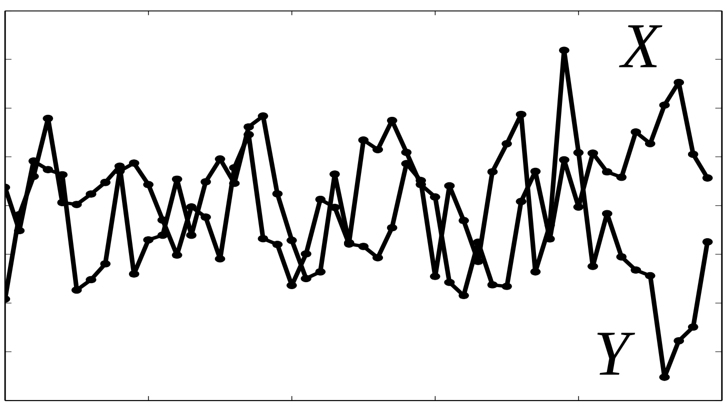
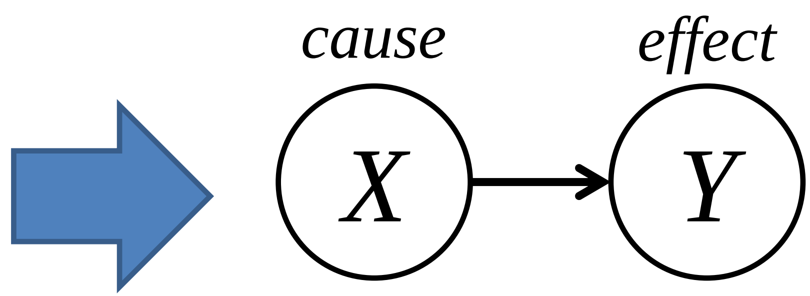
Causal Inference in Time Series via Supervised Learning

Yoichi Chikahara, Akinori Fujino
NTT Communication Science Laboratories
Kyoto, Japan

Causal inference in time series

- Given time series data
- Infer *causal relationships* between variables



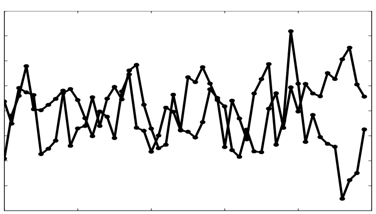


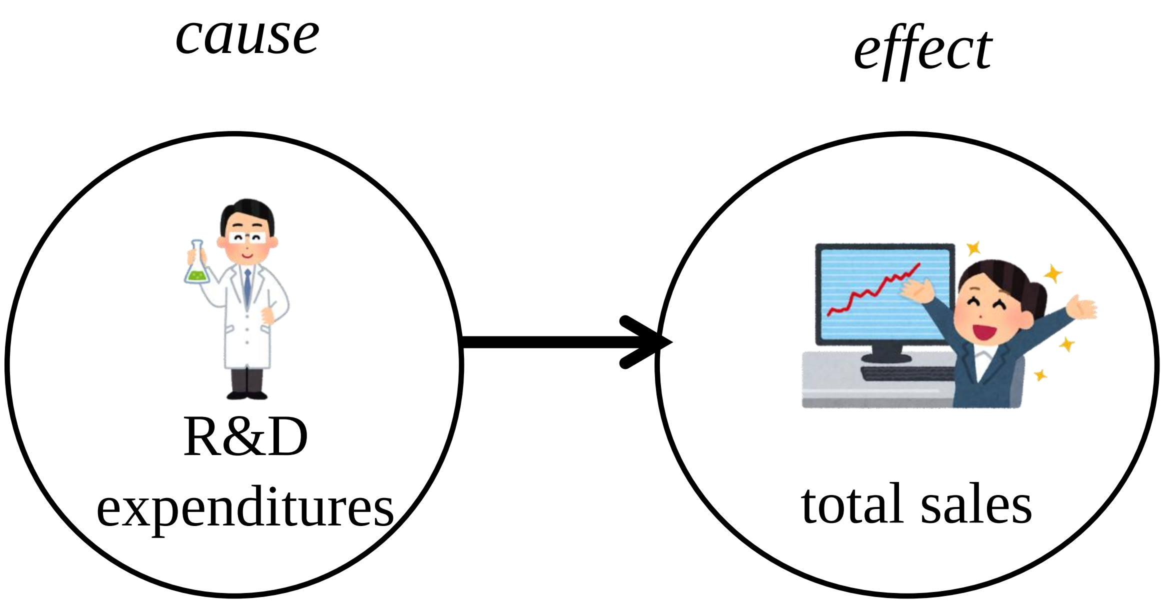
Input: Time Series Data

Output: *Causal Relationships*

Application: e.g., Economics

- Finding that R&D expenditures *influences* total sales is useful for companies





What is “causal relationship”?

A definition of temporal causality

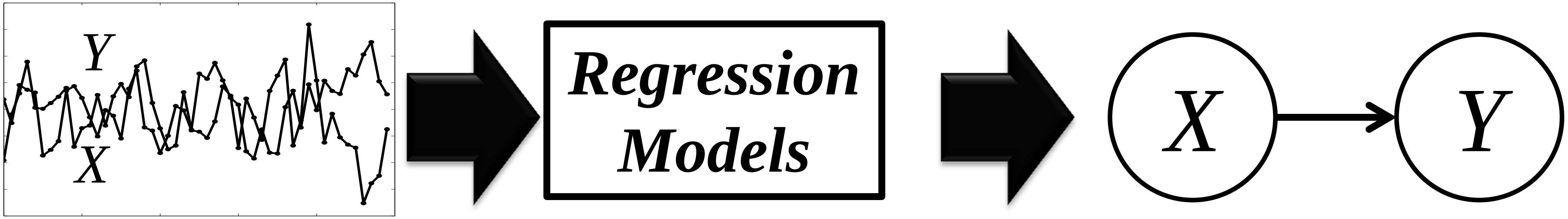
Granger causality [Granger1969]

X is the cause of *Y*

if the past values of *X* are **helpful in predicting** the future values of *Y*

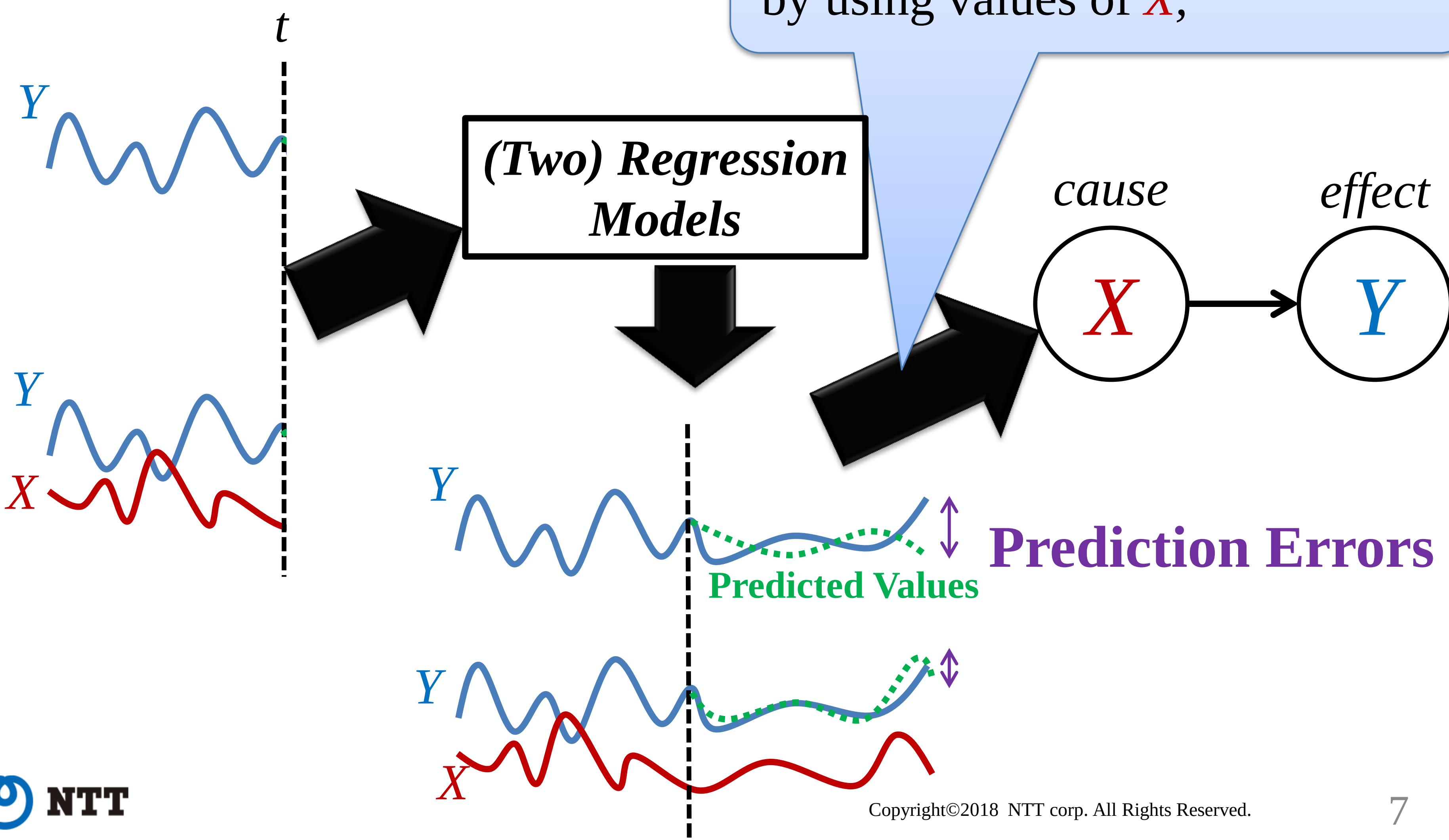
Clive W. J. Granger (1934-2009)

Existing Approach:
Using regression models



Existing approach:
Compare prediction errors with/without using values of *X*

If errors are significantly reduced by using values of *X*,



Weakness of existing approach

Which regression model should I use ?

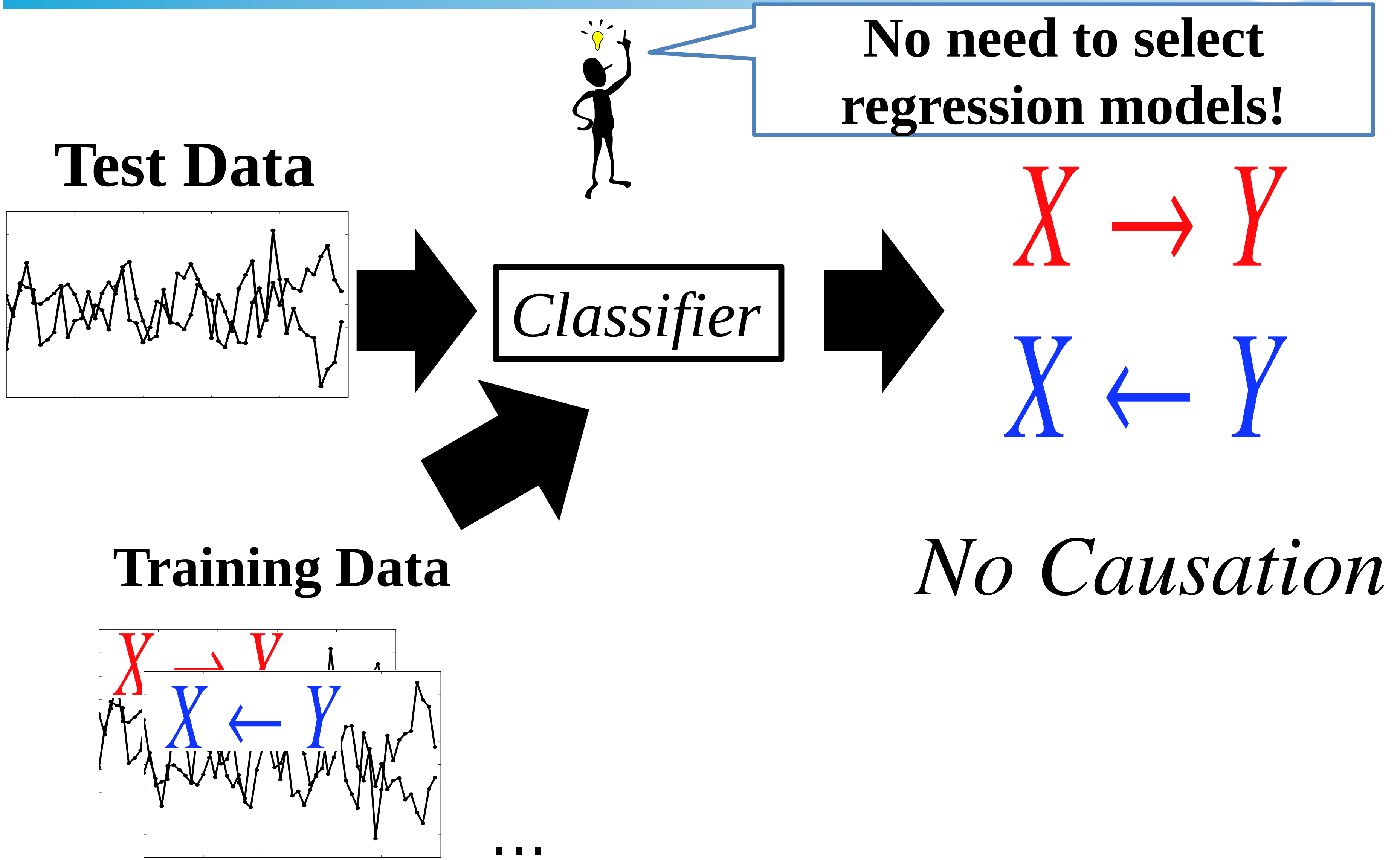
Regression Models

- **VAR models**
- **Gaussian Processes**
- **GAM**
- **…etc.**

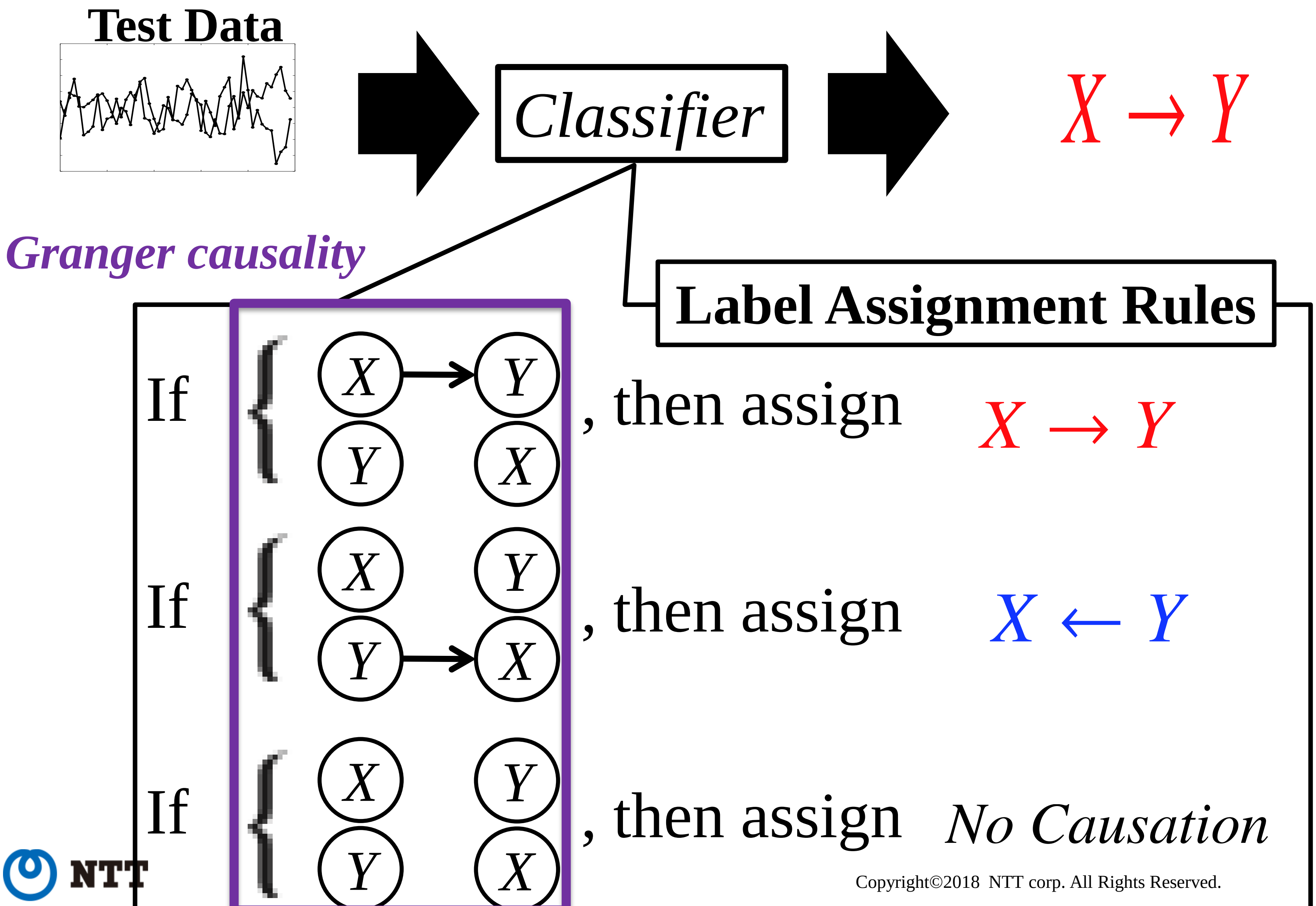
Misspecification of regression models leads to **low inference accuracy**

NTT

Our approach: Causal inference via supervised learning



Our goal:
Building a classifier that follows label assignment rules



Ideas for building classifier

The answer lies in Granger causality definition

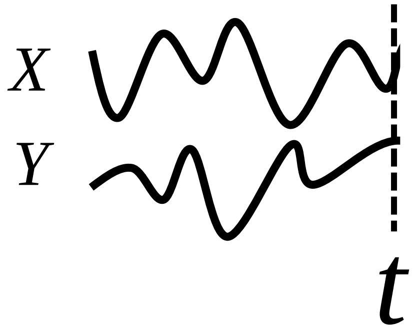
Revisiting definition of Granger causality

Granger causality defines that

cause *effect*

$$\begin{tikzcd} X \arrow[r] & Y \end{tikzcd} \quad \text{if} \quad P(Y_{t+1}|\underline{S_X, S_Y}) \neq P(Y_{t+1}|\underline{S_Y})$$

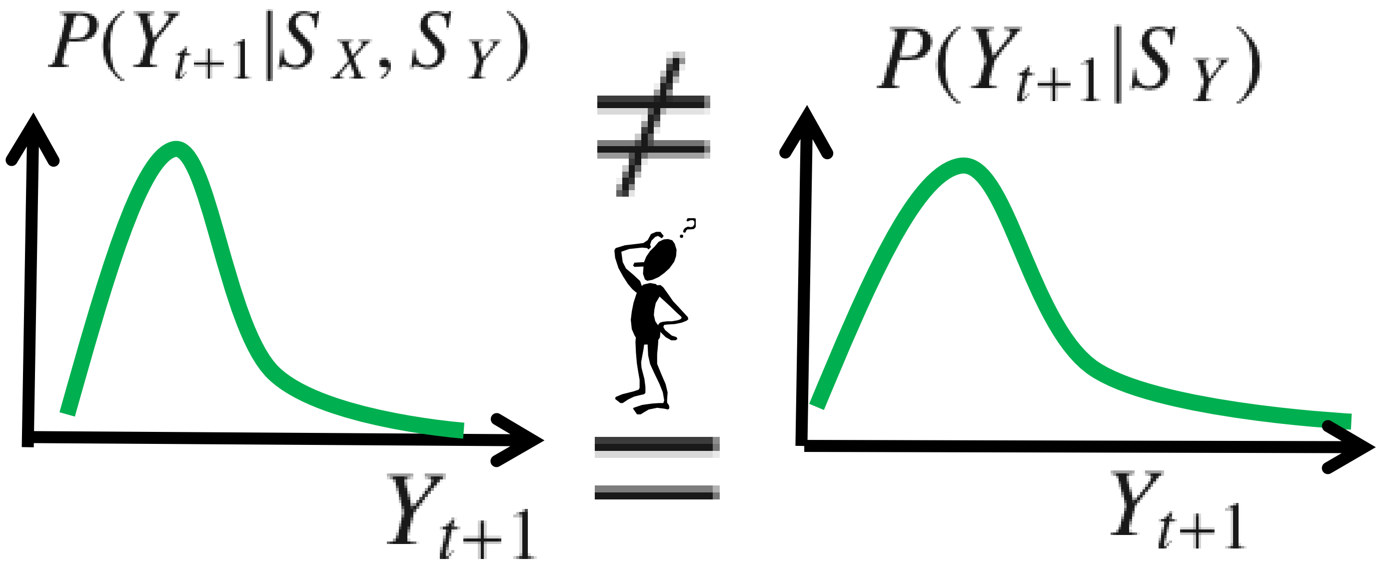
$$\begin{tikzcd} X & Y \end{tikzcd} \quad \text{if} \quad P(Y_{t+1}|\underline{S_X, S_Y}) = P(Y_{t+1}|\underline{S_Y})$$



$$S_X = \{x_1, \cdots, x_t\}$$

$$S_Y = \{y_1, \cdots, y_t\}$$

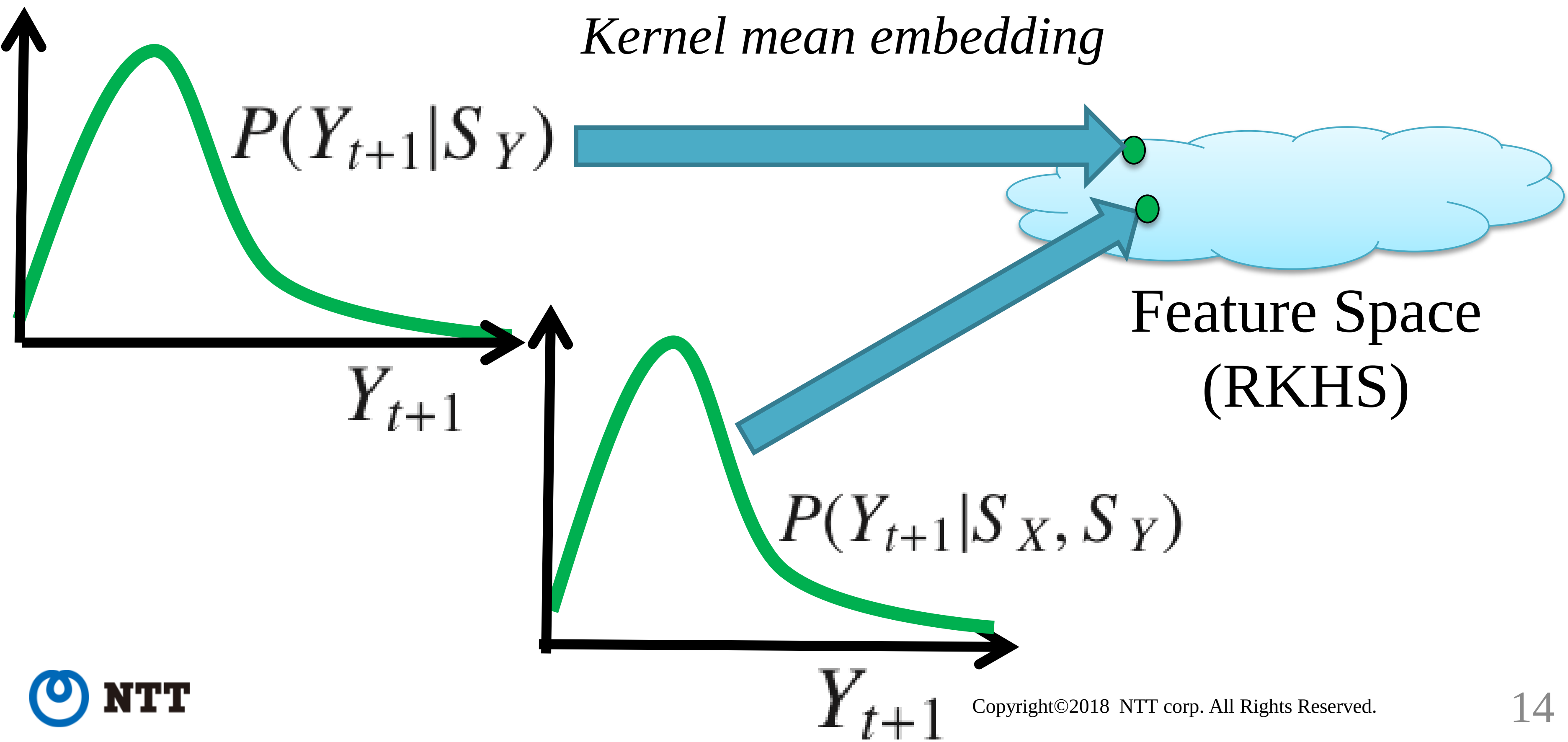
Whether or not **two distributions are equal** is important



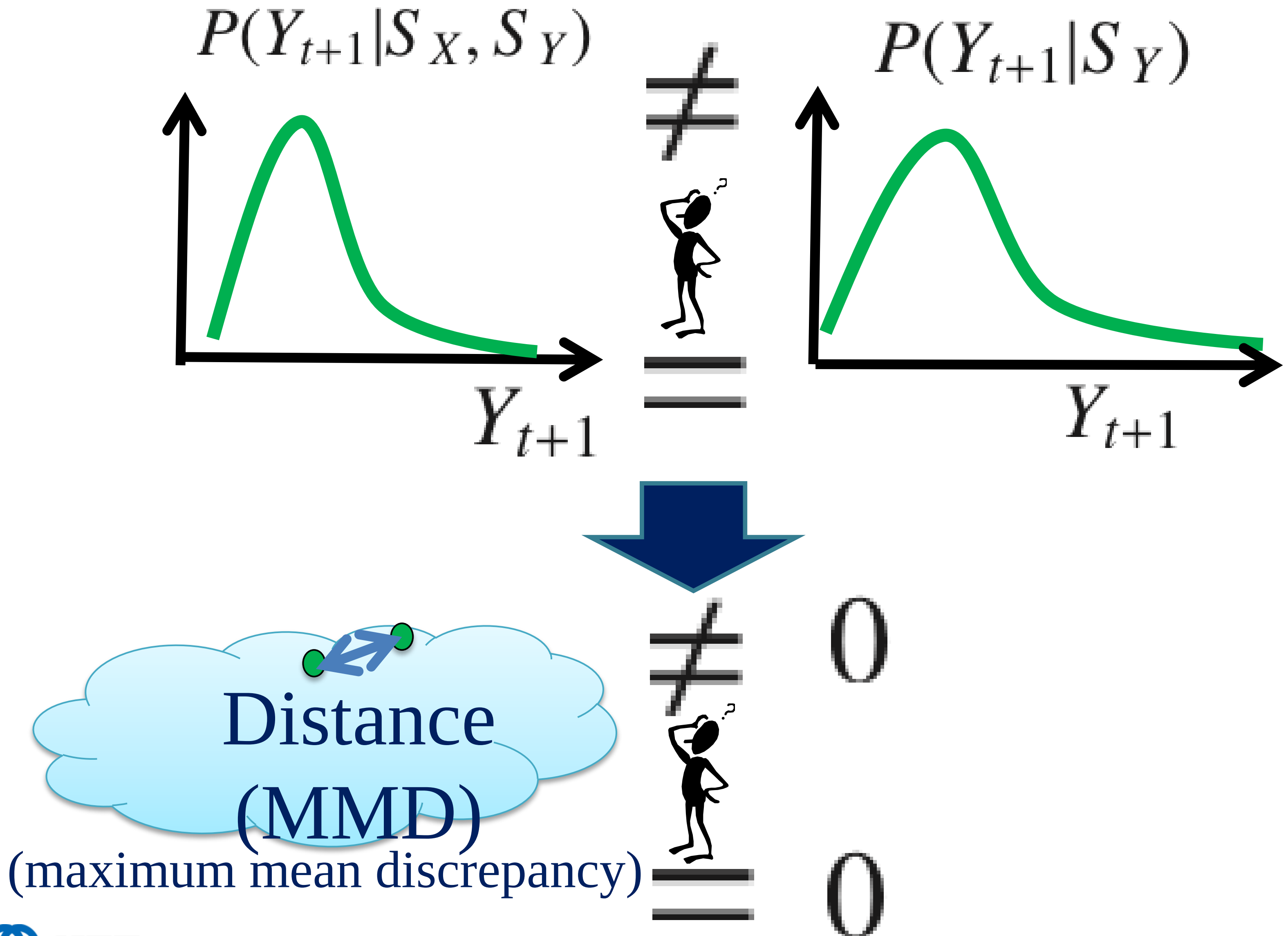
How can we determine whether or not two distributions are equal?

Representing features of distributions

- Using *kernel mean embedding* to map conditional distribution to a point in feature space



Whether or not **distance between points is zero** is important



Our goal:
Building a classifier for Granger causality identification

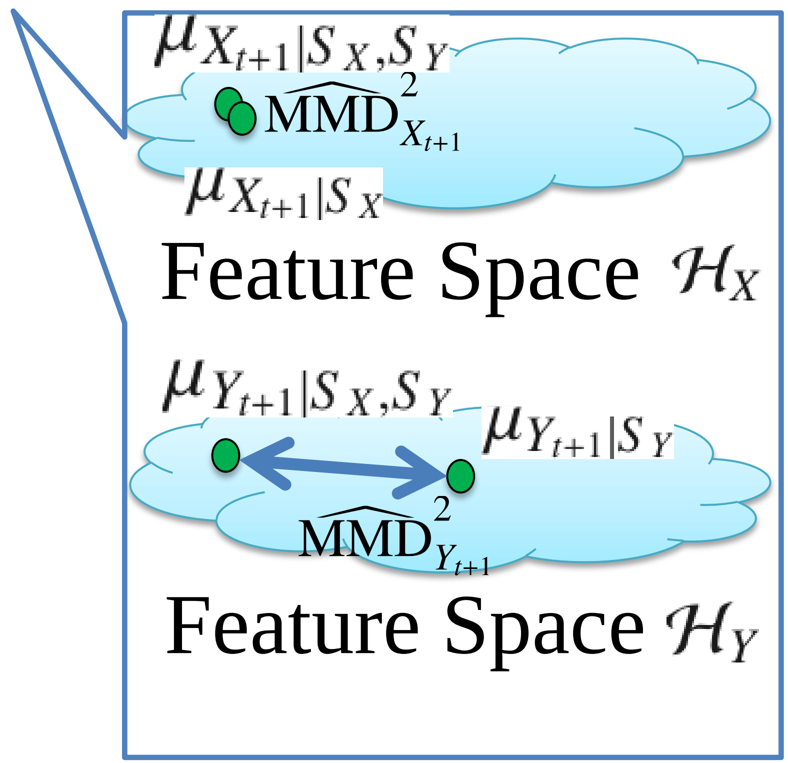
- By using MMDs, label assignment rules can be rephrased as follows:

$$\text{If } \begin{cases} \widehat{\mathrm{MMD}}^2_{X_{t+1}} = 0 \\ \widehat{\mathrm{MMD}}^2_{Y_{t+1}} \neq 0 \end{cases} \text{ then } \quad X \rightarrow Y$$

$$\text{If } \begin{cases} \widehat{\mathrm{MMD}}^2_{X_{t+1}} \neq 0 \\ \widehat{\mathrm{MMD}}^2_{Y_{t+1}} = 0 \end{cases} \text{ then } \quad X \leftarrow Y$$

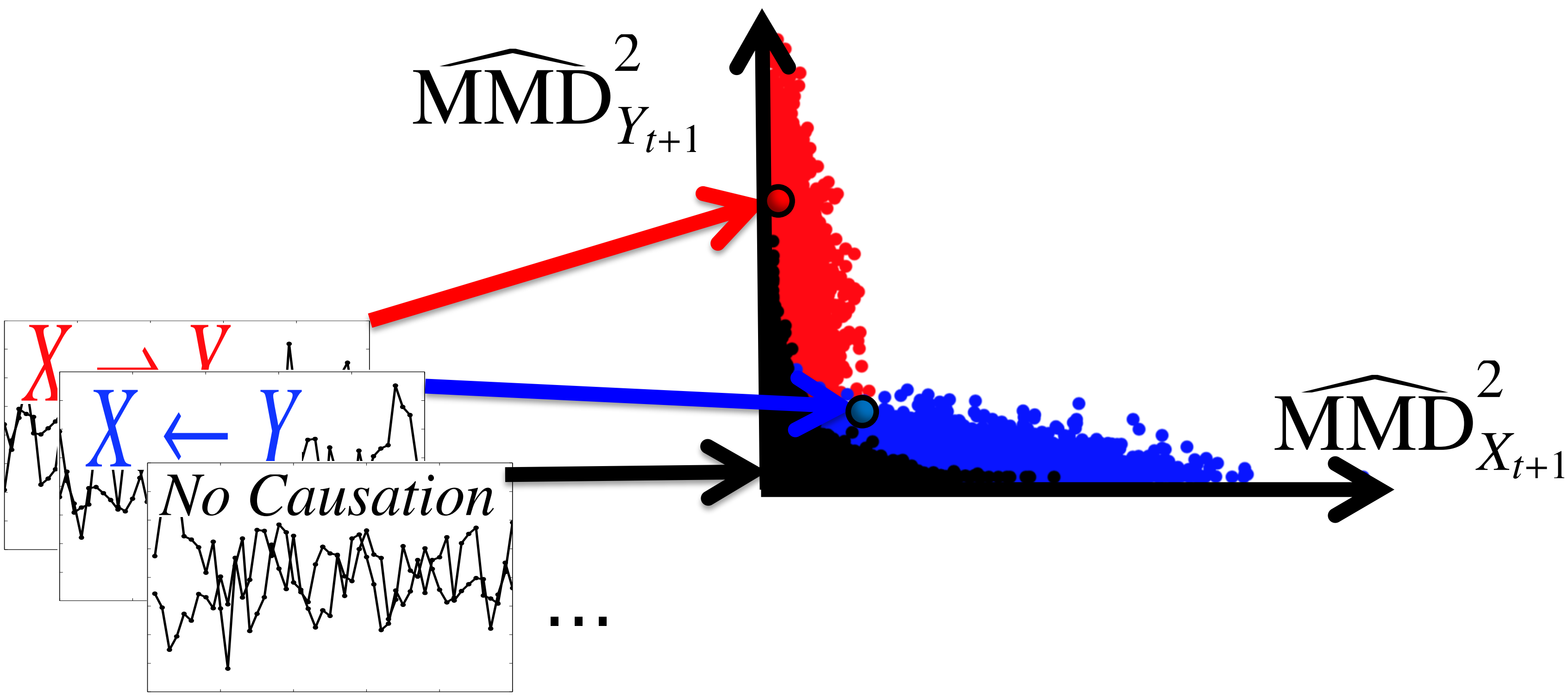
$$\text{If } \begin{cases} \widehat{\mathrm{MMD}}^2_{X_{t+1}} = 0 \\ \widehat{\mathrm{MMD}}^2_{Y_{t+1}} = 0 \end{cases}$$

then *No Causation*

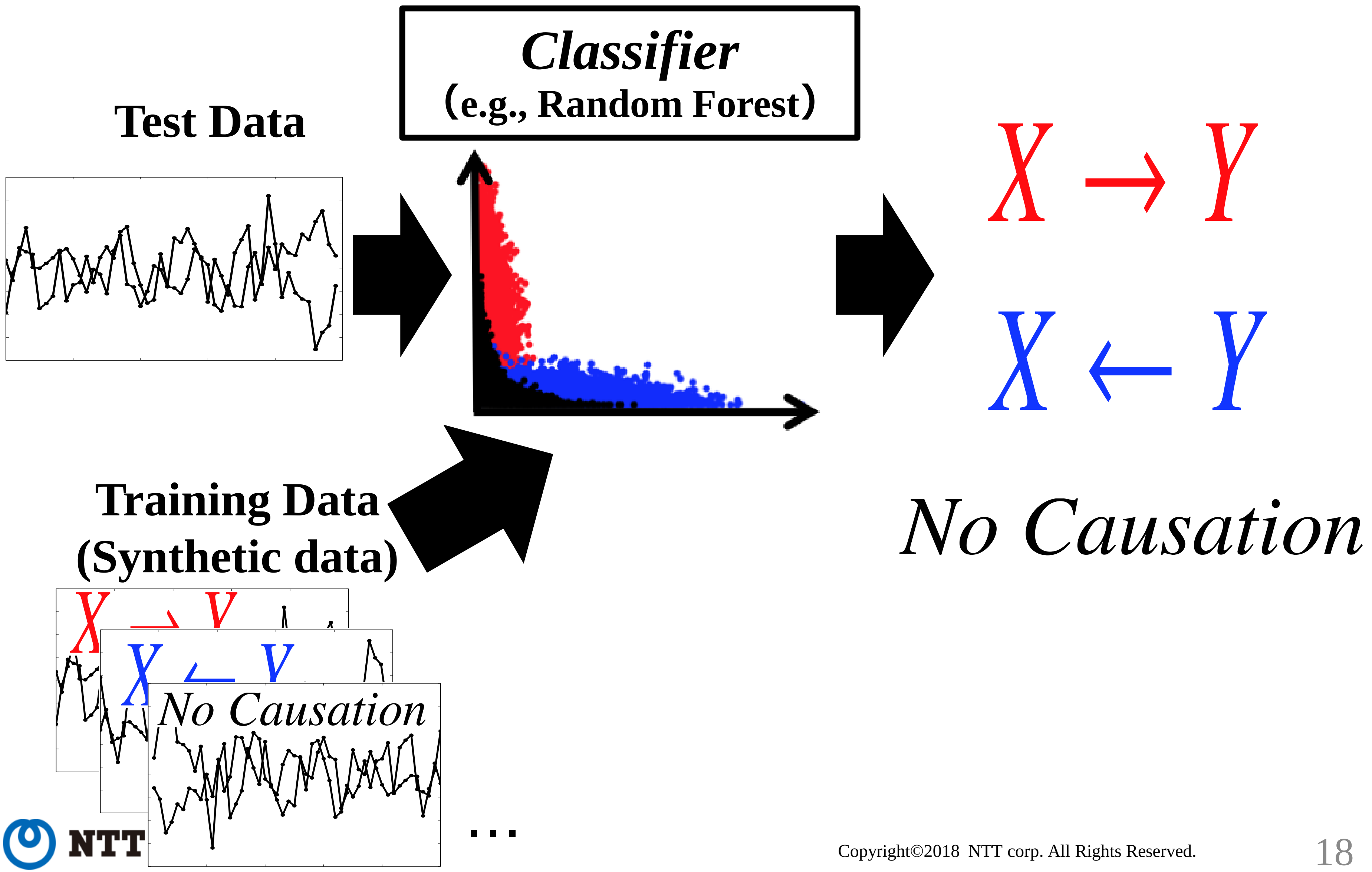


To do so: Utilizing MMDs as features for classification

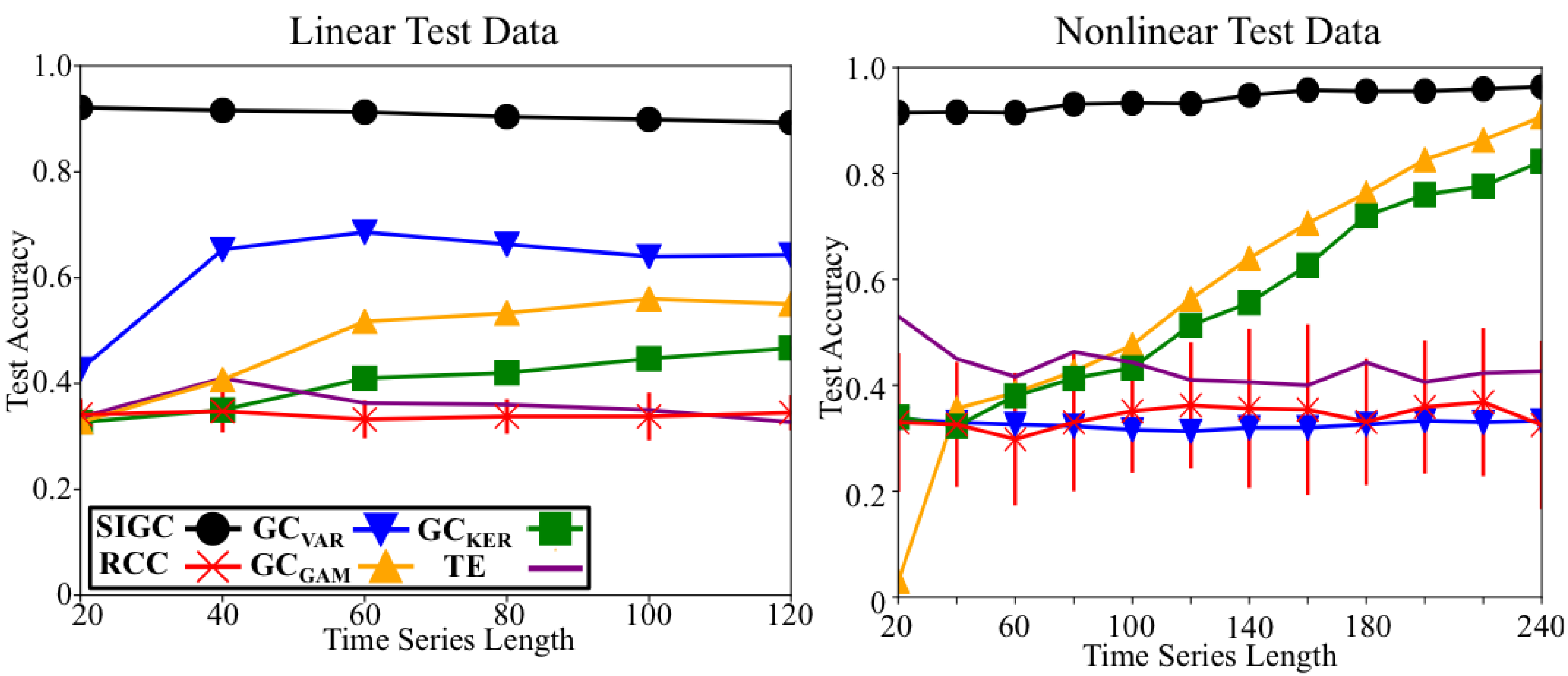
- By utilizing MMDs, we can obtain feature vectors that are sufficiently different depending on Granger causality



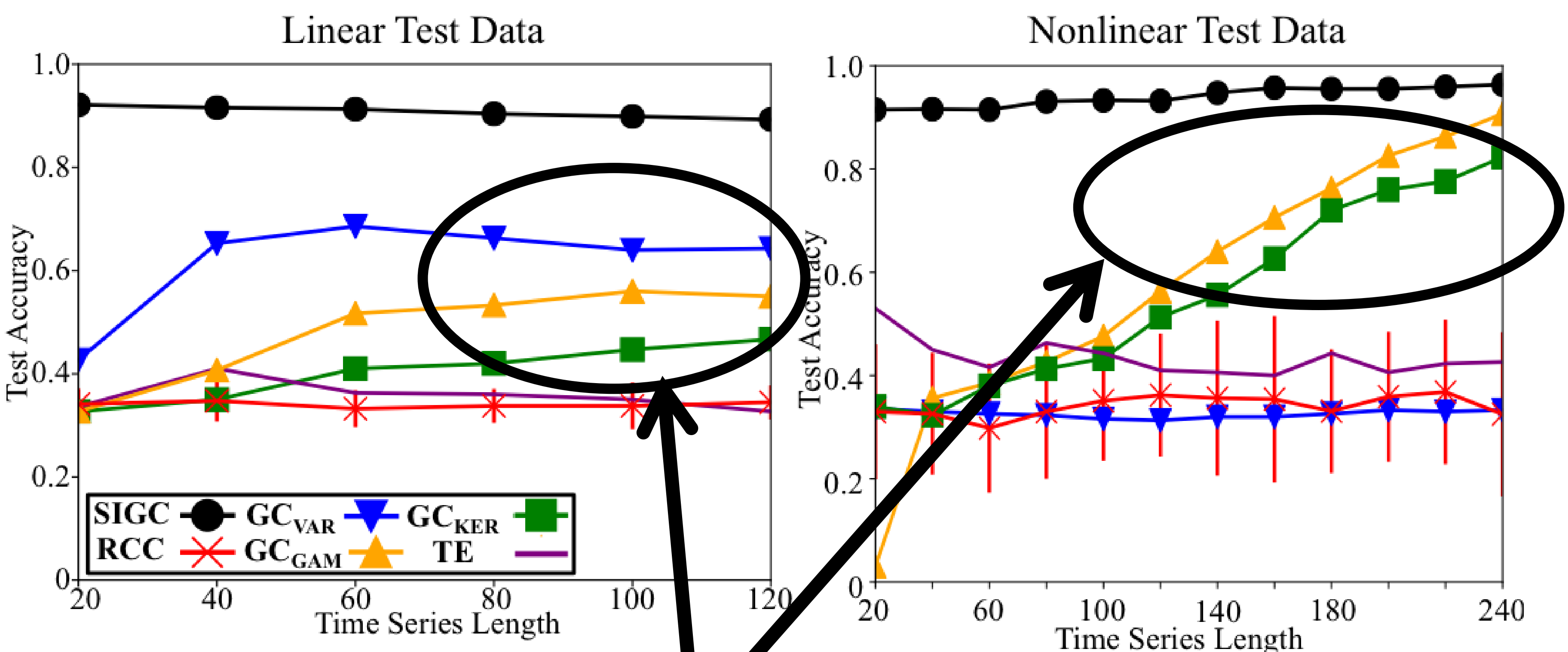
Experiments



Results on synthetic test data



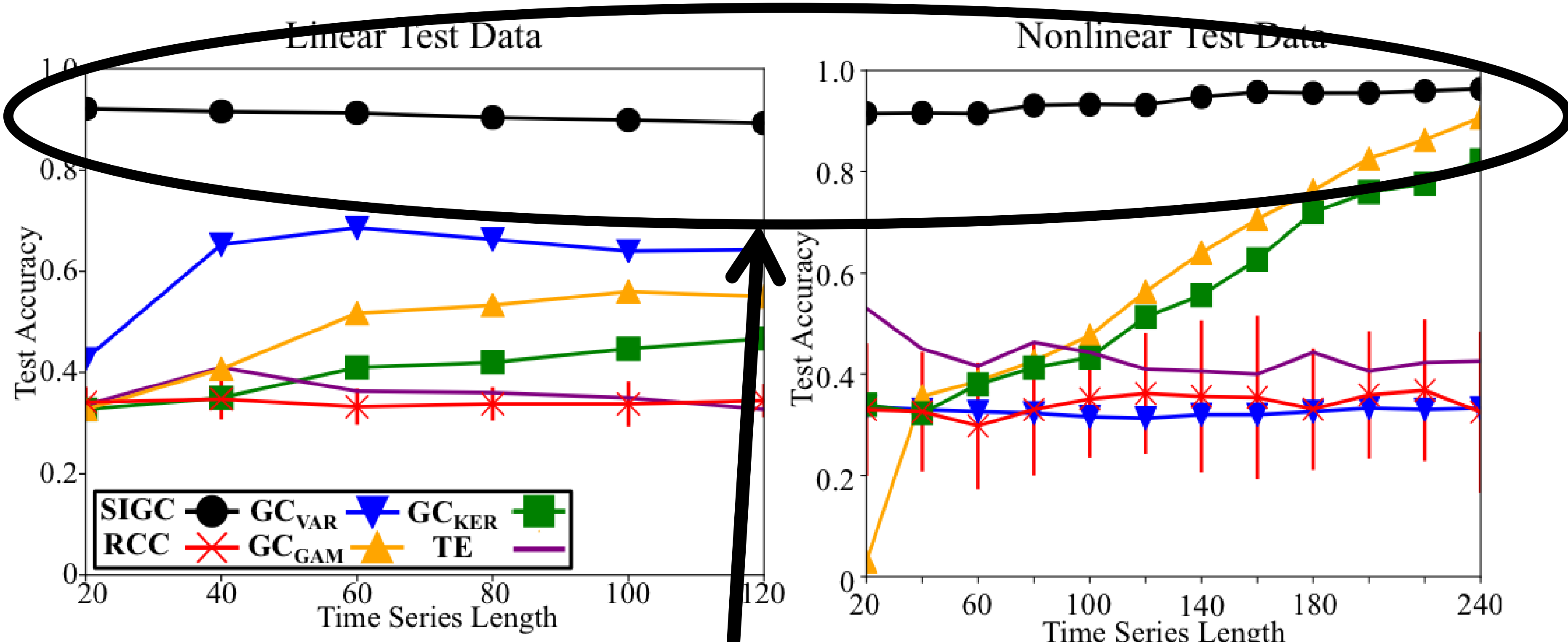
Results on synthetic test data



Existing Granger causality methods

Test accuracy strongly depended on the regression model

Results on synthetic test data



Proposed method
Performance was sufficiently good

Questions ?

Come and see **Poster #1571**
for more detailed information!