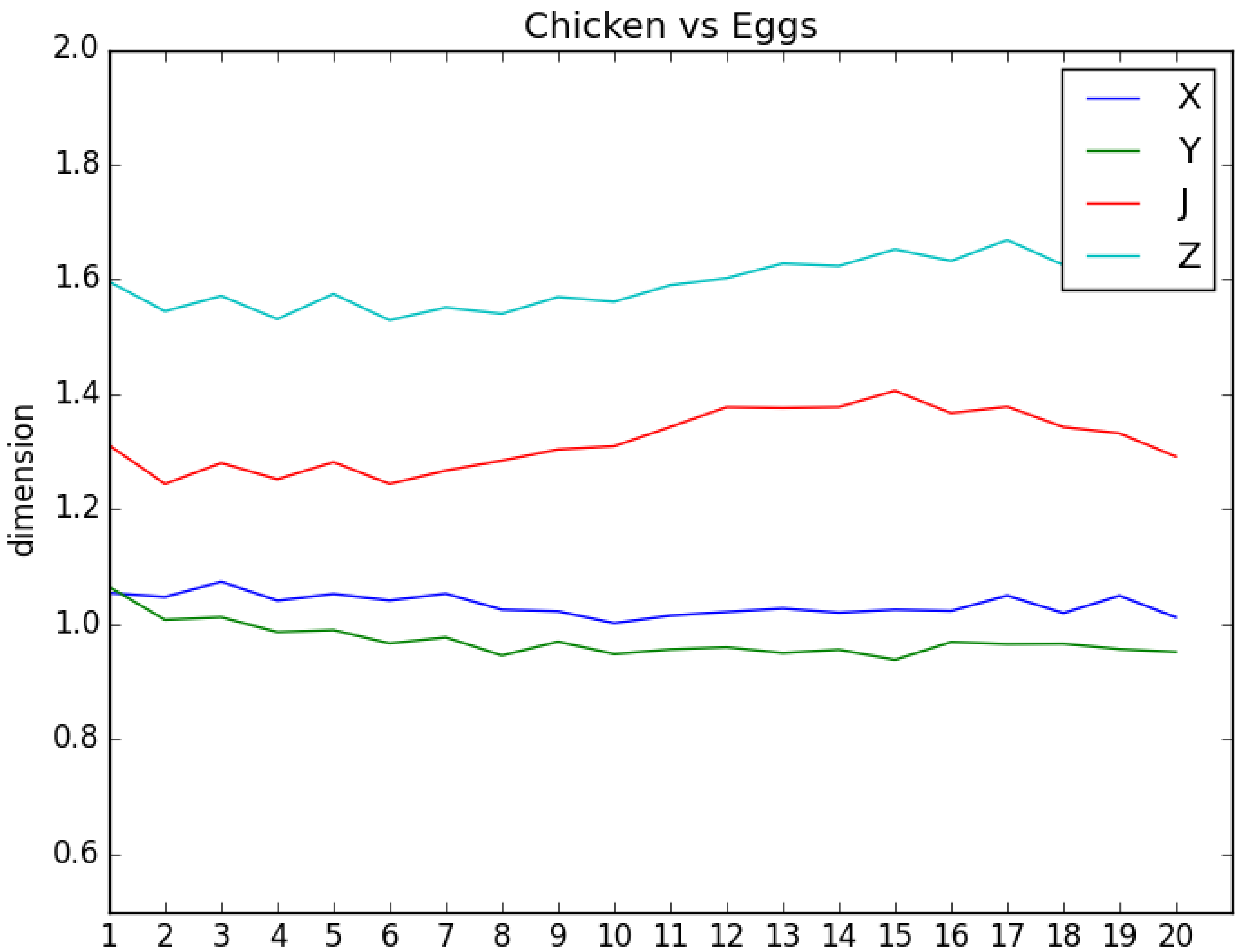
We perform the Granger causality tests using one to four lags. The number of lags in each equation is the same for eggs and chickens.

To conclude that one of the two "came first," we must find unidirectional causality from one to the other. In other words, we must reject the noncausality of the one to the other and at the same time fail to reject the noncausality of the other to the one. If either both cause each other or neither causes the other, the question will remain unanswered. The test results are presented in table 1. They indicate a clear rejection of the hypothesis that eggs do not Granger cause chickens. They provide no such rejection of the hypothesis that chickens do not Granger cause eggs. Therefore, we

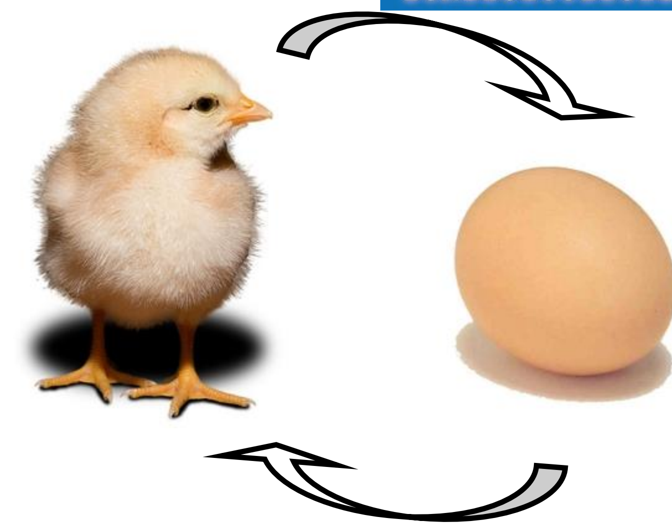
Therefore, we conclude that eggs came first

[1] Feige and Pearce describe and distinguish among the several Granger causality tests. The validity of our test statistic requires lack of serial correlation, homoskedasticity, and normality of the

Which one came first?



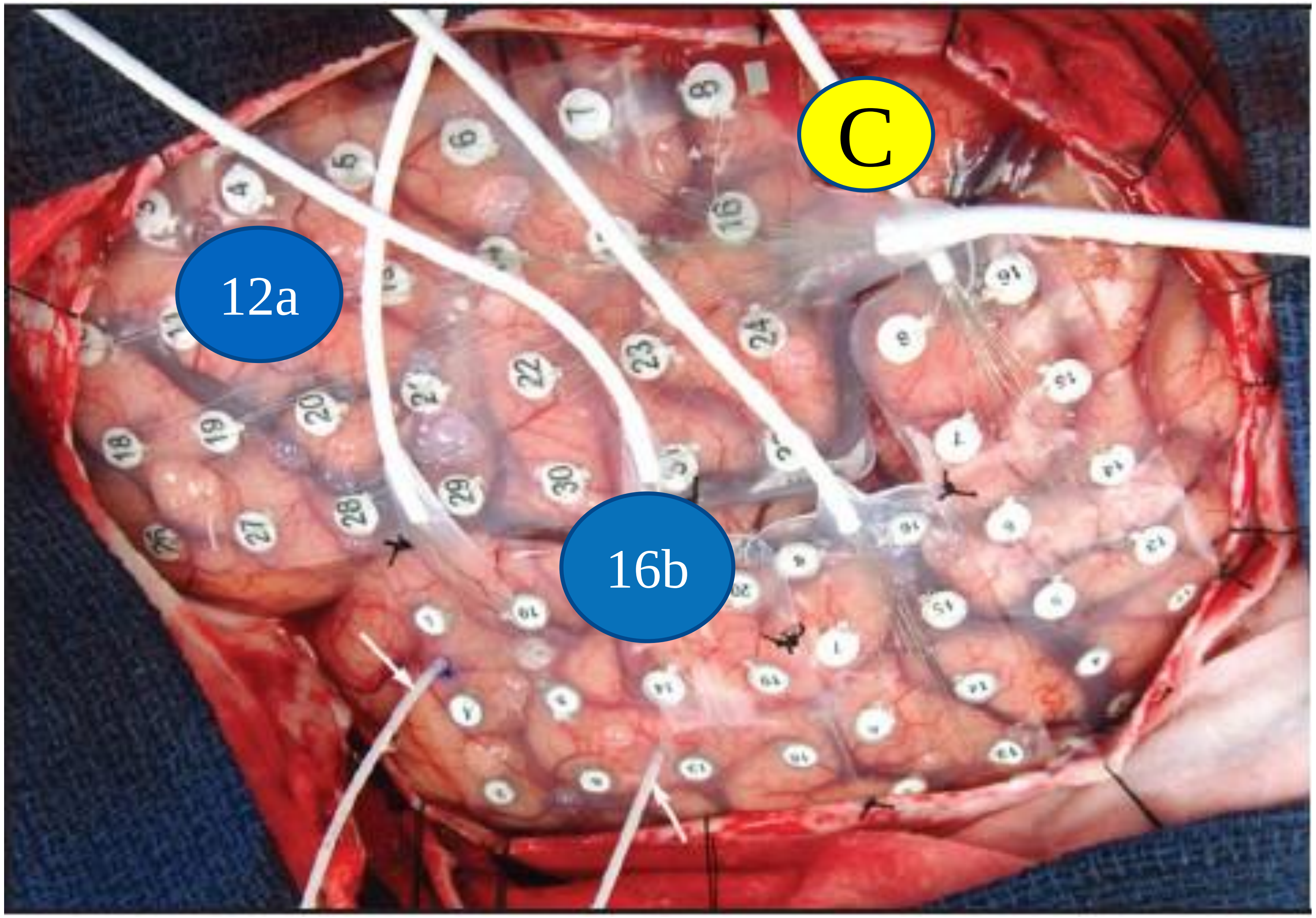
Size of neighborhood



Application to epilepsy focus detection

Application to epilepsy focus detection

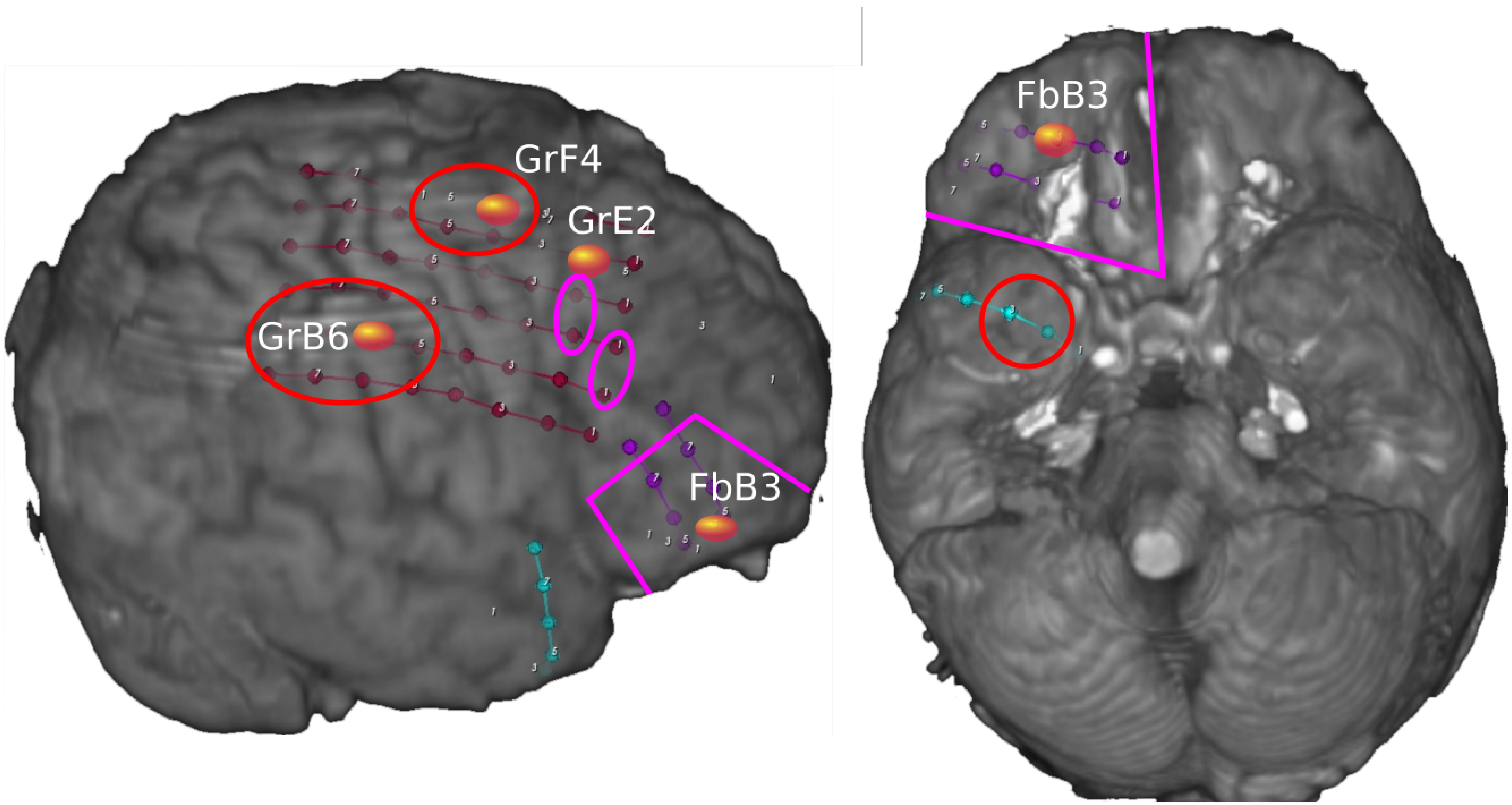
Which region is the source of the epileptic seizure?



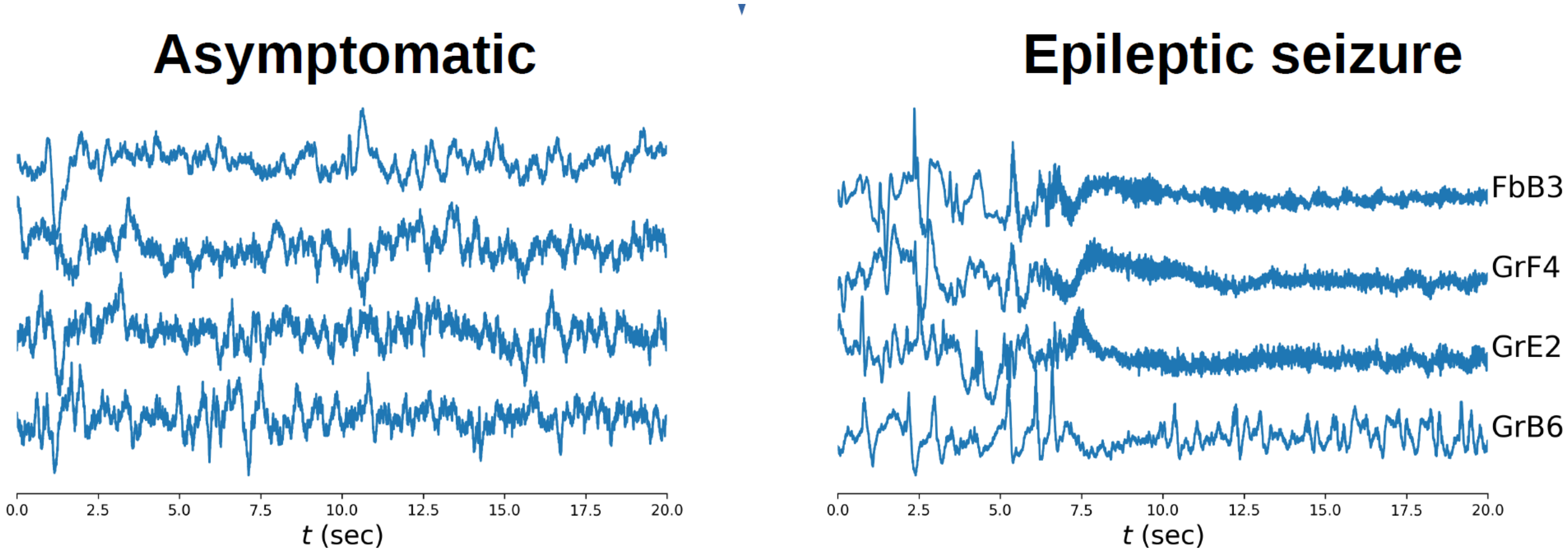
Shah XK, Mittal S. Invasive electroencephalography monitoring: Indications and presurgical planning. Xnn Indian Xcad Neurol 2014;17, Suppl S1:89-94

Application to epilepsy focus detection

Areas to be analysed



Application to epilepsy focus detection

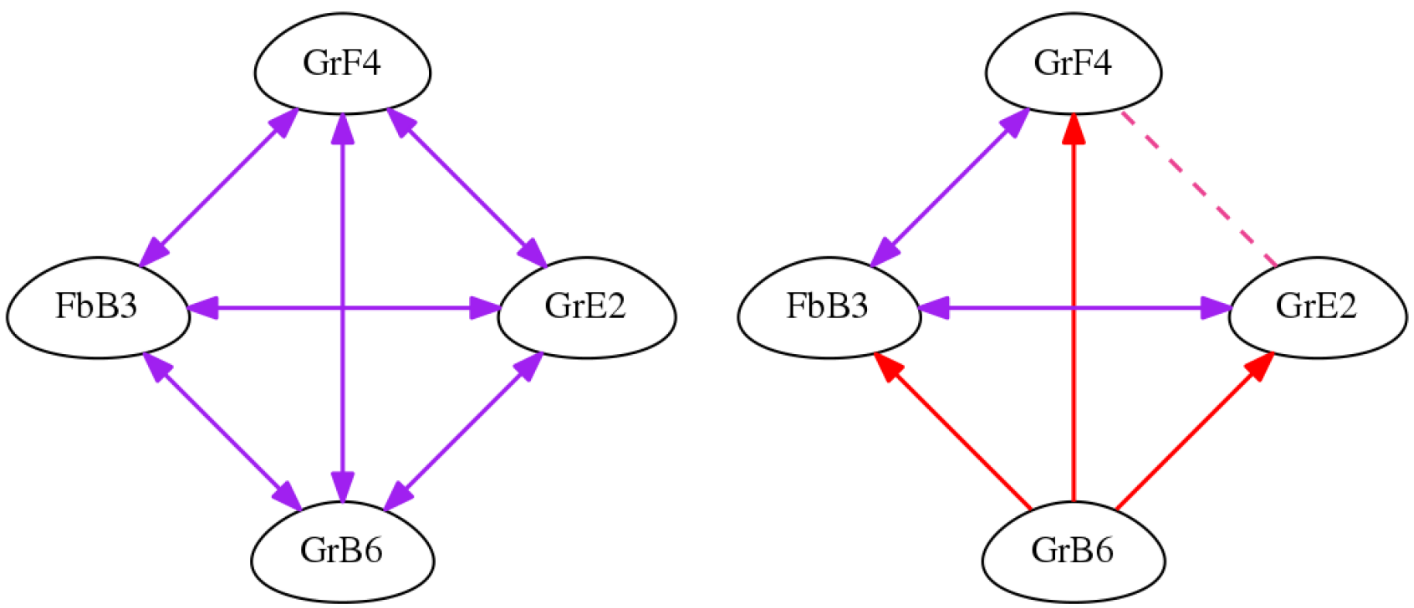


2 – Data preprocessing

Band-pass filtering (1-30 Hz)
Normalization

3 – Dimension-causality analysis

embedding dimension: 5
embedding delay: 11 step

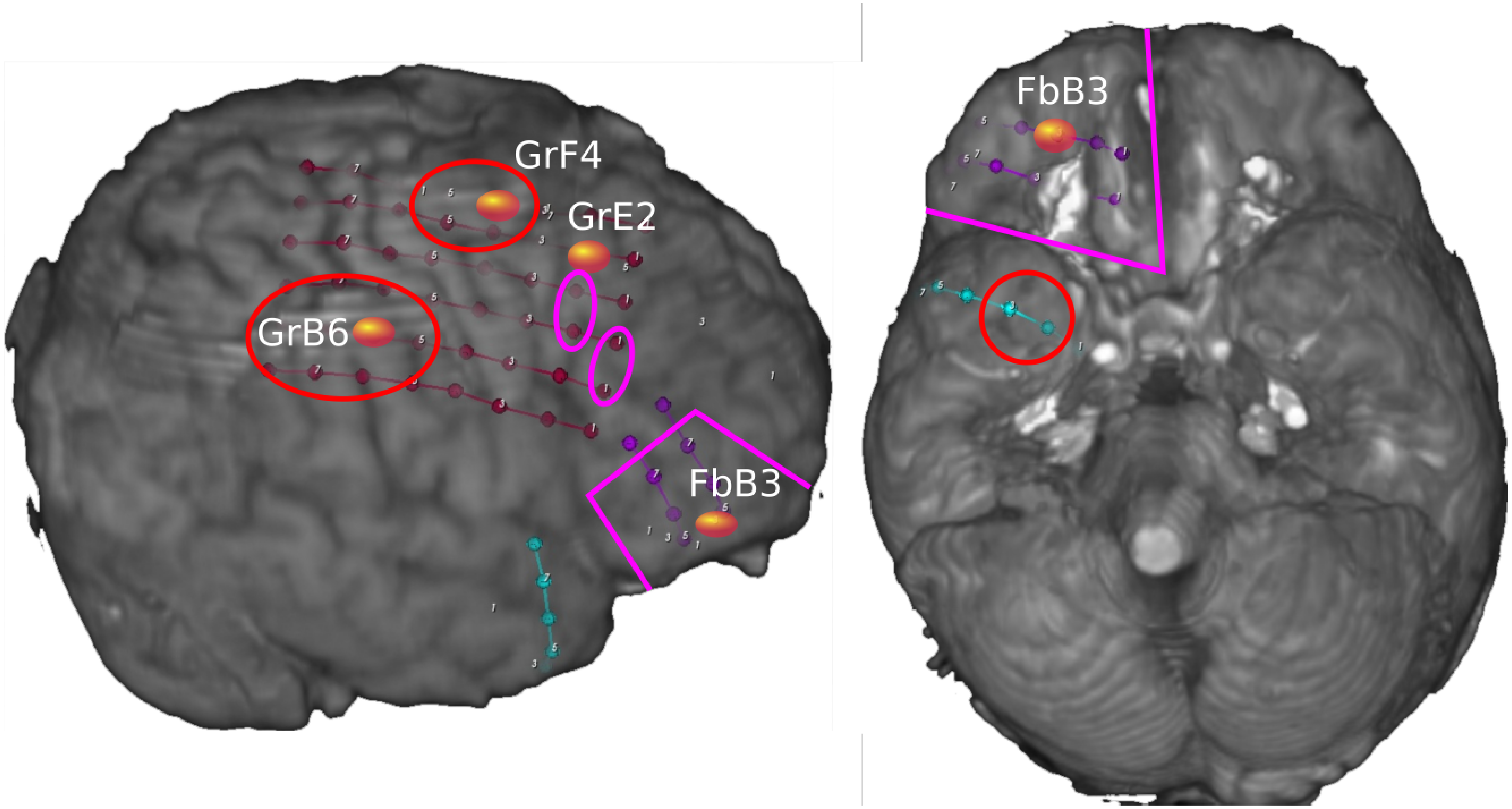


4 – Result

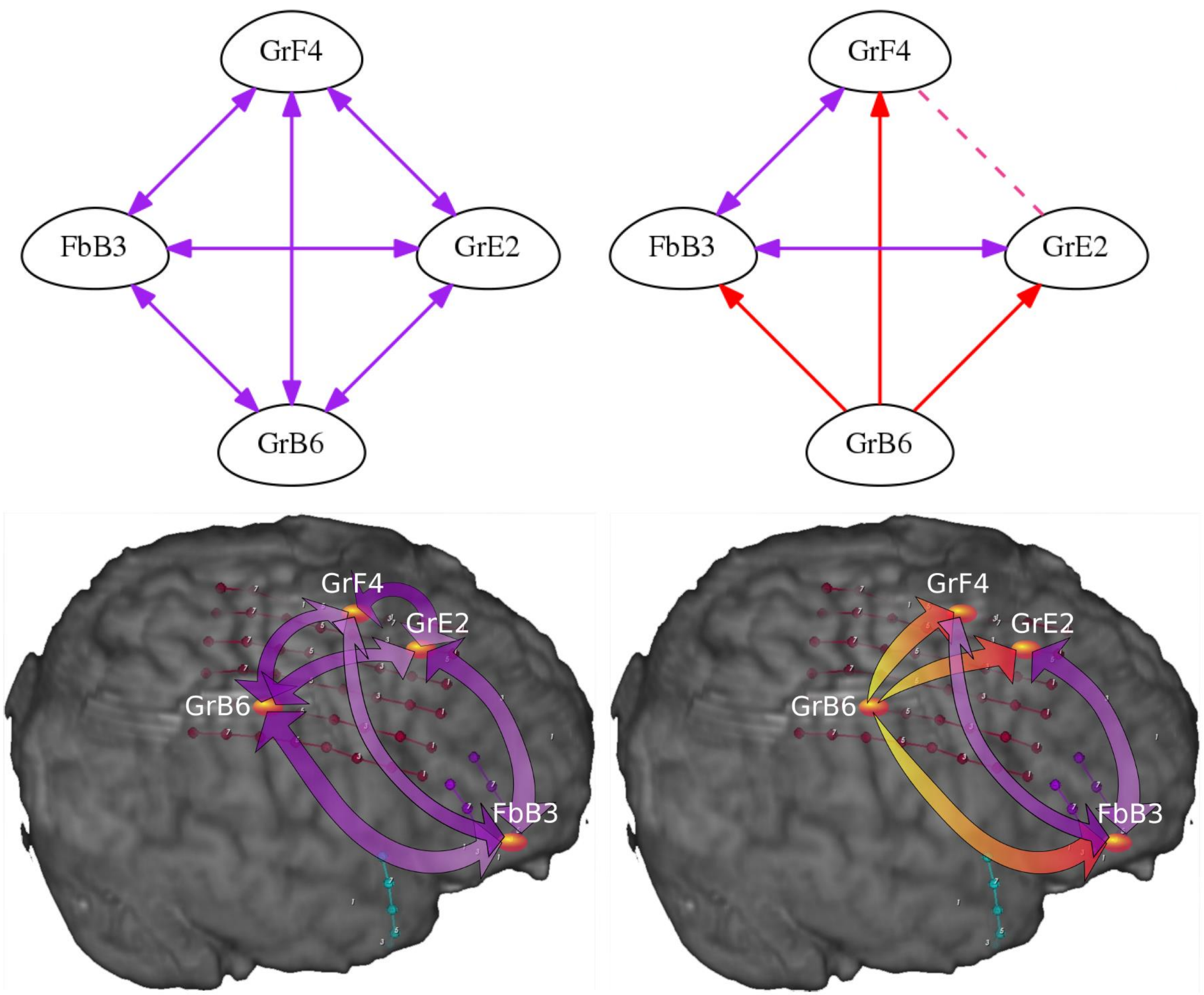
Our causality analysis showed that all the 4 area in question were mutually interconnected during normal, interictal activity, but the infero-temporal (GrB6) area became the dominant cause during seizure.

Application to epilepsy focus detection

Magenta areas have been removed



Application to epilepsy focus detection



Discussion

These results can be interpreted that, although, the resection of the large part of a highly interconnected epileptic network significantly reduced the seizure activity for a while, the un-touched primary cause transformed the remained tissue towards epilepsy and the seizures were restored.

Summary

Our method

- Detects and distinguish all causality relations

- Assigns probability to causality relations

Any suggestion for test data?

Thanks for the attention