A METHOD FOR FORECASTING FINANCIAL TIME SERIES BASED ON EMPIRICAL MODE DECOMPOSITION AND MANIFOLD-LEARNING

CANDIDATE AUTHOR : ANTONIO ROBERTO

UNIVERSITY OF
SALERNO

PROF. MARIO VENTO

UNIVERSITY OF
GRONINGEN

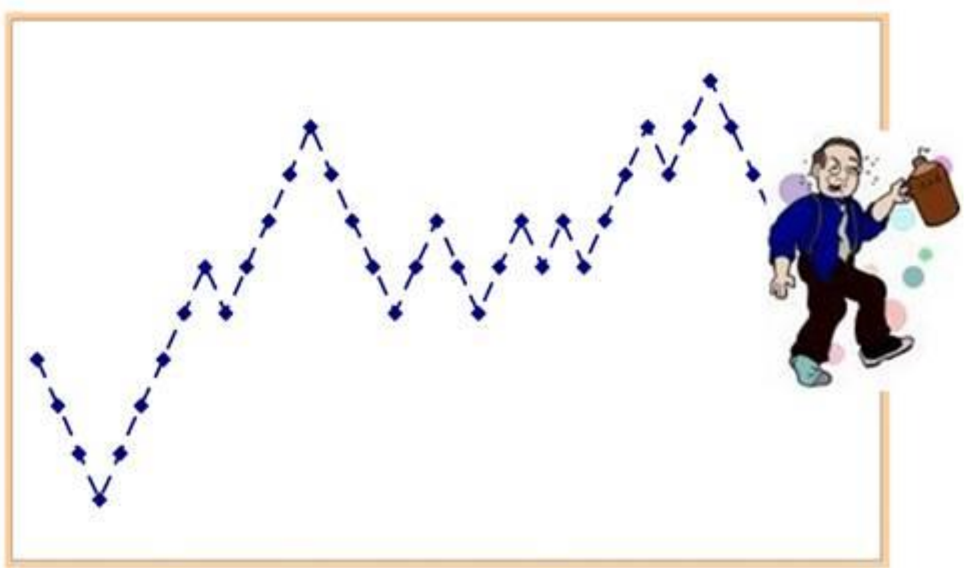
PROF. NICOLAI PETKOV

OUTLINE

- Problem statement
- Related work
- Normalization
- EMD
- Manifold Learning
- Results
- Conclusions

PROBLEM STATEMENT

PROBLEM STATEMENT



$$X(k+1) = X(k) + w(k)$$

- **Forecasting** problem
- Stock market prediction is considered a **challenging task** of time series prediction.
- **EMH**[1] states that stock market prices evolve according to a random walk and **cannot be forecasted**!

[1] J. Bouchaud, "The subtle nature of financial random walks", AIP, Chaos, 2005.

RELATED WORK

WHITENING & SCALING

- **Machine learning** techniques need large amount of **examples** to **generalize** well
- Training on samples of **different stocks** requires the **abstraction** from the specific time series
  - Whitening and scaling [2]
  - Non-stationarity

[2] S. F. Crone, J. Guajardo, and R. Weber, "The Impact of Preprocessing on Support Vector Regression and Neural Networks in Time Series Prediction", DMIN, 2006.

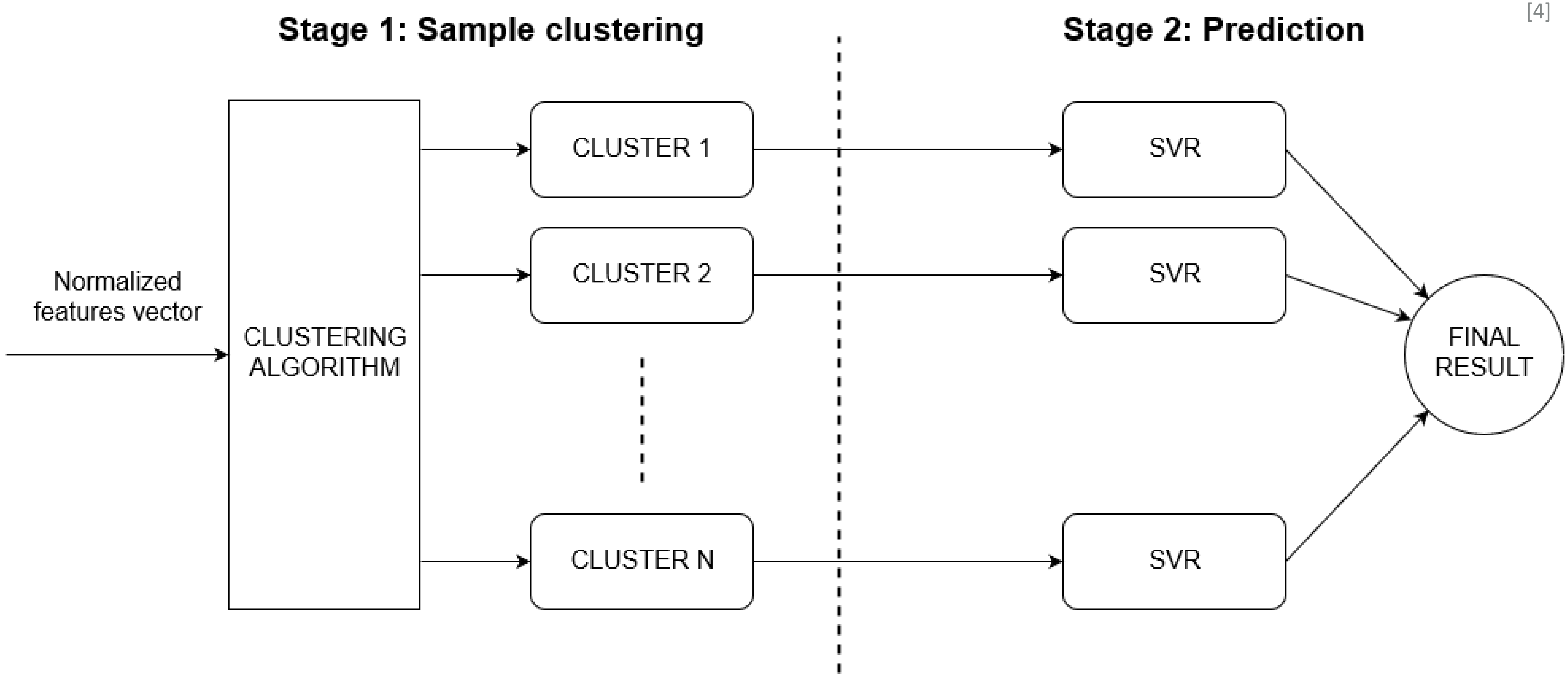
SIGNAL PROCESSING

- Transformation into **other domains**
  - Fourier and Wavelet
    - **Features extraction**
    - **Prediction** of each **component**
- **Symbolic** representation and **grammatical inference** [3]

[3] S. L. e. A. C. T. C. L. Giles, "Noisy Time Series Prediction using Recurrent Neural Networks and Grammatical Inference", Machine Learning, Kluwer Academic Publishers, vol. 44, pp. 161–183, 2001.

TWO STAGE ARCHITECTURE

[4]



[4] G. Galatro, "A kernel-based clustering approach for trends modelling in financial time series", Master's thesis, University of Salerno, Dec. 2018.

METHODS

DATASET

- 13 stocks in the banking field
- Fixed feature vector

Feature Vector	$(X_{-20} , X_{-15} , X_{-10} , X_{-5} , X_{-1} , X_0)$
Prediction Vector	$(X_3 , X_5 , X_{10} , X_{15} , X_{20})$

NORMALIZATION

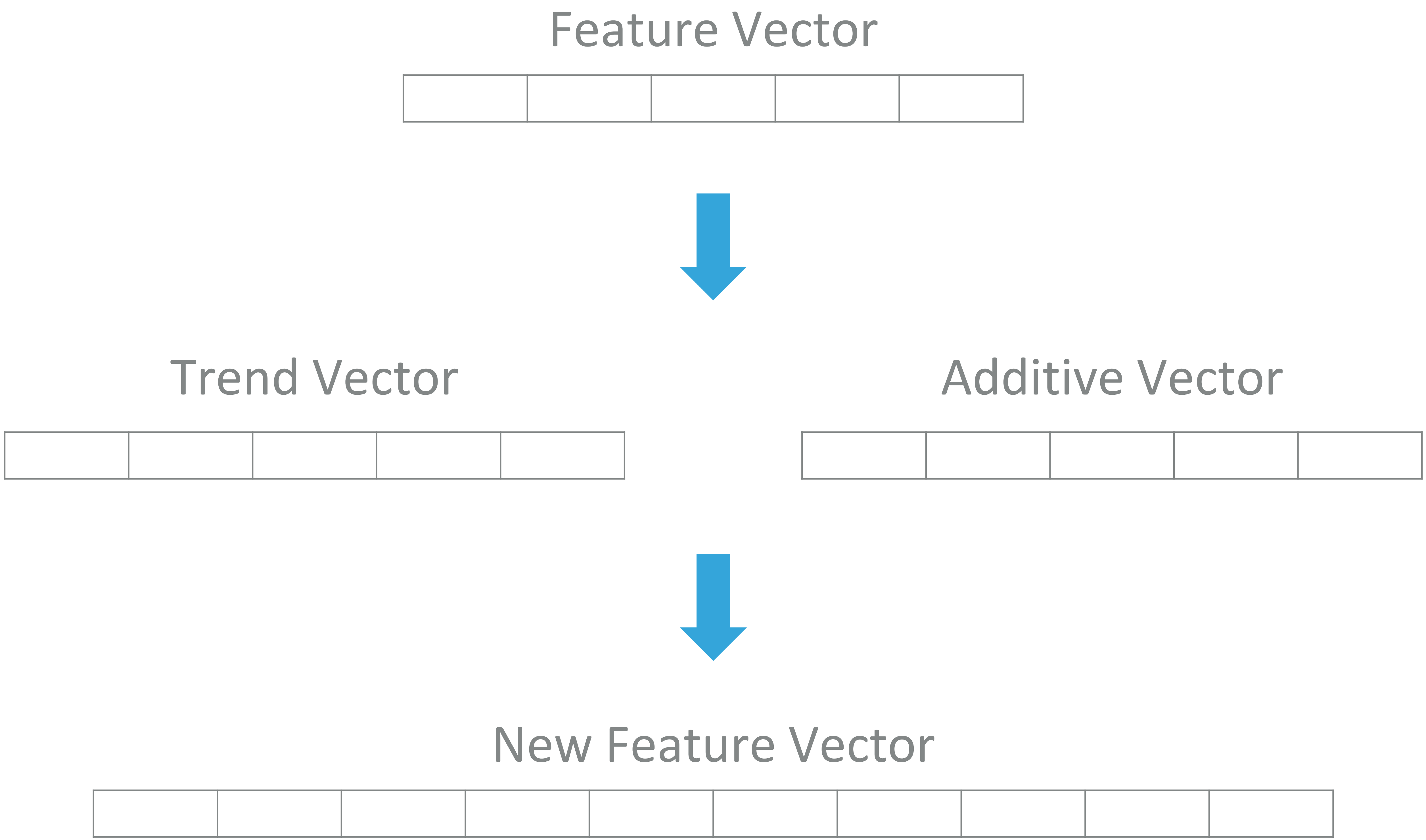
$$X'_i = \frac{X_i - X_0}{X_i} * 100$$

- Based on the **most recent price X(0)** in the features vector
- **Local** normalization
- Highlights **patterns** in terms of **relative changes**

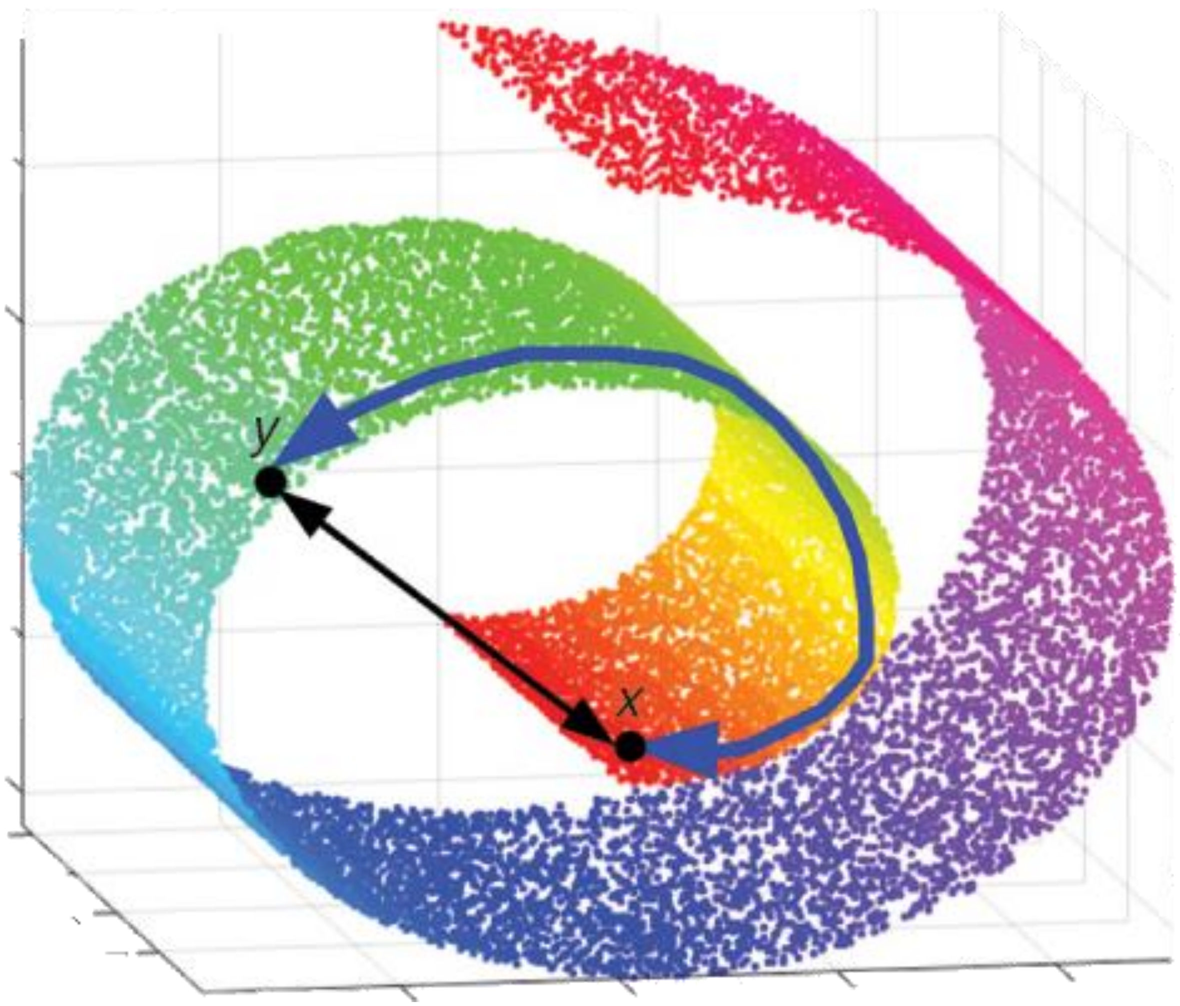
EMPIRICAL MODE DECOMPOSITION

- Decomposition in IMF (**Intrinsic mode functions**)
  - “An IMF is a function that has only one extreme between zero crossings, along with a mean value of zero.”
- Adaptive (Non-stationarity)
- Empirical

EMPIRICAL MODE DECOMPOSITION

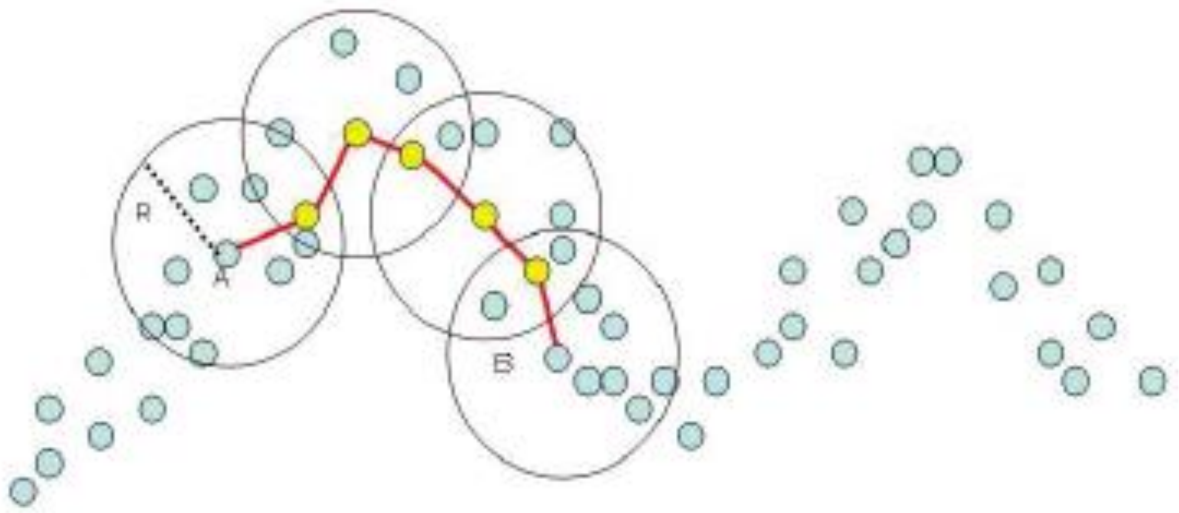


MANIFOLD



- A manifold is a **space** that are **locally Euclidean**
- Samples with **low Euclidean distance** may be **very different** from each other

ISOMAP – GEODESIC GRAPH



- Connections
  - **K-Neighbors** based
  - **Radius** based

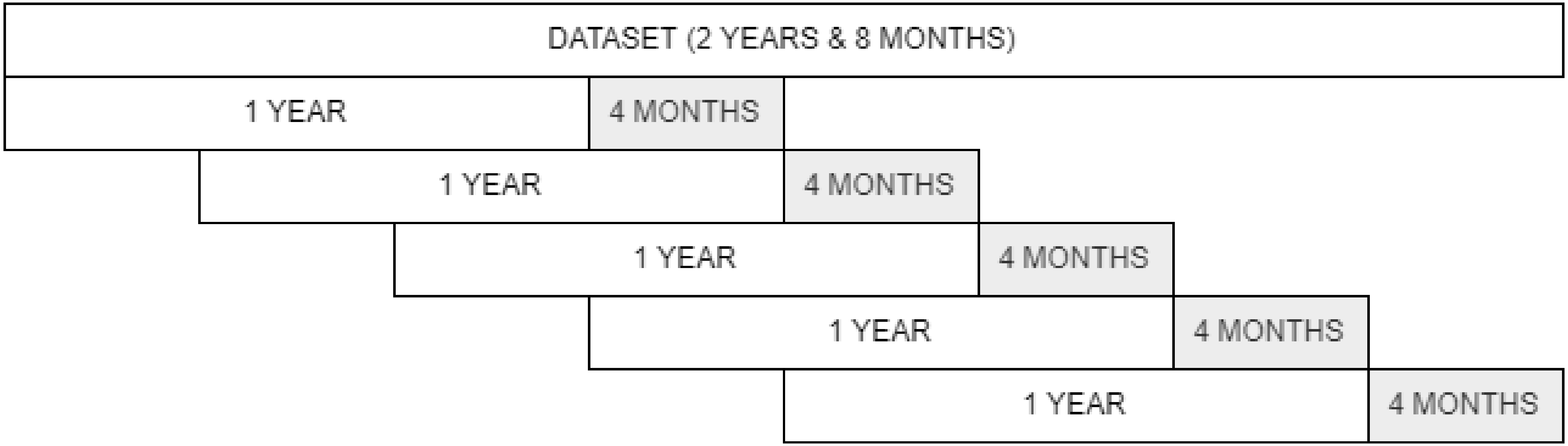
ISOMAP – PROJECTION

$$\Delta := \begin{pmatrix} \delta_{1,1} & \delta_{1,2} & \dots & \delta_{1,I} \\ \delta_{2,1} & \delta_{2,2} & \dots & \delta_{2,I} \\ \vdots & \vdots & & \vdots \\ \delta_{I,1} & \delta_{I,2} & \dots & \delta_{I,I} \end{pmatrix}$$

$$\min_{x_1, \ \dots \ , x_I} (||x_i - x_j|| - \delta_{i,j})^2$$

- **Geodesic** distance matrix
  - **Shortest path**
- **Non-linear** projection
  - **Multi-Dimensional Scaling**

PERFORMANCE ESTIMATION



- **R^2** related to the **EMH**
- **K-Fold** Cross Validation For **Time Series**
- P-Value as **degree of confidence**

RESULTS

EXPERIMENTS

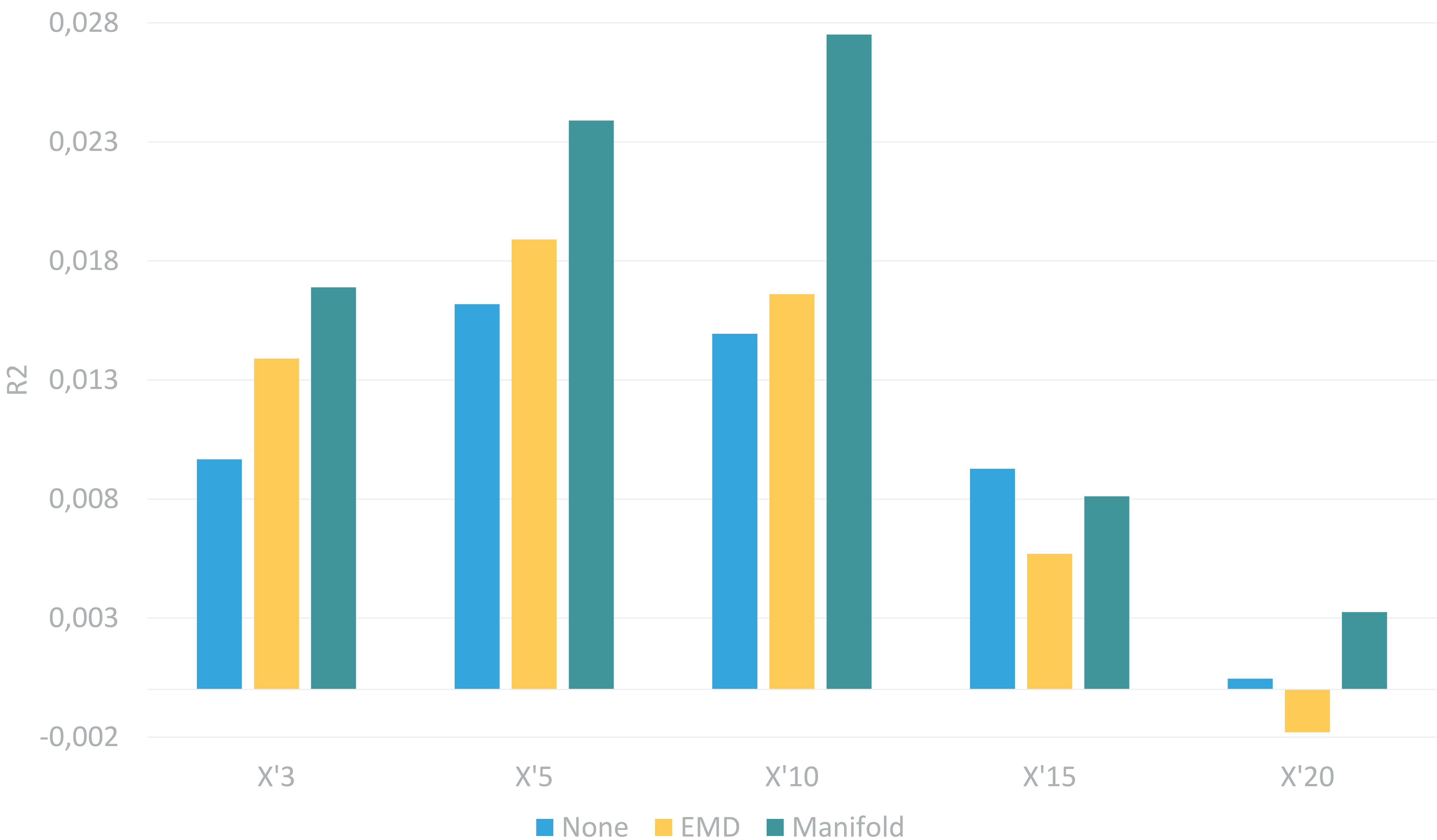
- Normalization
  - **Train** of SVR on sample of **each stock**
  - Normalization **related** to **S&P500**
- Effect of **preprocessing** techniques on the **two-stage architecture**
  - **Train** on sample of **different stocks**

NORMALIZATION

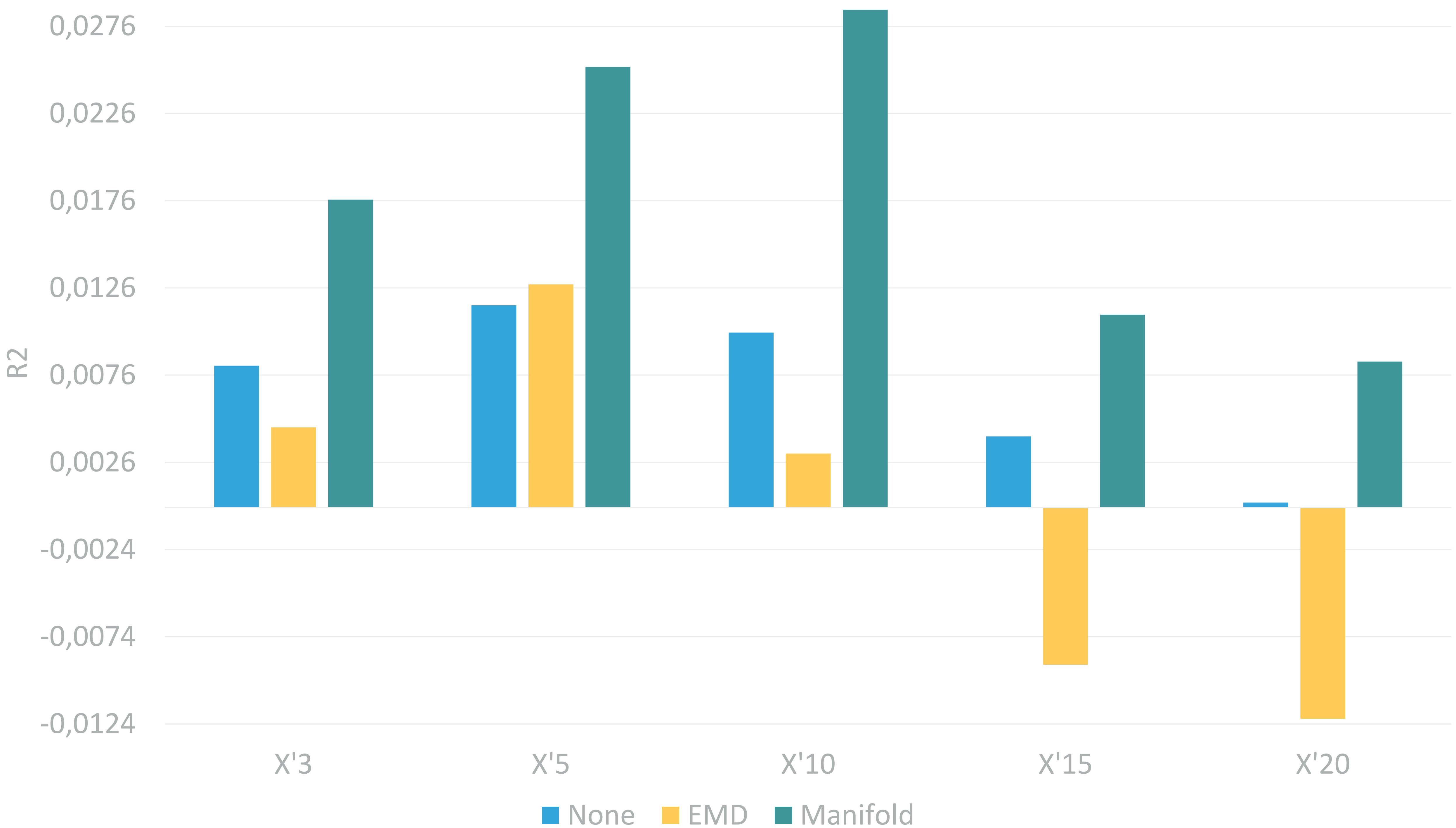
CONFIDENCE

Prediction	Normalized	Normalized with index relation	P-Value
X'_3	**0,00137**	-0,01241	0,0048
X'_5	**0,00415**	-0,03273	0,0002
X'_{10}	**-0,01127**	-0,07774	0,0002
X'_{15}	**-0,03741**	-0,11889	0,0006
X'_{20}	**-0,04559**	-0,12104	0,013

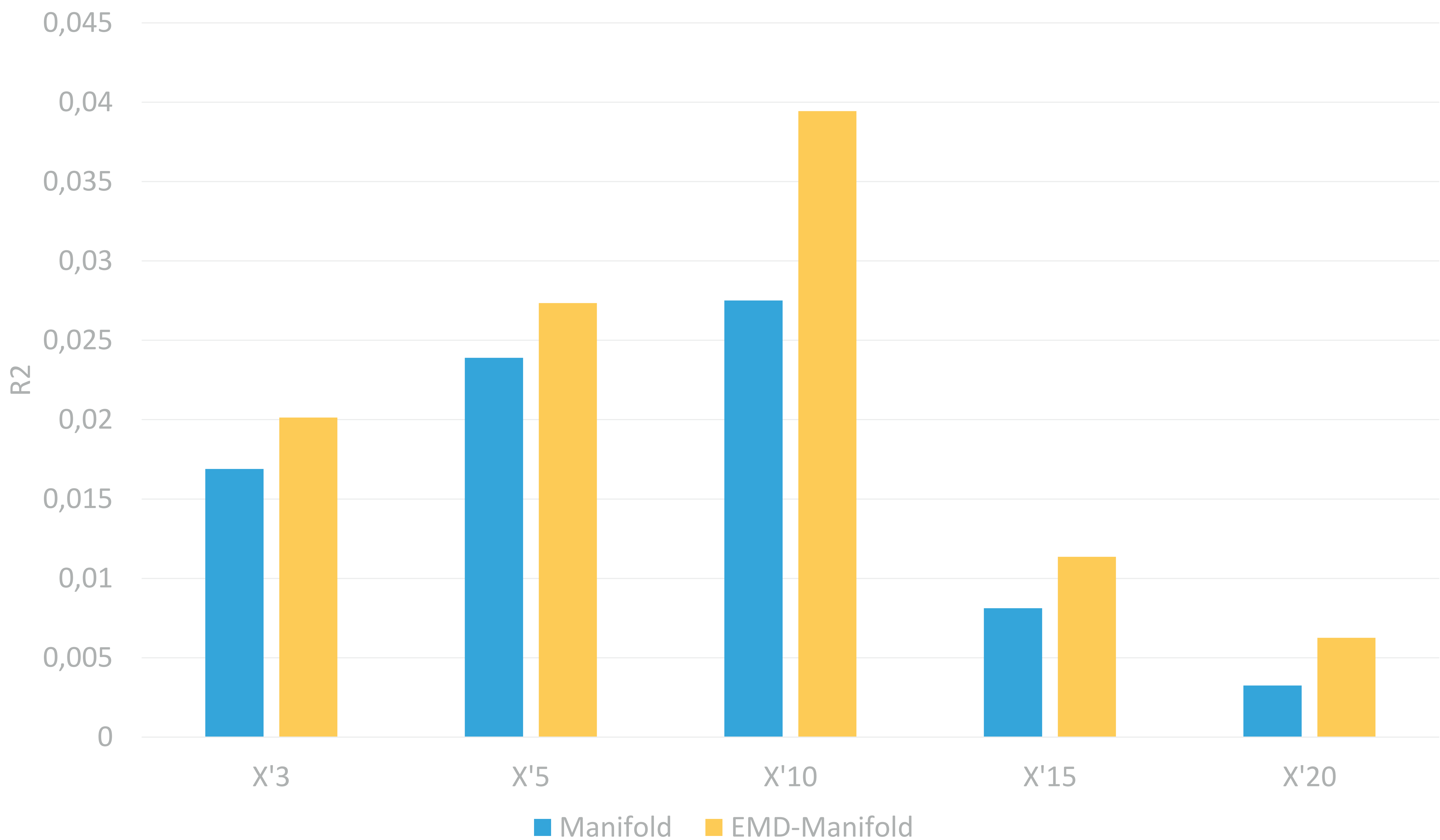
PREPROCESSING - KMEANS



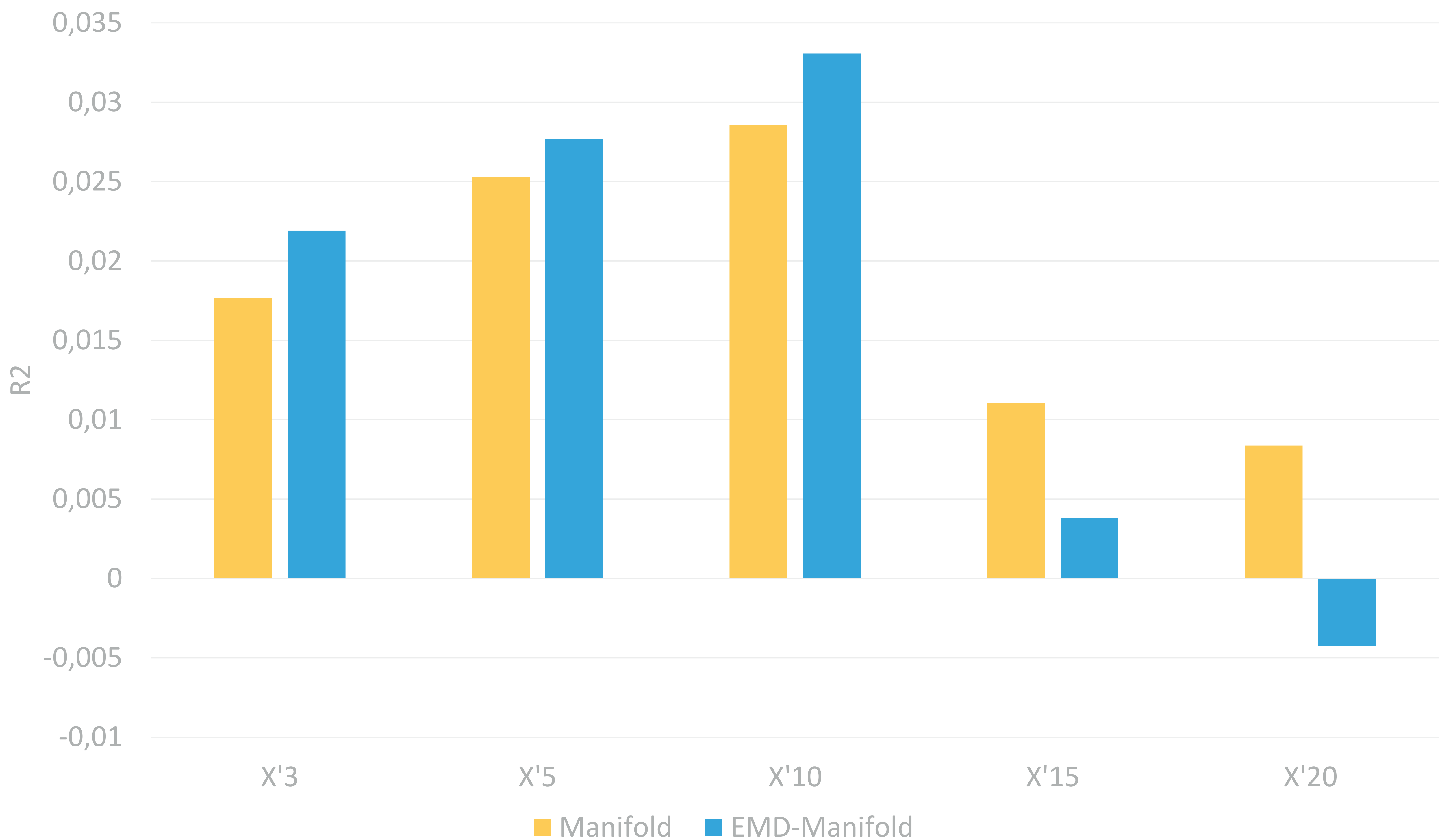
PREPROCESSING - EM



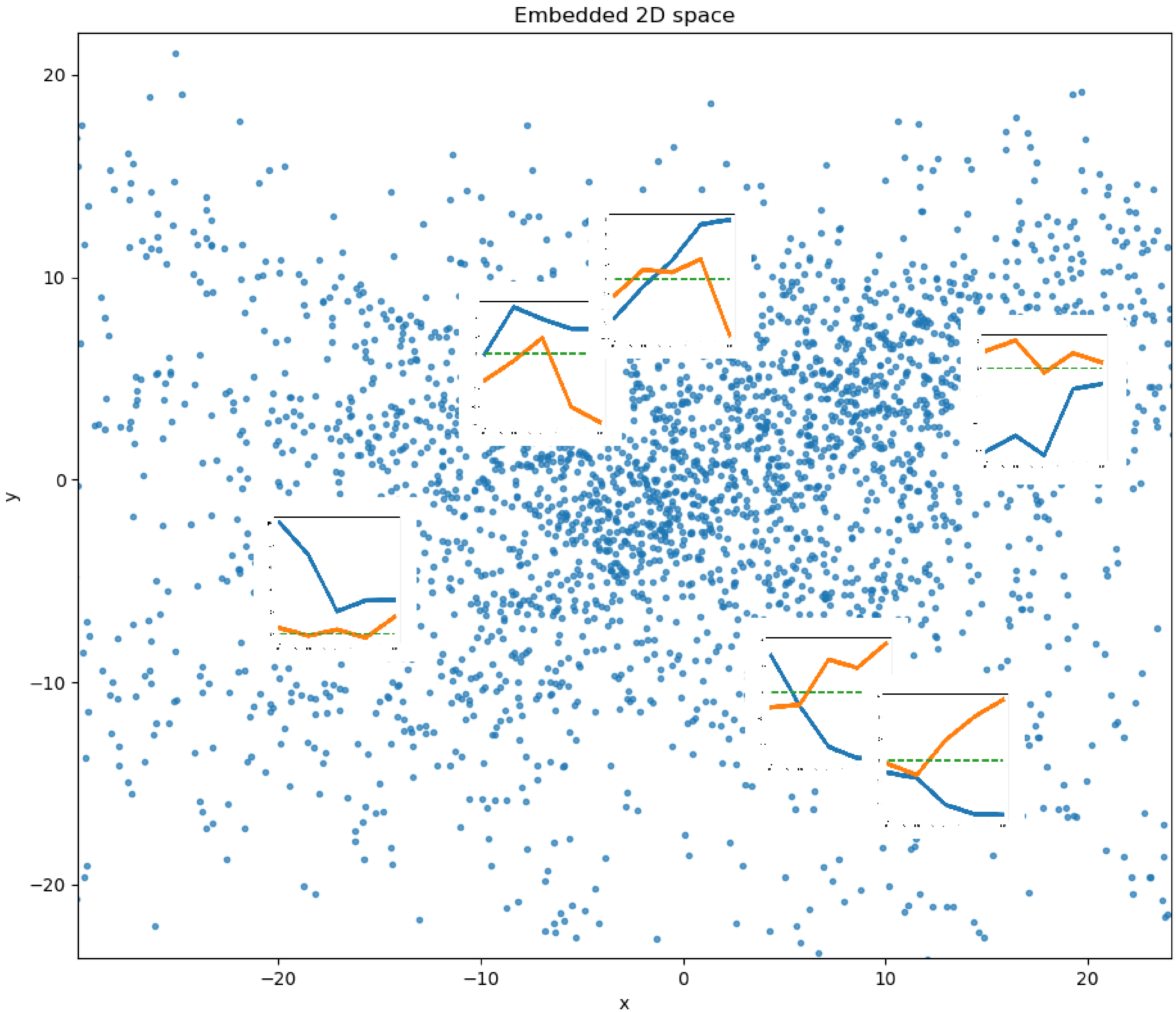
STACKED SOLUTION - KMEANS



STACKED SOLUTION - EM



DISCUSSION - MANIFOLD SPACE



CONCLUSIONS

CONCLUSIONS

- The use of **normalization** allows to highlights **trends** in term of relative change which **repeat over time** with **different absolute values** for the property of non-stationarity.
- The **stacked solution** can increase the performance of a two-stage architecture by reducing the **error contribute** due to the **non-stationarity** of the financial time series.

CONCLUSIONS

- Positive performance respect to the baseline
- Fast development

VS

- Fixed feature vector
- Hand-crafted features


LIVE DEMO

THANKS!
ANY QUESTIONS?