



Forecasting Regional Wind Production based on weather similarity and site clustering for the EEM20 Competition

Kévin Bellinguer, Valentin Mahler, **Simon Camal**, George Kariniotakis
MINES ParisTech - PSL University / ARMINES, Centre PERSEE,
Centre for processes, renewable energies and energy systems

Contacts:

kevin.bellinguer@mines-paristech.fr

valentin.mahler@mines-paristech.fr

simon.camal@mines-paristech.fr

International Symposium on Forecasting
Online, 26th October 2020



PERSEE is one of the 18 research centers of MINES ParisTech - ARMINES.

A leading research organization on renewables integration since '80s.

A long experience on the field of renewable energy forecasting and related applications are a cornerstone of PERSEE's activities since 1990

- Past projects: ANEMOS, ANEMOS.plus, SafeWind, REstable...
- Ongoing projects: Smart4RES, REgions, XFLEX Hydro...
- Industrial and academic collaborations on forecasting

Contribute to our survey to know your needs and interests about future forecasting developments on renewable energy



<https://ec.europa.eu/eusurvey/runner/02027011-b975-7860-1c35-4037eb359764>

I. The EEM20 competition

- 1) Objectives and available data
- 2) Modelling challenges

II. Our approach

- 1) Model overview
- 2) Focus on dimensionality reduction
- 3) Focus on feature engineering

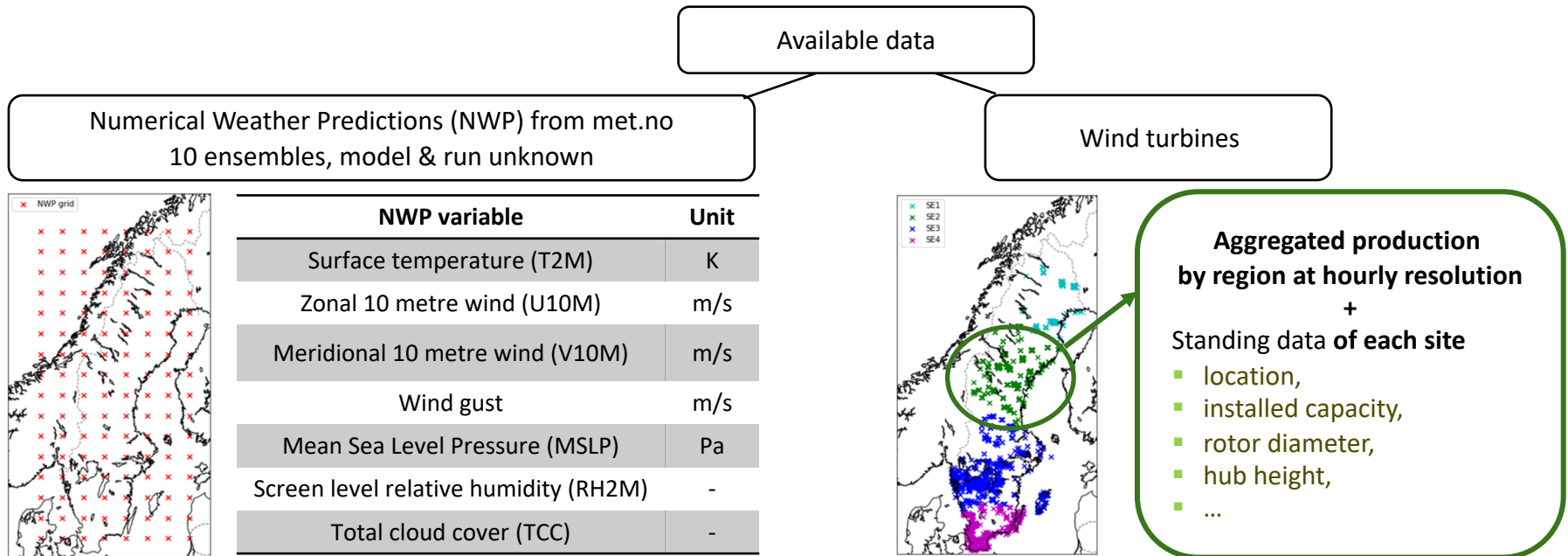
III. Results and discussion

IV. Conclusion

The EEM20 competition

 Forecast the aggregated hourly wind production in the four price regions of Sweden

- Day-ahead Probabilistic forecast (quantiles)
- 1 year of initial training set, 1 year for the full testing set split in 6 successive tasks



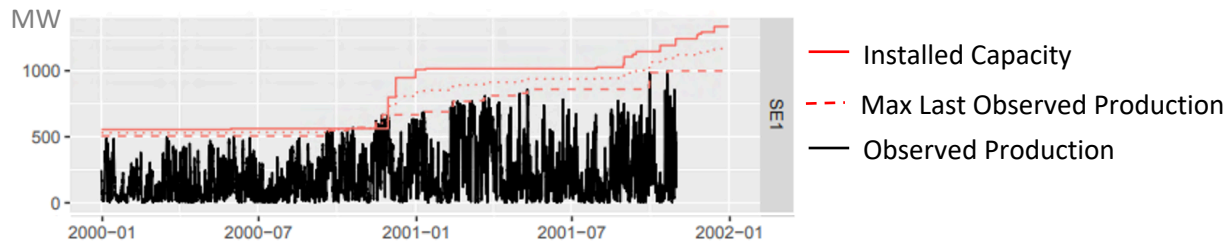
See also: <https://eem20.eu/forecasting-competition/>

- 10 NWP ensembles at national scale, 4000 turbines = **Large data problem**

↗ How to downscale to the **most relevant inputs?**

- The **installed capacity evolves** over time in all regions

 How to adapt the forecasting model to the **dynamic installed capacity?**



- Little is known about the contribution of turbines to the aggregated production



Consider **physics** of wind power generation to approximate better local conditions?



How to derive **direct forecasts of regional production** that are precise and reliable?

Dimensionality reduction

Selection of weather predictions within the price regions

Selection of a few key locations

Feature engineering

Mean of historical normalized production observed during similar weather situations

Weather-based features

Temporal features (hour, day of the week)

Regression model

Quantile Regression Forests (QRF)

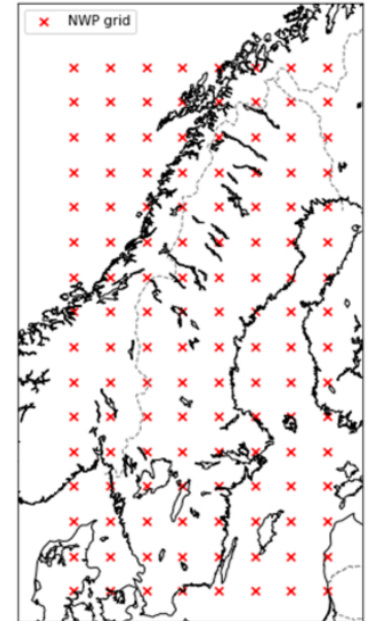
Target engineering

Denormalization of forecasted production

! Lots of available meteorological data (NWP)

- $7 \cdot 10 \cdot 169 \cdot 71 \approx 840,000$
physical quantities *ensembles* *x coordinates* *y coordinates* *NWP variables per hour*
- Around 10,000 hours in the training dataset

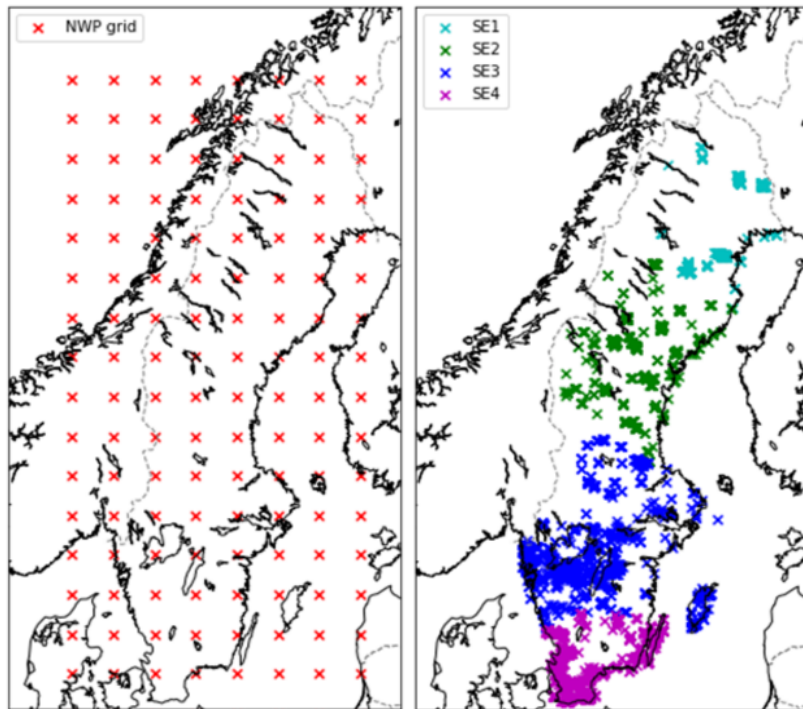
Ten-to-one downsampled version of the gridded NWP data



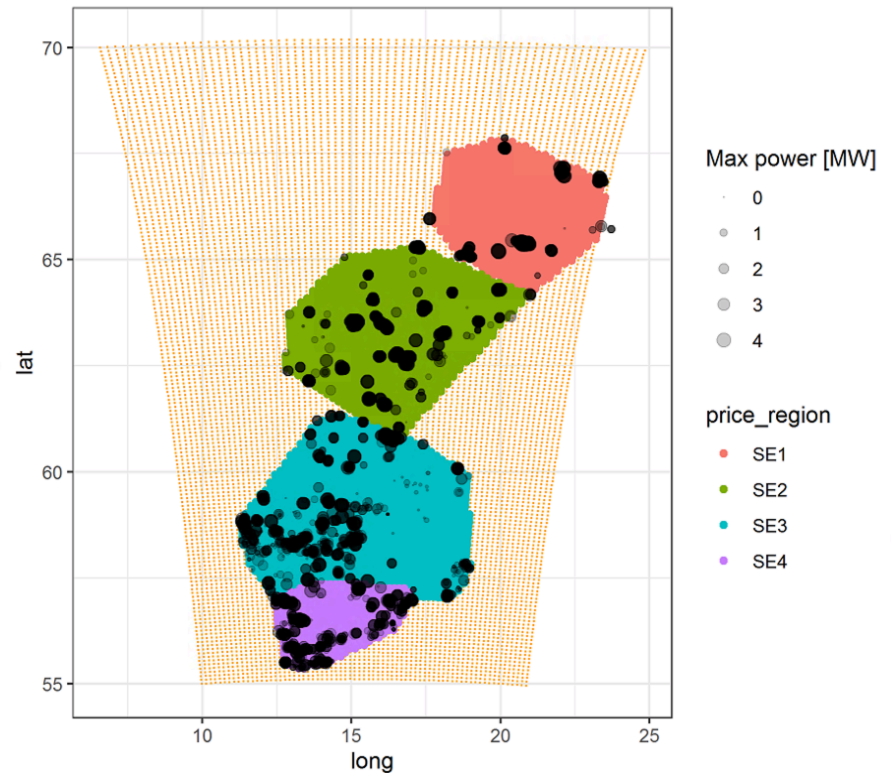
Dimension to handle	Modelling strategy
Physical quantities	➤ Focus on the most relevant variable (wind speed, temperature, ...)
Ensemble structure	➤ Consider only the mean of the 10 ensembles [Juban 2008]
Spatial extent	➤ Match NWP to price regions ➤ Select only a few key locations

Matching of NWP and price regions

From two disjointed datasets...



... to a selection of NWP points within the convex hulls defined by the turbines



Selection of a few key locations



Identify clusters of wind turbines and assign a unique point to each

Location and maximal power of wind turbines in a price region

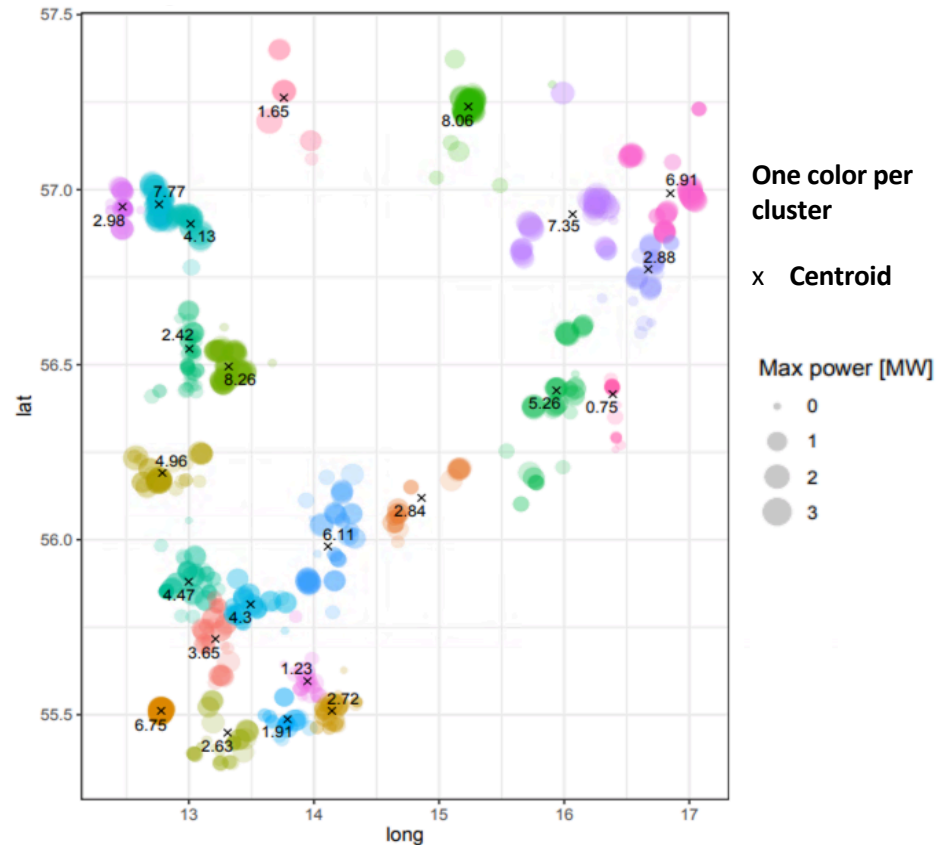


Define clusters using the weighted k-means approach



Keep only the center of each cluster (i.e. centroids)

Clusters of wind turbines for the southern swedish price region (SE4)



Reference: https://github.com/pararawendy/weighted_kmeans_R



Select past values of normalized production observed during similar weather conditions

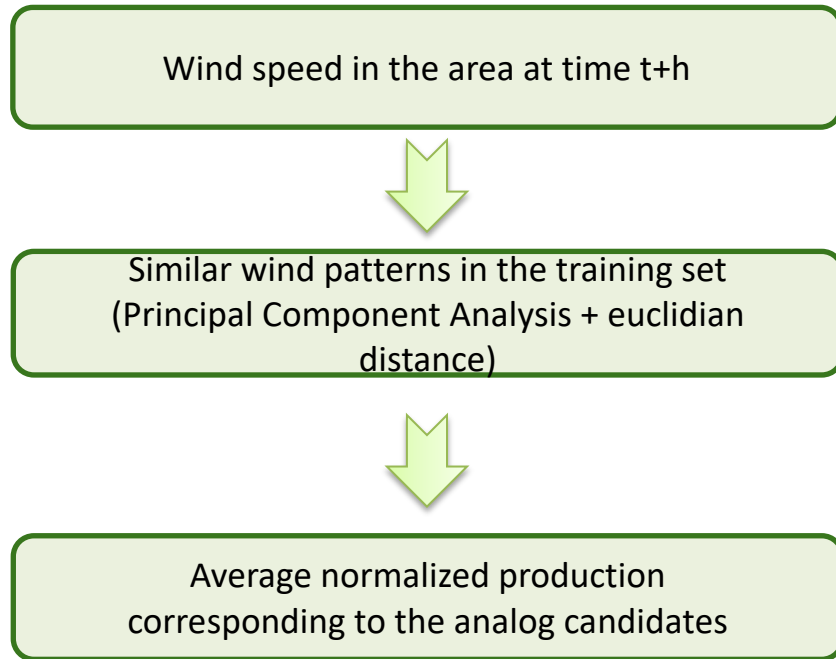
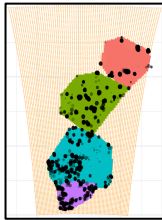
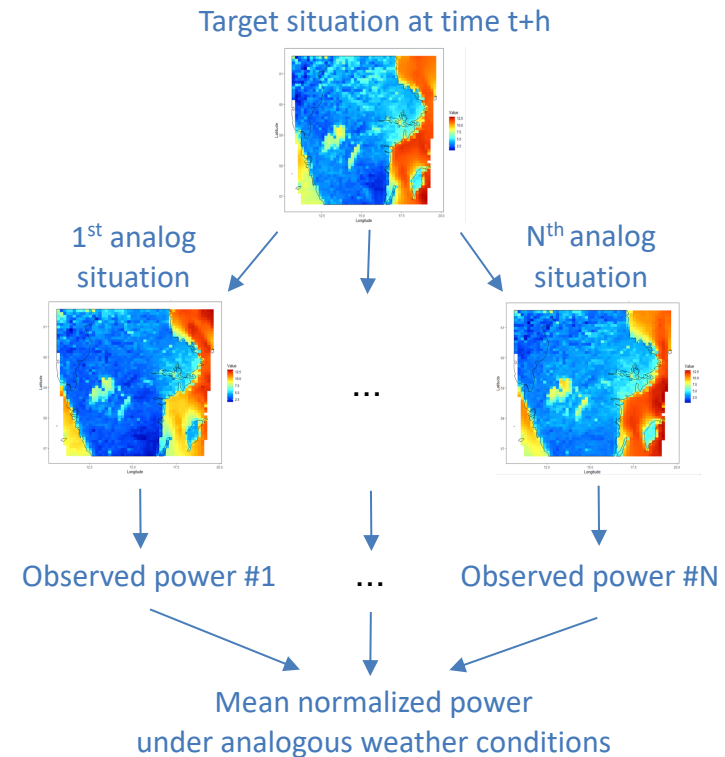
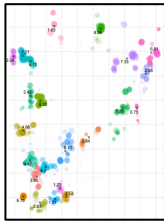


Illustration of the analog-based approach [Bellinguer 2020]





Explore the causes of wind power generation in order to create physically relevant features



Physical phenomena		Model features
Wind speed at higher altitude	➡	➤ Extrapolation of wind speed (Hellmann exponent)
Absolute wind energy	➡	➤ Maximal recoverable wind energy - cube of wind speed - air density
Temporal dependencies (inertia, actions of operators, ...)	➡	➤ Wind energy at t-1 and t+1
Spatial dependencies (non-isotropic phenomena, ...)	➡	➤ Wind orientation
Varying efficiency of the components (mechanical, electrical and electronic)	➡	➤ Temperature ➤ Relative humidity
Potential operational constraints (low demand, maintenance, ...)	➡	➤ Hour of the day ➤ Day of the week

- Evaluation of performances

- Score based on pinball loss function

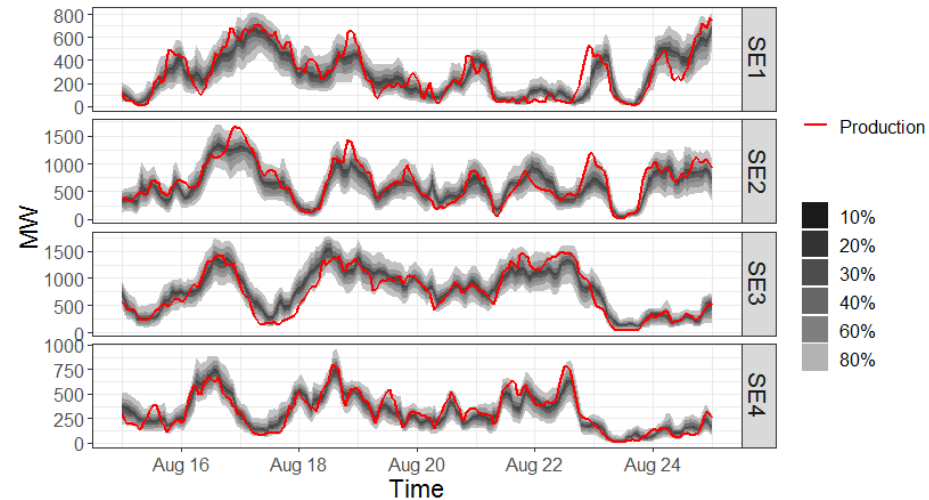
- An empirical approach

- Ongoing incremental improvements (only final solution presented)
 - Time-constrained competition
 - Further study and comparison to state-of-the-art are needed

- Backtesting of the best model

- Global improvement of the forecasting performances

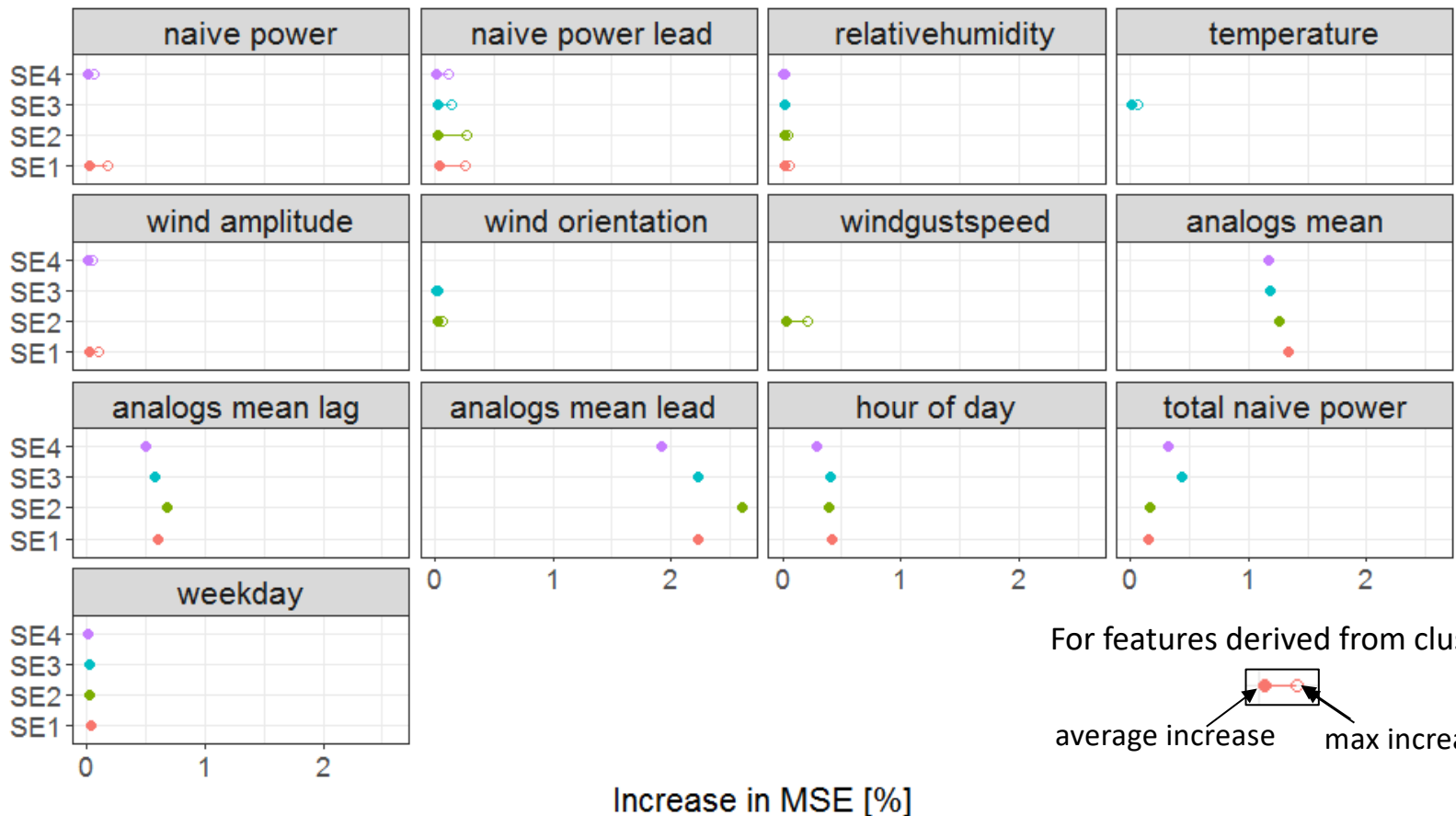
Example of predictions on 4 zones



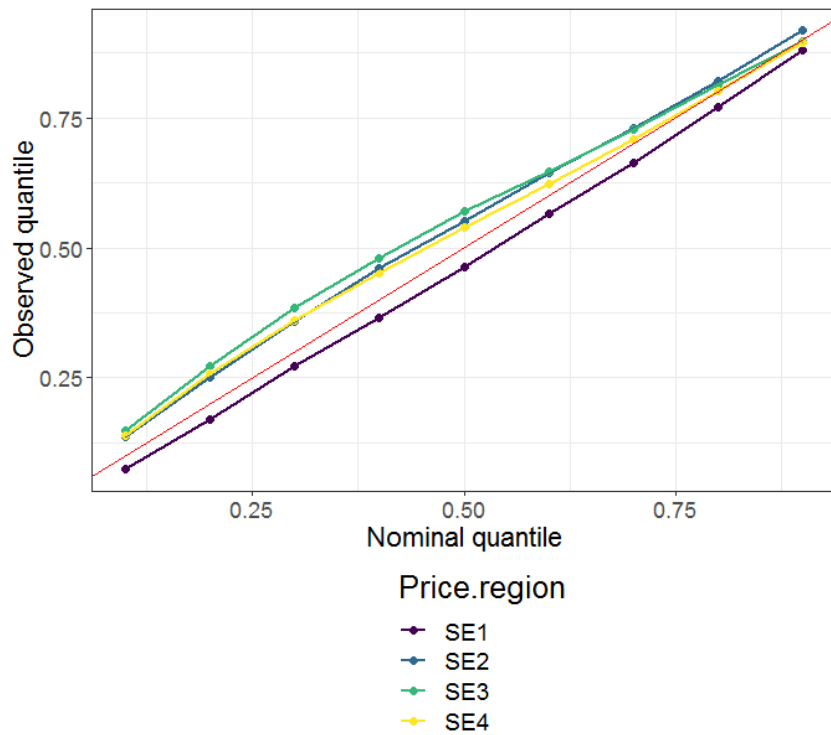
	Task 1 <i>Jan-Feb</i>	Task 2 <i>Mar-Apr</i>	Task 3 <i>Mai-Jun</i>	Task 4 <i>Jul-Aug</i>	Task 5 <i>Sep-Oct</i>	Task 6 <i>Nov-Dec</i>	Score EEM20
Submitted forecasts	58,96	52,59	38,39	34,73	42,92	55,97	44,92
Forecasts with latest model	58,36	52,11	37,56	33,07	43,03	55,97	44,35

■ Evaluation of performances

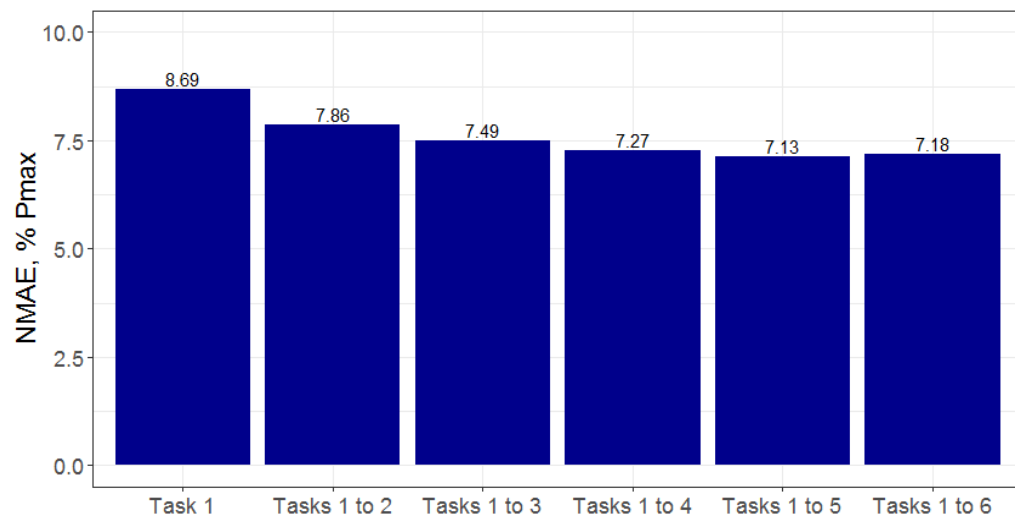
- Importance quantified by increase in MSE when a given feature is replaced by a random permutation of its values on out-of-bag samples
- Features among the 10 more important for each Region at Task 6 (around 300 features per model)



Reliability diagram



Evolution of the NMAE throughout the competition



Conclusion

- Our strategy: incremental model updates led to improvements
- A promising model for regional wind production forecasts
- Strengths of the model
 - Potential for real-life applications (open data, simple and quite robust model, reasonable computing time, ...)
 - Potential application to other regions and/or other renewable energy sources (PV, run-of-river hydro)
- Some perspectives
 - Further ideas will be developed in an upcoming paper

Thank you for your attention!

Acknowledgments

- La Compagnie nationale du Rhône (CNR)
for supporting the PhD thesis of Kévin Bellinguer
- ADEME and ARMINES
for supporting the PhD thesis of Valentin Mahler
- The European Commission
for supporting Smart4RES project
managed by Simon Camal

