Towards Multivariate and Multi-step-ahead Time Series Forecasting: A Machine Learning Approach

Ph.D. Thesis - Public Defense

Candidate: Jacopo De Stefani
Supervisor: Prof. Gianluca Bontempi

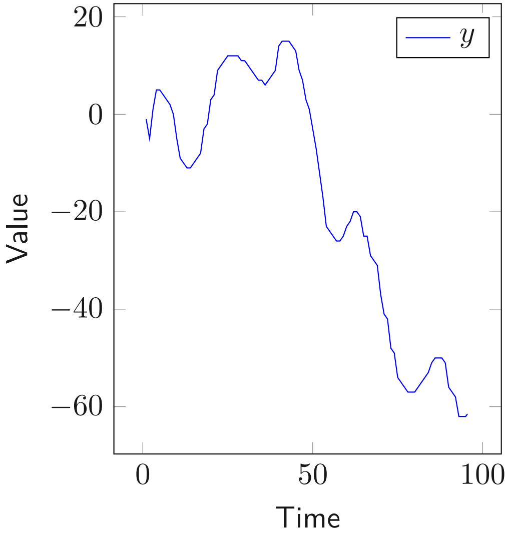
Université Libre de Bruxelles

A.A. 2021-2022
Friday 25th February, 2022

Introduction

ULB

Univariate time series



t	$\mathbf{y}_t$
0	-1
1	5
2	1
3	5
4	5
5	4
6	3
7	2
⋮	⋮
N	y_N

Univariate time series forecasting

UNIVERSITÉ LIBRE DE BRUXELLES

Definition

Given a univariate time series $\{y_1, \cdots, y_T\}$ comprising T observations, forecast the next H observations $\{y_{T+1}, \cdots, y_{T+H}\}$ where H is the forecast horizon.

Hypotheses:

- **Autoregressive:** There is a dependency of the future values on the past value
- **Model error:** There is an unpredictable error in the model

Univariate time series forecasting

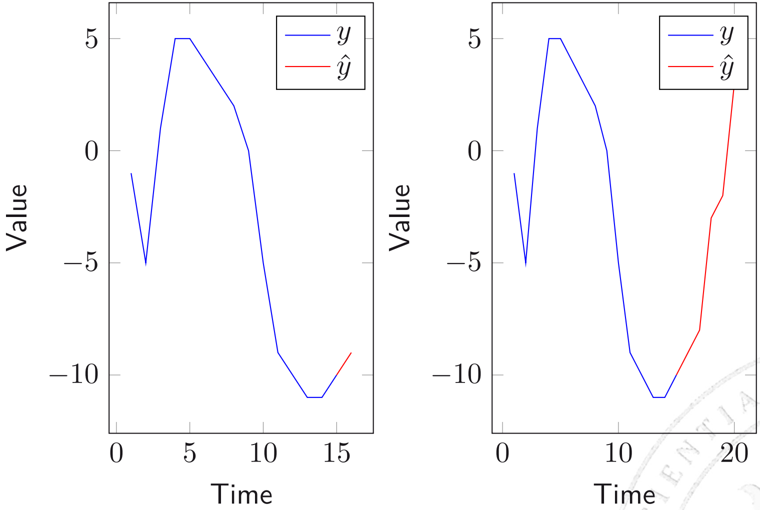
Definition

Given a univariate time series $\{y_1, \cdots, y_T\}$ comprising T observations, forecast the next H observations $\{y_{T+1}, \cdots, y_{T+H}\}$ where H is the forecast horizon.

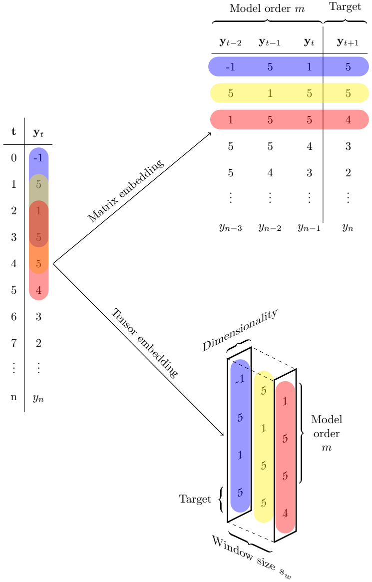
Hypotheses:

- **Autoregressive:** $y_t = f(y_{t-1}, \cdots, y_{t-d}) + \varepsilon_t$ with lag order d
- **Model error:** a stochastic iid model with $\mu_\varepsilon = 0$ and $\sigma_\varepsilon^2 = \sigma^2$ $\forall t$

One-step-ahead vs Multiple-step-ahead



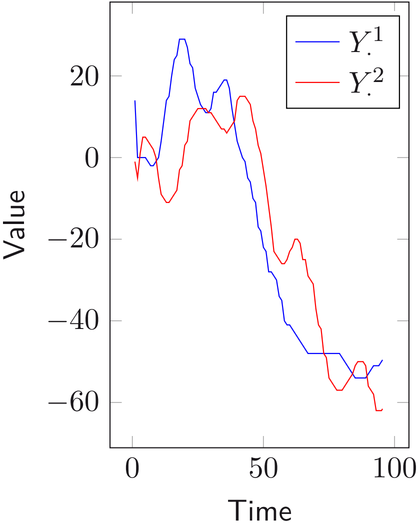
Embedding process - 2D vs 3D embedding



Embedding process - Terminology

$$
\begin{array}{r}
\\ \\ \\
\text{Sample}\left\{\vphantom{x_{11}}\right.
\end{array}
\begin{array}{ccccc|c}
\multicolumn{5}{c}{\overbrace{\hspace{12em}}^{\text{Features/Variables}}} & \overbrace{\hspace{2em}}^{\text{Target}} \\
\mathbf{X}_1 & \cdots & \mathbf{X}_i & \cdots & \mathbf{X}_{n'} & \mathbf{y} \\
\hline
x_{11} & \cdots & x_{1i} & \cdots & x_{1n'} & y_1 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
x_{k1} & \cdots & x_{ki} & \cdots & x_{kn'} & y_k \\
\hdashline
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
x_{N1} & \cdots & x_{Ni} & \cdots & x_{Nn'} & y_N
\end{array}
\begin{array}{l}
\\ \\
\left.\vphantom{\begin{array}{c}x\\x\\x\\x\end{array}}\right\}\text{Training set } \mathcal{D}_{train} \\
\left.\vphantom{\begin{array}{c}x\\x\end{array}}\right\}\text{Testing set } \mathcal{D}_{test}
\end{array}
$$

Multivariate time series

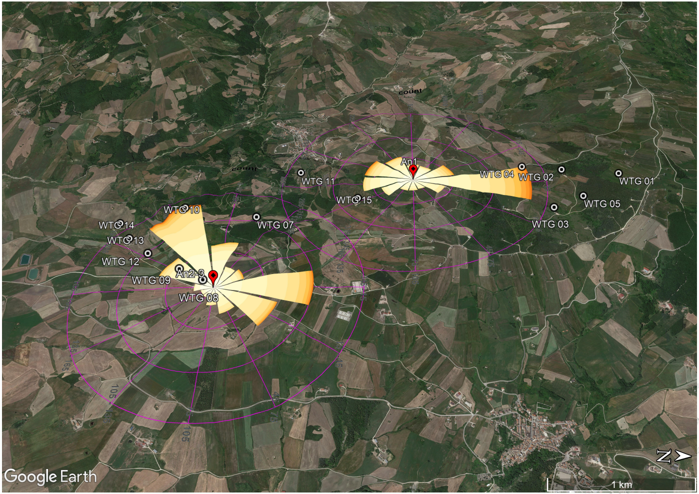


$\longrightarrow$

t	$\mathbf{Y}_t^1$	$\mathbf{Y}_t^2$
1	14	-1
2	0	-5
3	0	1
4	0	5
5	0	5
6	-1	4
7	-2	3
$\vdots$	$\vdots$	$\vdots$
N	Y_N^1	Y_N^2

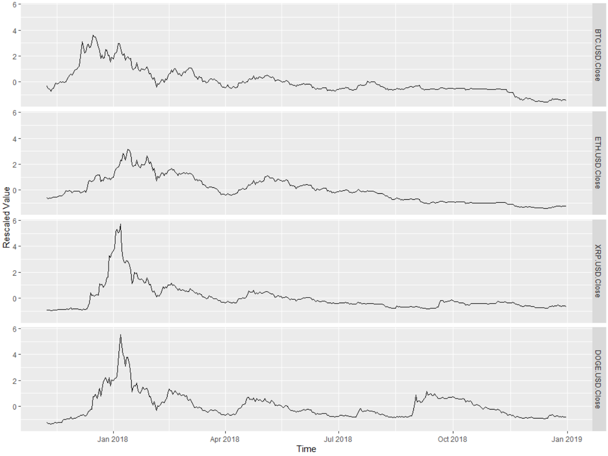
Multivariate time series - Wind Power Forecasting

Multivariate time series - Wind Power Forecasting

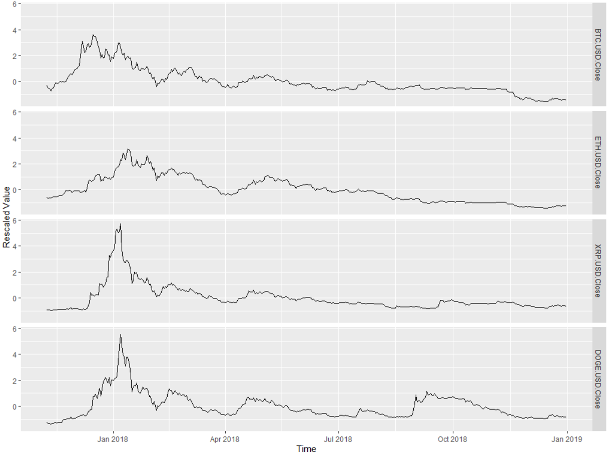


- Collaboration with UniSannio (Dr. De Caro and Prof. Vaccaro)
- Italian and Australian Wind Farms

Multivariate time series - Volatility Forecasting

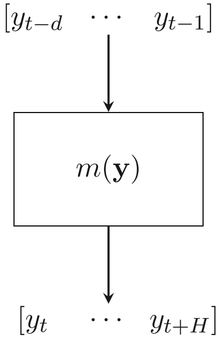


Multivariate time series - Volatility Forecasting



- Collaboration with Atos-Wordline (Dr. Caelen and Dr. Hattab)
- CAC40 and Cryptocurrency data

Univariate Multi-step-ahead forecasting strategies - prof. Ben Taieb PhD Thesis



1 TS Input
1 TS Output

Recursive strategy

- Single model one-step-ahead recursively employed H times.

Direct strategy

- H models one-step-ahead, one for each forecasting horizon.

Joint strategy

- Single model, producing a vector of H predictions at once

Contributions

Thesis Contributions

- Overview of the **state-of-the-art** for multivariate problems (Chapter 2 - Sections 2.3.4 and 2.3.5).
- Formalization of the multivariate and multi-step-ahead forecasting problem (Chapter 3).
- Two novel forecasting multivariate and multi-step-ahead strategies: DFML and SMURF-ES (Chapter 3), based on traditional machine learning approaches.
- The empirical assessment of the DFML (Chapter 4) and SMURF-ES (Chapter 5) strategies on several real-life challenging forecasting tasks

Thesis Contributions

- Overview of the state-of-the-art for multivariate problems (Chapter 2 - Sections 2.3.4 and 2.3.5).
- **Formalization of the multivariate and multi-step-ahead forecasting problem** (Chapter 3).
- Two novel forecasting multivariate and multi-step-ahead strategies: DFML and SMURF-ES (Chapter 3), based on traditional machine learning approaches.
- The empirical assessment of the DFML (Chapter 4) and SMURF-ES (Chapter 5) strategies on several real-life challenging forecasting tasks

Thesis Contributions

- Overview of the state-of-the-art for multivariate problems (Chapter 2 - Sections 2.3.4 and 2.3.5).
- Formalization of the multivariate and multi-step-ahead forecasting problem (Chapter 3).
- **Two novel** forecasting multivariate and multi-step-ahead strategies: DFML and SMURF-ES (Chapter 3), based on traditional machine learning approaches.
- The empirical assessment of the DFML (Chapter 4) and SMURF-ES (Chapter 5) strategies on several real-life challenging forecasting tasks

Thesis Contributions

- Overview of the state-of-the-art for multivariate problems (Chapter 2 - Sections 2.3.4 and 2.3.5).
- Formalization of the multivariate and multi-step-ahead forecasting problem (Chapter 3).
- Two novel forecasting multivariate and multi-step-ahead **strategies: DFML and SMURF-ES** (Chapter 3), based on traditional machine learning approaches.
- The empirical assessment of the DFML (Chapter 4) and SMURF-ES (Chapter 5) strategies on several real-life challenging forecasting tasks

Thesis Contributions

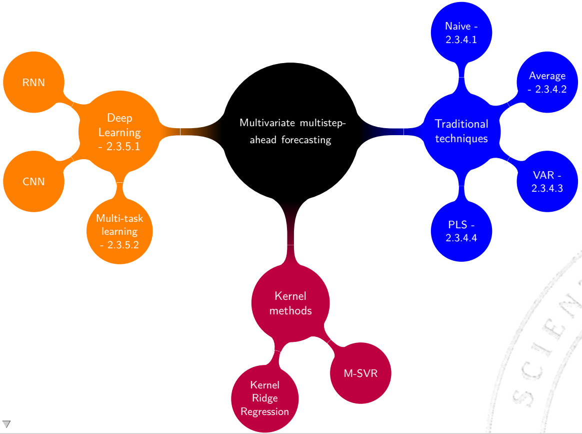
- Overview of the state-of-the-art for multivariate problems (Chapter 2 - Sections 2.3.4 and 2.3.5).
- Formalization of the multivariate and multi-step-ahead forecasting problem (Chapter 3).
- Two novel forecasting multivariate and multi-step-ahead strategies: DFML and SMURF-ES (Chapter 3), based on traditional machine learning approaches.
- The **empirical assessment of the DFML** (Chapter 4) and **SMURF-ES** (Chapter 5) strategies **on** several **real-life challenging forecasting tasks**

Thesis Contributions

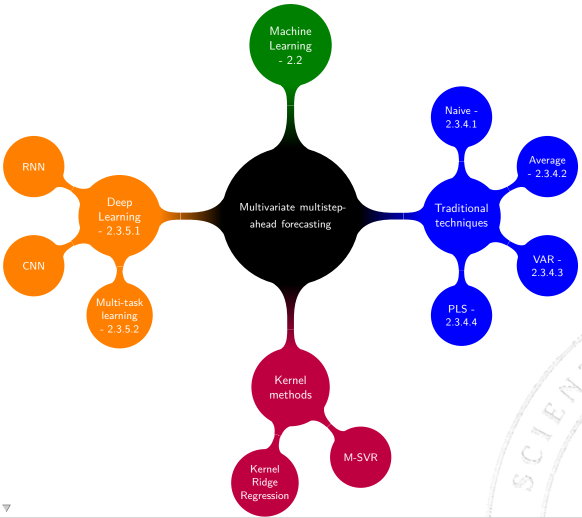
- Overview of the **state-of-the-art** for multivariate problems (Chapter 2 - Sections 2.3.4 and 2.3.5).
- **Formalization of the multivariate and multi-step-ahead forecasting problem** (Chapter 3).
- **Two novel** forecasting multivariate and multi-step-ahead **strategies: DFML and SMURF-ES** (Chapter 3), based on traditional machine learning approaches.
- The **empirical assessment of the DFML** (Chapter 4) and **SMURF-ES** (Chapter 5) strategies **on** several **real-life challenging forecasting tasks**

State-of-the-art

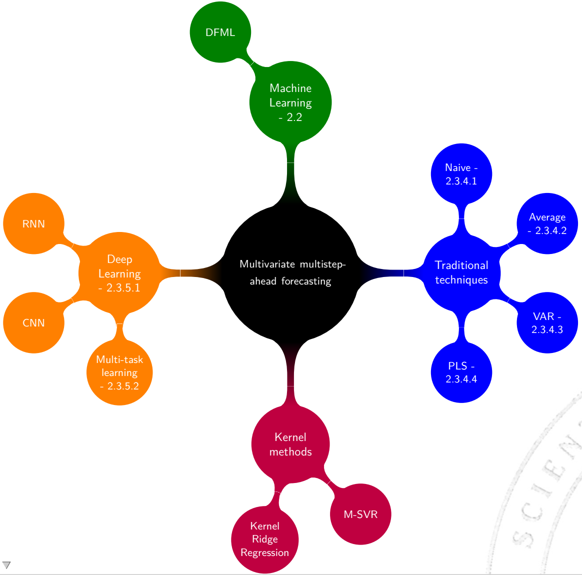
Literature review - Why Machine Learning?



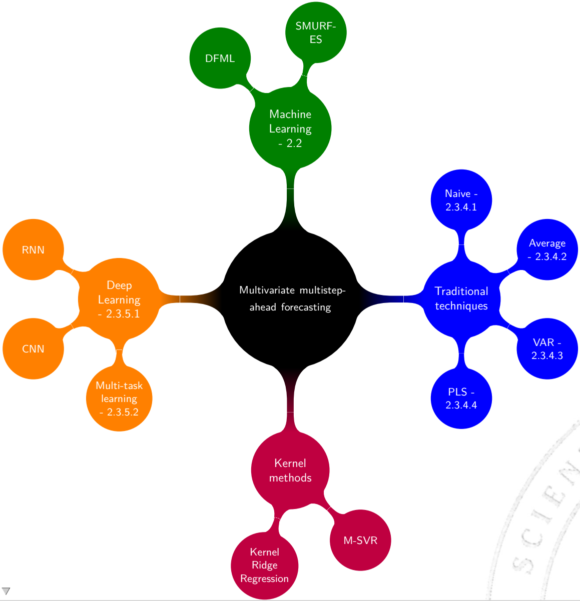
Literature review - Why Machine Learning?



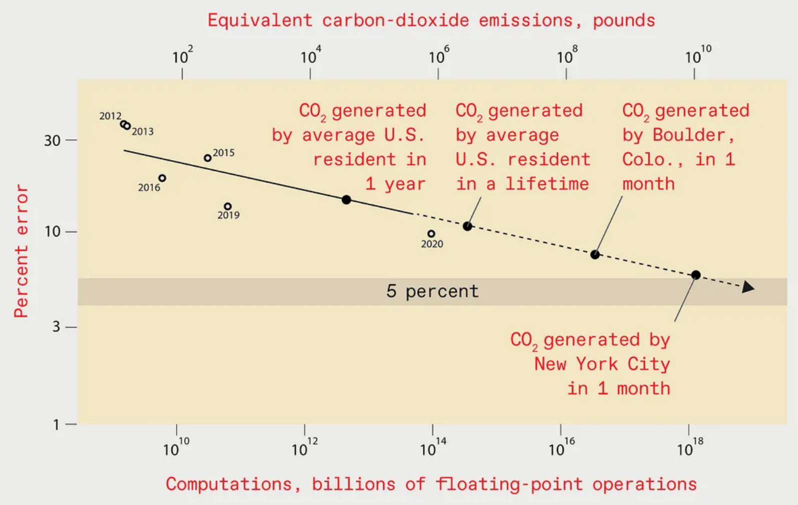
Literature review - Why Machine Learning?



Literature review - Why Machine Learning?



Literature review - Deep Learning Issues



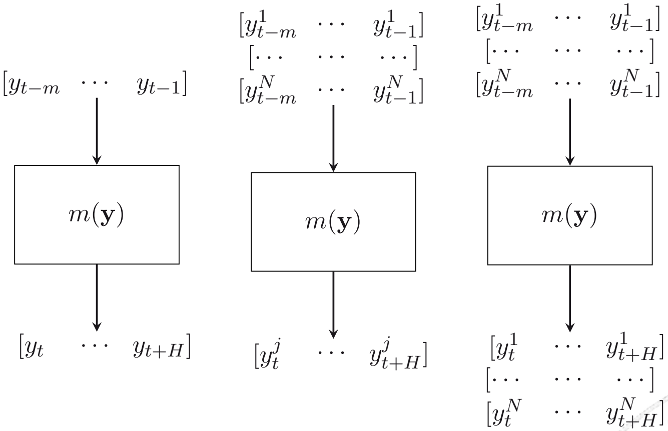
The white dots represent model sizes and performances of state-of-the-art models, with the corresponding years, while the black dots indicate potential future scenarios. - Source: (Deep Learning's Diminishing Returns 2021)

Contribution - Problem formalization

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

From univariate to multivariate

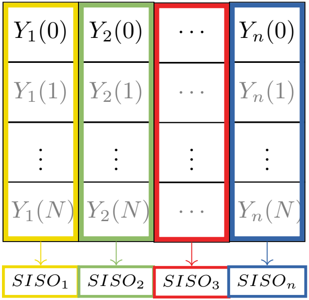


1 TS Input
1 TS Output
SISO

N **TS input**
1 TS output
MISO

M **TS input**
M **TS output**
MIMO

From univariate to multivariate - An overview

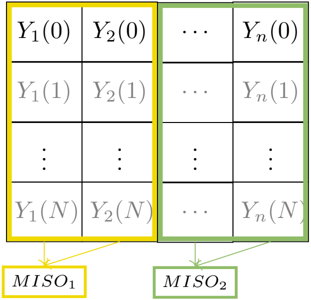


Multivariate - SISO tasks decomposition - Section 3.1.1

$$\begin{cases} Y_{t+1}[1] & = f_1^{(NAR)}(Y_t[1], \ldots, Y_{t-m+1}[1]) + e_{t+1}[1] \\ \vdots & \\ Y_{t+1}[n] & = f_n^{(NAR)}(Y_t[n], \ldots, Y_{t-m+1}[n]) + e_{t+1}[n] \end{cases}$$

where each function $f_i^{(NAR)} : \mathbb{R}^m \mapsto \mathbb{R}$, $i = 1, \ldots, n$, represents a non-linear autoregressive model NAR with model order m.

From univariate to multivariate - An overview



Multivariate - MISO tasks decomposition - Section 3.1.1

$$
\begin{cases}
Y_{t+1}[1] & = f_1^{(NARX)}(Y_t[1], \ldots, Y_{t-m+1}[1], \ldots, \\
 & \quad Y_t[n], \ldots, Y_{t-m+1}[n_i]) + e_{t+1}[1] \\
\vdots & \\
Y_{t+1}[n] & = f_n^{(NARX)}(Y_t[1], \ldots, Y_{t-m+1}[1], \ldots, \\
 & \quad Y_t[n], \ldots, Y_{t-m+1}[n]) + e_{t+1}[n]
\end{cases}
$$

where each function $f_i^{(NARX)} : \mathbb{R}^{m \times n} \mapsto \mathbb{R}$, $i = 1, \ldots, n$, represents a non-linear autoregressive model with model order m and up to n external regressors.

From univariate to multivariate - An overview

$Y_1(0)$	$Y_2(0)$	$\cdots$	$Y_n(0)$
$Y_1(1)$	$Y_2(1)$	$\cdots$	$Y_n(1)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$Y_1(N)$	$Y_2(N)$	$\cdots$	$Y_n(N)$

$MIMO$

Multivariate - One-step-ahead MIMO - Section 3.1.2

$$\mathbf{Y}_{t+1} = F\left(\mathbf{Y}_{t-d}, \mathbf{Y}_{t-d-1}, \ldots, \mathbf{Y}_{t-d-m+1}\right) + \mathbf{E}_{t+1}$$

$F : \mathbb{R}^{n\times m} \mapsto \mathbb{R}^n$ takes the embedding vectors of size m of the n time series as input and produces the one-step-ahead forecasts.

From univariate to multivariate - An overview

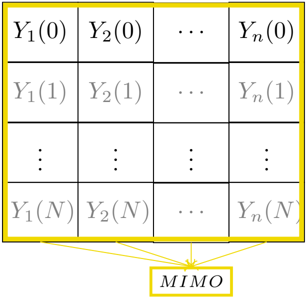
$Y_1(0)$	$Y_2(0)$	$\cdots$	$Y_n(0)$
$Y_1(1)$	$Y_2(1)$	$\cdots$	$Y_n(1)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$Y_1(N)$	$Y_2(N)$	$\cdots$	$Y_n(N)$

$\boxed{MIMO}$

Multivariate - Multi-step-ahead MIMO - Iterated - Section 3.1.2

$$\mathbf{Y}_{t+1} = F_I(\mathbf{Y}_t, \ldots, \mathbf{Y}_{t-m+1}) + \mathbf{E}_{t+1}$$

A single one-step-ahead MIMO model $F_I : \mathbb{R}^{n \times m} \mapsto \mathbb{R}^n$ is iteratively re-used H times to produce the desired forecasts.



Multivariate - Multi-step-ahead MIMO - Direct - Section 3.1.2

$$\mathbf{Y}_{t+h} = F_h(\mathbf{Y}_t, \ldots, \mathbf{Y}_{t-m+1}) + \mathbf{E}_{t+h}$$

H separate MIMO models $F_h : \mathbb{R}^{n\times m} \mapsto \mathbb{R}^n$, each one producing the h-step ahead forecasts $\forall h \in \{1, \cdots, H\}$

From univariate to multivariate - An overview

$Y_1(0)$	$Y_2(0)$	$\cdots$	$Y_n(0)$
$Y_1(1)$	$Y_2(1)$	$\cdots$	$Y_n(1)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$Y_1(N)$	$Y_2(N)$	$\cdots$	$Y_n(N)$

$MIMO$

Multivariate - Multi-step-ahead MIMO - MIMO/Joint - Section 3.1.2

$$\begin{bmatrix} \mathbf{Y}_{t+H} & \cdots & \mathbf{Y}_{t+1} \end{bmatrix} = F_{JM}(\mathbf{Y}_t, \ldots, \mathbf{Y}_{t-m+1}) + \mathbf{E}$$

$F_{JM} : \mathbb{R}^{n \times m} \mapsto \mathbb{R}^{H \times n}$ being the model jointly providing H-step ahead forecasts for all the n time series.

Contribution - Forecasting strategies

DFML - Dynamic Factor Machine Learner

Dynamic Factor Models - Prior art

- Generalized Dynamic Factor Models are a well-known multivariate econometric model (Stock and Watson, 2010):

$$\mathbf{Y}_{t+1} = \mathbf{W}\mathbf{Z}_{t+1} + \mathbf{E}_{Y,t+1}$$

$$\mathbf{Z}_{t+1} = \sum_{i=0}^{m-1} (\mathbf{A}_{t-i}\mathbf{Z}_{t-i}) + \mathbf{E}_{Z,t+1}$$

- **Key idea:** A small number of (latent) factors: $\mathbf{Z}_t$ account for the dynamics of the original series $\mathbf{Y}_t$

Dynamic Factor Models - Prior art

- Generalized Dynamic Factor Models are a well-known multivariate econometric model (Stock and Watson, 2010):

$$\mathbf{Y}_{t+1} = \mathbf{W}\mathbf{Z}_{t+1} + \mathbf{E}_{Y,t+1}$$

$$\mathbf{Z}_{t+1} = \sum_{i=0}^{m-1} (\mathbf{A}_{t-i}\mathbf{Z}_{t-i}) + \mathbf{E}_{Z,t+1}$$

- **Key idea:** A small number of (latent) factors: $\mathbf{Z}_t$ account for the dynamics of the original series $\mathbf{Y}_t$
- Two subproblems:
  - Factor estimation
  - Factor forecasting

Dynamic Factor Models - Prior art

- Generalized Dynamic Factor Models are a well-known multivariate econometric model (Stock and Watson, 2010):

$$\mathbf{Y}_{t+1} = \mathbf{W}\mathbf{Z}_{t+1} + \mathbf{E}_{Y,t+1}$$
$$\mathbf{Z}_{t+1} = \sum_{i=0}^{m-1}(\mathbf{A}_{t-i}\mathbf{Z}_{t-i}) + \mathbf{E}_{Z,t+1}$$

- **Key idea:** A small number of (latent) factors: $\mathbf{Z}_t$ account for the dynamics of the original series $\mathbf{Y}_t$
- Two subproblems:
  - Factor estimation $\Rightarrow$ PCA (Linear)
  - Factor forecasting

Dynamic Factor Models - Prior art

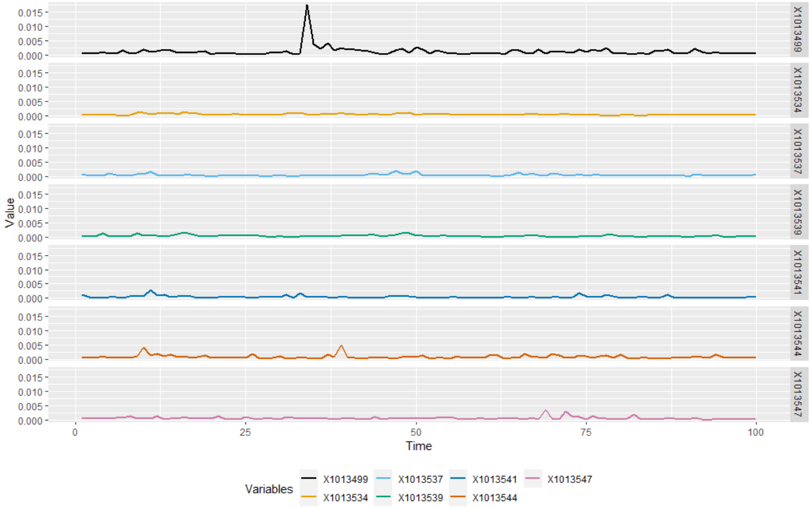
- Generalized Dynamic Factor Models are a well-known multivariate econometric model (Stock and Watson, 2010):

$$\mathbf{Y}_{t+1} = \mathbf{W}\mathbf{Z}_{t+1} + \mathbf{E}_{Y,t+1}$$

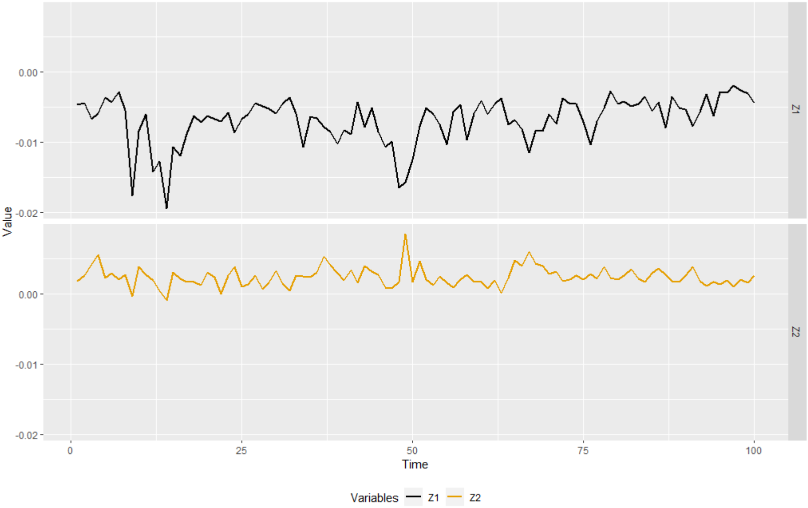
$$\mathbf{Z}_{t+1} = \sum_{i=0}^{m-1} (\mathbf{A}_{t-i}\mathbf{Z}_{t-i}) + \mathbf{E}_{Z,t+1}$$

- **Key idea:** A small number of (latent) factors: $\mathbf{Z}_t$ account for the dynamics of the original series $\mathbf{Y}_t$
- Two subproblems:
  - Factor estimation
  - Factor forecasting $\Rightarrow$ VAR

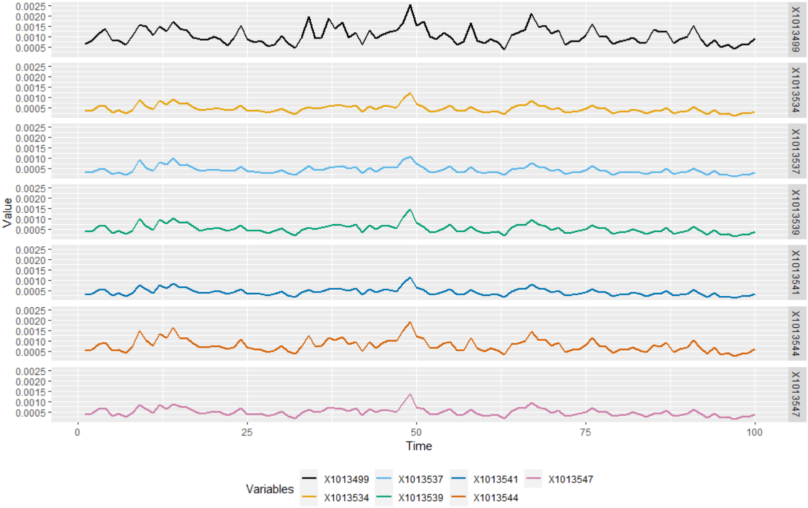
DFML strategy - PCA - Original series



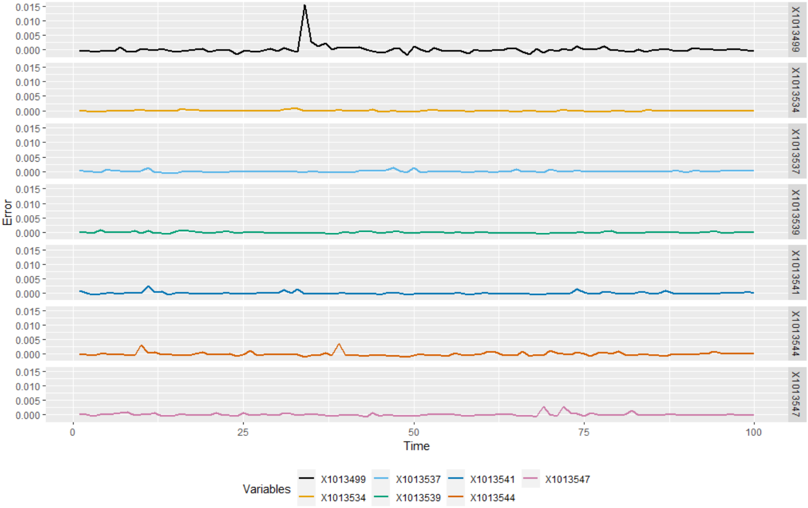
DFML strategy - PCA - Factors



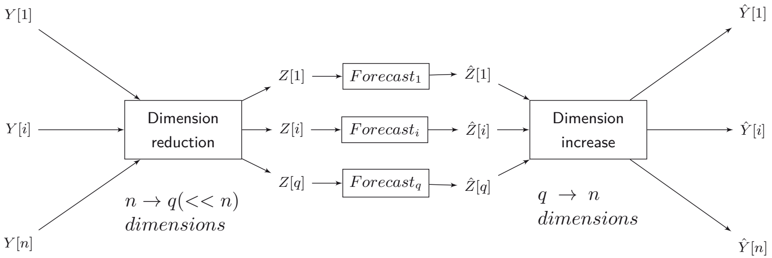
DFML strategy - PCA - Reconstructed series



DFML strategy - PCA - Reconstruction error

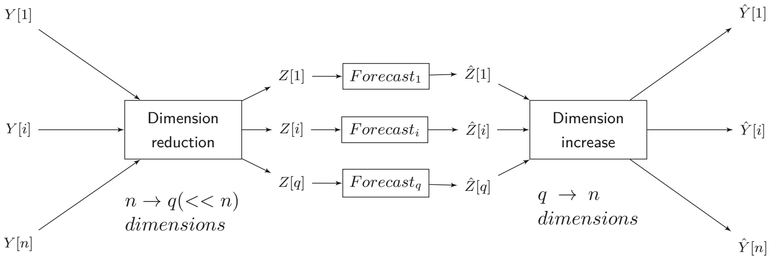


DFML strategy - Section 3.2



Factor estimation	Factor forecast
PCA	ES
Autoencoder (Shallow, Deep)	Theta
Recurrent autoencoder (LSTM,GRU)	Combined
	VAR
	Lazy Learning (Direct, Recursive, Joint)
	Gradient boosting (Direct, Recursive)

DFML strategy - Section 3.2



	Factor estimation	Factor forecast
DFM	PCA	VAR
DFML	PCA/NN/RNN	Statistical/ML-based

DFML contribution - Extensions

- **Incremental factor estimation - Section 3.2.4.1**
  - Introduction of incremental factor estimation techniques (i.e. various implementations of online PCA - Section 2.4.1.1)
- **Automatic hyperparameter search strategy - Section 3.2.4.2**
  - Automatic selection of the best combination of number of factors q and multi-step-ahead strategy a based on out-of-sample error.

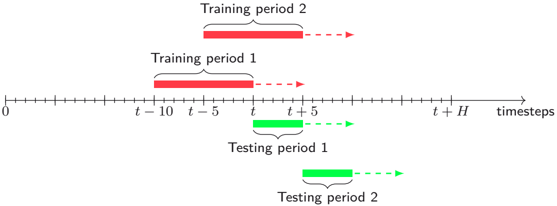
DFML - Experimental assessment

Experimental Assessment - Section 4.3 - 4.4 - 4.5

Assessment	Dataset	N	n
Factor estimation and forecasting	Electricity	26304	321
	Traffic	17544	862
	OBU Mobility Data	1416	389
Volatility forecasting	CAC40	1645	40
	Cryptocurrencies	495	291
Automatic hyperparameter search and iterative factor estimation	Synthetic	1500	20, 50, 100, 200, 400, 1000
	Earth Surface Temperature	1625	100, 200
	Volatility	1489	40

DFML contributions - Experimental Assessment - Section 4.3 - 4.4 - 4.5

- **Rolling window approach:** (Tashman, 2000)



- **Error metric:** Naive Normalized Mean Squared Error:

$$\text{NNMSE} = \frac{1}{n}\sum_{j=1}^{n} \text{NNMSE}[j]$$

$$\text{NNMSE}[j] = \frac{\frac{1}{H}\sum_{h=1}^{H}(\mathbf{Y}_{T+h}[j]-\hat{\mathbf{Y}}_{T+h}[j])^2}{\frac{1}{H}\sum_{h=2}^{H}(\mathbf{Y}_{T+h}[j]-\mathbf{Y}_{T+h-1}[j])^2}$$

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

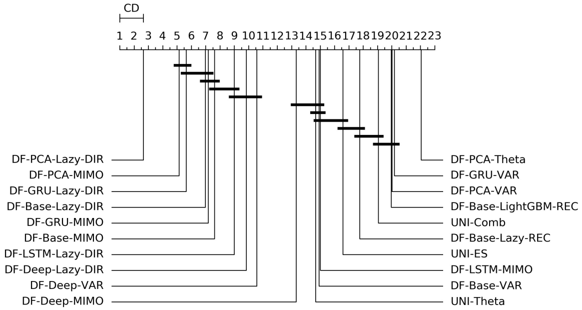
Factor estimation and forecasting assessment - Experimental Setup - Section 4.3

- **Strategy configuration:**

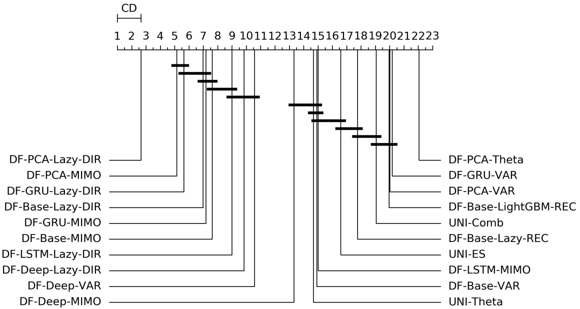
Factor estimation	Factor forecast
PCA	ES
Autoencoder (Shallow, Deep)	Theta
Recurrent autoencoder (LSTM,GRU)	Combined
	VAR
	Lazy Learning (Direct, Recursive, Joint)
	Gradient boosting (Direct, Recursive)

- **Forecasting horizon:** $H \in \{4, 6, 12, 24\}$ (hours).
- **Statistical assessment:** (Demšar, 2006) of the results of Friedman statistical test (with post-hoc Nemenyi test). The methods are ordered according to their performance from left to right (the leftmost the best), while the black bar connects methods that are not significantly different (at $p = 0.05$).

Factor estimation and forecasting assessment - Traffic Data - Accuracy - Section 4.3



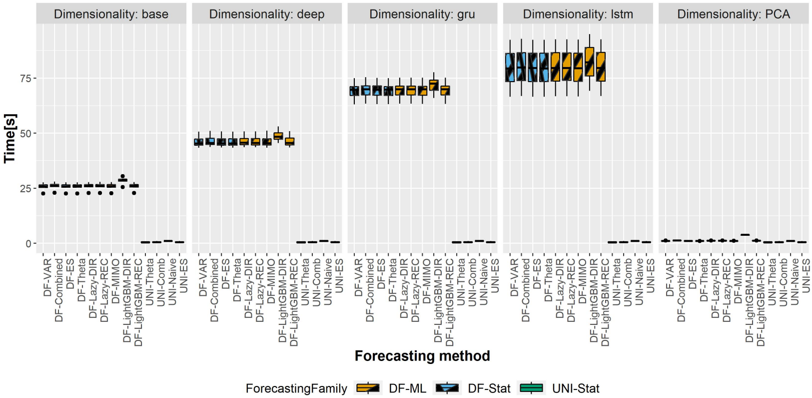
Factor estimation and forecasting assessment - Traffic Data - Accuracy - Section 4.3



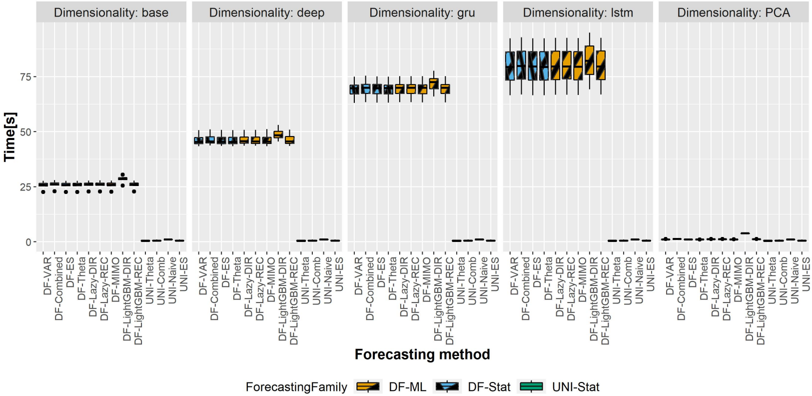
- DFML (DF-) approaches outperform univariate techniques (UNI)

- Non-linear factor estimations (Base, Deep, LSTM, GRU) obtain competitive performances
- PCA + Lazy is consistently among the top performers

Factor estimation and forecasting assessment - Traffic Data - Computational time - Section 4.3



Factor estimation and forecasting assessment - Traffic Data - Computational time - Section 4.3



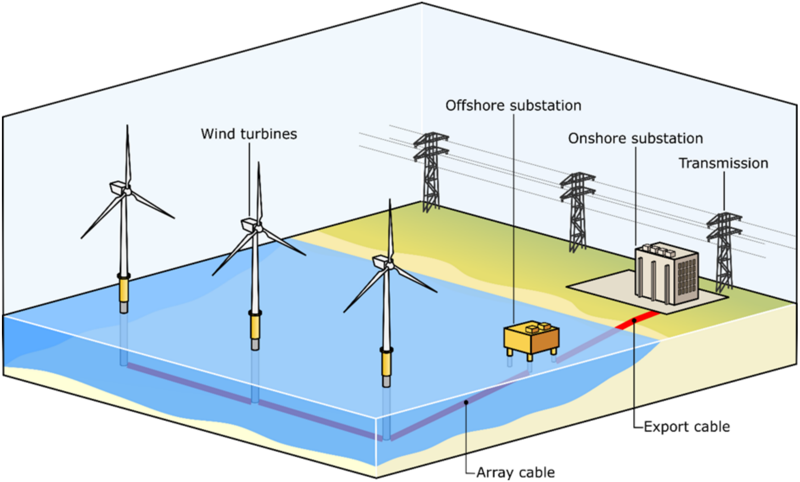
- UNI approaches are computationally inexpensive techniques
- For similar accuracies, non-linear factor estimation techinques have higher computational cost wrt to linear ones

Concluding remarks

- Linear factor estimation techniques in combination with lazy learning or VAR for forecasting (corresponding to DF-PCA-Lazy/DFML and DF-PCA-VAR/DFM, according to the experiments) are consistently among the top performers across all the considered datasets.
- Non-linear factor estimation techniques in combination with both data-driven and model-driven technique have comparable forecasting performances, but higher computational costs.
- Although often negleted in the scientific literature, univariate techniques (ES/Holt-Winters) are often tough competitor to beat.

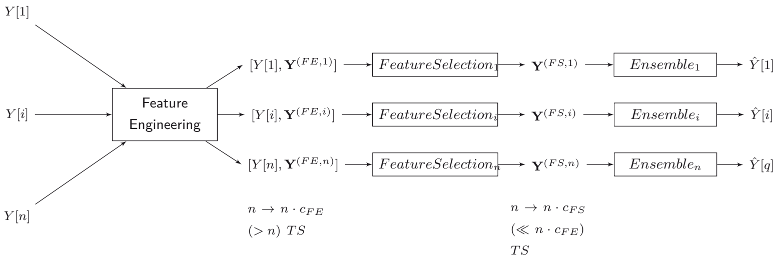
SMURF-ES - Selective Multivariate to Univariate Reduction through Feature Engineering and Selection

SMURF-ES Strategy - Motivation



Source: Bhattacharya S, Biswal S, Aleem M, Amani S, Prabhakaran A, Prakhya G, Lombardi D, Mistry HK. Seismic Design of Offshore Wind Turbines: Good, Bad and Unknowns. Energies. 2021; 14(12):3496. https://doi.org/10.3390/en14123496

SMURF-ES strategy - Feature Engineering

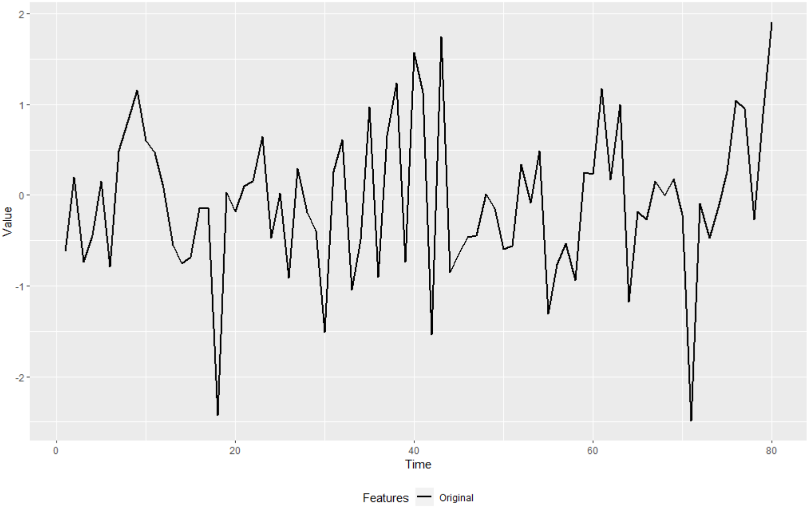


Feature Engineering - Section 3.3.1 - 3.3.2

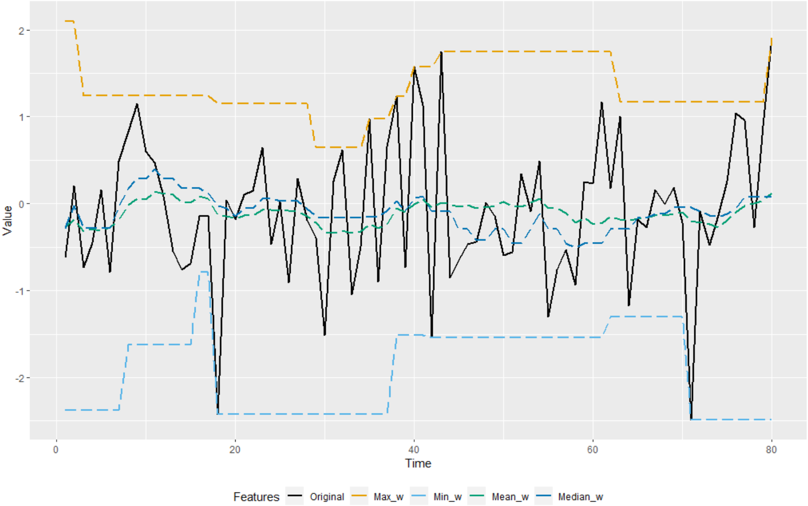
$$1 \rightarrow c_{FE}$$

- **Statistical**
- **Quantiles**
- **Custom**

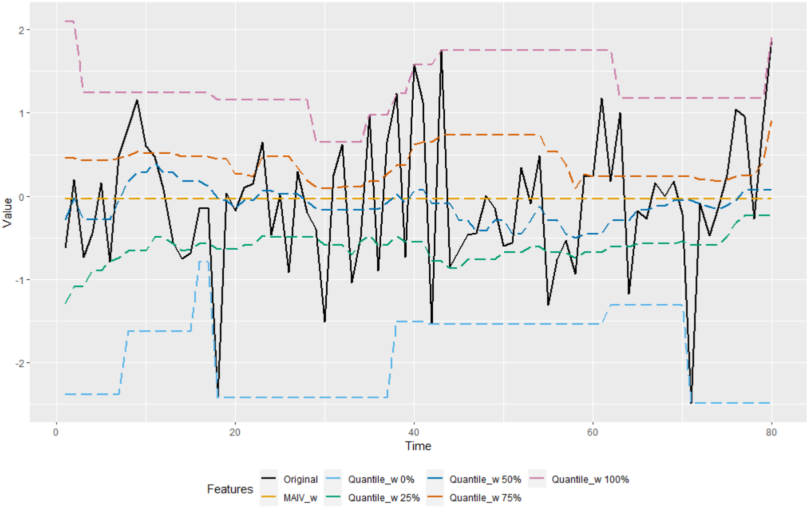
SMURF-ES strategy - Feature Engineering - Original



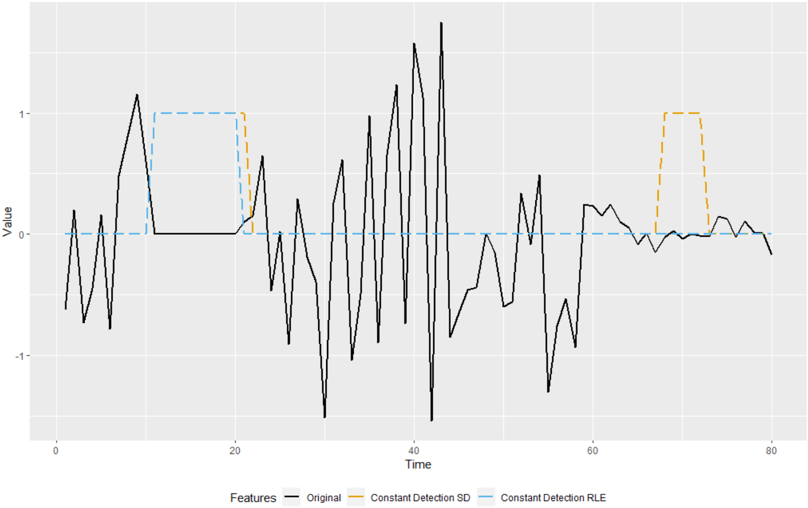
SMURF-ES strategy - Feature Engineering - Statistical



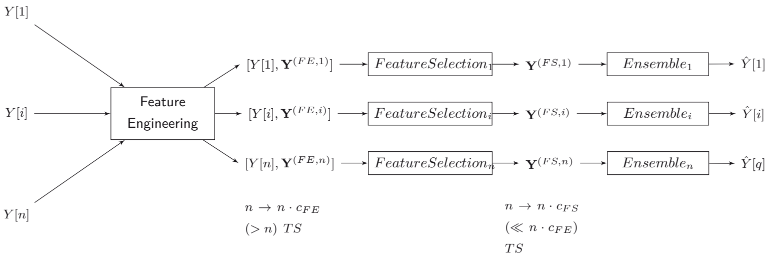
SMURF-ES strategy - Feature Engineering - Quantiles



SMURF-ES strategy - Feature Engineering - Custom



SMURF-ES strategy - Feature Selection



Feature Selection - Section 3.3.3

$$c_{FE} \rightarrow c_{FS} \ll c_{FE}$$

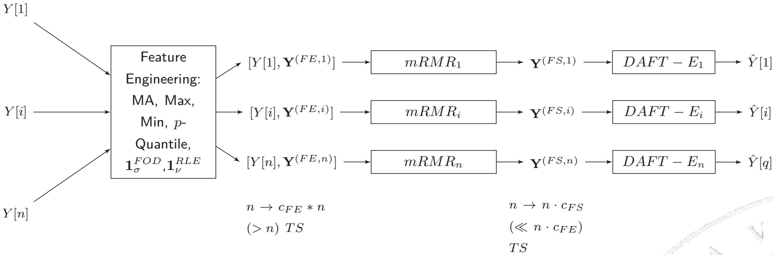
- **PCA:** Feature transformation
- **mRMR:** Feature selection

SMURF-ES - Ensemble techniques - Section 3.3.4

Model-driven	Data-driven	SMURF-ES Ensembles
Naive (Persistence)	Feed-forward ANN	Av-GBM-RF
Average	GBM	Av-Average-RF
	Random Forest (RF)	Av-GBM-Average
	Lazy Learning (LL)	Av-SVM-GBM
	SVM	Av-SVM-Average
		Ad-ANN-Average
		Ad-SVM-Average
		Ad-GBM-Average
		Ad-RF-Average
		Ad-Average-RF-SVM-GBM
		DAFT-E (Ad.RF-LL-Naive)

- **Av** Simple average ensembling
- **Ad:** Adaptive, forecasting-error-based ensembling

SMURF-ES strategy implementation - DAFT-E - Section 3.3.5.2



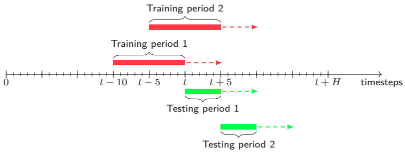
SMURF-ES - Experimental assessment

DAFT-E assessment - Experimental Setup - Section 5.4

- **Strategy configuration:**

Feature engineering	Feature selection	Ensemble
MA Max, Min p-quantile $\mathbf{1}_{\sigma}^{FOD}, \mathbf{1}_{v}^{RLE}$	mRMR	DAFT-E (Ad.RF-LL-Naive)

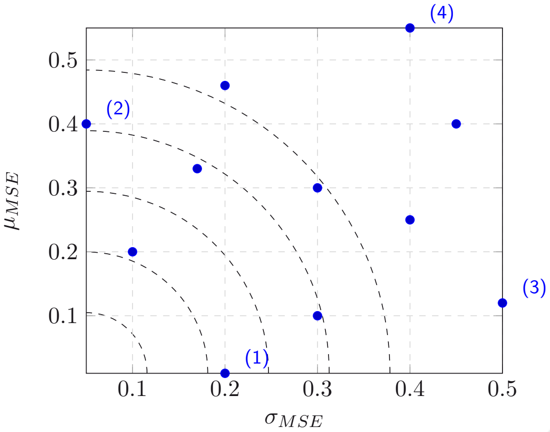
- **Rolling window approach:** (Tashman, 2000)



- **Forecasting horizon:** $H \in \{1, \cdots, 12\}$ (hours).
- **Error metric:** MSE reduction ratio (nMSE) between the ω-th model and the Naive baseline for the kth trial and the hth forecasting horizon:

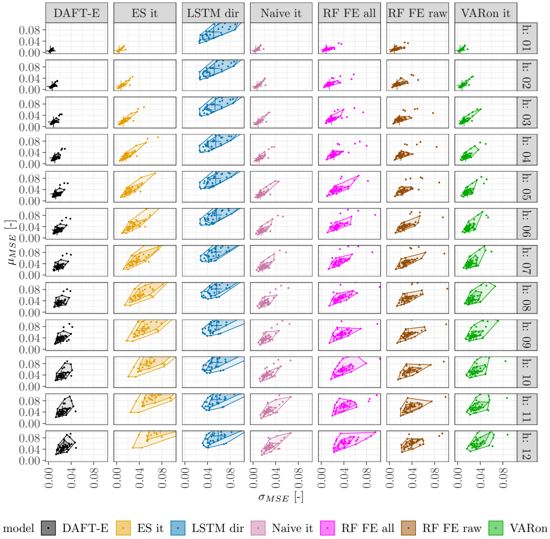
$$\mathrm{nMSE}^{(\omega,k,h)} = (\mathrm{MSE}^{(\omega,k,h)} / \mathrm{MSE}^{(Naive,k,h)}) - 1$$

DAFT-E assessment - Results - Section 5.4



Example of bivariate plot $\sigma_{\mathrm{ERR}} - \mu_{\mathrm{ERR}}$, for with the selected error metric being MSE. The highlighted points represent configuration with low bias and low variance (1), large bias and low variance (2), large variance and low bias (3) and high bias and high variance (4).

DAFT-E assessment - Results - Australian Case Study - Section 5.4



- Comparison between DAFT-E, internal algorithms and SOTA.
- DAFT-E shows a reduced variability in its predictions (smaller bagplot area) and reduced MSE with respect to competitors.

Concluding remarks

- Both implementations (DAF-E and DAFT-E) show the effectiveness of heterogeneous ensembles with a dynamic combination rule for time series forecasting on a challenging problem such as wind power forecasting.
- The introduction in DAFT-E of a memory component (in the form of forgetting factors) further improves the forecasting performances and reduces variability of the forecasts.
- In DAFT-E, using the same model family, trained on different sets of input features, seems to be beneficial to forecasting performance.

Conclusions

Conclusions

- Algorithms and methodologies from the traditional machine learning literature have been often neglected in the multivariate and multistep-ahead forecasting literature.
- The prevalent approaches are either from the econometric (VARMA) or the deep learning domains (RNN, CNN, Transformers).
- Nevertheless, traditional machine learning approaches can produce competitive results when employed in tailored strategies
- An added benefit is the reduced computational complexity (and consequently energy consumption) to obtain comparable forecasting performances.

Research perspectives

- Further study of the application of neural techniques within the strategies
- Development of hierarchical extensions of the proposed techniques for problems with an inherent hierarchical structure.
- Streaming/Online extensions of the strategies for Big Data applications

Publications - DFML I

- Gianluca Bontempi, Yann-Aël Le Borgne, and Jacopo De Stefani (2017). "A Dynamic Factor Machine Learning Method for Multi-Variate and Multi-Step-Ahead Forecasting". In: *Proceedings of DSAA 2017, the 4th IEEE International Conference on Data Science and Advanced Analytics 2017*
  DSAA 2017 - Honorable Mention Research Paper Award
- Jacopo De Stefani, Yann-Aël Le Borgne, Olivier Caelen, Dalila Hattab, and Gianluca Bontempi (Aug. 31, 2018). "Batch and Incremental Dynamic Factor Machine Learning for Multivariate and Multi-Step-Ahead Forecasting". In: *International Journal of Data Science and Analytics*. ISSN: 2364-4168. DOI: 10.1007/s41060-018-0150-x. URL: https://doi.org/10.1007/s41060-018-0150-x (visited on 09/12/2018)

ULB

Publications - DFML II

- Jacopo De Stefani and Gianluca Bontempi (2021). "Factor-Based Framework for Multivariate and Multi-Step-Ahead Forecasting of Large Scale Time Series". In: *Frontiers in Big Data* 4, p. 75. ISSN: 2624-909X. DOI: 10.3389/fdata.2021.690267. URL: https://www.frontiersin.org/article/10.3389/fdata.2021.690267 (visited on 09/10/2021)

UNIVERSITÉ LIBRE DE BRUXELLES

Publications - SMURF-ES

- Fabrizio De Caro, Jacopo De Stefani, Gianluca Bontempi, Alfredo Vaccaro, and Domenico Villacci (Oct. 18, 2020). "Robust Assessment of Short-Term Wind Power Forecasting Models on Multiple Time Horizons". In: *Technology and Economics of Smart Grids and Sustainable Energy* 5.1, p. 19. ISSN: 2199-4706. DOI: 10.1007/s40866-020-00090-8. URL: https://doi.org/10.1007/s40866-020-00090-8 (visited on 07/27/2021)
- Fabrizio De Caro, Jacopo De Stefani, Alfredo Vaccaro, and Gianluca Bontempi (2021). "DAFT-E : Feature-based Multivariate and Multi-step-ahead Wind Power Forecasting". In: *IEEE Transactions on Sustainable Energy*, pp. 1–1. DOI: 10.1109/TSTE.2021.3130949 - **Shared first authorship**

Additional publications developed I

- **European Patent**
  - Jacopo De Stefani, Gianluca Bontempi, Olivier Caelen, and Dalila Hattab (Jan. 3, 2019a). “System and Method for Managing Risks in a Process”. Pat. WO2019002582 (A1). Worldline. **Classifications: IPC:** G06Q10/04; G06Q50/04, **CPC:** G06Q10/04 (EP); G06Q50/04 (EP); Y02P90/30 (EP). URL: https://worldwide.espacenet.com/publicationDetails/biblio?FT=D&date=20190103&DB=&locale=en_EP&CC=WO&NR=2019002582A1&KC=A1&ND=5 (visited on 07/27/2021)
- **Appendix A - Volatility forecasting**
  - Jacopo De Stefani, Olivier Caelen, Dalila Hattab, and Gianluca Bontempi (n.d.). “Machine Learning for Multi-Step Ahead Forecasting of Volatility Proxies”. In: *ECML PKDD 2017 Workshops - MIning DAta for financial applicationS (MIDAS 2017)* **MIDAS 2017 - Best Paper Award**

ULB

Additional publications developed II

- Jacopo De Stefani, Olivier Caelen, Dalila Hattab, Yann-Aël Le Borgne, and Gianluca Bontempi (2019b). “A Multivariate and Multi-Step Ahead Machine Learning Approach to Traditional and Cryptocurrencies Volatility Forecasting”. In: *ECML PKDD 2018 Workshops - MIning DAta for financial applicationS (MIDAS 2018)*. Ed. by Carlos Alzate, Anna Monreale, Livio Bioglio, Valerio Bitetta, Ilaria Bordino, Guido Caldarelli, Andrea Ferretti, Riccardo Guidotti, Francesco Gullo, Stefano Pascolutti, Ruggero G. Pensa, Celine Robardet, and Tiziano Squartini. Lecture Notes in Computer Science. Cham: Springer International Publishing, pp. 7–22. ISBN: 978-3-030-13463-1. DOI: 10.1007/978-3-030-13463-1_1

- **AutoML**

Additional publications developed III

- Gian Marco Paldino, Jacopo De Stefani, Fabrizio De Caro, and Gianluca Bontempi (2021). "Does AutoML Outperform Naive Forecasting?" In: *Engineering Proceedings* 5.1 (1), p. 36. DOI: 10.3390/engproc2021005036. URL: https://www.mdpi.com/2673-4591/5/1/36 (visited on 07/27/2021)

Software development

The work presented in this thesis has led to the development of multiple R packages, publicly available on Github:

- `UEMTS`: Univariate Error Measures for Time Series forecasting - `https://www.github.com/jdestefani/UEMTS`, implementing the forecasting error measures analyzed in (Hyndman and Koehler, 2006).
- `MEMTS`: Multivariate Error Measures for Time Series forecasting - `https://www.github.com/jdestefani/MEMTS`, proving the multivariate extensions of the forecasting error measures analyzed in (Hyndman and Koehler, 2006), employed for the experimental assessments in Chapter 4 and 5.
- `MM4Benchmark`: Multivariate extension of the Benchmarks used for the M4 competition - `https://github.com/jdestefani/MM4Benchmark`, implementing the most common multivariate benchmark, employed in the experimental assessments in Chapter 4 and 5.
- `ExtendedDFML`: Extended DFML framework - `https://www.github.com/jdestefani/ExtendedDFML`, providing an implementation of the DFML strategy with both linear/non-linear factor estimation and model-driven/data-driven factor forecasting techniques.

Acknowledgements

Special thanks to:

Financial support & Scientific Collaboration

Scientific collaboration

... and to all the collaborators throughout the years.

Thank you for your attention! Any questions/comments?

Scan the QR code to access the thesis

References

- Bollerslev, Tim (1986). "Generalized autoregressive conditional heteroskedasticity". In: *Journal of econometrics* 31.3, pp. 307–327 (cit. on p. 99).
- Bontempi, Gianluca, Yann-Aël Le Borgne, and Jacopo De Stefani (2017). "A Dynamic Factor Machine Learning Method for Multi-Variate and Multi-Step-Ahead Forecasting". In: *Proceedings of DSAA 2017, the 4th IEEE International Conference on Data Science and Advanced Analytics 2017* (cit. on p. 76).

Bibliography II

- De Caro, Fabrizio, Jacopo De Stefani, Gianluca Bontempi, Alfredo Vaccaro, and Domenico Villacci (Oct. 18, 2020). "Robust Assessment of Short-Term Wind Power Forecasting Models on Multiple Time Horizons". In: *Technology and Economics of Smart Grids and Sustainable Energy* 5.1, p. 19. ISSN: 2199-4706. DOI: 10.1007/s40866-020-00090-8. URL: https://doi.org/10.1007/s40866-020-00090-8 (visited on 07/27/2021) (cit. on p. 78).
- De Caro, Fabrizio, Jacopo De Stefani, Alfredo Vaccaro, and Gianluca Bontempi (2021). "DAFT-E : Feature-based Multivariate and Multi-step-ahead Wind Power Forecasting". In: *IEEE Transactions on Sustainable Energy*, pp. 1–1. DOI: 10.1109/TSTE.2021.3130949 (cit. on p. 78).

Bibliography III

De Stefani, Jacopo and Gianluca Bontempi (2021). "Factor-Based Framework for Multivariate and Multi-Step-Ahead Forecasting of Large Scale Time Series". In: *Frontiers in Big Data* 4, p. 75. ISSN: 2624-909X. DOI: 10.3389/fdata.2021.690267. URL: https://www.frontiersin.org/article/10.3389/fdata.2021.690267 (visited on 09/10/2021) (cit. on p. 77).

Bibliography IV

- De Stefani, Jacopo, Gianluca Bontempi, Olivier Caelen, and Dalila Hattab (Jan. 3, 2019a). “System and Method for Managing Risks in a Process”. Pat. WO2019002582 (A1). Worldline. **Classifications: IPC:** G06Q10/04; G06Q50/04, **CPC:** G06Q10/04 (EP); G06Q50/04 (EP); Y02P90/30 (EP). URL: https://worldwide.espacenet.com/publicationDetails/biblio?FT=D&date=20190103&DB=&locale=en_EP&CC=WO&NR=2019002582A1&KC=A1&ND=5 (visited on 07/27/2021) (cit. on p. 79).
- De Stefani, Jacopo, Olivier Caelen, Dalila Hattab, and Gianluca Bontempi (n.d.). “Machine Learning for Multi-Step Ahead Forecasting of Volatility Proxies”. In: *ECML PKDD 2017 Workshops - MIning DAta for financial applicationS (MIDAS 2017)* (cit. on p. 79).

Bibliography V

De Stefani, Jacopo, Olivier Caelen, Dalila Hattab, Yann-Aël Le Borgne, and Gianluca Bontempi (2019b). “A Multivariate and Multi-Step Ahead Machine Learning Approach to Traditional and Cryptocurrencies Volatility Forecasting”. In: *ECML PKDD 2018 Workshops - MIning DAta for financial applicationS (MIDAS 2018)*. Ed. by Carlos Alzate et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, pp. 7–22. ISBN: 978-3-030-13463-1. DOI: 10.1007/978-3-030-13463-1_1 (cit. on p. 80).

Bibliography VI

De Stefani, Jacopo, Yann-Aël Le Borgne, Olivier Caelen, Dalila Hattab, and Gianluca Bontempi (Aug. 31, 2018). "Batch and Incremental Dynamic Factor Machine Learning for Multivariate and Multi-Step-Ahead Forecasting". In: *International Journal of Data Science and Analytics*. ISSN: 2364-4168. DOI: 10.1007/s41060-018-0150-x. URL: https://doi.org/10.1007/s41060-018-0150-x (visited on 09/12/2018) (cit. on p. 76).

Demšar, Janez (2006). "Statistical comparisons of classifiers over multiple data sets". In: *The Journal of Machine Learning Research* 7, pp. 1–30 (cit. on p. 52).

Garman, Mark B and Michael J Klass (1980). "On the Estimation of Security Price Volatilities from Historical Data". In: *Journal of business*, pp. 67–78 (cit. on pp. 97, 98).

ULB UNIVERSITÉ LIBRE DE BRUXELLES

Bibliography VII

Hansen, Peter R and Asger Lunde (2005). "A forecast comparison of volatility models: does anything beat a GARCH (1, 1)?" In: *Journal of applied econometrics* 20.7, pp. 873–889 (cit. on p. 99).

Hyndman, Rob J. and Anne B. Koehler (2006). "Another look at measures of forecast accuracy". In: *International Journal of Forecasting* 22.4, pp. 679 –688. ISSN: 0169-2070. DOI: https://doi.org/10.1016/j.ijforecast.2006.03.001. URL: http://www.sciencedirect.com/science/article/pii/S0169207006000239 (cit. on p. 82).

ULB

Bibliography VIII

Paldino, Gian Marco, Jacopo De Stefani, Fabrizio De Caro, and Gianluca Bontempi (2021). "Does AutoML Outperform Naive Forecasting?" In: *Engineering Proceedings* 5.1 (1), p. 36. DOI: 10.3390/engproc2021005036. URL: https://www.mdpi.com/2673-4591/5/1/36 (visited on 07/27/2021) (cit. on p. 81).

Stock, J.H. and M.W. Watson (2010). "Dynamic Factor Models". In: *Oxford Handbook of Economic Forecasting*. Ed. by M.P. Clements and D.F. Hendry. Oxford University Press (cit. on pp. 38–41).

Bibliography IX

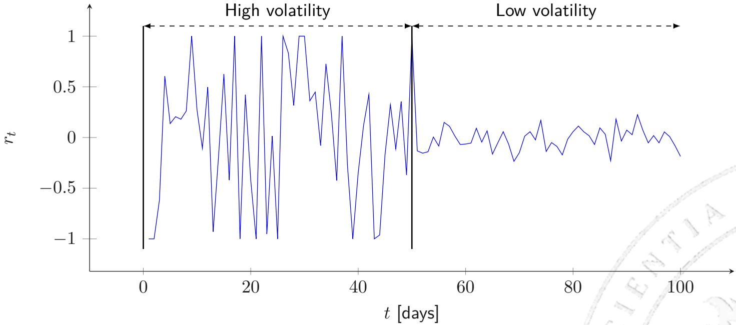
Tashman, Leonard J. (2000). “Out-of-Sample Tests of Forecasting Accuracy: An Analysis and Review”. In: *International Journal of Forecasting* 16.4, pp. 437 –450. ISSN: 0169-2070. DOI: https://doi.org/10.1016/S0169-2070(00)00065-0. URL: http://www.sciencedirect.com/science/article/pii/S0169207000000650 (cit. on pp. 51, 69, 100).

Appendix

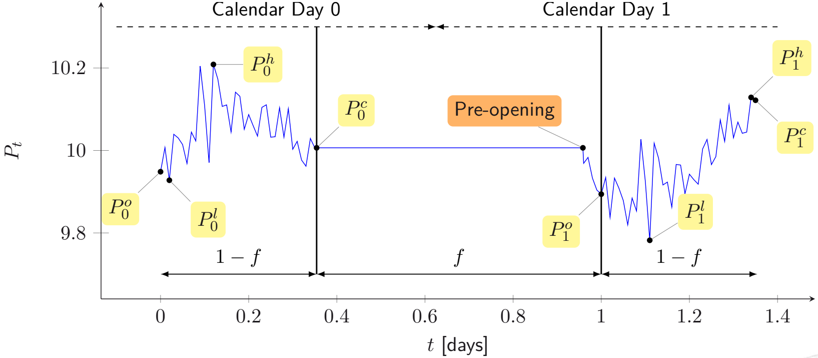
What is volatility?

Definition

Volatility is a statistical measure of the dispersion of returns for a given security or market index.



A closer look on data - Volatility measures



Volatility proxies (1) - Garman and Klass, 1980

- **Closing prices**

$$\hat{\sigma}_0(t) = \left[\ln\left(\frac{P_{t+1}^{(c)}}{P_t^{(c)}}\right)\right]^2 = r_t^2 \tag{1}$$

- **Opening/Closing prices**

$$\hat{\sigma}_1(t) = \underbrace{\frac{1}{2f} \cdot \left[\ln\left(\frac{P_{t+1}^{(o)}}{P_t^{(c)}}\right)\right]^2}_{\text{Nightly volatility}} + \underbrace{\frac{1}{2(1-f)} \cdot \left[\ln\left(\frac{P_t^{(c)}}{P_t^{(o)}}\right)\right]^2}_{\text{Intraday volatility}} \tag{2}$$

- **OHLC prices**

$$\hat{\sigma}_2(t) = \frac{1}{2\ln 4} \cdot \left[\ln\left(\frac{P_t^{(h)}}{P_t^{(l)}}\right)\right]^2 \tag{3}$$

$$\hat{\sigma}_3(t) = \underbrace{\frac{a}{f} \cdot \left[\ln\left(\frac{P_{t+1}^{(o)}}{P_t^{(c)}}\right)\right]^2}_{\text{Nightly volatility}} + \underbrace{\frac{1-a}{1-f} \cdot \hat{\sigma}_2(t)}_{\text{Intraday volatility}} \tag{4}$$

Volatility proxies (2) - Garman and Klass, 1980

- **OHLC prices**

$$u = \ln\left(\frac{P_t^{(h)}}{P_t^{(o)}}\right) \qquad d = \ln\left(\frac{P_t^{(l)}}{P_t^{(o)}}\right) \qquad c = \ln\left(\frac{P_t^{(c)}}{P_t^{(o)}}\right) \tag{5}$$

$$\hat{\sigma}_4(t) = 0.511(u-d)^2 - 0.019[c(u+d) - 2ud] - 0.383c^2 \tag{6}$$

$$\hat{\sigma}_5(t) = 0.511(u-d)^2 - (2\ln 2 - 1)c^2 \tag{7}$$

$$\hat{\sigma}_6(t) = \underbrace{\frac{a}{f} \cdot \log\left(\frac{P_{t+1}^{(o)}}{P_t^{(c)}}\right)^2}_{\text{Nightly volatility}} + \underbrace{\frac{1-a}{1-f} \cdot \hat{\sigma}_4(t)}_{\text{Intraday volatility}} \tag{8}$$

Volatility proxies (3)

- **GARCH (1,1) model** - Hansen and Lunde, 2005

$$\sigma_t^G = \sqrt{\omega + \sum_{j=1}^{p} \beta_j (\sigma_{t-j}^G)^2 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2}$$

where $\varepsilon_{t-i} \sim \mathcal{N}(0,1)$, with the coefficients $\omega, \alpha_i, \beta_j$ fitted according to Bollerslev, 1986.

- **Sample standard deviation**

$$\sigma_t^{SD,n} = \sqrt{\frac{1}{n-1} \sum_{i=0}^{n-1} (r_{t-i} - \bar{r})^2}$$

where

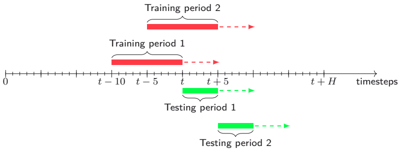
$$r_t = \ln\left(\frac{P_t^{(c)}}{P_{t-1}^{(c)}}\right) \qquad \bar{r_n} = \frac{1}{n} \sum_{j=t-n}^{t} r_j$$

DFML Extensions assessments - Experimental Setup - Section 4.5

- **Strategy configuration:**

Factor estimation	Factor forecast
PCA	Lazy Learning (Direct, Recursive, Joint)
Autoencoder (Shallow)	VAR

- **Rolling window approach:** (Tashman, 2000)



- **Forecasting horizon:** $H \in \{2, 5, 10, 25, 20, 50\}$ (days).
- **Error metric:** Naive Normalized Mean Squared Error:

$$\text{NNMSE} = \frac{1}{n}\sum_{j=1}^{n} \text{NNMSE}[j]$$

$$\text{NNMSE}[j] = \frac{\frac{1}{H}\sum_{h=1}^{H}(\mathbf{Y}_{T+h}[j]-\hat{\mathbf{Y}}_{T+h}[j])^2}{\frac{1}{H}\sum_{h=2}^{H}(\mathbf{Y}_{T+h}[j]-\mathbf{Y}_{T+h-1}[j])^2}$$

DFML Extensions assessments - Results - Section 4.5

Rolling		$DFML_{PC}$						DFM	
		Direct		Iter		MIMO			
n	H	Batch	Online	Batch	Online	Batch	Online	Batch	Online
100	2	**0.106**	**0.106**	0.110	0.107	0.110	0.109	0.109	**0.106**
100	5	**0.118**	**0.118**	0.144	0.206	0.125	0.128	0.140	0.137
100	10	**0.118**	**0.118**	0.178	0.173	0.127	0.123	0.151	0.148
100	15	0.114	**0.112**	0.175	0.172	0.120	0.120	0.165	0.153^{+}
100	20	0.112	**0.111**	0.170	0.168	0.118	0.120	0.182	0.166^{+}
100	50	**0.111**	**0.111**	0.173	0.173	0.117	0.112^{+}	0.307	0.282^{+}
200	2	0.195	0.195	0.199	0.197	0.198	0.195	**0.124**	0.135^{+}
200	5	0.274	0.266	0.345	0.337	0.281	0.269	**0.176**	0.184
200	10	0.259	0.253	0.388	0.369^{+}	0.278	0.268	0.195	**0.191**
200	15	0.242	0.237	0.399	0.388^{+}	0.254	0.244	0.211	**0.206**
200	20	0.269	0.262	0.412	99.062	0.270	0.269	0.226	**0.225**
200	50	0.253	**0.245**	0.434	43.889	0.266	0.256	0.290	0.285^{+}

Table B.6 - Earth Surface Temperature series: NMSE (averaged over all the continuation sets) of the different forecasting methods.Comparison between batch and iterative PCA. The superscript $^{+}$ denotes a significant improvement (pv=0.05) while using the iterative PCA calculation instead of the batch one. The bold notation is used to identify the best method for each horizon.

What evidence do you have that your approaches are competitive with DL ?

- **DFML**
  - Volatility assessment (Section 4.4) and Extensions assessment (Section 4.5) compared the DFML approach with RNN (Elman and LSTM) networks, showing the outperformance of the DFML approach on the considered settings
- **SMURF-ES**
  - DAFT-E (Section 5.4) outperformed a LSTM network (whose hyperparameters were optimized through a grid search).
- In both cases, the proposed strategies allow to achieve comparable (if not better) results to DL techniques with a reduced computational cost.