Wind power predictions from nowcasts to 4-hour forecasts: a learning approach with variable selection

Dimitri Bouche[1] Rémi Flamary[4] Florence d'Alché-Buc[1] Riwal Plougonven[3] Marianne Clausel[2] Jordi Badosa[3] Philippe Drobinski[3]

[1] LTCI, Télécom Paris, IP Paris, [2] Université de Lorraine, CNRS, IECL
[3] LMD-IPSL, Ecole polytechnique, IP Paris, ENS-PSL Université
[4] CMAP, Ecole Polytechnique, IP Paris

Short term and local wind forecasting

Objective: predict the wind speed (ultimately wind power)...

- **locally** (at a wind farm location)
- in the **short term** (up to few hours ahead)
- at high temporal resolution (every 10 min)

Challenges

- **Numerical weather prediction models** (NWPM) at desired time/space resolution: too costly
- Statistical methods using only **past local information**: costly if complex (deep) models are used (**one model per site**)

Downscaling with statistical learning

Downscaling : combine using statistical learning...

- NWPM's outputs (low resolution but readily available)
- Past local data collected by turbines' sensors.

Rationale

- NWPM: valuable information for the evolution of the atmosphere on larger scale
- Past local data: site specific information

Wind power prediction: direct/indirect

Two main approaches for wind power prediction:

Direct approach: Model predicts directly local wind power

Indirect approach:

1. Model predicts the local wind speed
2. Predictions are passed through an estimated or a theoretical power curve.

Overview

In the context of **downscaling with statistical learning**

Variable selection

- many available variables from NWPM and several local variables
- Linear or nonlinear dynamics ?
- Essential variables ?

Which methods perform well ?

- Linear vs nonlinear ?
- Simplest method ?

Indirect or direct prediction ?

Dataset

Wind farms data from **Zéphyr ENR**'s [Dupré et al., 2020]:

- 6 wind farms (we study 5 of those: BM, BO, MP, RE, VE)
- Between 2 and 3 years of historical data
- 10 min resolution
- Between 3 and 6 wind turbines

NWPM outputs from the **European Centre for Medium-Range Weather Forecasts (ECMWF)**

- Spatial resolution: 16km $\times$ 16km
- Time resolution: $1h$

Wind farms locations

Figure: Cartography of the studied farms, BM (A), BO (B), MP (C), RE (D), VE (E)

Prediction scheme

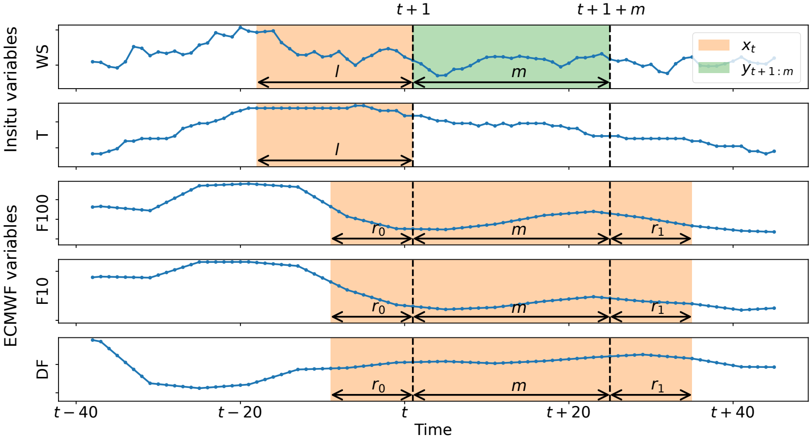

Figure: Downscaling prediction scheme

Prediction scheme (2)

Input space

- $\mathcal{X} = \mathbb{R}^q$ for some $q \in \mathbb{N}$
- q depends on chosen variables and time windows

Training data

- $(\mathrm{x}_t, \mathrm{y}_{t+1:m})_{t=1}^n \in (\mathcal{X} \times \mathbb{R}^m)^n$ for some $n \in \mathbb{N}$,
- with $\mathrm{y}_{t+1:m} := (y_{t+1+m_0})_{m_0=1}^m$.

Learning problem

Consider

- A model $f_w : \mathcal{X} \to \mathcal{Y}$ parametrized by parameters $w \in \mathcal{W}$
- A regularization function $\Omega : \mathcal{W} \longrightarrow \mathbb{R}$

Generic learning problem

$$\min_{w \in \mathcal{W}} \frac{1}{n} \sum_{t=1}^{n} \| f_w(x_t) - y_{t+1:m} \|_2^2 + \lambda \Omega(w).$$

Variable selection

What do we put in x_t ?

- Variable selection: which variables ?
- Temporal selection: how much past (or future) information ?

Which dependencies can be detected ?

- Linear: account only for linear dependencies only
- Nonlinear: account for a rich variety of dependencies

Linear variable selection

We consider two techniques

- A greedy forward approach (add the variables one by one)
- The LASSO [Tibshirani, 1996] (sparsity-promoting penalty)

LASSO problem

$$\min_{\mathrm{w}\in\mathcal{W}, b\in\mathbb{R}} \frac{1}{n}\sum_{t=1}^{n}(\mathrm{w}^{\mathsf{T}}\mathrm{x}_t + b - y_{t+1+m})^2 + \lambda\|\mathrm{w}\|_1,$$

LASSO selection (WS)

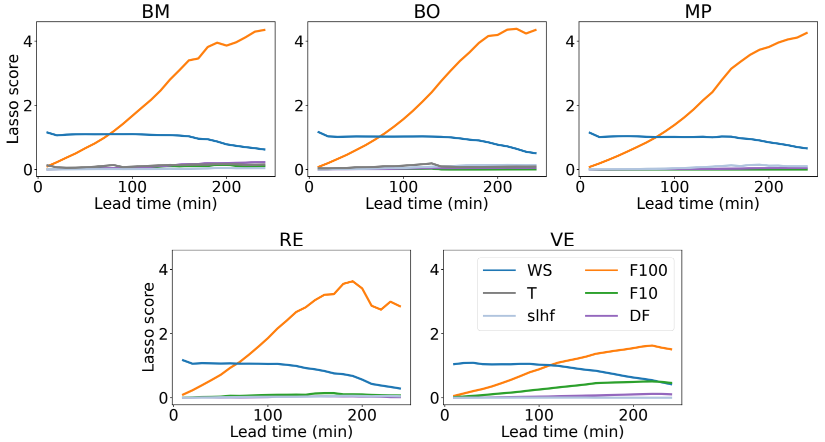

Figure: LASSO selection: Wind speed as target variable

LASSO selection (PW)

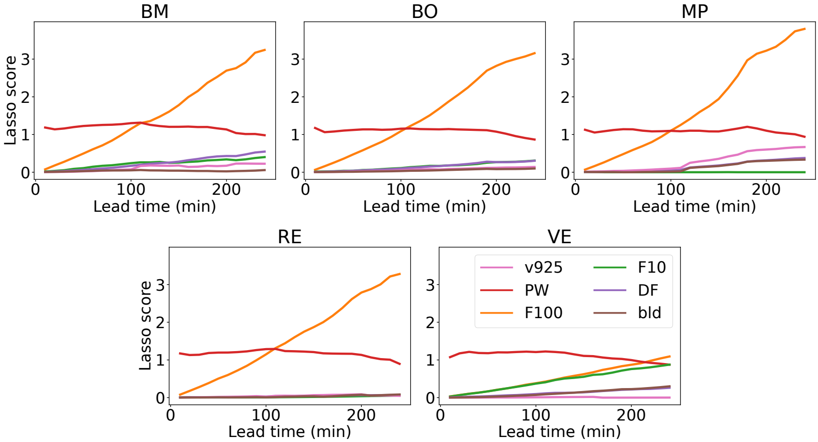

Figure: LASSO selection: Wind power as target variable

HSIC

Backward selection using HSIC (BAHSIC) [Song et al., 2012]

Consider

- An input p. d. kernel $k : \mathcal{X}^2 \longrightarrow \mathbb{R}$
- An output p. d. kernel $g : (\mathbb{R}^m)^2 \longrightarrow \mathbb{R}$

HSIC estimator [Gretton et al., 2008]

$$\widehat{\text{HSIC}} := \frac{1}{n^2}\text{Trace}(HKHG),$$

With

- $H := \frac{1}{n}(I - \mathbf{1}\mathbf{1}^\top)$ the centering matrix
- $K \in \mathbb{R}^{n\times n}$ and $G \in \mathbb{R}^{n\times n}$ the kernel matrices

Nyström HSIC

Nyström approximation [Drineas and Mahoney W., 2005]

- Random indices $\{i_1, ..., i_p\}$ and $\{i'_1, ..., i'_{p'}\}$ from $[\![n]\!]$.
- K_{np}, K_{pp}, $G_{np'}$, $G_{p'p'}$ corresponding sub kernel matrices

Centered Nyström features

$$\widehat{\Phi} := HK_{np}K_{pp}^{-\frac{1}{2}},$$
$$\widehat{\Psi} := HG_{np'}G_{p'p'}^{-\frac{1}{2}},$$

Nyström HSIC estimator [Zhang et al., 2018]

$$\widetilde{\text{HSIC}} := \|\frac{1}{n}\widehat{\Phi}^{\top}\widehat{\Psi}\|_F^2 \ .$$

HSIC importance evolution (WS)

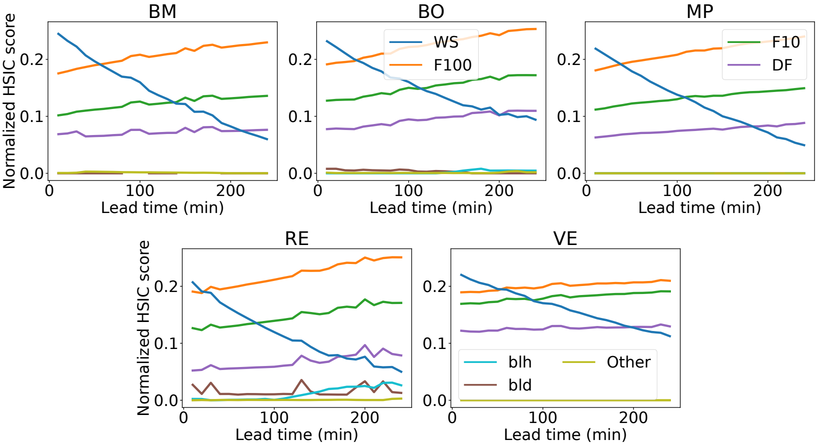

Figure: HSIC selection: Wind speed as target variable

HSIC importance evolution (PW)

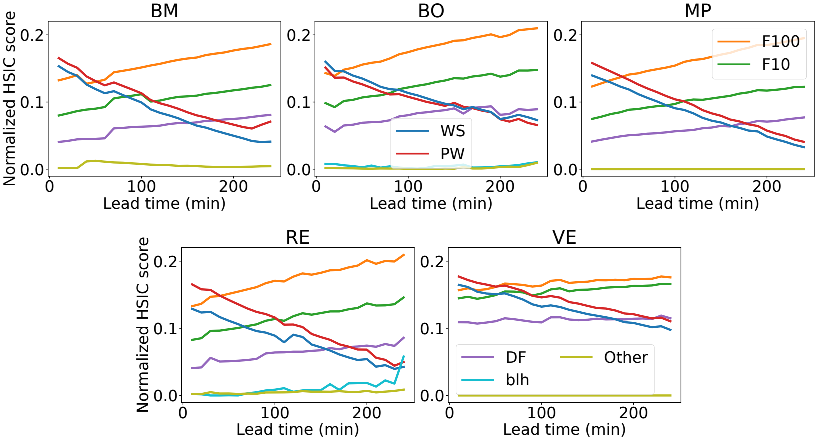

Figure: HSIC selection: Wind power as target variable

Benchmarks

- **LASSO** (linear, built in variable selection)
- **OLS f-stepwise**: Forward stepwise OLS (linear, greedy variable selection)
- **Nyström KRR**: Kernel ridge regression with Nyström approximation (nonlinear, variable selection through BAHSIC)
- **XgBoost** (nonlinear, variable selection through BAHSIC)
- **Feedforward deep neural network** (nonlinear, variable selection through BAHSIC)
- **Persistence**
- **ECMWF**

Overall performance (WS)

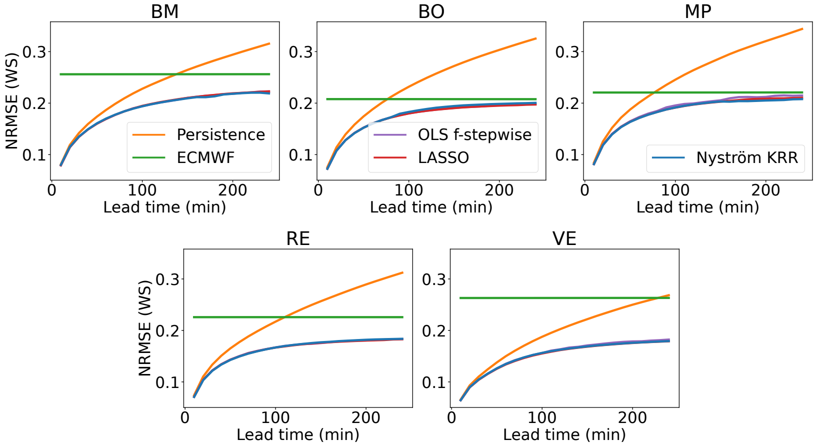

Figure: NRMSE evolution for best performing methods, WS as target

Example of estimated power curve

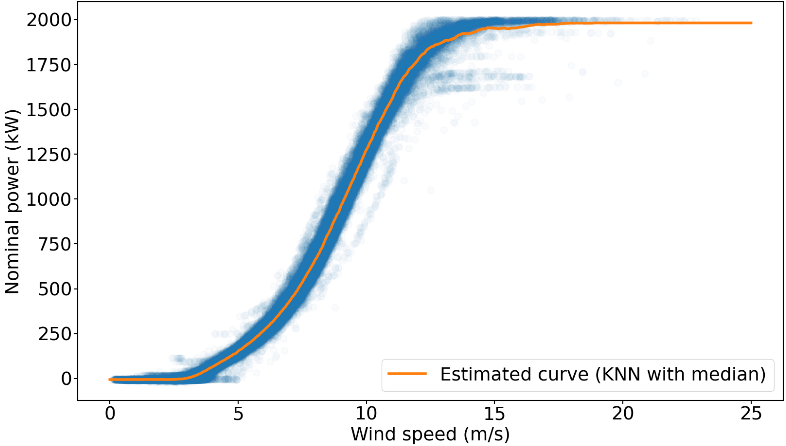

Figure: Estimated power curve for BO

Overall performance (PW)

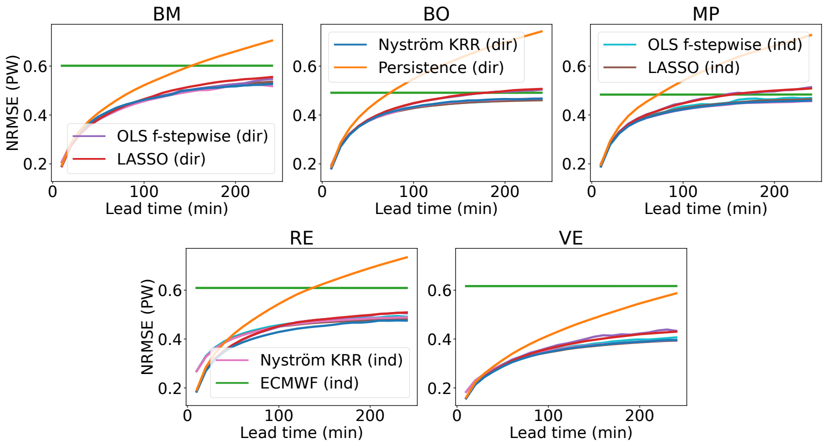

Figure: NRMSE evolution for best performing methods, PW as target

Quantitative comparison (WS)

Method (average rank)	BM	BO	MP	RE	VE
LASSO (1.6)	1.14	**0.13**	0.17	**0.14**	**0.15**
Nyström KRR (2.0)	**1.07**	0.36	**0.07**	0.19	0.21
OLS f-stepwise (2.4)	1.12	0.16	0.48	0.17	0.34
Feedforward NN (4.4)	1.54	0.87	0.61	1.18	1.34
XG-Boost (4.6)	1.83	1.25	1.64	0.73	0.97
ECMWF (6.4)	7.83	3.63	3.76	6.65	11.37
Persistence (6.6)	5.61	6.72	7.03	6.7	4.5

Table: Average NRMSE degradation w.r.t. best predictor for wind speed prediction ($\times 10^{-2}$)

Quantitative comparison (PW)

Type	Method (average rank)	BM	BO	MP	RE	VE
Direct	**Nyström KRR (2.2)**	3.93	1.16	0.63	**0.36**	**0.69**
Indirect	Nyström KRR (2.8)	**3.46**	1.15	**0.55**	2.88	1.2
Indirect	LASSO (3.0)	3.96	**0.77**	1.08	2.62	0.86
Indirect	OLS f-stepwise (4.0)	3.62	0.83	1.53	3.32	1.61
Direct	XG-Boost (4.4)	4.54	2.36	1.5	1.7	1.8
Direct	OLS f-stepwise (5.6)	4.07	2.98	3.56	2.33	3.0
Direct	LASSO (6.0)	4.83	3.12	3.46	2.44	2.45
Direct	Persistence (8.8)	12.0	15.45	14.91	14.16	9.88
Direct	Feedforward NN (9.6)	25.06	7.8	5.09	20.62	18.67
Indirect	Persistence (9.8)	13.13	15.86	15.22	16.29	10.85
Indirect	ECMWF (9.8)	18.85	8.63	8.02	19.64	28.53

Table: Average NRMSE degradation w.r.t. best predictor for wind power prediction ($\times 10^{-2}$)

Questions ?

Thank you !

Paper available on Arxiv:

D. Bouche, R. Flamary, F. d'Alché Buc, R. Plougonven,

M. Clausel, J. Badosa, and P. Drobinski. Wind power predictions from nowcasts to 4-hour forecasts: a learning approach with variable selection, 2022.
https://arxiv.org/abs/2204.09362

Selected bibliography I

D. Bouche, R. Flamary, F. d'Alché Buc, R. Plougonven, M. Clausel, J. Badosa, and P. Drobinski. Wind power predictions from nowcasts to 4-hour forecasts: a learning approach with variable selection, 2022. https://arxiv.org/abs/2204.09362.

P. Drineas and M. Mahoney W. On the nystrom method for approximating a gram matrix for improved kernel-based learning. *Journal of Machine Learning Research*, 6:2153–2175, 2005.

A. Dupré, P. Drobinski, B. Alonzo, J. Badosa, C. Briard, and R. Plougonven. Sub-hourly forecasting of wind speed and wind energy. *Renewable Energy*, 145:2373–2379, 2020.

A. Gretton, K. Fukumizu, C. Teo, L. Song, B. Schölkopf, and A. Smola. A kernel statistical test of independence. In *Advances in Neural Information Processing Systems*, volume 20, 2008.

Selected bibliography II

L. Song, A. Smola, A. Gretton, J. Bedo, and K. Borgwardt. Feature selection via dependence maximization. *Journal of Machine Learning Research*, 13:1393–1434, 2012.

R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996.

Q. Zhang, S. Filippi, A. Gretton, and D. Sejdinovic. Large-scale kernel methods for independence testing. *Statistics and Computing*, 28, 1 2018.