

Egyetemi Hallgatói Képviselőt
Egyetemi BME ösztöndíj
Előadás adatlap
2022/23. tanév

Adatok:

Név:	Markovics Dávid
Neptun-kód:	CNEJCK
Képzés kódja:	2N-ME0
Téma kódja:	A
Beszámolási időszak:	2022.09.01. - 2023.01.31.

Előadásra vonatkozó adatok:

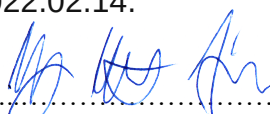
A megfelelő bekarikázandó:

HAZAI	NEMZETKÖZI
-------	------------

Előadás pontos címe:	Photovoltaic power forecasting using machine learning
Előadás nyelve:	angol
Előadás dátuma ¹ :	2022-10-26
Előadás helye ² :	Dublin, Írország
Előadás elérhetősége (amennyiben van):	
Rendezvény megnevezése:	AEE Europe 2022
Témavezető neve:	Dr. Mayer Martin János
Oktatási szervezeti egysége:	GPK/EGR

	Szerző(k)	Szerző(k) munkájának %-os megoszlása ³	Szerző(k) aláírása
1.	Markovics Dávid	100%	
2.			
3.			
4.			
5.			
6.			
7.			

Budapest, 2022.02.14.


.....
témavezető aláírása


.....
pályázó aláírása

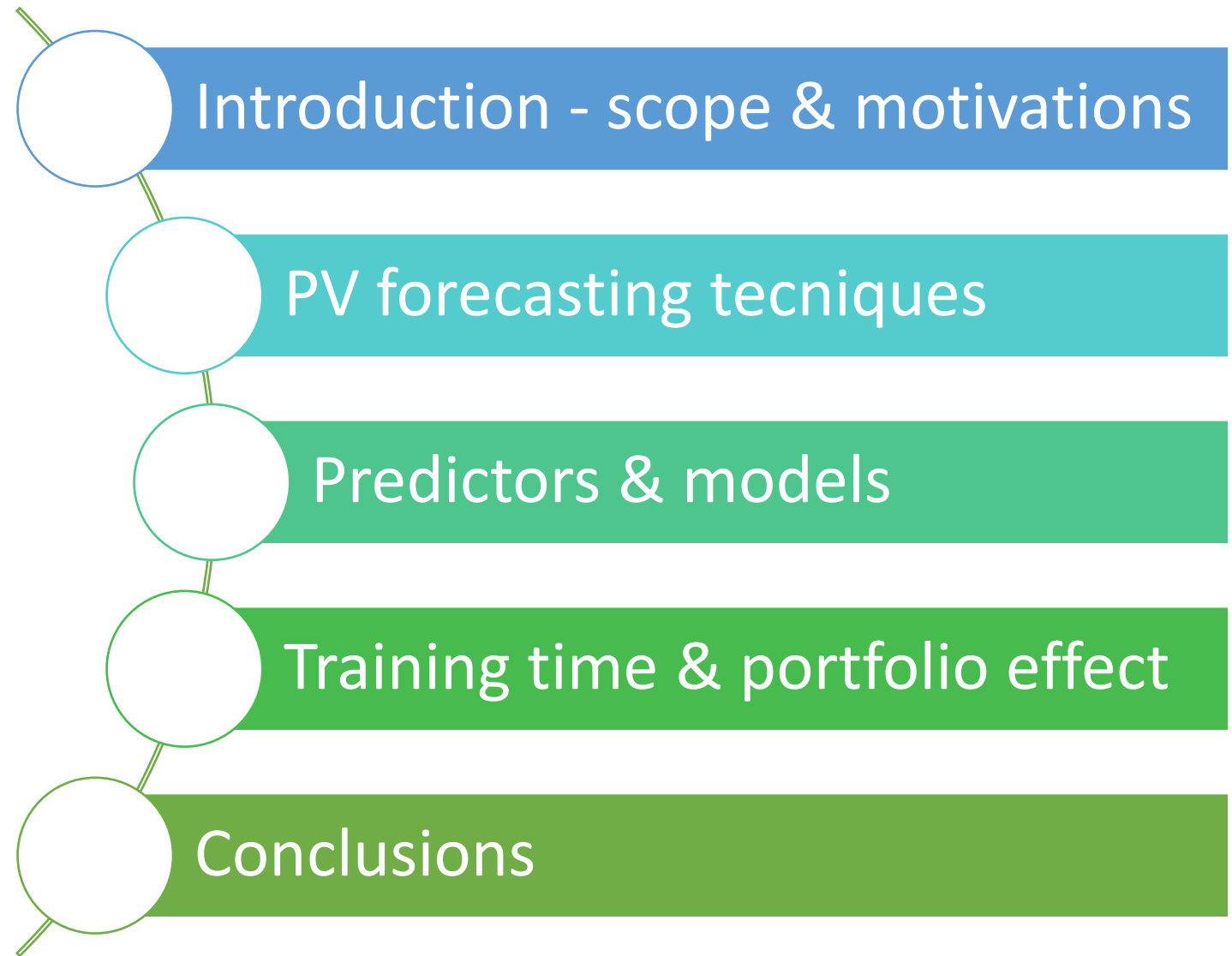
- 1 éééé.hh.nn. – éééé.hh.nn. formátum
- 2 helység, ország
- 3 Témavezető igazolja, aláírásával hitelesíti.

Photovoltaic power forecasting using machine learning

Dávid Markovics

Consultant professor: Dr. Martin János Mayer

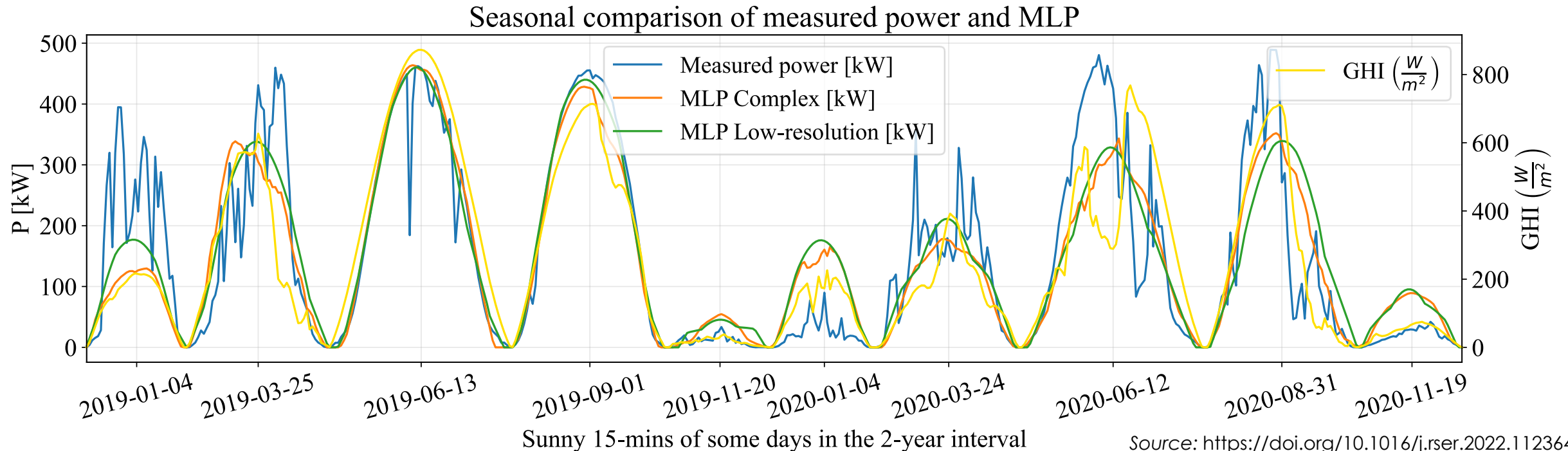
Overview



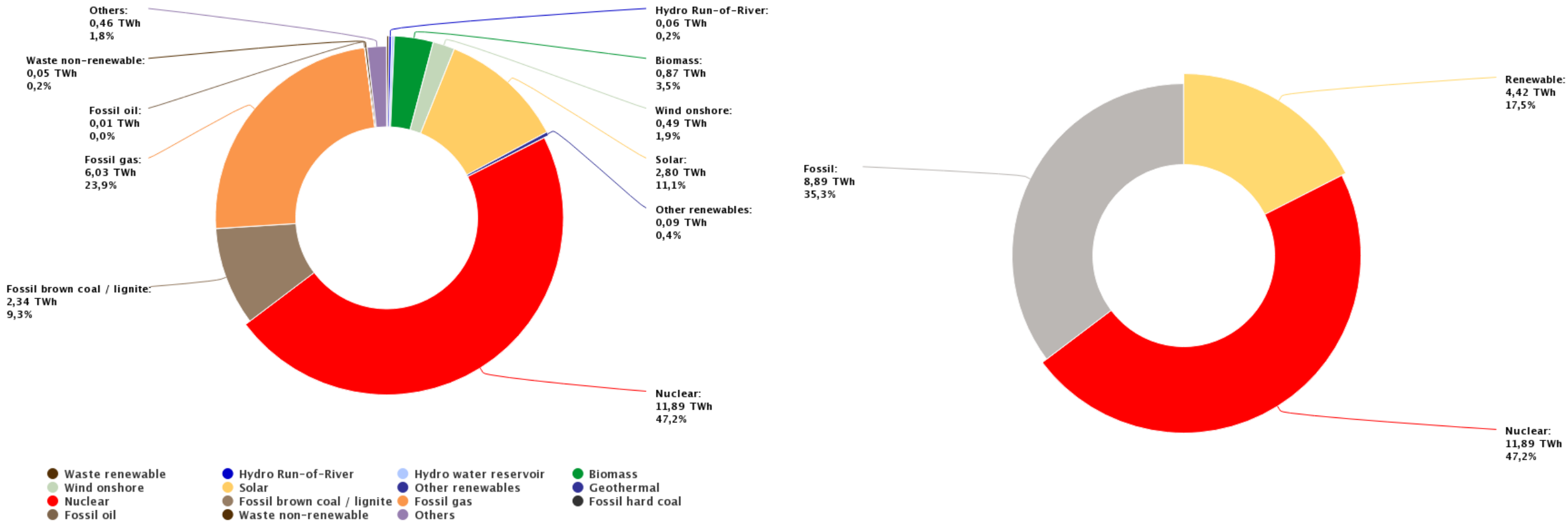
Nature of the problem - intraday and seasonal fluctuation

Main issues of PV integration:

1. Seasonal (declination angle) → strategic problem
2. Intraday availability (azimuth angle) → strategic & balancing problem (DFD)
3. Schedule error → balancing problem

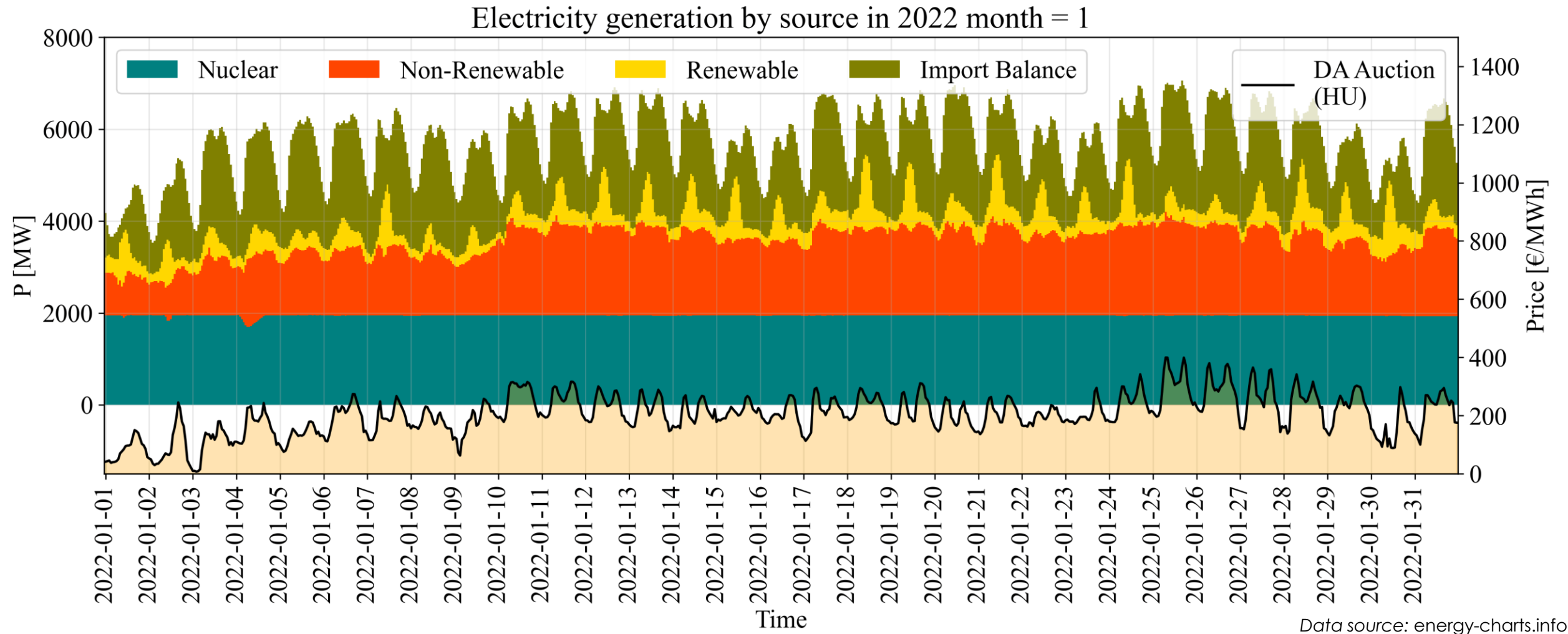


Position of PV in the HU power generation



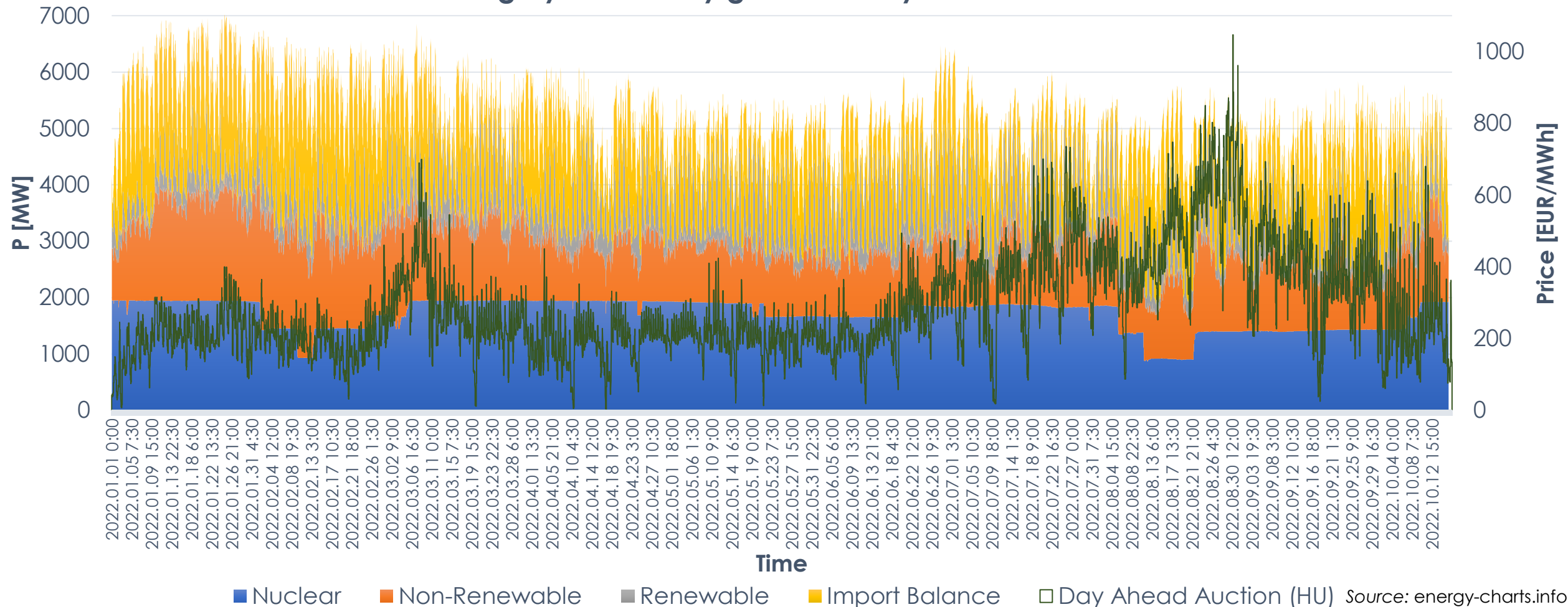
Source: https://energy-charts.info/charts/energy_pie/chart.htm?l=en&c=HU

Position of PV in the HU power generation



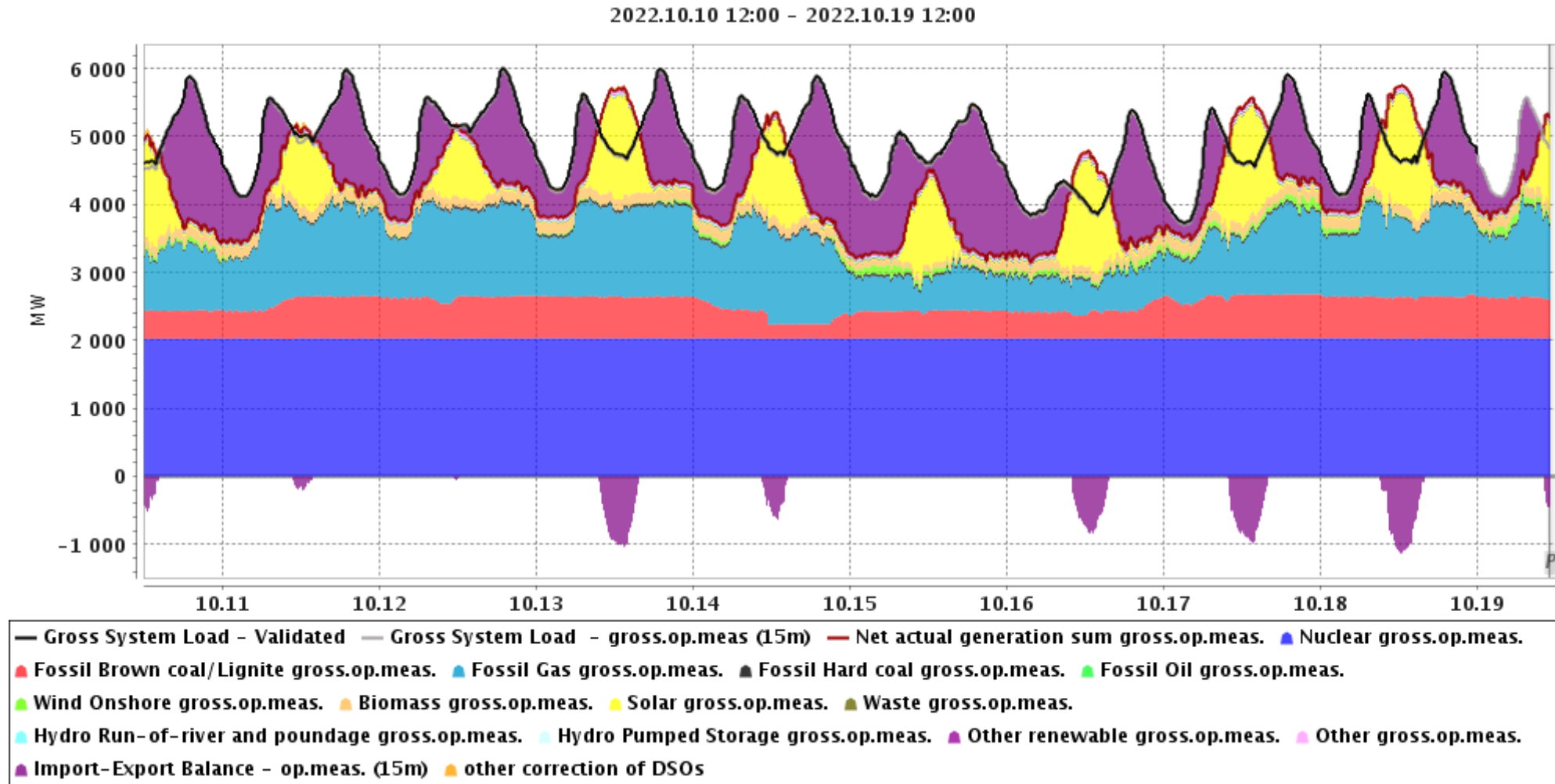
Position of PV in the HU power generation

Hungary's electricity generation by source in 2022



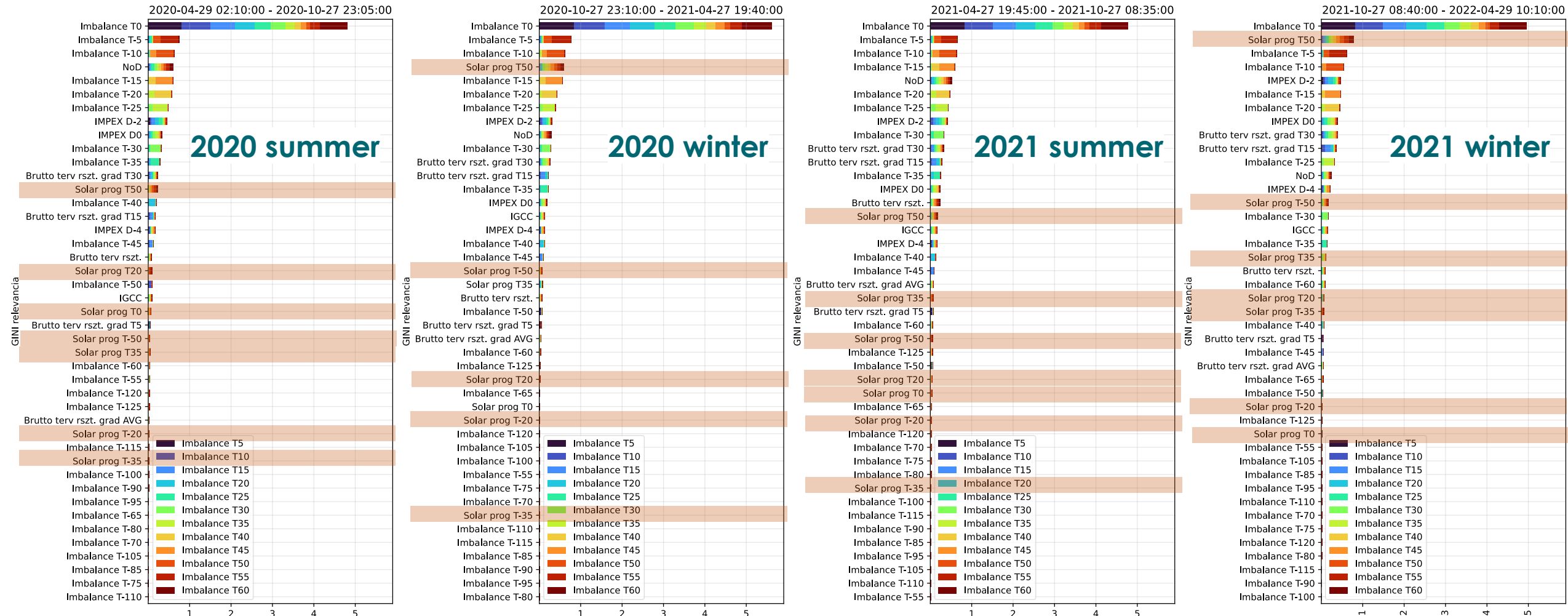
Position of PV in the HU power generation

- Industrial:
 - ~2.2 GW
- Domestic:
 - ~1.5 GW
- Gross demand:
 - ~4-7 GW



Source: data publication of MVIR Ltd. (Hungarian TSO)

Relevance increase of solar prognosis in short-term system-level imbalance forecasting



Approaches for PV forecasting

Physical modelling

Pros:

- Well-published model chain concepts
- No previous data needed
- Same accuracy level from the first day (except from post-tuning)

Cons:

- Sensitive for design data and it's quality
- Not able to decrease weather forecast errors
- Feedback on local features (weather, shadowing) is harder

Data-driven methods ANN, SVM, RF, XGBoost, k-NN etc.

Pros:

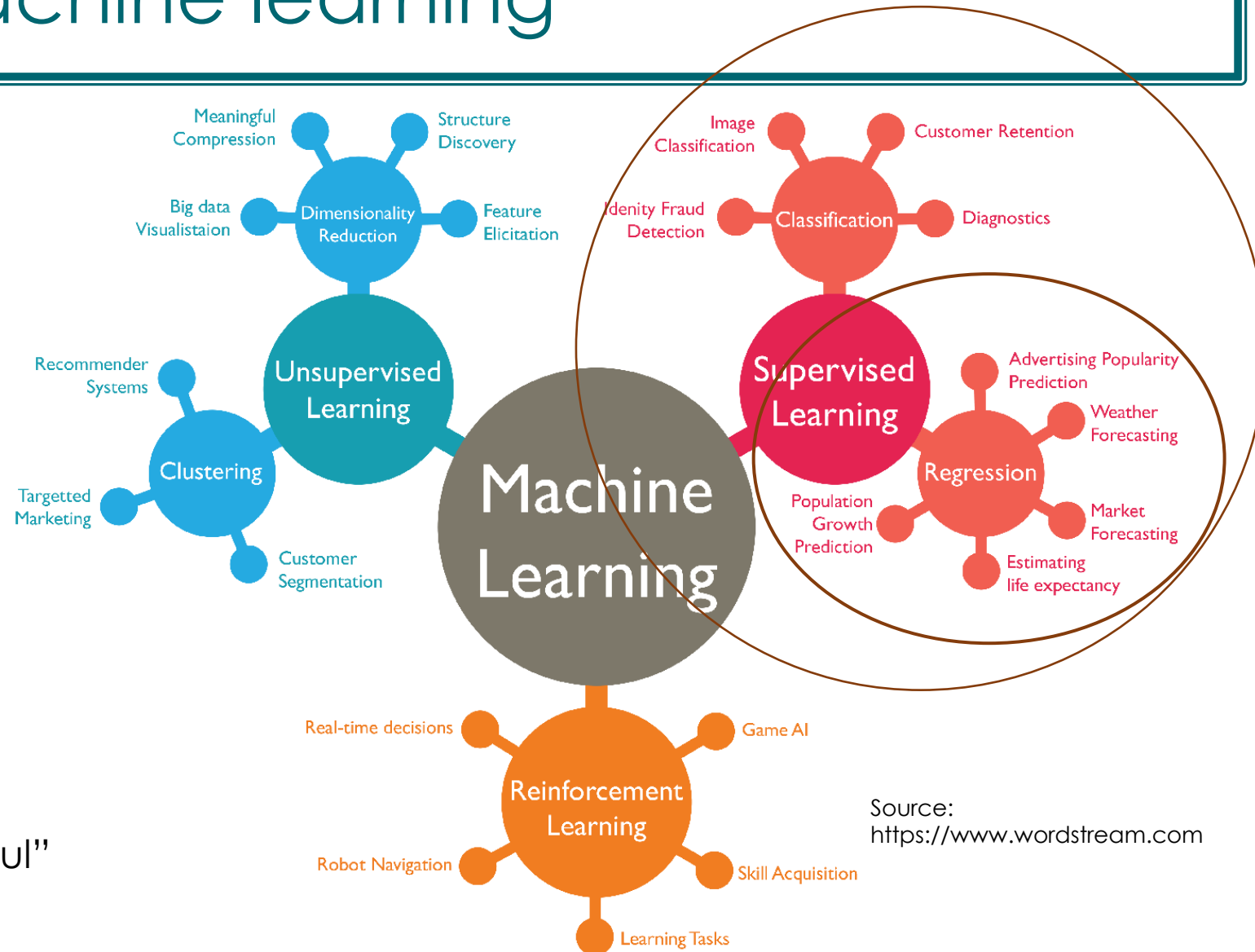
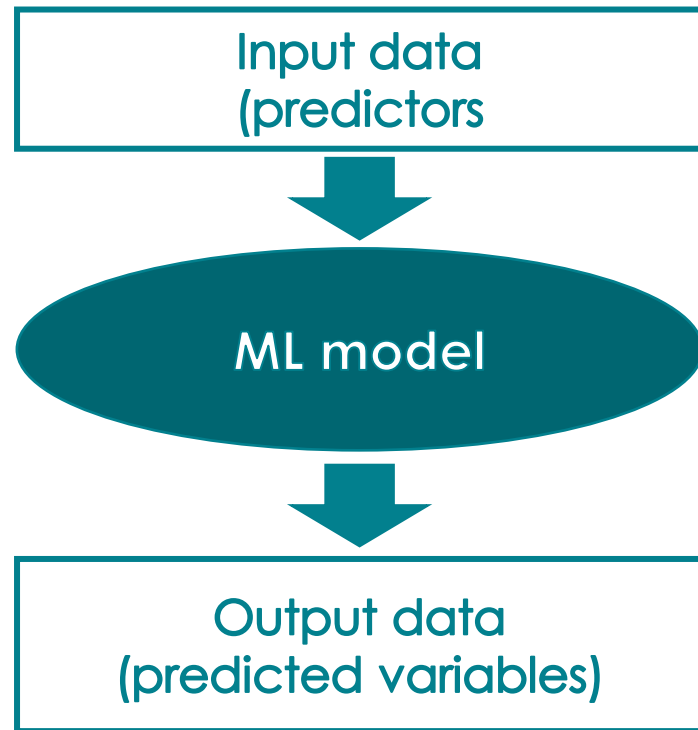
- No need for design data
- Able to cope with local features (unique weather, shadowing)
- Capable of decreasing weather forecast inaccuracy

Cons

- Previous data needed
 - the same weather model !
 - undisturbed production !!
 - feed-in tariff self regulating
- Post-processing needed

Rule nr.1.: *garbage in := garbage out*

Machine learning



Source:
<https://www.wordstream.com>

„All models are wrong, but some are useful”
George Box

General concepts & issues

The **target function** of PV forecasting is driven by regulative environment

- Imbalance payoff **cost or accuracy?**
- Hard to generalize
 - **Difficult to compare** results of studies

Market competition → closed data

- Lack of **compareability and repeatability**

Interdisciplinary field

- Three main approaches:
 - **Data scientist** → pro algorithm, but lack of practical evaluation
 - **Energy professionals** → practical thinking but lack of programming skills
 - **Meteorologists** → deep knowledge about weather drivers, although might not the sharpest in energy and data science

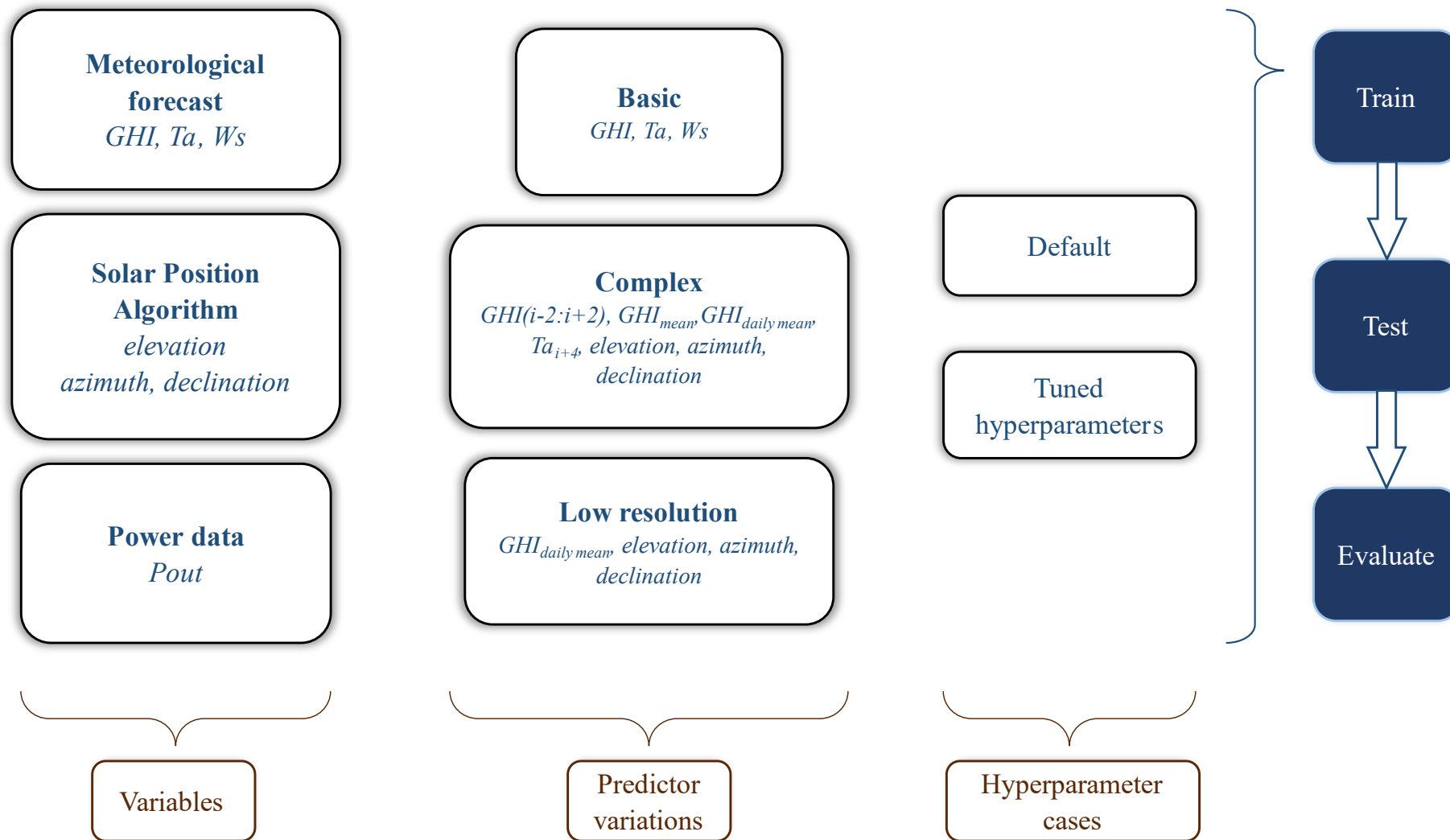
	Input	Hyperp.	Linear models												Kernel Ridge	Support Vector Machines	Neural Network	Neighbors	Decision Tree	Ensembles						
			LR	Lasso	Ridge	EN	Lars	OMP	BR	ARD	PAR	RAN-SAC	TR	Huber	KR	SVM	MLP	KNN	DT	RF	ET	ABR	GBR	XG-Boost	LGBM	Cat-Boost
nRMSE	Basic	default	52.6	94.0	52.6	78.1	52.6	53.3	52.6	52.6	81.2	54.0	52.8	53.0	52.6	53.2	52.1	59.6	72.9	56.2	56.3	56.5	52.1	55.6	53.6	54.2
		tuned	52.6	52.6	52.6	53.4	52.6	52.5	52.6	52.5	52.7	53.8	52.8	52.6	51.5	52.5	51.5	52.8	52.7	52.0	51.7	53.2	51.7	51.8	52.4	51.7
	Complex	default	49.4	94.0	49.4	72.6	58.4	52.8	49.4	49.4	77.1	56.7	49.9	50.1	49.4	46.7	46.7	53.1	63.9	48.7	47.7	51.8	46.7	50.6	48.4	48.8
		tuned	49.4	49.7	49.4	49.8	49.5	49.4	49.4	49.4	49.4	50.3	49.9	49.4	45.3	46.6	45.4	46.4	48.2	46.1	45.9	49.7	46.5	46.4	46.7	46.1
	Low resolution	default	52.1	94.0	52.1	84.8	52.1	65.7	52.1	52.1	79.9	56.1	51.2	52.2	52.1	47.2	46.2	54.2	64.6	52.7	51.4	55.5	48.0	52.2	49.9	50.4
		tuned	52.1	52.6	52.1	55.1	52.6	63.2	52.1	52.1	51.5	55.3	51.4	52.1	46.0	46.6	46.2	47.1	51.3	49.1	47.8	57.3	48.2	48.3	48.8	47.4
nMAE	Basic	default	37.4	82.7	37.4	68.8	37.4	38.2	37.4	37.4	59.9	36.3	36.2	35.8	37.4	35.1	36.5	40.9	48.7	39.0	39.1	44.9	36.6	38.5	37.2	37.6
		tuned	37.4	37.7	37.5	40.4	37.4	37.4	37.4	37.4	37.5	36.4	36.2	36.8	36.2	37.1	36.3	38.0	36.9	36.4	36.5	38.9	36.0	36.8	36.7	36.6
	Complex	default	35.3	82.7	35.3	63.7	40.9	37.8	35.3	35.3	56.0	37.9	34.3	34.3	35.3	27.9	30.8	32.7	38.8	31.0	30.7	39.5	30.9	32.6	30.9	31.5
		tuned	35.3	35.7	35.3	35.9	35.4	35.3	35.3	35.3	35.0	34.2	34.3	34.9	29.6	30.8	29.6	30.4	31.8	30.2	30.2	36.4	30.8	30.8	30.9	30.5
	Low resolution	default	39.8	82.7	39.8	74.8	39.8	49.8	39.8	39.8	58.8	39.5	38.2	38.5	39.8	29.2	30.4	34.3	39.9	33.5	33.1	43.2	32.8	34.0	32.4	32.9
		tuned	39.8	40.3	39.8	43.4	40.1	48.0	39.8	39.8	38.7	39.6	38.3	39.6	31.0	30.3	31.0	32.2	34.9	33.4	33.0	42.5	33.1	33.7	34.8	32.5
Corr.	Basic	default	82.9	-1.1	82.9	82.1	82.9	82.3	82.9	82.9	61.3	82.5	82.9	82.9	82.9	83.0	83.3	78.1	69.6	80.3	80.2	81.1	83.2	80.8	82.2	81.8
		tuned	82.9	82.9	82.9	82.6	82.9	82.9	82.9	82.9	82.9	82.4	82.9	82.9	83.7	83.0	83.6	82.8	82.8	83.3	83.5	82.5	83.5	83.4	83.0	83.5
	Complex	default	85.1	-1.1	85.1	82.6	78.7	82.7	85.1	85.1	65.3	81.0	84.9	84.7	85.1	87.2	86.8	83.2	76.7	85.6	86.2	84.5	86.8	84.5	85.8	85.5
		tuned	85.1	84.9	85.1	84.8	85.0	85.1	85.1	85.1	85.3	84.6	84.9	85.0	87.6	86.9	87.6	87.0	85.9	87.2	87.3	85.2	86.9	86.9	86.8	87.1
	Low resolution	default	83.6	-1.1	83.6	70.9	83.6	71.5	83.6	83.6	66.6	81.0	84.0	83.2	83.6	86.8	87.1	82.3	76.1	83.1	83.9	81.7	86.0	83.5	84.9	84.6
		tuned	83.6	83.3	83.6	82.1	83.2	73.9	83.6	83.6	84.1	81.1	83.9	83.5	87.2	86.9	87.1	86.6	83.8	85.3	86.2	79.2	85.8	85.8	85.5	86.4
	Basic	default	35.7	-15.0	35.7	4.4	35.7	34.8	35.7	35.7	0.6	34.0	35.4	35.2	35.7	34.9	36.3	27.1	10.8	31.3	31.1	30.8	36.3	32.0	34.5	33.7
		tuned	35.7	35.7	35.7	35.7	35.7	35.7	35.7	35.7	35.7	34.2	35.4	35.7	37.1	35.7	37.0	35.5	35.6	36.5	36.8	34.9	36.7	36.7	35.9	36.8
		default	39.7		35.7																					

- 24 machine learning models tested for day-ahead photovoltaic power forecasting
- **Kernel ridge** and **multilayer perceptron** are the overall most accurate models
- **Predictor selection** is even more important than model selection
- **Hyperparameter optimization** is essential for the highest accuracy
- Up to 13.9% RMSE improvement over a baseline linear regression model

Review study

<https://doi.org/10.1016/j.rser.2022.112364>

Predictor selection



What is the optimal training time? - Methodology

Methodology:

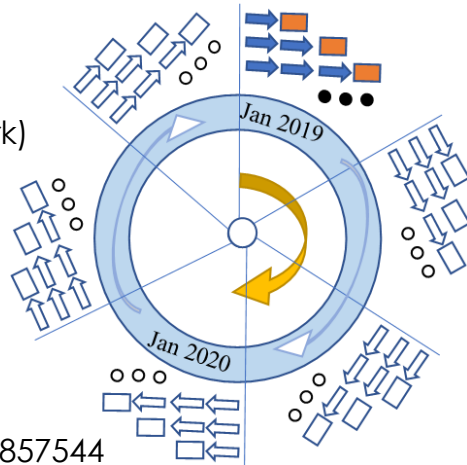
- 51 train-time cases: 2 → 102 weeks
- 52 different start-time (start of operation)
- 1 PV park
- ID and DA forecast

Data:

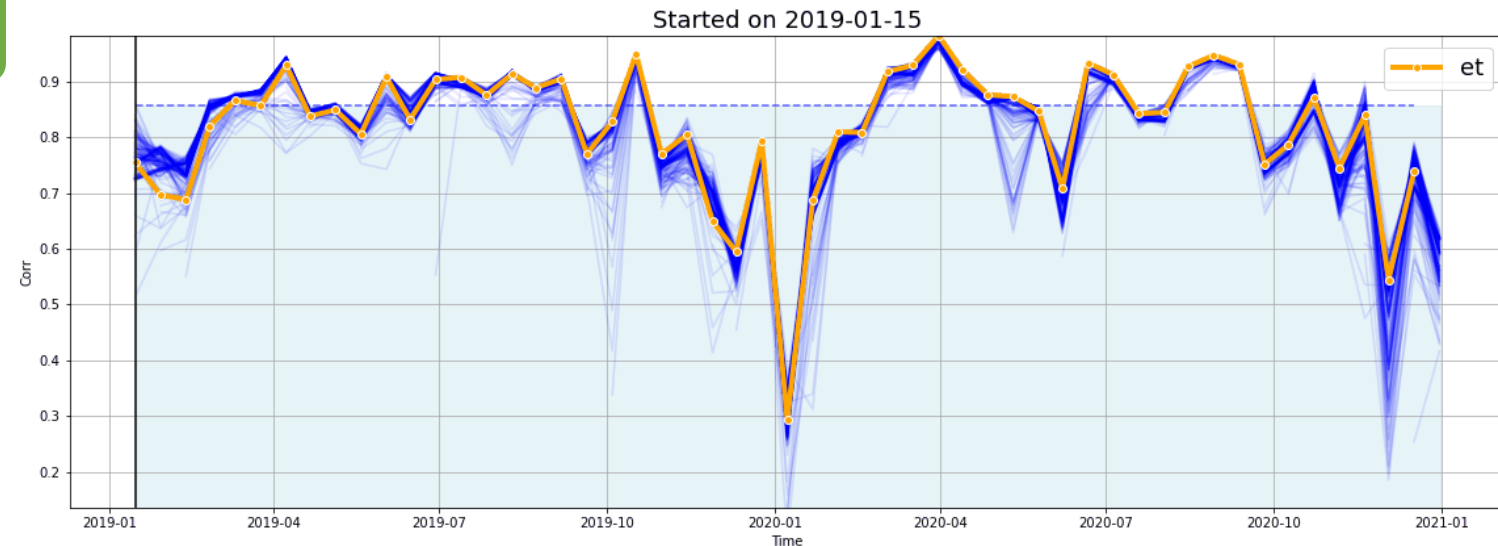
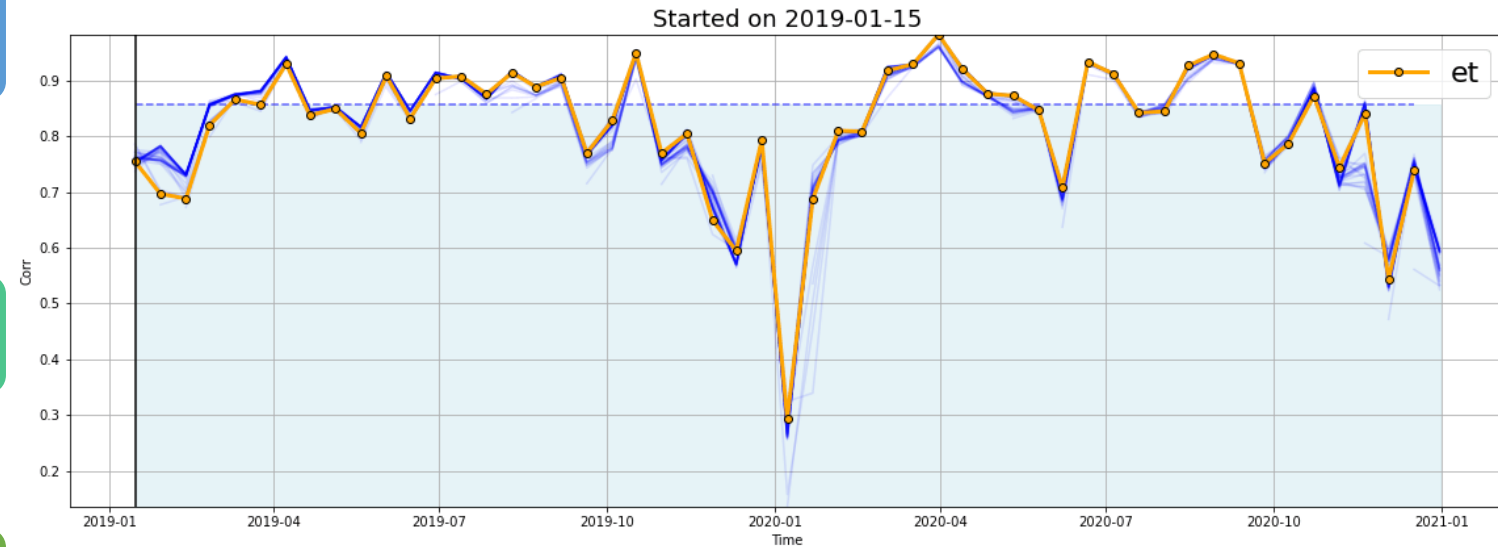
- 2 years, NWP weather forecast
- Solar angles from SPA
- real measured output power of a 500kW plant

Models

- ExtraTrees
- XGBoost
- LightGBM
- (linear regression as benchmark)



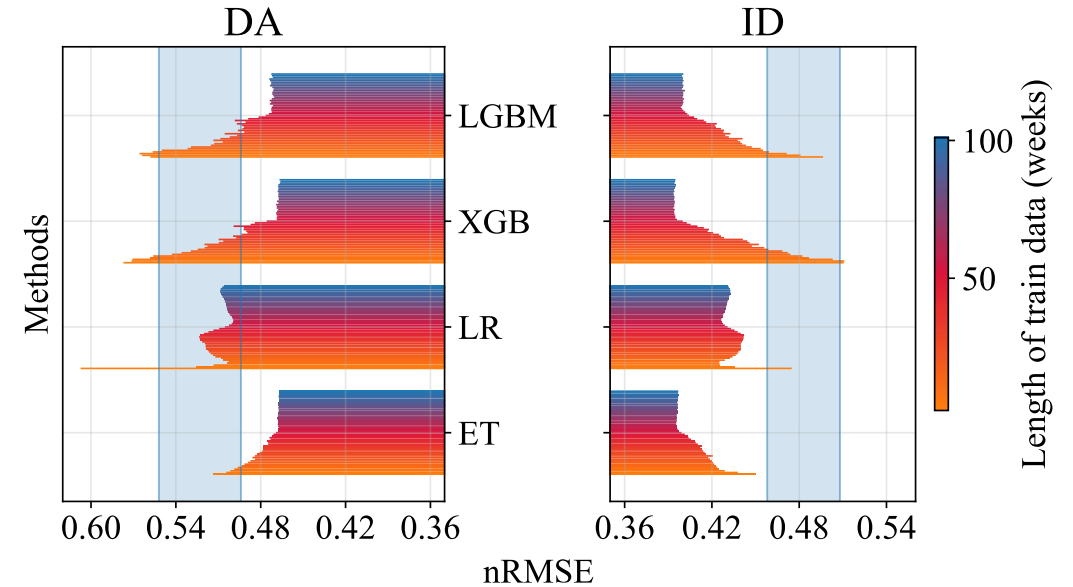
Source: 10.1109/IYCE54153.2022.9857544



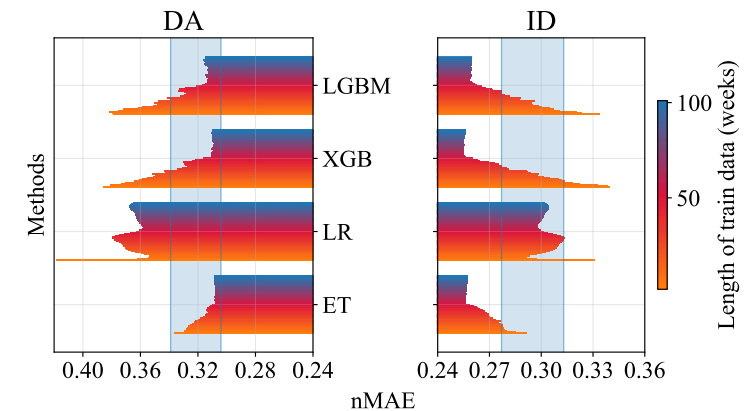
What is the optimal training time? - Results

	Train time	Horizon	ET	LR	XG-Boost	Light-GBM
nRMSE [%]	2-week	DA	51.4	60.7	57.7	55.8
		ID	45.0	47.5	51.1	49.6
	52-week	DA	46.9	50.5	47.6	47.6
		ID	39.9	43.1	40.0	40.4
	102-week	DA	46.7	50.6	46.7	47.3
		ID	39.7	43.1	39.5	40.0
nMAE [%]	2-week	DA	33.6	41.9	38.6	37.9
		ID	29.2	33.1	34.0	33.4
	52-week	DA	31.0	36.3	31.6	31.7
		ID	25.9	30.1	26.0	26.3
	102-week	DA	30.8	36.5	31.0	31.5
		ID	25.8	30.2	25.6	26.0
Skill [%]	2-week	DA	37.3	25.9	29.6	31.9
		ID	41.9	38.7	34.1	35.9
	52-week	DA	42.7	38.3	41.9	41.8
		ID	48.5	44.4	48.4	47.8
	102-week	DA	43.0	38.2	43.0	42.3
		ID	48.8	44.4	49.0	48.3
Weeks needed to reach the best physical forecast						
Base metric	nRMSE	DA	12	-	36	32
		ID	2	4	20	10
	nMAE	DA	-	-	-	-
		ID	18	-	40	40

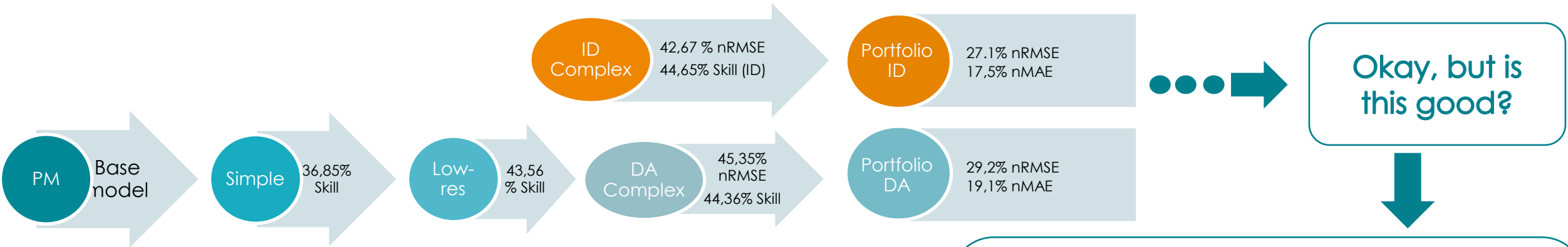
Average nRMSE results with increasing the training time



Average nMAE results with increasing the training time



Portfolio-level (16 parks), DA/ID



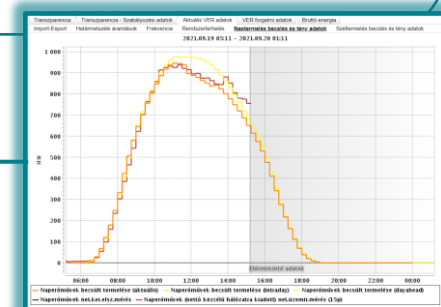
Results with portfolio-level sum production of 16 PV parks

nRMSE - Komplex_da	34.22	29.2	29.8	29.99	29.67	29.87	29.29
nRMSE - Komplex_id	32.55	27.1	27.74	27.66	27.42	27.76	27.26
nMAE - Komplex_da	24.96	19.1	18.56	19.64	19.75	19.97	19.24
nMAE - Komplex_id	23.69	17.59	17.2	17.95	18.07	18.31	17.58
	lr	kr	svm	knn	rf	gbr	mlp

Color scale for nRMSE: 27.5 to 32.5
Color scale for nMAE: 17.5 to 22.5

nMAE ($\text{kWh}_{\text{KE}} / \text{kWh}_{\text{PV}}$)
Industrial scale PV in Hungary
(2020.01. – 2021.06.):

- Solar prog. (intraday): **11,48 %**
- Solar prog(dayahead): **12.31 %**
- Net TSO measurement (15p): **13,13 %**



What can we use the forecast

A good forecast means good predictability therefore a lower risk level. Accurate planning allows for cost optimization, the utilization of which depends on the motivations.

As a PV company

- Dayahead scheduling (~T+24h)
- Intraday schedule modification (~T+1.5h)
- mFRR potential (T+30m)

As a portfolio manager

- Self-regulation
 - By changing market position (~T+30m)
 - By other assets (~T+15m)

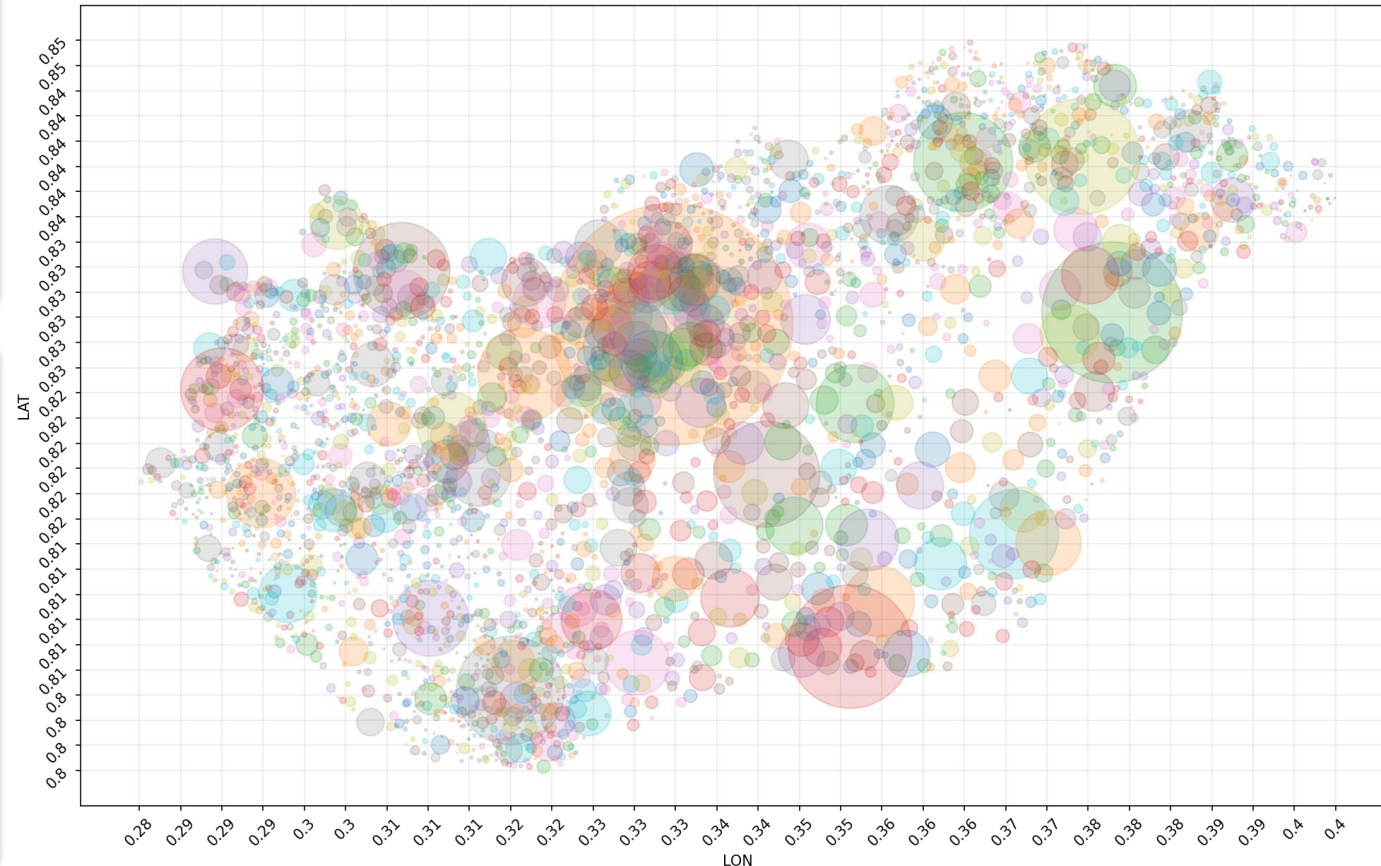
As a TSO

- Load division
 - Load forecasting (ID&DA)
- Reserve management (bids)
- Imbalance forecasting (T+15m)

As a DSO

- DSO-level load forecasting
- Load flows & state estimation
- Domestic consumption & residual curve forecasting for better trading
 - Missing element: household-sized small power plants (HSPP)!

Distribution of ~1.5 GW HSPP in Hungary



Conclusions & final thoughts

Predictor selection is the most important

The most important variables:

- GHI forecasts surrounding the observed timestamp,
- P(T-90mins)
- solar angles

Data quality & post processing

Make sure that the data is clean!
Post-processing is always necessary.

Suggested models:
ExtraTrees, MLP,
Kernel Ridge

ExtraTrees – accurate, reached physical models (same weather data) in 12/2W DA/ID
Kernel Ridge is accurate, but has high computational demand
XGBoost and LGBM → needs more attention to proper regularization and hyperparameter tuning

Not recommended
to use more than 52-
54 weeks for training

Extending the dataset increases computational demand with no significant improvement

Thank you for your attention!

ACKNOWLEDGMENT

- I hereby express my gratitude to Gabriella Szépszó and Mihály Szűcs from the Hungarian Meteorological Service for providing the AROME numerical weather prediction data, and Norbert Péter from the MVM Green Generation Ltd. for the PV production data. The research reported in this paper and carried out at BME has been supported the ÚNKP-22-2 New National Excellence Program.

For any further questions, please contact by:

- *markovics.david@eszk.org*



Used
metrics

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (f_i - P_i)^2} \quad (1)$$

$$nRMSE = \frac{RMSE}{\frac{1}{N} \sum_{i=1}^N P_i} \quad (2)$$

$$MAE = \frac{\sum_{i=1}^N |f_i - P_i|}{N} \quad (3)$$

$$nMAE = \frac{MAE}{\frac{1}{N} \sum_{i=1}^N P_i} \quad (4)$$

$$s = 1 - \frac{RMSE_{forecast}}{RMSE_{persistence}} \quad (5)$$

Photovoltaic power forecasting using machine learning

Dávid Markovics

Budapest University of Technology and Economics

AEE Europe 2022

Abstract

Photovoltaic (PV) energy production is under great development, the installed capacity is rising exponentially worldwide, gaining higher share from the total portfolio of sources. According to EU energy strategy, PV technology plays an important role in accomplishing the various climate protection purposes. According to these trends, the topic of power forecasting and grid integration (balancing) of weather-based intermittent energy sources, without risking the safety of supply, became widely researched in the past years. Most of the papers in this topic are driven by the aim of improving the solar irradiation prognosis, however my research is focused on predicting PV power production from meteorological forecasts, for which I use machine learning techniques trained on past set of weather and real power data provided by industrial partners. My predictions are made on both day-ahead (24-48h) and intraday (00-24h) time horizons. In the presentation, I am going to introduce the structure of forecasting methods, the results and comparison of different machine learning algorithms, and some key issues from the practical side, like bias and verification.

