PHAROS
THE GREEK AI FACTORY

Forecasting Renewable Energy Production

Anestis Ampatzidis, MSc
Vasileios Kochliaridis, PhD
Vlahavas Ioannis, Prof.

ARISTOTLE UNIVERSITY OF THESSALONIKI

Introduction

The climate change challenge: Shift towards clean energy and rapid growth of photovoltaic (PV) systems (Renewable Energy Sources - RES)

The core problem: Solar energy generation is characterized by high stochasticity and variability

Weather dependency: Performance is directly affected by uncontrollable factors (solar irradiance, temperature, cloud cover)

Impact on the grid: Creates instability and significant difficulty in balancing electricity supply and demand

Purpose of research

Research focus: Development of a day-ahead forecasting system for solar energy generation

Approach: A wide spectrum of models was evaluated, ranging from simple statistical methods to neural networks and foundation models, to identify the most effective forecasting solution

Dataset: Creation of a custom dataset (UK photovoltaics + ERA5/ECMWF weather data)

Substantial impact: Improving smart grid efficiency and reducing financial penalties for solar energy producers

Dataset – Data source

The generation data was provided by Hugging Face

The weather data was provided by ECMWF API

UK PV Dataset

UK PV Dataset

About

1

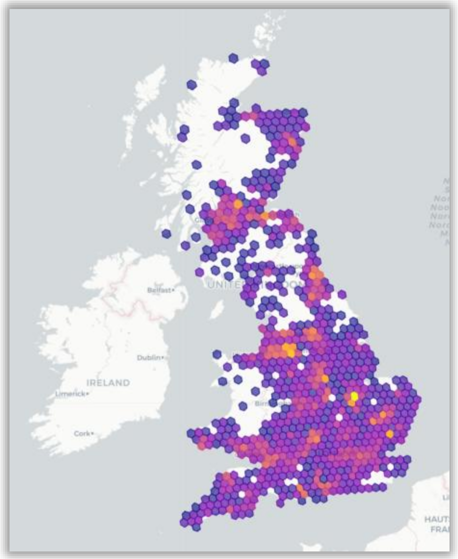
General information

- 30.000 installations
- Range: 2010-2025
- Frequency: 30 minutes

2

Installations	Bin edges (capacity)
473	[0, 1]
4809	(1, 2]
13203	(2, 3]
11854	(3, 4]
247	(4, 10]
158	(10, 50]
6	(50, 100]
2	(100, 150]
2	(150, 200]
3	(200, 250]

3



1 PV system

4,500 PV systems

UK PV Dataset

Generation Data

1 Features

ss_id

The Sheffield Solar ID of the system

datetime_GMT

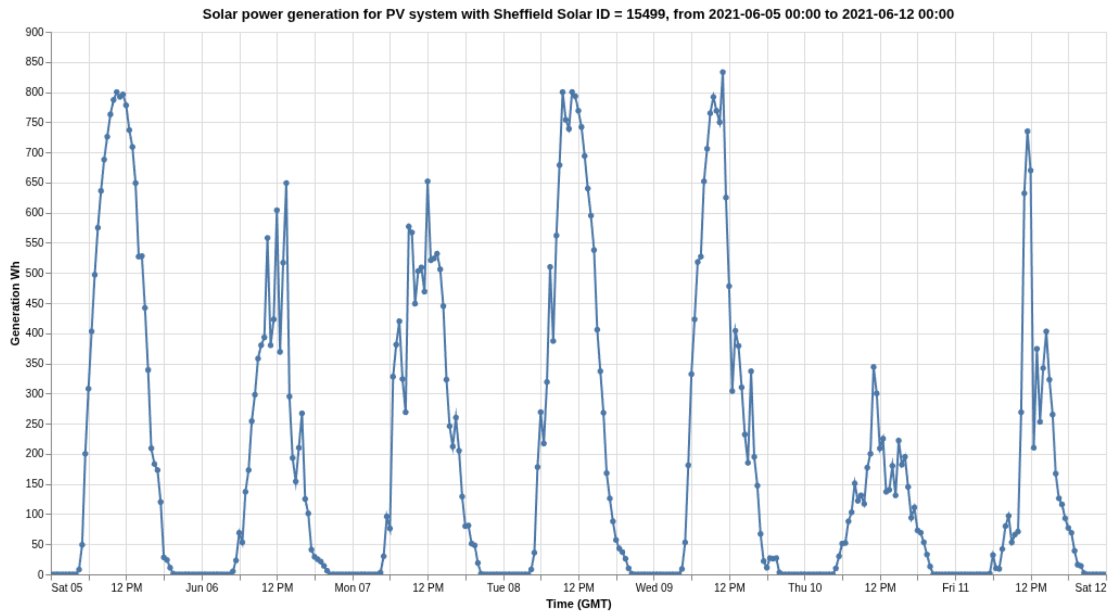
The datetime of the recording in Greenwich Mean Time (UTC+0)

generation_Wh

The energy generated in the period (in watt hours) at the given timestamp for the given system

Metadata:
latitude, longitude, orientation, tilt, kWp etc.

2 Generation plot



Introduction
Dataset
Data Preprocessing
Feature Engineering
Evaluation Settings
Models
Results
Current Focus
Conclusions

Dataset – Data source

ECMWF

Dataset – Data source

Data

Range:
2010-2025

ECMWF

Frequency:
1 hour

Features

Introduction
Dataset
Data Preprocessing
Feature Engineering
Evaluation Settings
Models
Results
Current Focus
Conclusions

Dataset – Data source

Range:
2010-2025

Data

ECMWF

Frequency:
1 hour

Features

ERA5 Land

- Temperature
- Total Precipitation
- Surface Solar Radiation Downwards
- U - component of wind
- V - component of wind

ERA5 Single

- Total cloud cover
- Low cloud cover
- Medium cloud cover
- High cloud cover

Final Dataset

Merging Hugging Face data with weather data – challenges

Merging different sources:
Weather grid vs PV coordinates

Different data frequency (30 minutes/1 hour)

Final Dataset

Merging Hugging Face data with weather data – challenges

Merging different sources: Weather grid vs PV coordinates

Step 1: Spatial Matching

Applied **Nearest Neighbor** to map each PV system to the closest weather grid node

Different data frequency (30 minutes/1 hour)

Introduction
Dataset
Data Preprocessing
Feature Engineering
Evaluation Settings
Models
Results
Current Focus
Conclusions

Final Dataset

Merging Hugging Face data with weather data – challenges

Merging different sources: Weather grid vs PV coordinates

Step 1: Spatial Matching

Applied **Nearest Neighbor** to map each PV system to the closest weather grid node

Different data frequency (30 minutes/1 hour)

Step 2: Temporal alignment

Replicated hourly weather data across both 30-minute intervals within each hour to align with the generation data (e.g., 10:00 AM weather applies to both 10:00 and 10:30 timestamps)

Data Preprocessing

Data analysis → Filter negative values → Completeness check

Filter negative values

Removal of erroneous time logs flagged as problematic by the dataset creators (Hugging Face repository)

Completeness check

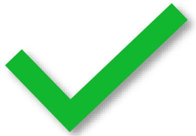
Ensured exactly 48 records per day by discarding any incomplete days

Data Preprocessing

Introduction
Dataset
Data Preprocessing
Feature Engineering
Evaluation Settings
Models
Results
Current Focus
Conclusions

Filter negative values → Completeness check → Advanced quality checks

Removal of erroneous time logs flagged as problematic by the dataset creators (Hugging Face repository)



Ensured exactly 48 records per day by discarding any incomplete days

Implemented a strict 5-step daily filtering process to eliminate anomalies

Data Preprocessing

Advanced quality checks

An entire day was discarded if it triggered any of the following conditions

Night-time production > 0.5 Wh

Dead sensor: Maximum daily production ≈ 0

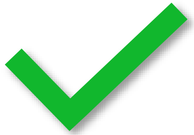
Sync check: > 2 hours gap between peak production and peak irradiance (GHI)

Sudden generation drops: Unexplained, sharp drops in generation during active daylight hours

Extreme outliers: Negative production values or physically impossible efficiency rates

Data Preprocessing

Completeness check → **Advanced quality checks** → **Downsampling**



Ensured exactly 48 records per day by discarding any incomplete days

Implemented a strict 5-step daily filtering process to eliminate anomalies

The data was downsampled from a 30-minute to a 1-hour resolution, but both versions were retained for the experiments

Feature Engineering

Additional features

Correlation matrix

Feature selection

Scaling

Feature Engineering

Additional features

- **Irradiance:** Converted SSRD (J/m^2) to GHI (W/m^2)
- **Solar geometry:** Calculated solar zenith, solar azimuth, and angle of incidence
- **Temperature:** Added the temperature rate of change

Correlation matrix

Feature selection

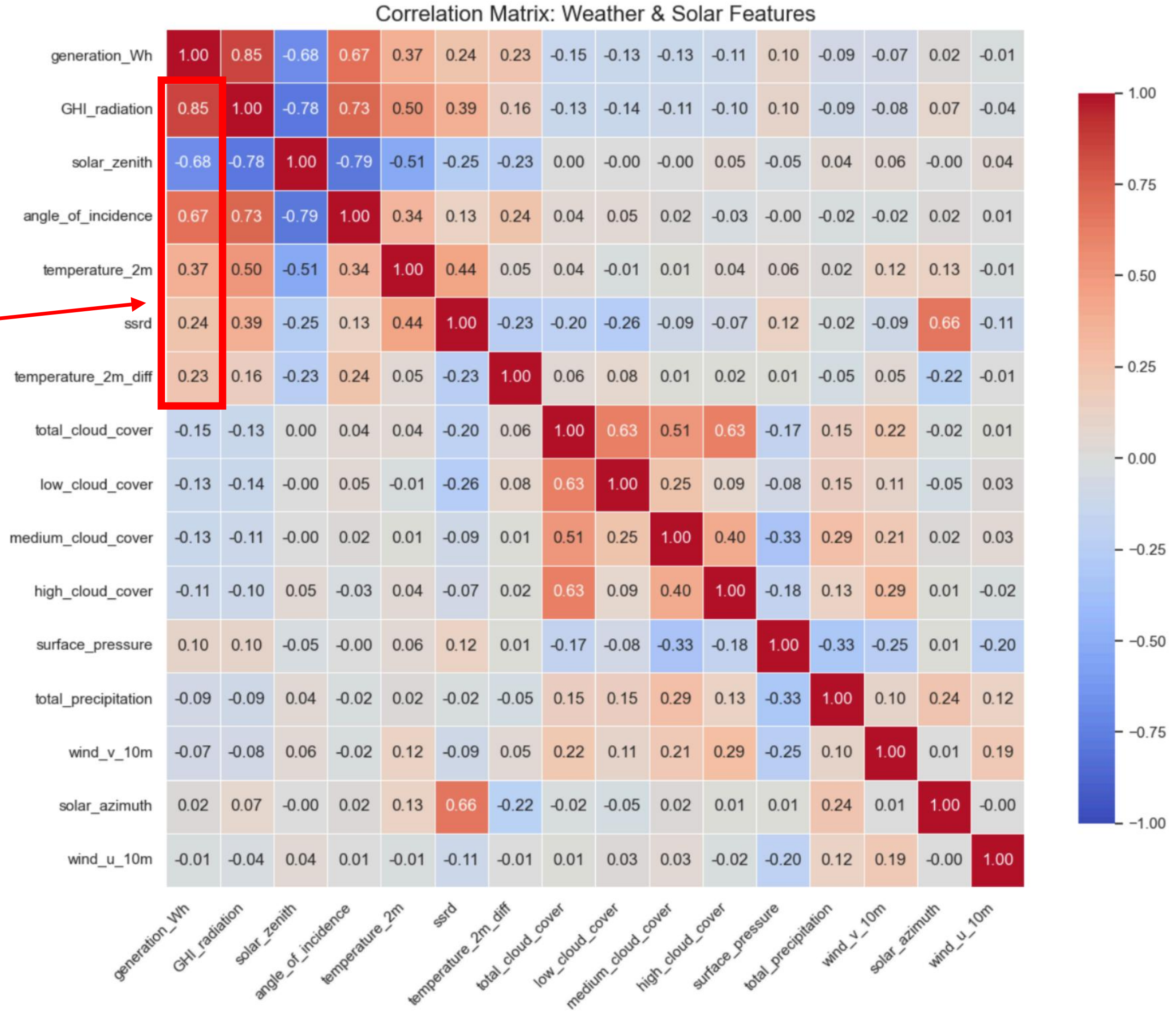
Scaling

Correlation matrix

Feature Engineering

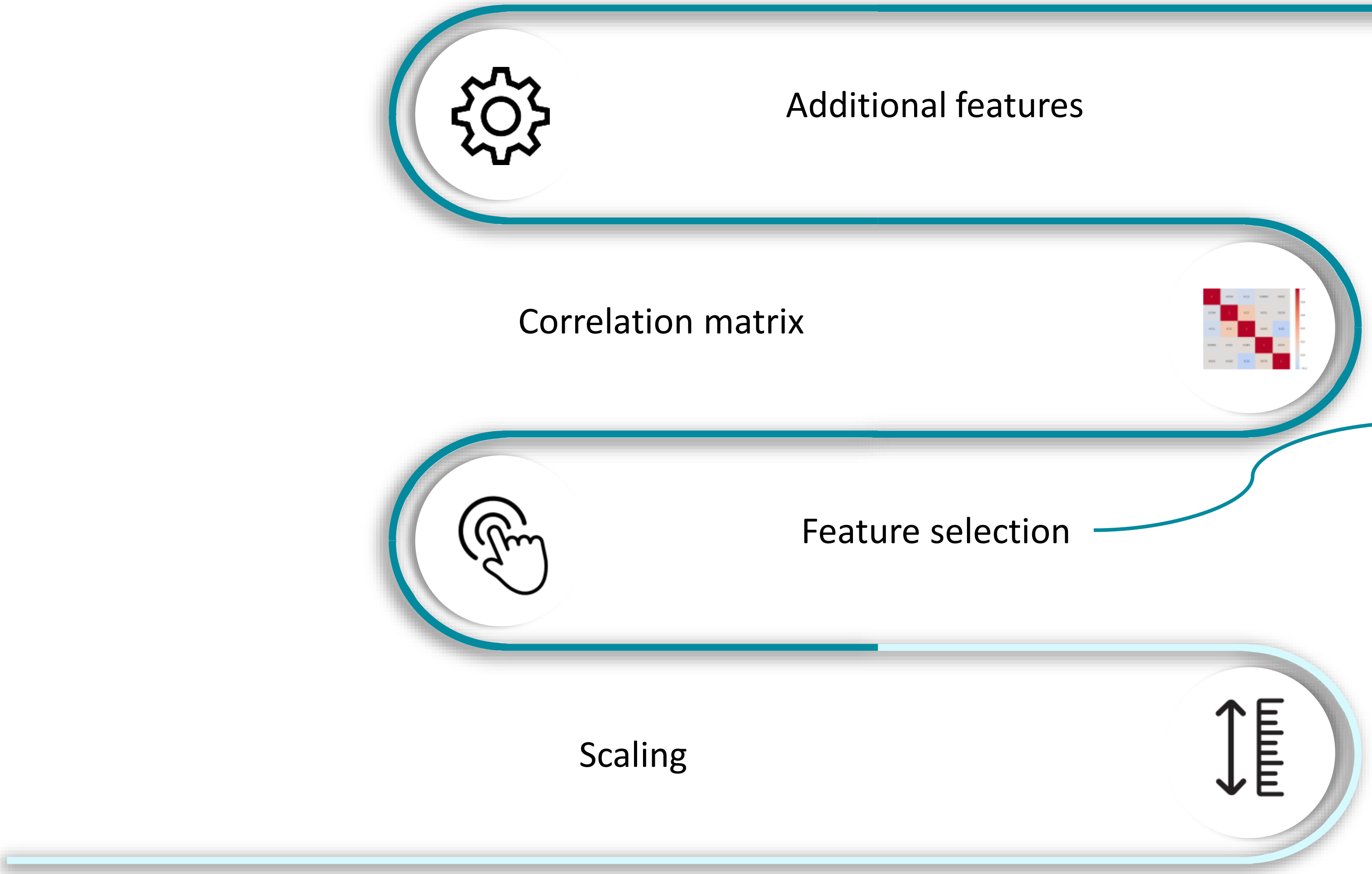
Top features:

1) GHI radiation
2) Solar_zenith
3) Angle_of_incidence
4) SSRD
5) Temperature_2m
6) Temperature_2m_diff



Feature selection

Feature Engineering



- **Dynamic features:** Historical production (generation_Wh) & meteorological/solar parameters (GHI, temperature, cloud cover, solar_zenith, solar_azimuth, angle_of_incidence, temperature_diff)
- **Static features:** Metadata (orientation, tilt, kWp)

Feature Engineering

- **Input normalization:** Z-score standardization is applied to all input features
- **Target normalization:** Energy generation is expressed per kWp, enabling direct comparison across systems of different sizes

Additional features

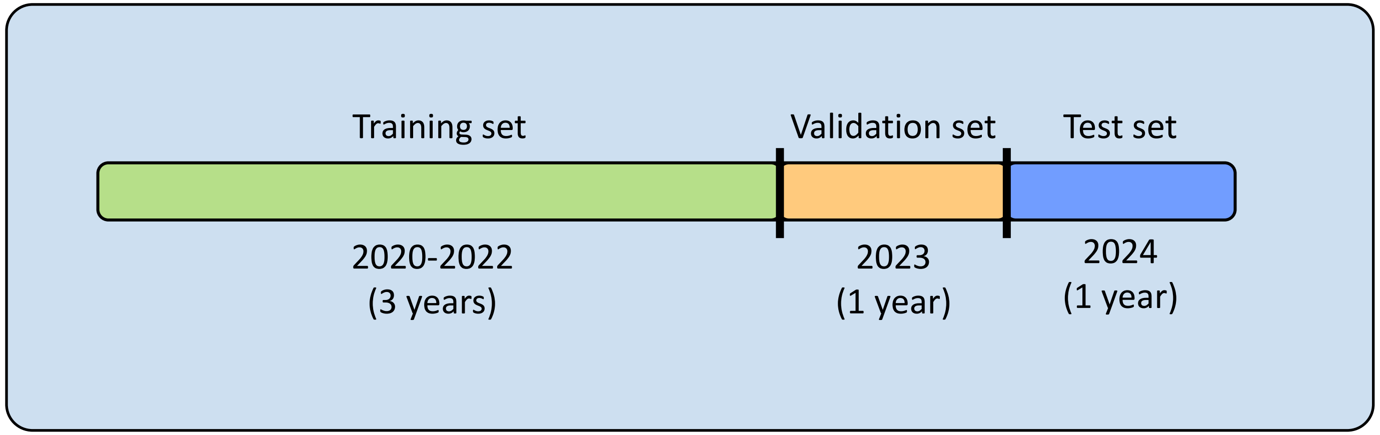
Correlation matrix

Feature selection

Scaling

Evaluation Settings

Installations used: 100



A strictly chronological, rather than spatial, data split ensures that every PV station is included in the training, validation, and test datasets

1

Step-level evaluation
(24/48 steps)

2

Daily sum evaluation

Metrics

1

Step-level evaluation
(24/48 steps)

2

Daily sum evaluation

1. Mean Absolute Error (MAE)
2. Root Mean Square Error (RMSE)
3. Coefficient of determination (R^2)
4. Weighted Mean Absolute Percentage Error (wMAPE)
5. Average Daily Error
6. Average Relative Error

Models

1 Baselines

2 Classic ML

3 Neural Networks

4 Foundation Models

ARIMA, SARIMA
- Type: Univariate
- History: 14 days

Models

1

2

3

4

Baselines

ARIMA, SARIMA
- Type: Univariate
- History: 14 days

Classic ML

XGBoost
- Type: Multivariate
- History: 14 days
- Role: feature importance

Neural Networks

Foundation Models

Models

1 | 2 | 3 | 4

Baselines

ARIMA, SARIMA
- Type: Univariate
- History: 14 days

Classic ML

XGBoost
- Type: Multivariate
- History: 14 days
- Role: feature importance

Neural Networks

TCN, BiLSTM, N-HITS
- Type: Multivariate
- History: 14 days

Foundation Models

Models

1

Baselines

ARIMA, SARIMA
- Type: Univariate
- History: 14 days

2

Classic ML

XGBoost
- Type: Multivariate
- History: 14 days
- Role: feature importance

3

Neural Networks

TCN, BiLSTM, N-HITS
- Type: Multivariate
- History: 14 days

4

Foundation Models

Chronos, Tiny Time Mixers (TTM), Lag-Llama, TimeGPT

Models

Chronos

•**Type:** Univariate
•**History (context window):** 512 steps (approx. 10.5 days for 30-minute data)
•**Pre-training approach:** Trained by treating time series as sequences of tokens, based on the T5 language model

Lag-Llama

•**Type:** Univariate
•**History (context window):** 1200 steps (approx. 25 days)
•**Pre-training approach:** Pre-trained on a wide range of heterogeneous data from the Monash time series repository

TimeGPT

•**Type:** Multivariate
•**History (context window):** 1008 steps (approx. 21 days for 30-minute data)
•**Pre-training approach:** Trained on over 100 billion data points from economics, finance, and weather data

TTM

•**Type:** Multivariate
•**History (context window):** 512 steps (approx. 10.5 days for 30-minute data)
•**Pre-training approach:** Trained with a focus on low computational requirements (based on an MLP architecture instead of a Transformer)

1 — Baselines

ARIMA, SARIMA
- Type: Univariate
- History: 14 days

2 — Classic ML

XGBoost
- Type: Multivariate
- History: 14 days
- Role: feature importance

3 — Neural Networks

TCN, BiLSTM, N-HITS
- Type: Multivariate
- History: 14 days

4 — Foundation Models

Chronos, Tiny Time Mixers (TTM), Lag-Llama, TimeGPT

Results – Statistical Models

	Model	MAE	RMSE	R^2	wMAPE
Frequency: 30 minutes	ARIMA	34.6808	63.8450	0.5086	67.23%
	SARIMA	**31.0401**	**63.0046**	**0.5214**	**60.17%**
Frequency: 1 hour	ARIMA	67.2692	123.6967	0.5238	65.20%
	SARIMA	**55.1978**	**109.6956**	**0.6255**	**53.50%**

SARIMA outperforms ARIMA across both time scales (30min & 1H)

Notes:

- The hourly scale (1H) performs better due to noise smoothing
- Metrics MAE / RMSE are increased (≈ doubled) in 1H: Expected behavior, reflecting the larger scale of aggregated energy

These models serve as the essential **baselines** for evaluating the subsequent Deep Learning algorithms

Results – Classic ML

	Model	MAE	RMSE	R^2	wMAPE
Frequency: 30 minutes	XGBoost	26.1183	51.6282	0.6786	50.63%
Frequency: 1 hour	XGBoost	50.1112	99.2538	0.6934	48.57%

XGBoost shows a clear improvement **compared to the statistical models**. Specifically, in 1-hour mode, **wMAPE improved by ≈ 5%** and R^2 improved by 0.07

Results – Neural Networks

	Model	MAE	RMSE	R^2	wMAPE
Frequency: 30 minutes	**TCN**	**16.0903**	**36.6814**	**0.8380**	**31.15%**
	BiLSTM	16.2529	36.4373	0.8401	31.46%
	N-HITS	17.8225	39.2755	0.8140	34.55%
Frequency: 1 hour	TCN	29.9709	67.0011	0.8604	29.01%
	BiLSTM	**29.3421**	**66.0079**	**0.8645**	**28.40%**
	N-HITS	33.6287	72.1772	0.8379	32.59%

TCN and BiLSTM demonstrate nearly identical performance, clearly **outperforming N-HITS** across both time resolutions (30min and 1H)

Results – Foundation Models

Frequency: 30 minutes

Model	MAE	RMSE	R^2	wMAPE
Chronos-ZS	30.9040	68.3061	0.4341	60.34%
Chronos-FT	27.8823	59.7449	0.5670	54.44%
TTM-ZS	33.4445	57.1504	0.6038	65.30%
TTM-FT	26.2306	54.3479	0.6417	51.21%
Lag-Llama-ZS	34.8717	65.0875	0.4813	68.92%
Lag-Llama-FT	23.5158	52.9775	0.6564	46.48%
Time-GPT-ZS	**19.7306**	**35.3812**	**0.8482**	**38.52%**
Time-GPT-FT	**16.0800**	**30.5759**	**0.8866**	**31.40%**

Worst model (zero-shot): Lag-Llama
Worst model (fine-tuned): Chronos

Impressive reduction in wMAPE from 68.92% (ZS) to 46.48% (FT)

Best model (zero-shot): Time-GPT
Best model (fine-tuned): Time-GPT

Results – Foundation Models

Frequency: 1 hour

Model	MAE	RMSE	R^2	wMAPE
Chronos-ZS	63.5596	138.6178	0.4082	60.90%
Chronos-FT	51.7679	109.3718	0.6316	49.61%
TTM-ZS	54.4169	102.3420	0.6774	52.14%
TTM-FT	50.9432	102.8635	0.6741	48.81%
Lag-Llama-ZS	54.3312	112.7788	0.6005	53.43%
Lag-Llama-FT	47.0874	104.0540	0.6599	46.30%
Time-GPT-ZS	**32.2455**	**60.5034**	**0.8873**	**30.90%**
Time-GPT-FT	**25.8414**	**49.0476**	**0.9259**	**24.76%**

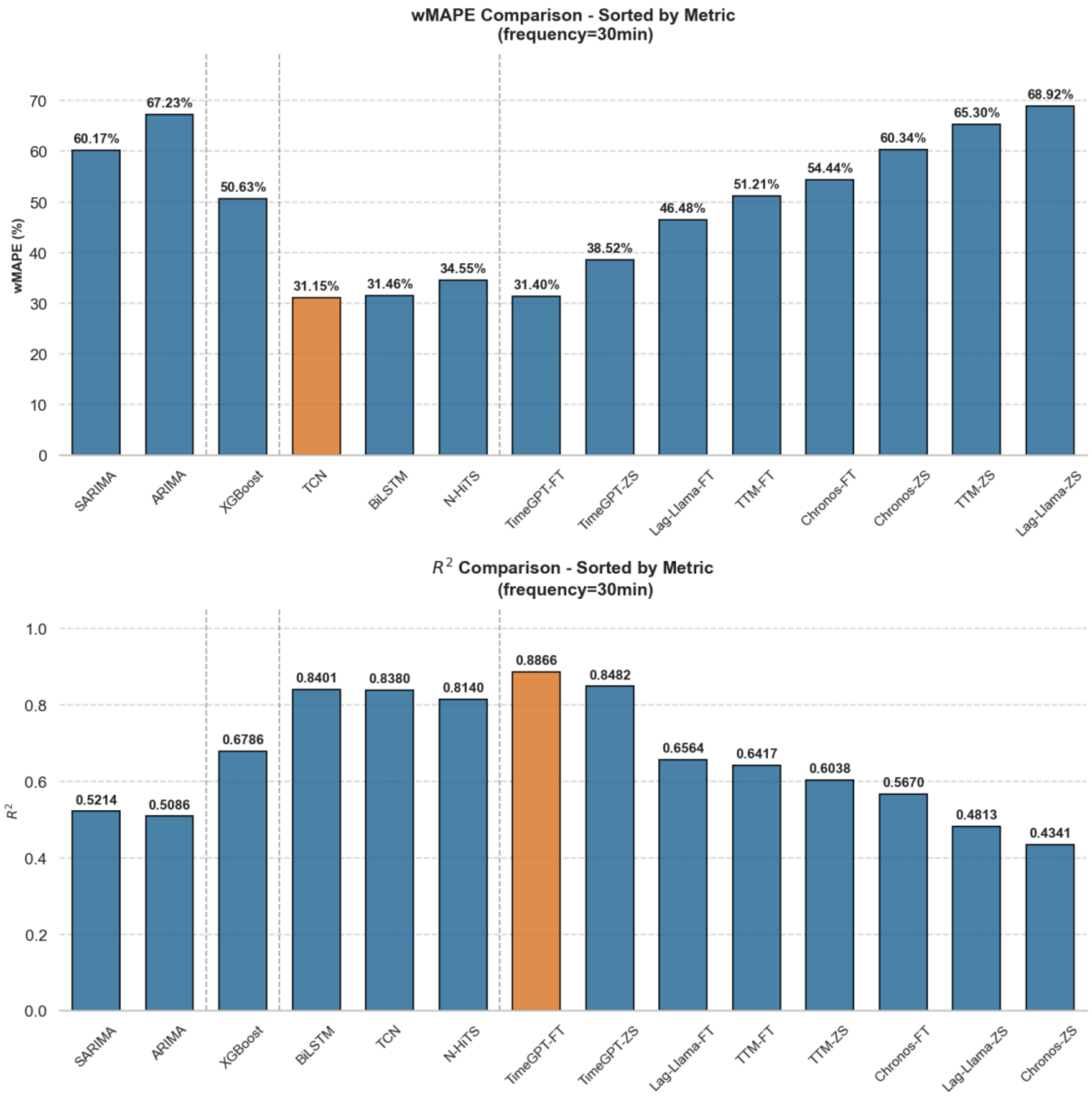
Worst model (zero-shot): Chronos
Worst model (fine-tuned): Chronos

Best model (zero-shot): Time-GPT
Best model (fine-tuned): Time-GPT

Results

Frequency: 30 minutes

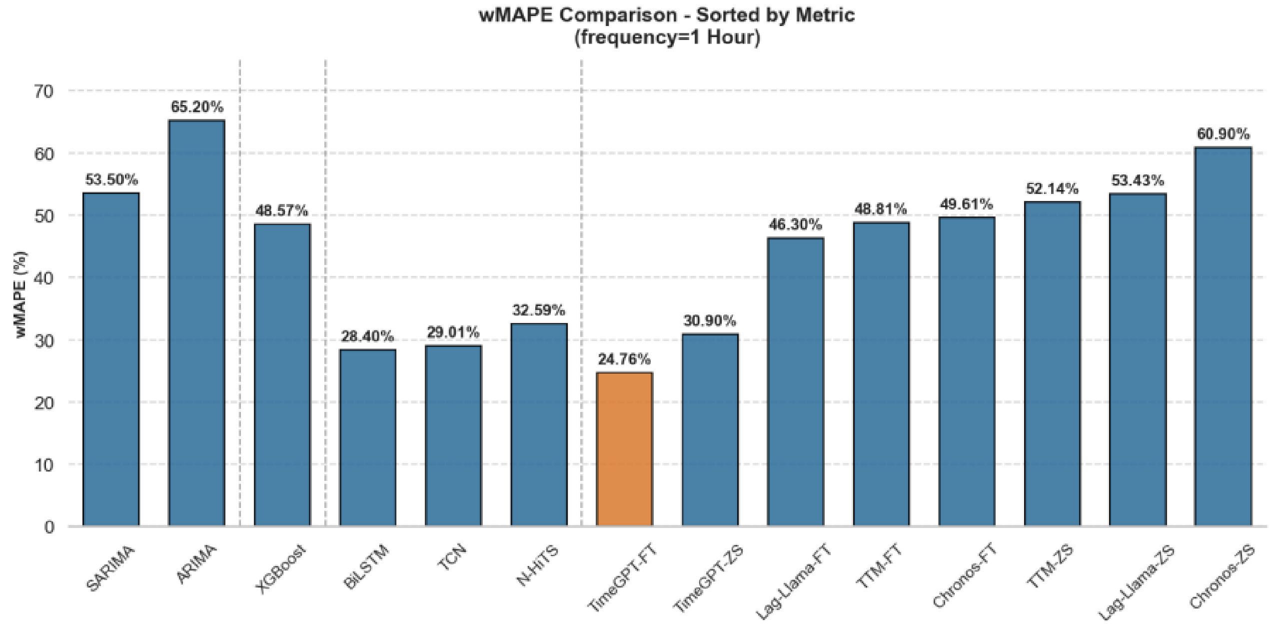
- TCN achieves the lowest **wMAPE** at 31.15%
- TimeGPT-FT achieves the highest $\boldsymbol{R^2}$ score at 0.8866
- **Overall best model: TimeGPT-FT**. Although TCN has a slight advantage in wMAPE, TimeGPT-FT is extremely close (31.40%) and its significant lead in R^2 makes it the strongest overall performer

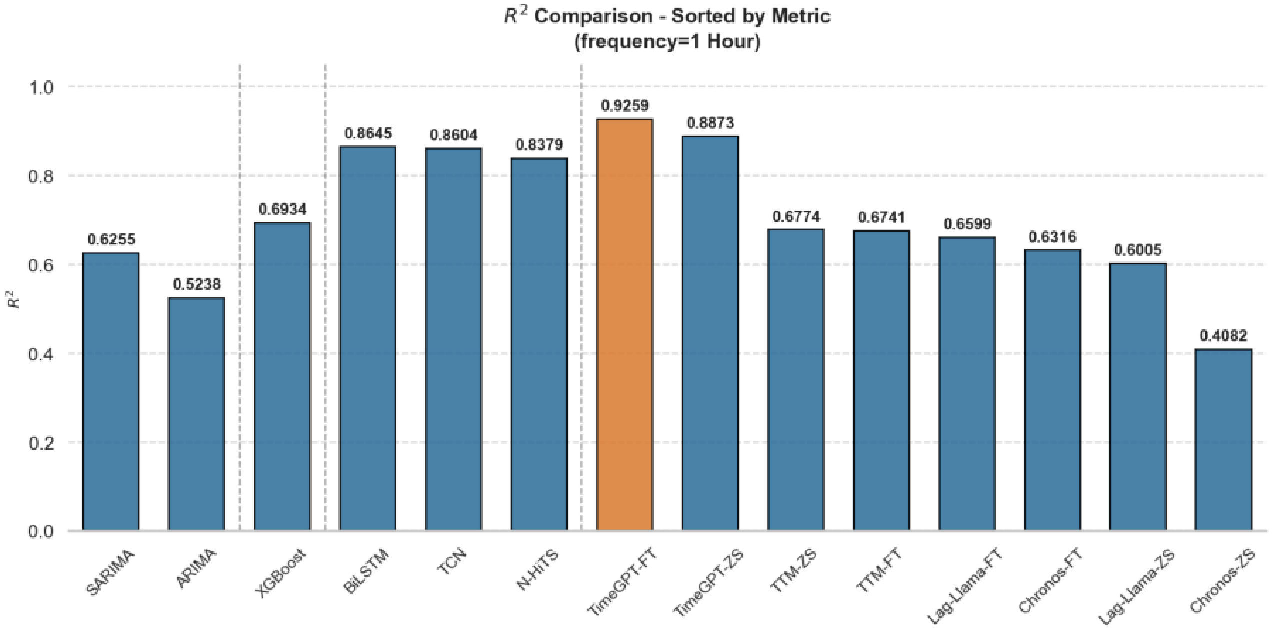


Results

Frequency: 1 hour

Overall best model: TimeGPT-FT achieved the best scores, it is clearly the top model for the 1-hour resolution





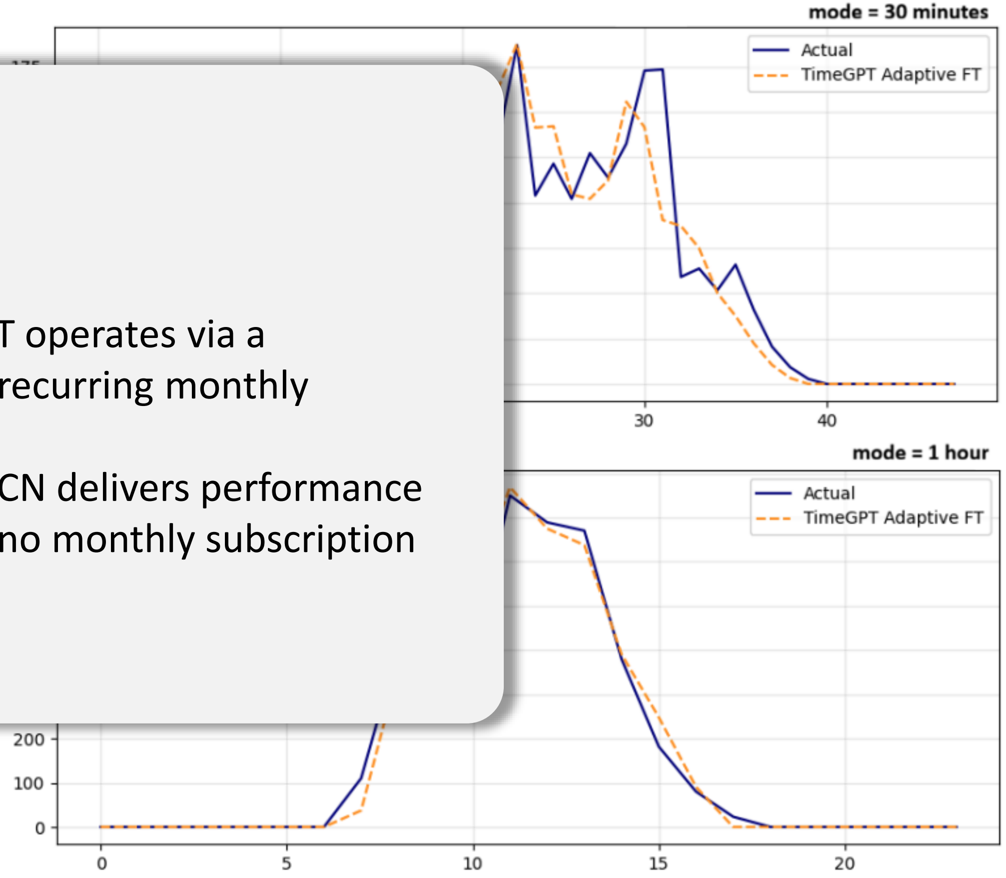
Introduction
Dataset
Data Preprocessing
Feature Engineering
Evaluation Settings
Models
Results
Current Focus
Conclusions

Results – Daily

	Model	Avg Daily	Avg Daily Error	Avg Relative
Frequency: 30 minutes	TCN			
	Time-GPT-			
Frequency: 1 hour	TCN			
	Time-GPT-			

- **Cost consideration:** TimeGPT operates via a commercial API, requiring a recurring monthly subscription
- **Cost-effective alternative:** TCN delivers performance very close to TimeGPT, with no monthly subscription costs

- The daily aggregated metrics fully validate the superior performance of TimeGPT
- The plot illustrates TimeGPT-FT's predictions across randomly selected days



Introduction
Dataset
Data Preprocessing
Feature Engineering
Evaluation Settings
Models
Results
Current Focus
Conclusions

Next steps

Step 1: Scaling up

Step 2: Cross-country testing

Step 3: Fine-tuning & optimization

Next steps

Step 1: Scaling up

Expanding the training of the initial model (developed on UK data) from a subset of 100 systems to the complete pool of 30.000 available PV installations

Step 2: Cross-country testing

Step 3: Fine-tuning & optimization

Next steps

Step 1: Scaling up

Expanding the training of the initial model (developed on UK data) from a subset of 100 systems to the complete pool of 30.000 available PV installations

Step 2: Cross-country testing

We will attempt to deploy and test the UK-trained model directly in the Greek environment to evaluate its performance and assess how well it generalizes to a new geographic region.

Step 3: Fine-tuning & optimization

Next steps

Step 1: Scaling up

Expanding the training of the initial model (developed on UK data) from a subset of 100 systems to the complete pool of 30.000 available PV installations

Step 2: Cross-country testing

We will attempt to deploy and test the UK-trained model directly in the Greek environment to evaluate its performance and assess how well it generalizes to a new geographic region.

Step 3: Fine-tuning & optimization

To achieve the highest predictive accuracy, we will either train a new model from scratch using local data or fine-tune the existing UK model with Greek datasets, comparing both approaches to identify the optimal solution.

New dataset

New dataset

43
Installations

Range:
2016-current

Greek dataset

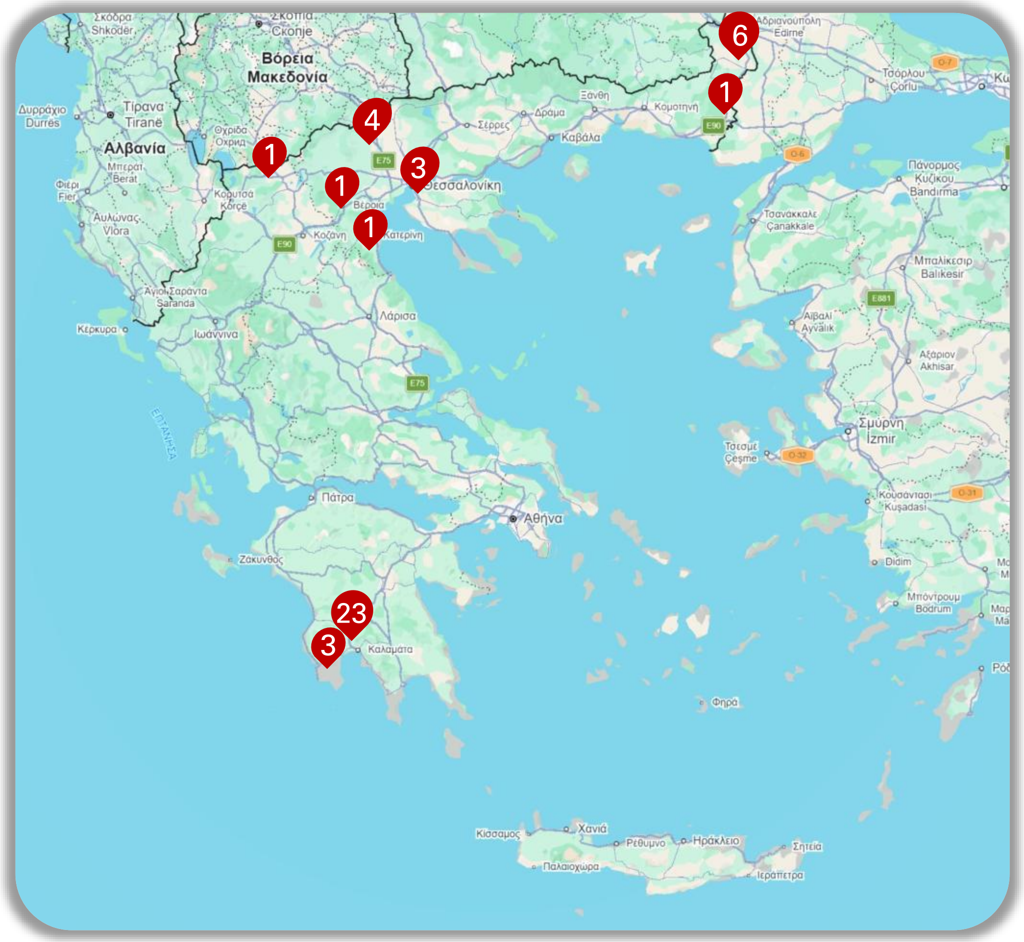
Frequency:
15 min/ 1 hour

OpenWeather

Weather data

New dataset

Locations of installations



Introduction
Dataset
Data Preprocessing
Feature Engineering
Evaluation Settings
Models
Results
Current Focus
Conclusions

Current Focus

Following comprehensive research, our current focus is the development of a day-ahead forecasting solution, designed to add value for both energy producers and aggregators

Under the European electricity market framework, energy producers are required to declare their day-ahead generation output with high accuracy. Any deviations from these declarations result in **financial penalties**

As direct market participation is often technically complex and economically prohibitive for independent producers (e.g., small-scale photovoltaic operators), an **aggregator** (FOSE) provides essential market access by undertaking the following:

- Aggregating multiple power plants into a unified portfolio
- Forecasting generation output with high accuracy
- Executing strategic bids (day-ahead, intra-day & balancing)
- Mitigating financial risks for individual producers

Conclusions

Data Complexity

Solar power data is very complicated and constantly changing. Traditional models aren't enough, so we need to use advanced AI models, such as neural networks, to predict accurately.

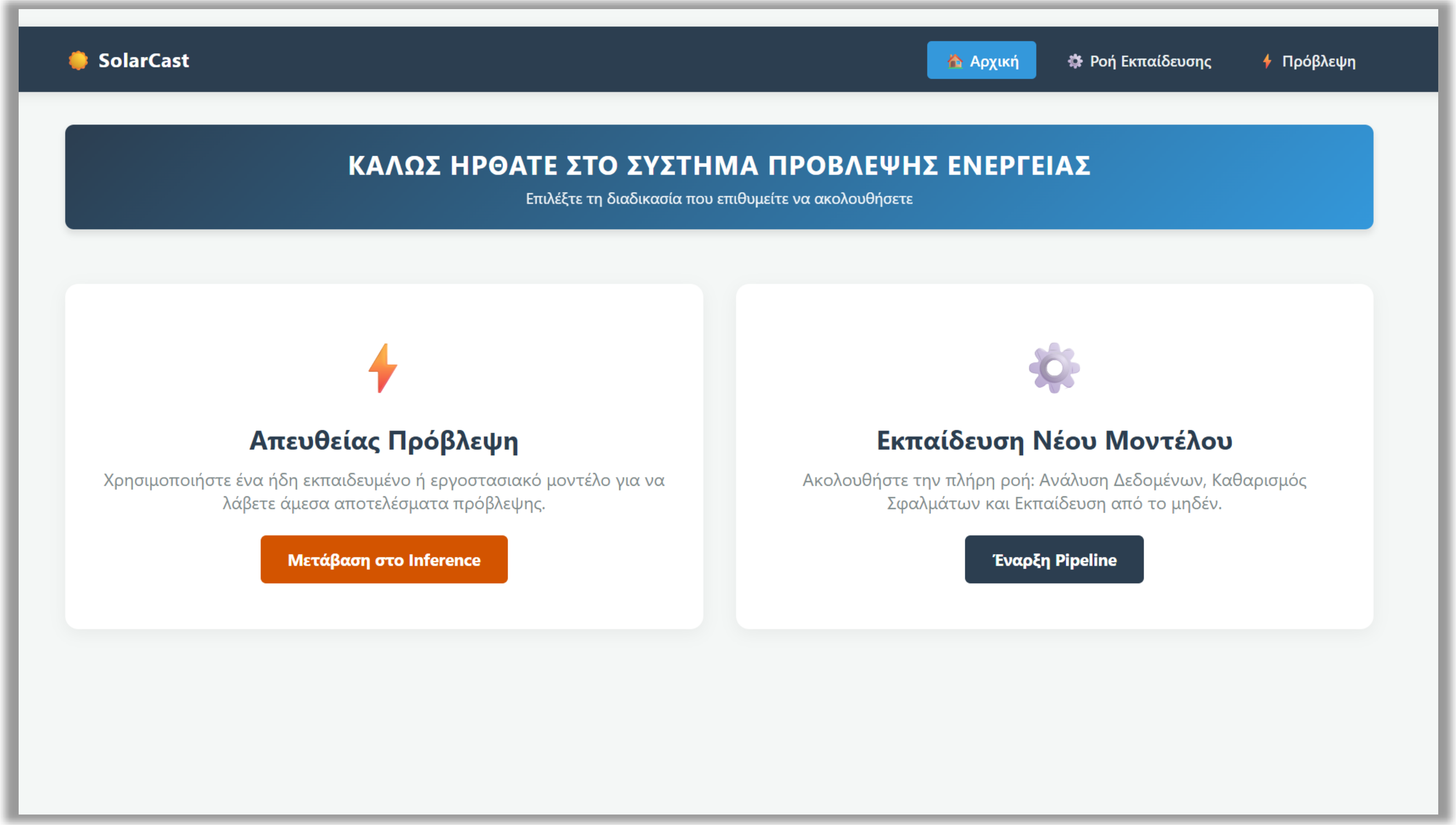
Balancing the Grid

Accurate daily predictions stop grid failures and make it easier to match electricity supply with demand.

Financial & Practical Benefits

By reducing prediction mistakes, this system helps the smart grid work more efficiently and saves solar producers from paying heavy fines.

UI Demo



Introduction
Dataset
Data Preprocessing
Feature Engineering
Evaluation Settings
Models
Results
Current Focus
Conclusions

PHAROS

THE GREEK AI FACTORY

Thank you!

Any questions?

PHAROS

THE GREEK AI FACTORY

Forecasting Renewable Energy Production

Anestis Ampatzidis, AI Consultant
Vasileios Kochliaridis, PhD
Vlahavas Ioannis, Prof.

ARISTOTLE UNIVERSITY OF THESSALONIKI