solute

People just want numbers – How to fairly compare and interpret forecasts with a benchmarking framework for performance evaluation

IEA Wind Task 51 Workshop

13-09-2022

Juan Manuel González Sopeña

Background

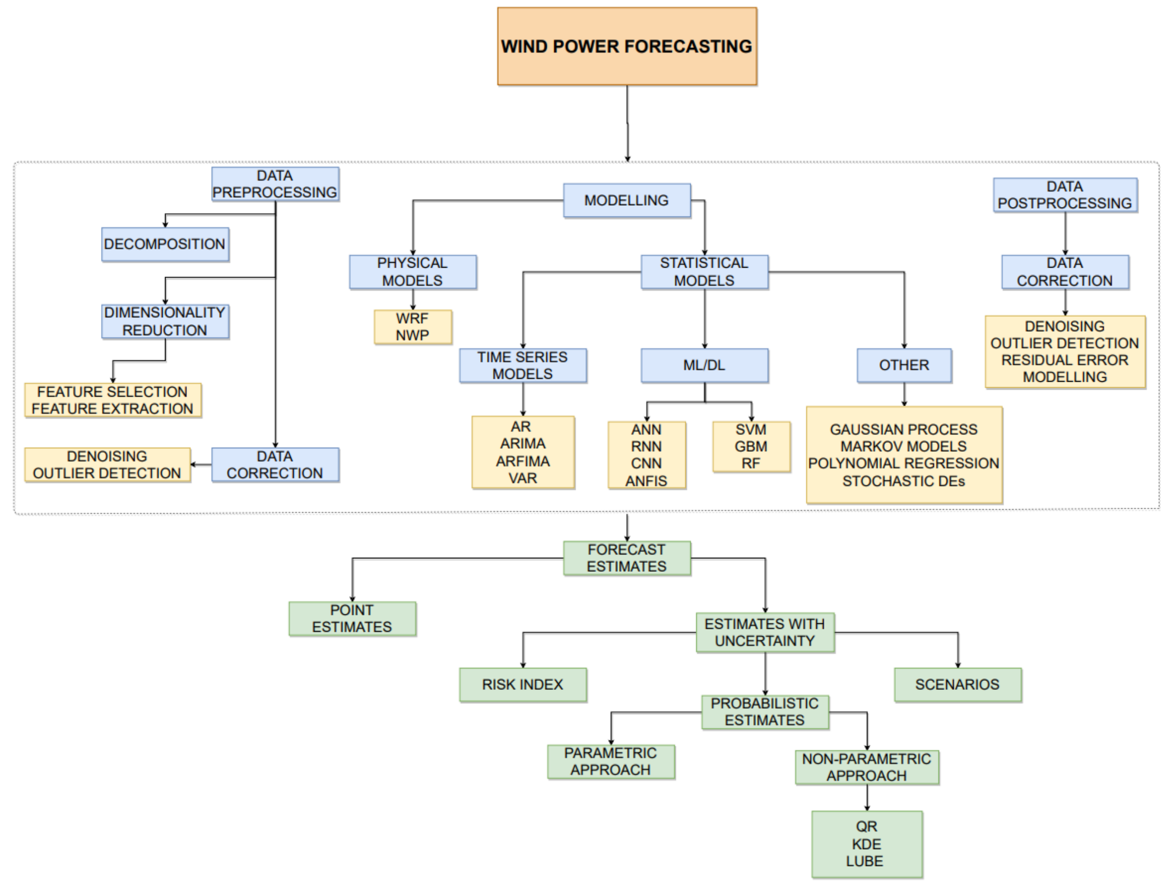
- PhD researcher on wind power forecasting at Trinity College Dublin.

- Project Engineer at Solute -> Development of energy forecasting tools (https://aphelion.com.es/)

How are wind power forecasts modelled?

Hundreds of publications on wind power forecasting published every year

-> and everybody claims their models to be the best!



How is this possible?

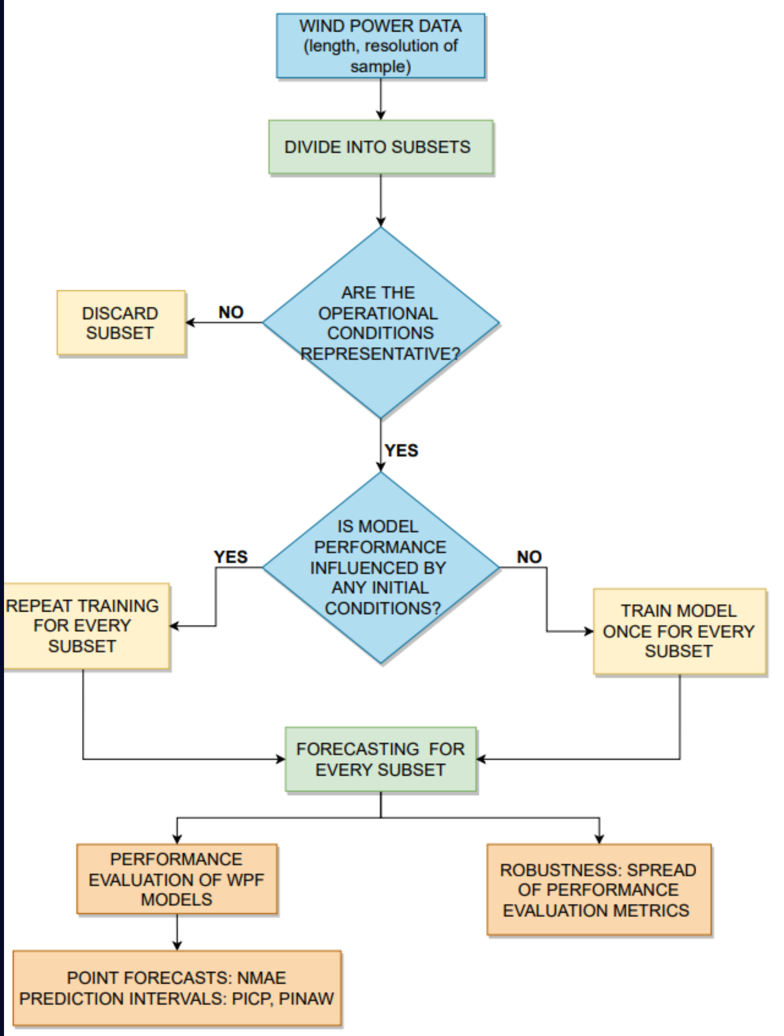
1. Lack of standards to evaluate wind power forecasting models.
2. Lack of understanding to develop an appropriate experimental design.
3. Lack of understanding to select performance evaluation metrics.
4. Datasets might not be representative (e. g., testing periods too short).
5. Lack of details to reproduce the experiment.

NEED OF DEVELOPING BENCHMARKS!!

Why do we need benchmarks?

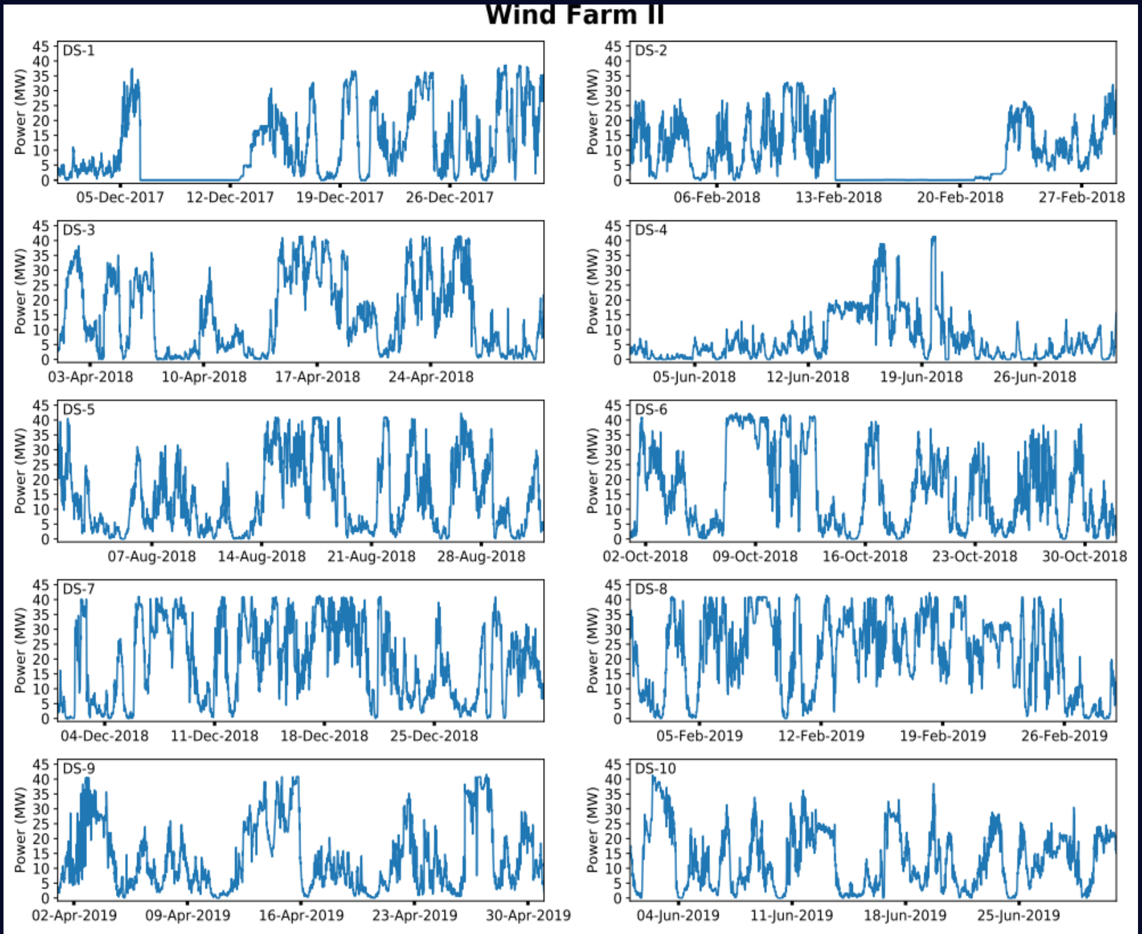
Benchmarking framework basic principles

- Fair, representative
- Prediction horizon
- Historical data available
- Representativeness of the operational conditions
- Standardized performance evaluation metrics


WIND POWER DATA (length, resolution of sample)
DIVIDE INTO SUBSETS
ARE THE OPERATIONAL CONDITIONS REPRESENTATIVE?
NO
DISCARD SUBSET
YES
IS MODEL PERFORMANCE INFLUENCED BY ANY INITIAL CONDITIONS?
YES
REPEAT TRAINING FOR EVERY SUBSET
NO
TRAIN MODEL ONCE FOR EVERY SUBSET
FORECASTING FOR EVERY SUBSET
PERFORMANCE EVALUATION OF WPF MODELS
ROBUSTNESS: SPREAD OF PERFORMANCE EVALUATION METRICS
POINT FORECASTS: NMAE
PREDICTION INTERVALS: PICP, PINAW

Example: benchmark for very short-term forecasting

- SCADA data from two Irish wind farms.
- Turbine-level, recorded at 10-minute resolution.
- Maximize representativeness of the benchmark.



Example: benchmark for very short-term forecasting

- 21 models based on decomposition algorithms and artificial intelligence
- Ideally, this should be extended to other methodologies (e. g., vector autoregression)
- Extension to probabilistic representations of forecasts

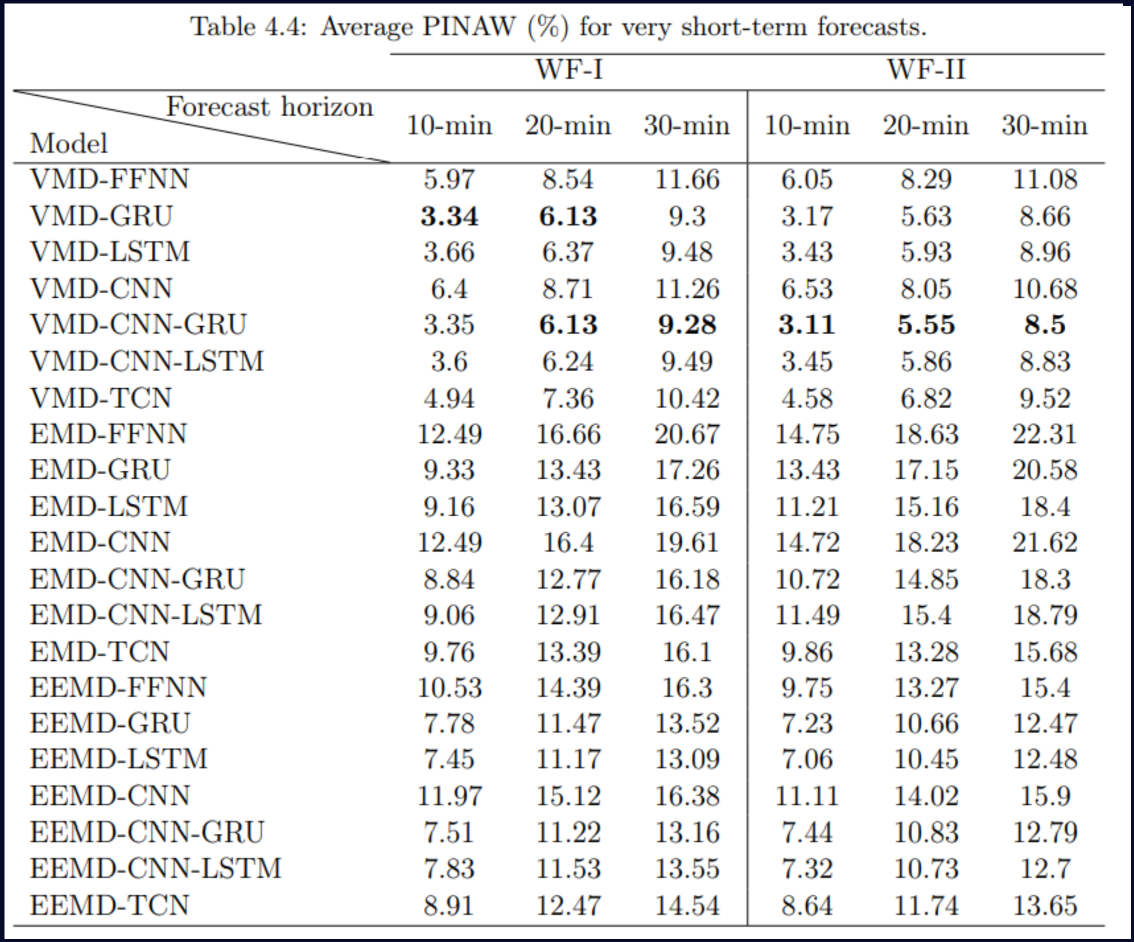


Table 4.4: Average PINAW (%) for very short-term forecasts.

Forecast horizon / Model	WF-I			WF-II		
	10-min	20-min	30-min	10-min	20-min	30-min
VMD-FFNN	5.97	8.54	11.66	6.05	8.29	11.08
VMD-GRU	**3.34**	**6.13**	9.3	3.17	5.63	8.66
VMD-LSTM	3.66	6.37	9.48	3.43	5.93	8.96
VMD-CNN	6.4	8.71	11.26	6.53	8.05	10.68
VMD-CNN-GRU	3.35	**6.13**	**9.28**	**3.11**	**5.55**	**8.5**
VMD-CNN-LSTM	3.6	6.24	9.49	3.45	5.86	8.83
VMD-TCN	4.94	7.36	10.42	4.58	6.82	9.52
EMD-FFNN	12.49	16.66	20.67	14.75	18.63	22.31
EMD-GRU	9.33	13.43	17.26	13.43	17.15	20.58
EMD-LSTM	9.16	13.07	16.59	11.21	15.16	18.4
EMD-CNN	12.49	16.4	19.61	14.72	18.23	21.62
EMD-CNN-GRU	8.84	12.77	16.18	10.72	14.85	18.3
EMD-CNN-LSTM	9.06	12.91	16.47	11.49	15.4	18.79
EMD-TCN	9.76	13.39	16.1	9.86	13.28	15.68
EEMD-FFNN	10.53	14.39	16.3	9.75	13.27	15.4
EEMD-GRU	7.78	11.47	13.52	7.23	10.66	12.47
EEMD-LSTM	7.45	11.17	13.09	7.06	10.45	12.48
EEMD-CNN	11.97	15.12	16.38	11.11	14.02	15.9
EEMD-CNN-GRU	7.51	11.22	13.16	7.44	10.83	12.79
EEMD-CNN-LSTM	7.83	11.53	13.55	7.32	10.73	12.7
EEMD-TCN	8.91	12.47	14.54	8.64	11.74	13.65

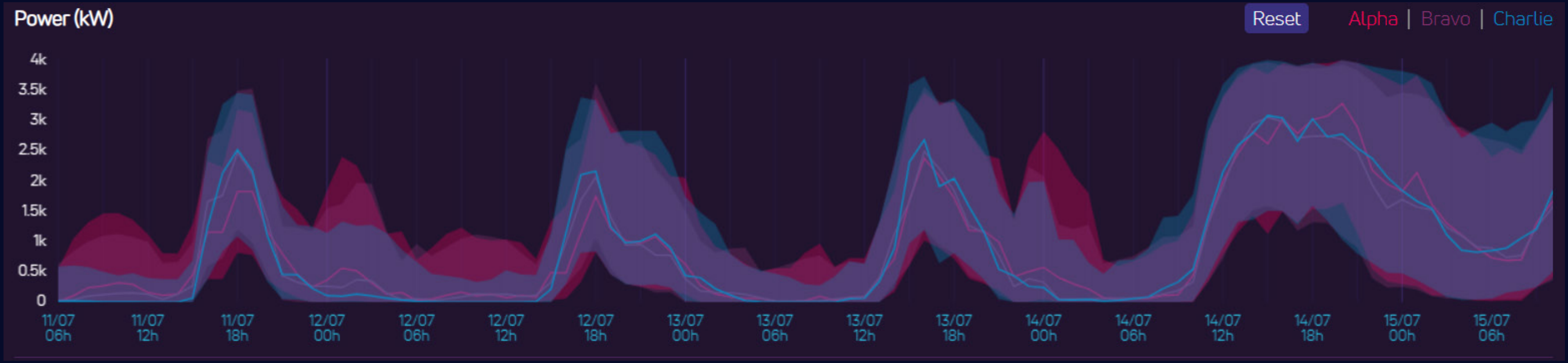
How can we keep improving a benchmark?

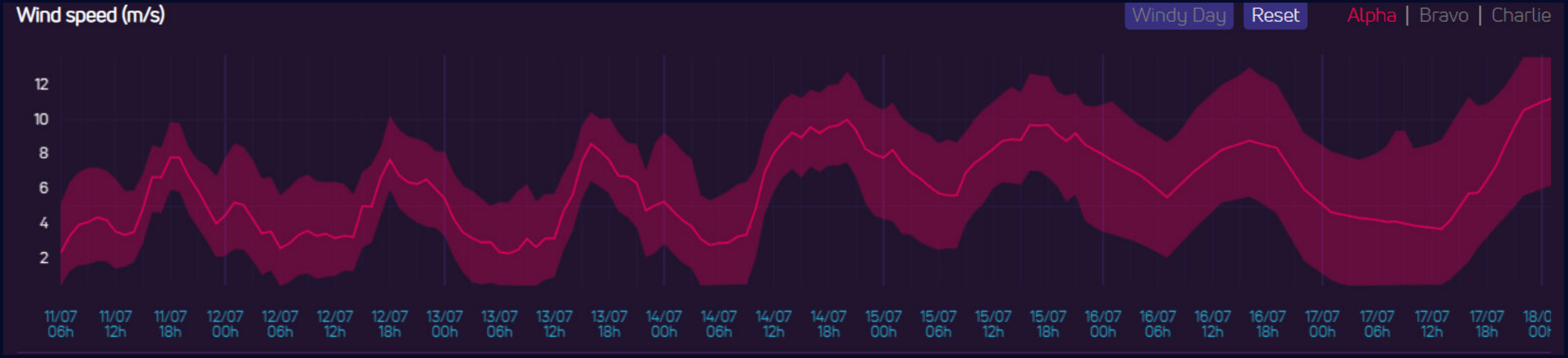
1. Keep implementing state-of-the-art methodologies for your specific case study.

2. If a specific pre-/post-processing technique is used, effects of selecting variants of these techniques (such as user-defined parameters).

3. Extension to other datasets and regions of interest (ideally open-source data!)

Aphelion Wind





solute

Thanks for your attention! – any questions?

juanmanuel.gonzalez@solute.es
+34 646 791 796

Avenida Cerro del Águila, 3
28703 San Sebastián de los Reyes
Madrid

Carretera de l’Hospitalet, 147
Edif. Lisboa 3ºA
08940 Cornellà de Llobregat
Barcelona

Avenida de la Investigación, s/n Edif. Parque
Científico y Tecnológico de Extremadura
06006 Badajoz
Badajoz

www.solute.es