RSE Ricerca Sistema Energetico | we move rsearch

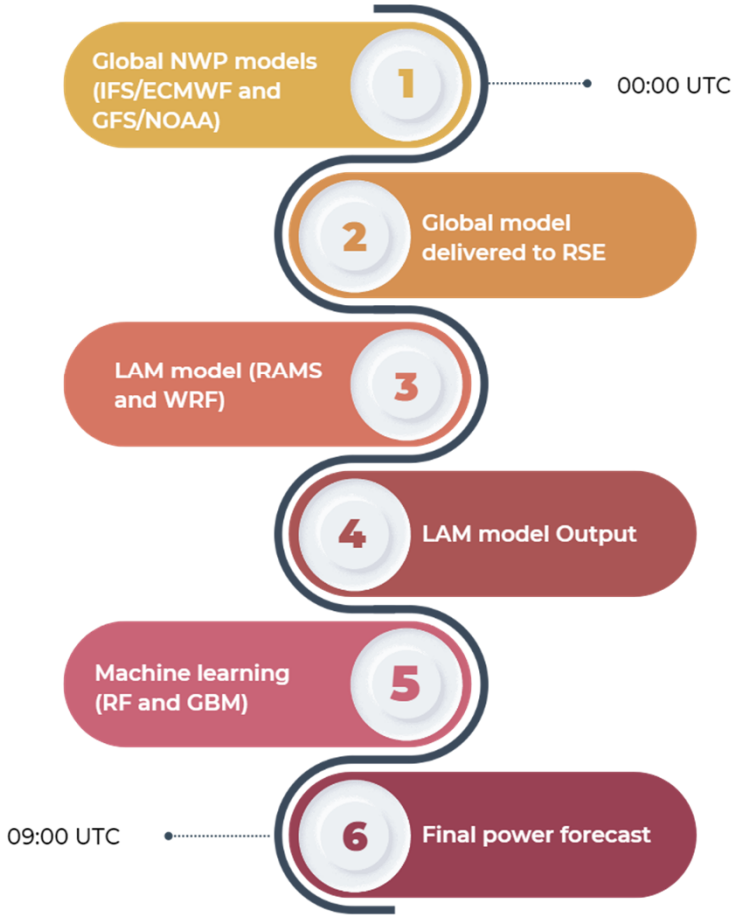
Discussion on improving efficiency in machine learning wind power forecast for operational purposes

Daniele Perona, Dario Ronzio, Elena Collino

Wind Farms

- 8 Wind Farms located in Southern Italy
- 3 days ahead hourly Forecast used in the Day-Ahead Energy Market
- Forecast is made using multiple machine learning models (like Random Forest and Gradient Boosting Machine)
- Forecast must be issued by 09:00 AM UTC

TIMELINE



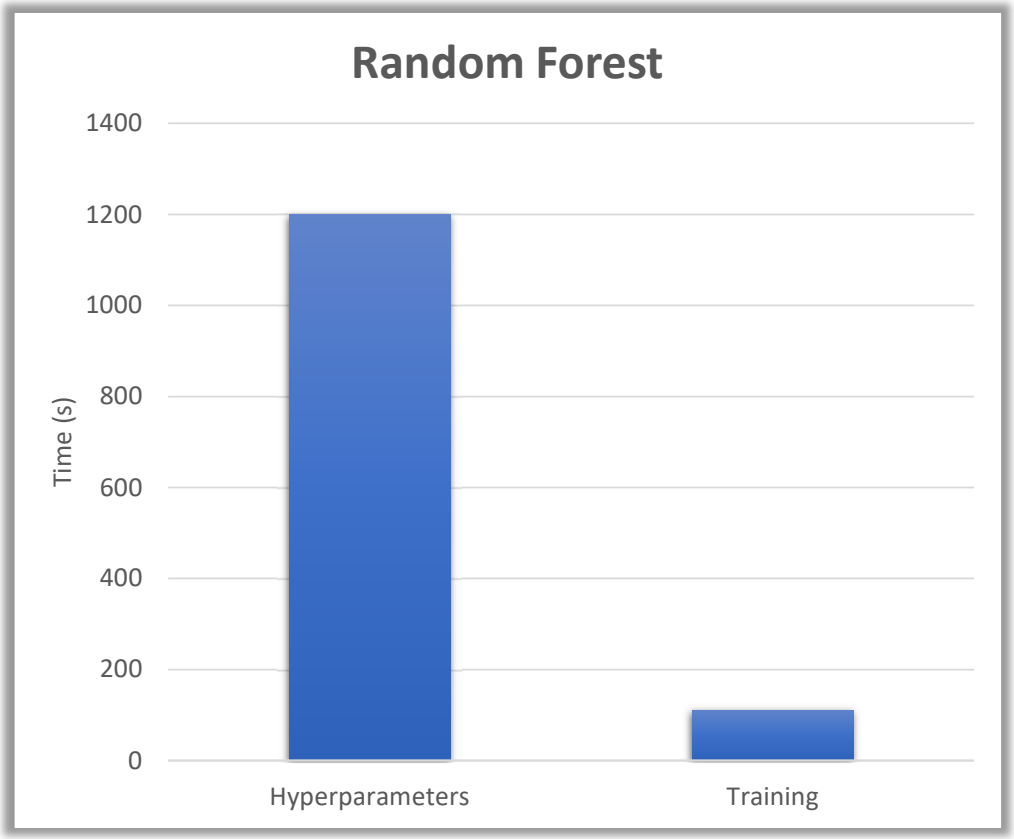
- The operational chain Start around 00:00 UTC
- 2 Global Models and 2 Local Area Models
- The only step of the operational chain that can be modified is the training of machine learning models

Random Forest	Gradient Boosting Machine
Runs in parallel	**Does not** run in parallel
Not significantly influenced by hyperparameters	Significantly influenced by hyperparameters

Machine Learning training dataset

7 Years of Training data (about 36,000 records for every variable).

- Time of day
- Temperature at 2m
- Sea-level pressure
- Wind direction
- Wind Intesity
- Solar zenith angle



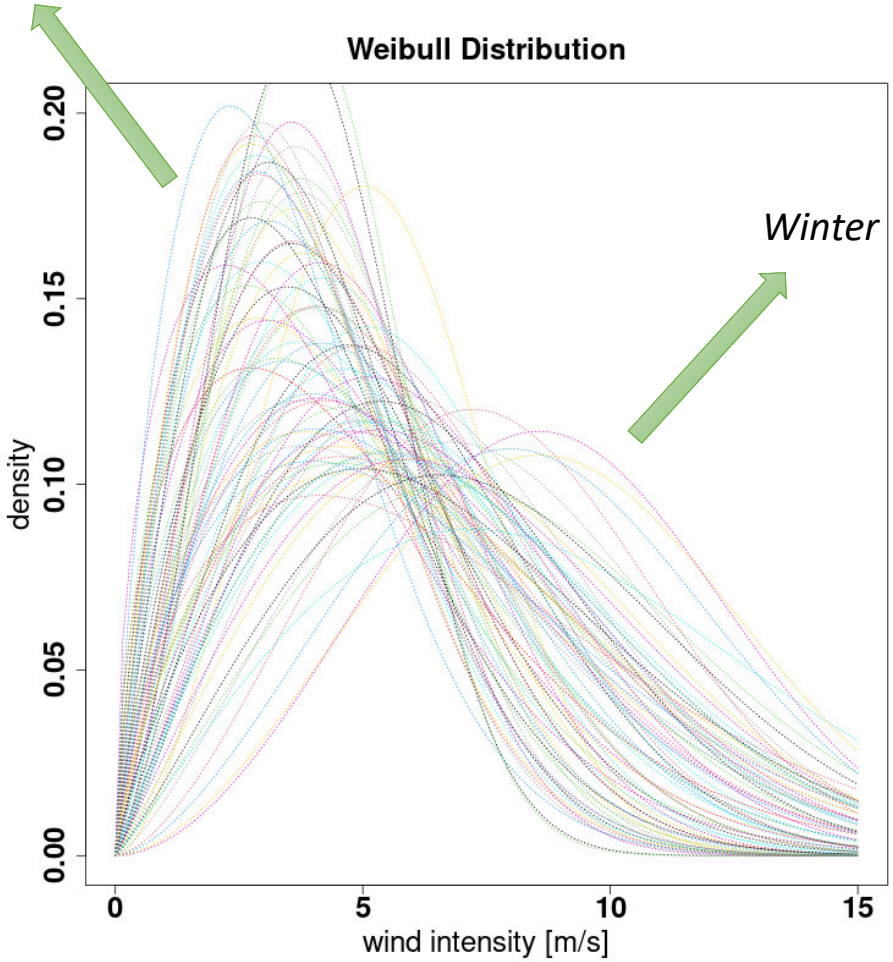
10 Intel(R) Xeon(R) Gold 5218R CPU @ 2.10GHz.

Hyperparameters search

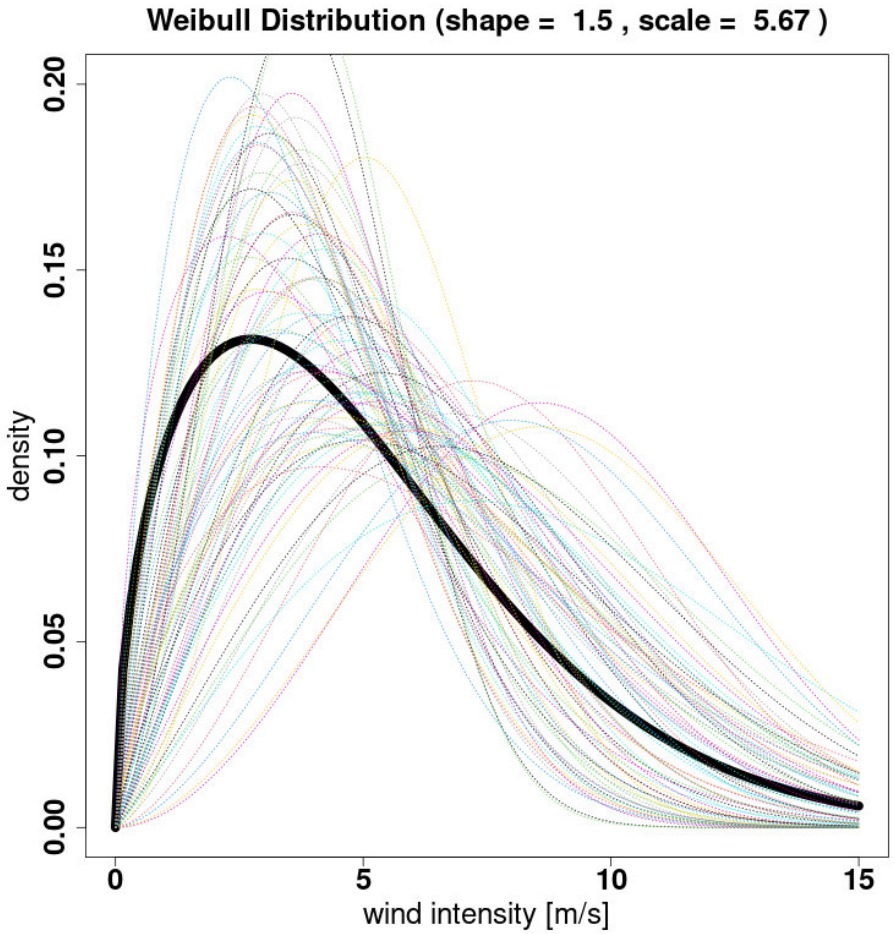
Training

Forecasting

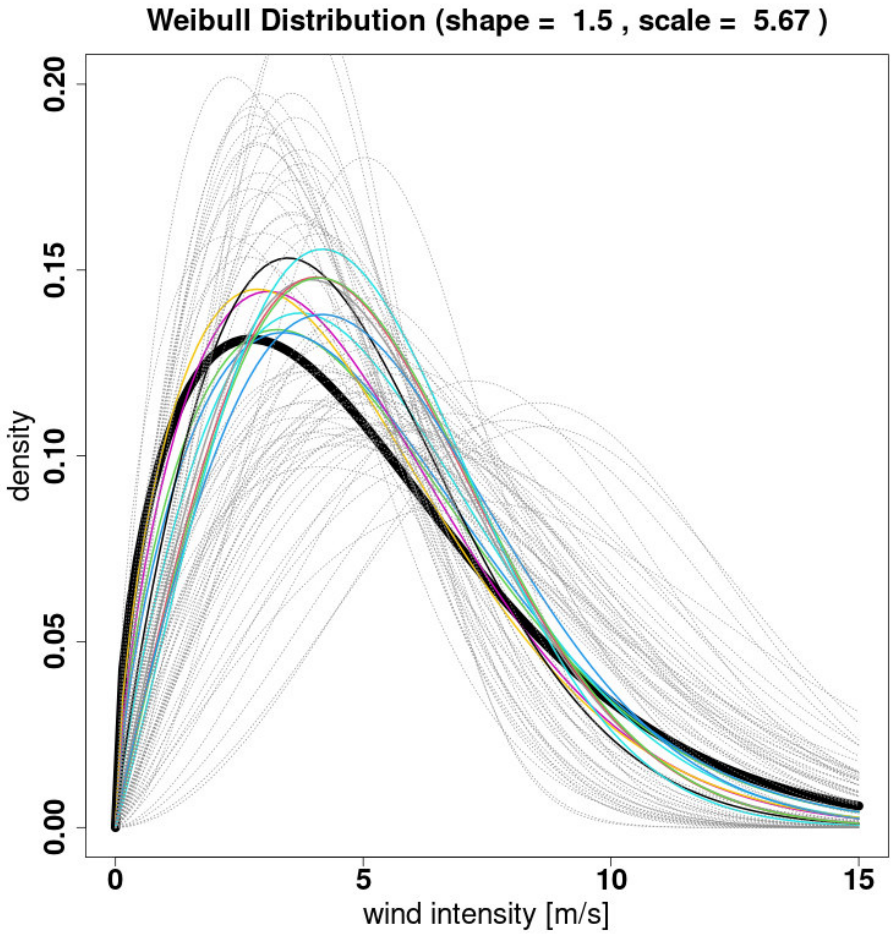
Summer



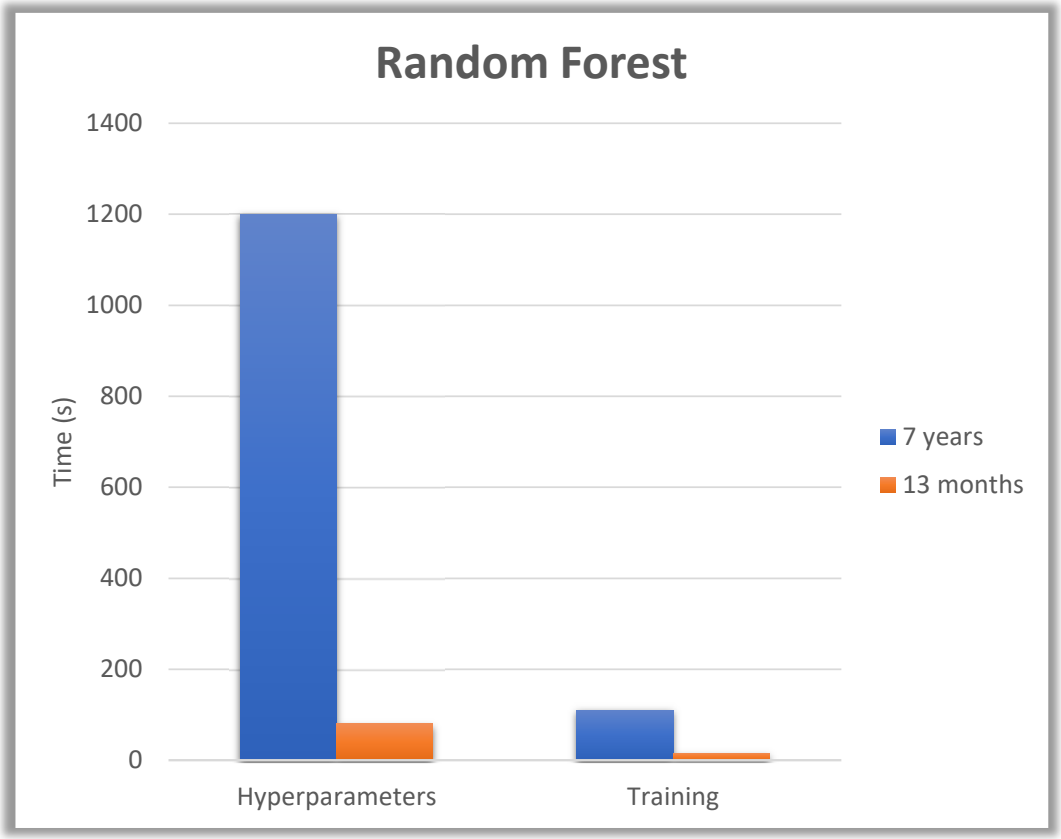
- High variability
- The wind speed exhibits seasonality patterns
- More wind power production during winter period



- Searching for 30-day periods in the past, where the predicted wind Weibull distributions exhibit similarity to the distributions observed in the last 30 days.
- The similarity between distributions is determined by calculating the Euclidean distance between the scale and shape parameters of the distributions



- Searching for 30-day periods in the past, where the predicted wind Weibull distributions exhibit similarity to the distributions observed in the last 30 days.
- The similarity between distributions is determined by calculating the Euclidean distance between the scale and shape parameters of the distributions

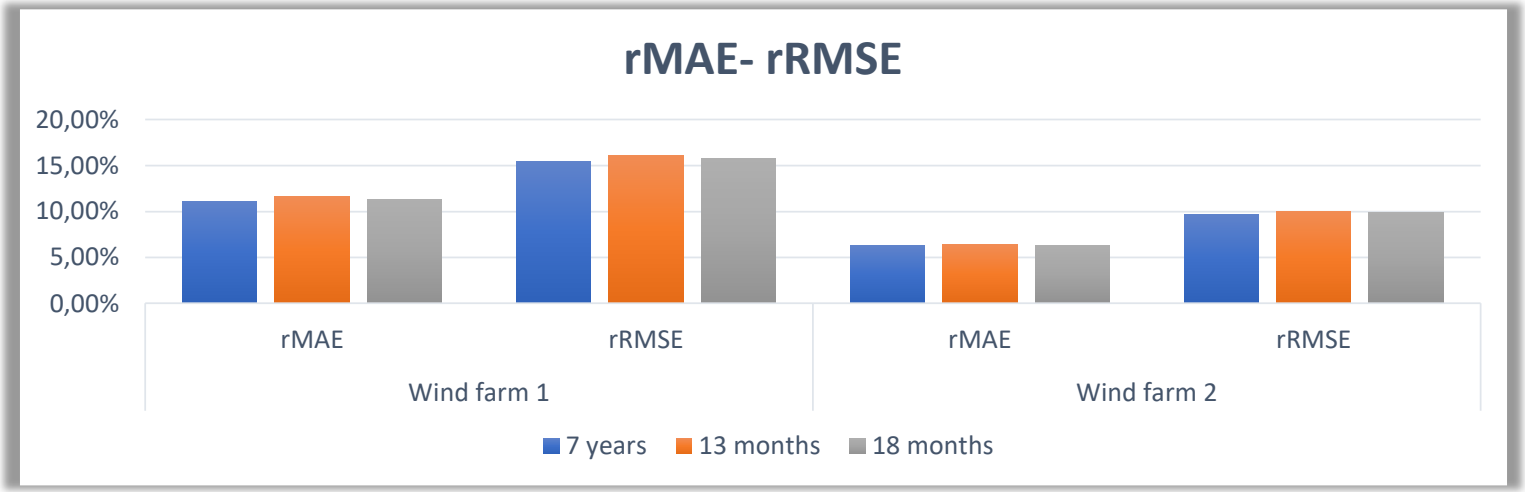


Hyperparameters search

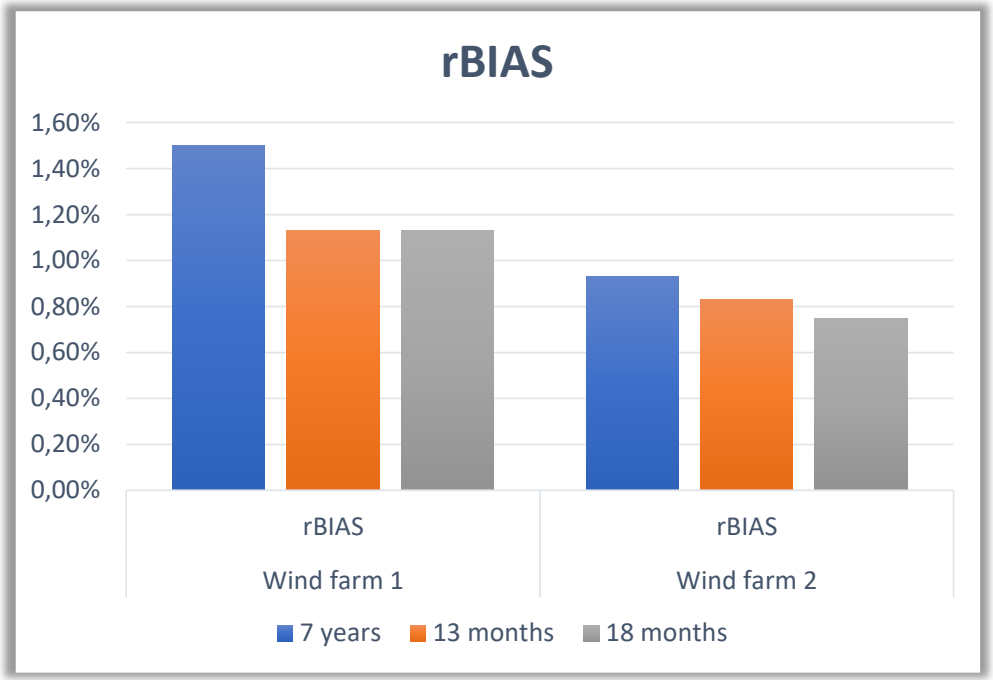
Training

Forecasting

10 Intel(R) Xeon(R) Gold 5218R CPU @ 2.10GHz.



	Wind farm 1				Wind farm 2			
	R	rMAE	rRMSE	rBIAS	R	rMAE	rRMSE	rBIAS
7 years	0.842	11.09%	15.46%	1.50%	0.898	6.26%	9.66%	0.93%
13 months	0.825	11.56%	16.13%	1.13%	0.890	6.40%	9.99%	0.83%
18 months	0.834	11.34%	15.76%	1.13%	0.892	6.32%	9.89%	0.75%



- BIAS is lower for shorter dataset
- It is possible to create an ensemble of elementary predictions by leveraging different meteorological models and post-processing techniques.
- These predictions can then be fed into an optimizer to generate a single prediction (using a Quantile Random Forest) along with an estimation of its reliability.

Conclusion

- The time constraints imposed by the day-ahead market need to be considered in the operational mode
- The machine learning models makes use of a carefully selected training dataset that includes a limited range of months (13 or 18), from 2017 onwards
- The selection of this training dataset offers several advantages:
  - Substantial reduction in computation time while maintaining performance
  - Enables daily hyperparameter tuning for enhanced optimization
  - Having lower biases facilitates the generation of an optimizer

RSE Ricerca Sistema Energetico | we move rsearch

Daniele Perona daniele.perona@rse-web.it

Dario Ronzio dario.ronzio@rse-web.it

Elena Collino elena.collino@rse-web.it