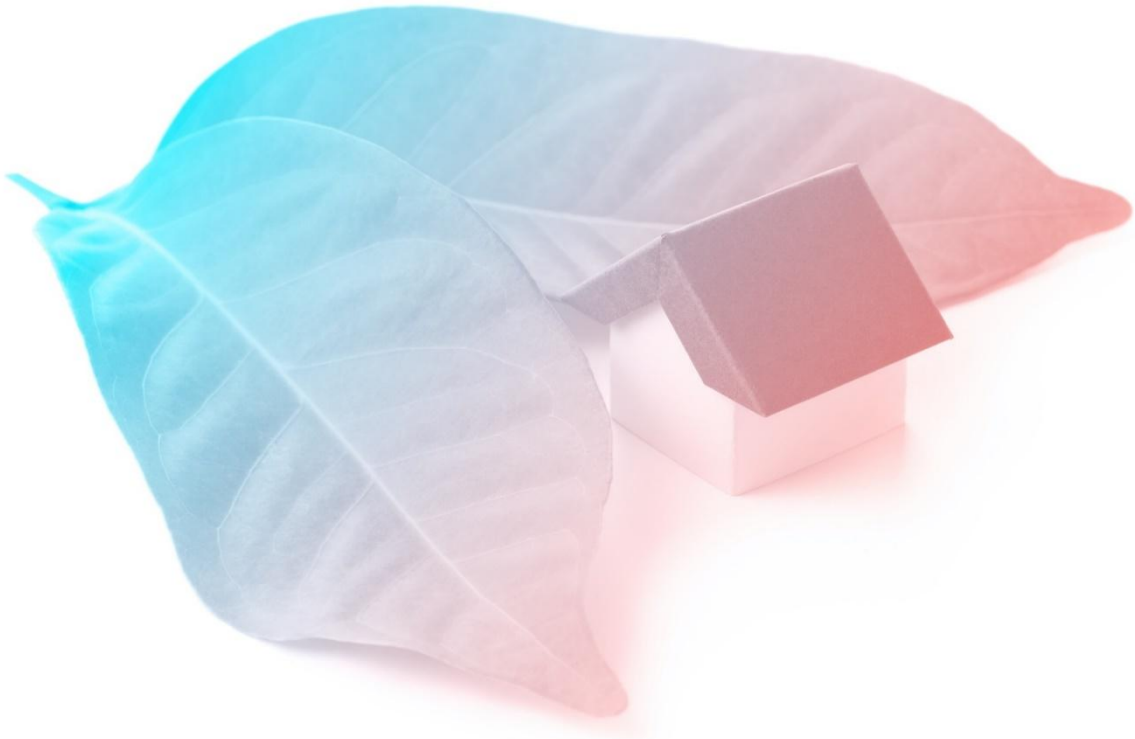
D5.3 Design of the forecasting modes



Authors: Luis Jimeno (CARTIF), Alberto Belda (CARTIF), Roberto Arnanz (CARTIF), Sofía Mulero (CARTIF), Aris Mystakidis (CERTH), Nikolaos Tarsounas (CERTH), Nikolaos Tsalikidis (CERTH)

This project has received funding from the European Union´s Horizon 2020 research and innovation programme under the grant agreement No 869821

D5.3 - Design of the forecasting modes

Summary
Deliverable D5.3 presents the design and validation of forecasting modes for Distributed Energy Resource (DER) generation and energy demand within the MiniStor Smart Home Energy Management System (SHEMS). For generation forecasting, Multi-Layer Perceptron (MLP) Artificial Neural Networks (ANNs) were trained to predict irradiance, photovoltaic (PV) electrical output and solar thermal power at the five European demo sites, achieving improved accuracy compared to raw meteorological forecasts while also highlighting limitations in complex climatic or operational contexts. For demand forecasting, a range of Machine Learning (ML) models was benchmarked, showing that ensemble methods such as Extra Trees (ETs) and Voting Regressor (VotRegr) provided robust performance, while Deep Learning (DL) and Transformer models showed potential but required further tuning. The forecasting modes developed here have been integrated into the SHEMS alongside the control and optimization strategies and IoT platform defined in deliverables D5.2 and D5.5, supporting MiniStor's objectives of maximizing renewable self-consumption, reducing costs and ensuring user comfort.

Deliverable Number	Work Package
D5.3	WP. 5
Lead Beneficiary	**Deliverable Author (s)**
CARTIF Technology Centre (CARTIF) Centre for Research and Technology, Hellas (CERTH)	Luis Jimeno (CARTIF), Alberto Belda (CARTIF), Roberto Arnanz (CARTIF), Sofía Mulero (CARTIF), Aris Mystakidis (CERTH), Nikolaos Tarsounas (CERTH), Nikolaos Tsalikidis (CERTH)
Beneficiaries	**Deliverable Reviewer (s)**
CERTH Psyctotherm Woodspring	Paschalis A. Gkaidatzis Apostolos Gkountas Zoltan Kovari
Planned Delivery Date	**Actual Delivery Date**
30/06/2025	29/08/2025

Type of deliverable	R	Report	X
	Demo	Demonstrator	
	PU	Public	X
	CO	Confidential	

Index

List of tables

List of figures

List of acronyms and abbreviations

AI	Artificial Intelligence
ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Artificial Neural Networks
API	Application Programming Interface
ARIMA	Autoregressive Integrated Moving Average
CAMS	Copernicus Atmosphere Monitoring Service
CDD	Cooling Degree Days
DER	Distributed Energy Resource
DHW	Domestic Hot Water
DL	Deep Learning
DNI	Direct Normal Irradiance
DT	Decision Tree (regressor)
ELM	Extreme Learning Machine
ET	Extra Trees
FFA	Firefly Algorithm
FFNN	Feed Forward backpropagation Neural Network
GHI	Global Horizontal Irradiance
HDD	Heating Degree Days
HGBR	Hist Gradient Boosting Regressor
HMM	Hidden Markov Model
HVAC	Heating, Ventilation and Air Conditioning
JSON	JavaScript Object Notation
LGBM	Light Gradient Boosting Machine
LSSVM	Least Square Support Vector Machine
LSTM	Long Short-Term Memory
LTLF	Long-term Load Forecasting
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MARS	Multivariate Adaptive Regression Splines
ME	Mean Error
MERRA-2	Modern-Era Retrospective analysis for Research and Applications
MS	MiniStor system
MSE	Mean Squared Error
ML	Machine Learning
MLP	Multi-Layer Perceptron
MLR	Multiple Linear Regression
MTLF	Medium-Term Load Forecasting
NRMSE	Normalized Root Mean Squared Error
PV	Photovoltaics
PVT	Photovoltaic-Thermal
R^2	R-squared
RBFN	Radial Basis Function Network
ReLU	Rectified Linear Unit
RF	Random Forest
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
SHEMS	Smart Home Energy Management System
SoDa	Solar radiation Data
ST	Solar Thermal
STLF	Short-Term Load Forecasting
SVM	Support Vector Machines
SVR	Support Vector Regression

TLR	Time-Lagged Recurrent Network
UTC	Universal Time Coordinate
VotRegr	Voting Regressor
VSTLF	Very Short-Term Load Forecasting
WA	Weatherbit API
WS	Weather Station
ZIVR	Zero Inflation Voting Regressor

Executive summary

Deliverable D5.3 “Design of the forecasting modes” presents the development and validation of forecasting approaches for renewable energy generation and energy demand within the MiniStor System (MS). These forecasting modes are a core component of the Smart Home Energy Management System (SHEMS), enabling proactive coordination between renewable generation, demand profiles, and operation.

For Distributed Energy Resource (DER) generation forecasting, Artificial Neural Networks (ANNs) based on Multi-Layer Perceptrons (MLPs) were trained to predict Global Horizontal Irradiance (GHI), photovoltaic (PV) electrical power, and solar thermal (ST) power across the five demonstration sites. The models achieved significant improvements compared to raw meteorological forecasts, particularly in reducing systematic biases and capturing nonlinear relationships among meteorological variables such as irradiance, temperature, wind, and cloud cover. Performance varied depending on local climate and system configuration: strong results were obtained in Thessaloniki and Santiago for PV generation, while limitations appeared in Sopron and Cork, and in thermal prediction where auxiliary devices and control logics introduced discontinuities. These findings confirm both the potential and the constraints of MLP-ANNs when applied to real hybrid solar systems.

For energy demand forecasting, multiple Machine Learning (ML) techniques were benchmarked and applied across demonstrators. The results showed that no single model was universally optimal. Tree-based and ensemble methods such as Extra Trees (ET) and Voting Regressor (VotRegr) delivered robust and consistent performance, often outperforming more complex Deep Learning (DL) models such as Long Short-Term Memory (LSTM) networks, particularly over longer forecasting horizons. Transformer-based models showed promise but suffered from degradation in accuracy with increasing prediction steps. Data quality and availability were critical factors, with strategies such as merging datasets, applying normalization, and focusing on recent stable data proving essential to achieve reliable models.

Overall, the forecasting modes developed in D5.3 provide validated predictive capabilities for both Distributed Energy Resource (DER) generation and demand within the MiniStor system framework. Their integration into the SHEMS, in connection with the control and optimization strategies defined in deliverables D5.2 and D5.5, enable the MiniStor platform to align renewable supply with demand, optimize the operation, and contribute to the project’s overarching objectives of maximizing renewable self-consumption, reducing energy costs, and ensuring occupant comfort.

1. Introduction

Deliverable D5.3, titled *Design of the forecasting modes*, presents the methodologies, implementation and validation of forecasting models for renewable energy generation and energy demand within the MiniStor system. These forecasting modules are a central component of the SHEMS, as they enable predictive optimization of DER generation deployed at the demonstration sites and their energy demand. By anticipating generation and demand patterns, the models provide the necessary inputs for proactive scheduling of resources, improving renewable self-consumption, reducing costs and enhancing comfort.

This deliverable builds on the requirements and system characterization defined in earlier work packages, particularly the SHEMS architecture, optimization and control strategies presented in D5.1, and consolidated in D5.2. It documents the development of two main forecasting streams: DER generation forecasting, led by CARTIF, and energy demand forecasting, led by CERTH. Together, these components ensure that the SHEMS has access to reliable short-term predictions for all relevant energy flows in the MiniStor system.

1.1. Scope and objectives of the deliverable

The scope of D5.3 is to design, train, validate and benchmark forecasting models tailored to the needs of the MS demonstration sites. The objectives of this deliverable are threefold:

- Develop DER generation forecasting models.
- Develop demand forecasting models.
- Provide validated forecasting modules for integration into SHEMS.

1.2. Structure of the deliverable

For clarity and completeness, the document is structured as follows:

- **Section 1 – Introduction:** presents the scope, objectives, structure and relations of the deliverable.
- **Section 2 – Forecasting techniques for energy demand and solar generation:** describes the ML and DL forecasting methodologies and evaluation metrics for both generation and demand forecasting.
- **Section 3 – PVT energy forecasting:** details the ANN-based models for irradiance, electrical and thermal power across the different demo sites, highlighting site-specific challenges and performance indicators.
- **Section 4 – Demand energy forecasting:** presents the development, evaluation and lessons learned from the demand prediction models at each demonstration site, including a comparison of algorithms and data treatments.
- **Section 5 – Conclusions:** summarizes the main findings, highlights strengths and limitations of the forecasting models, and outlines their role in supporting control and optimization within the SHEMS.

This structure ensures a coherent narrative from methodology to validation and integration, while maintaining parallel treatment of generation and demand forecasting streams.

1.3. Relation to other tasks and deliverables

Deliverable D5.3 is directly connected with several tasks within WP5 and across other work packages:

- **WP2 – MiniStor specifications and operation modes:** provides the system parameters (D2.4) that inform the forecasting requirements.
- **WP5 – EMS design and integration:** D5.3 delivers the forecasting modules that are integrated into the SHEMS architecture defined in D5.1 and D5.2. It also provides the predictive signals required for the optimization and scheduling strategies, and in general the forecasting models that have been integrated into the IoT platform (D5.4 and 5.5).
- **WP6 – Demonstration and monitoring:** the models developed in D5.3 are validated against data collected at the five European demo sites and their performance will be continuously assessed during the demonstration phase, in line with the KPIs defined in D6.1.

In summary, D5.3 provides the forecasting backbone of the MiniStor SHEMS, ensuring that both renewable energy availability and user demand are anticipated and aligned. Its outputs are key enablers for the proactive, optimized operation of hybrid renewable systems targeted by the project.

2. Forecasting techniques for energy demand and solar generation

This section introduces the main concepts and algorithms commonly used in energy demand and generation forecasting. It first analyses the different forecasting horizons and their applications, followed by a state-of-the-art review. Demand forecasting methods are discussed first, and subsequently techniques for modelling PV and ST generation. The most relevant algorithms are theoretically analysed to provide an overview of their main characteristics. The section also outlines the most commonly used parameters for evaluating forecasting model performance and concludes with a review of the features that most significantly affect the demand and generation forecasting problem.

2.1. Load forecasting horizons

The load forecasting problem can be addressed over different time horizons. This step is essential for the operation and planning of power generation, transmission and distribution:

- Very Short-Term Load Forecasting (VSTLF): provides projections ranging from a few minutes to an hour in advance, making it appropriate for ongoing surveillance.
- Short-Term Load Forecasting (STLF): ranges from a few hours up to one week ahead. This horizon is particularly relevant for system operators to regulate loads and is widely used in several smart grid applications.
- Medium-Term Load Forecasting (MTLF): covers periods from one week up to one year.
- Long-Term Load Forecasting (LTLF): spans from one year to ten years ahead. This type of forecast is useful to maintain a long-term balance between production and demand, considering external factors such as economic and social indicators.

2.2. State of the art in demand prediction algorithms

Demand prediction has been extensively analysed in the literature. In the short-term horizon, it enables planning of the energy required to reach a desired comfort level and target setpoint temperature. When combined with information on resource availability, it supports the definition of optimal operation modes. Approaches range from conventional statistical methods to Artificial Intelligence (AI)-based techniques. The most widely applied methods include simple and Multiple Linear Regression (MLR), Support Vector Machines (SVM), ANN and hybrid approaches combining two or more techniques.

- Regression: simple and multiple linear regression analyses are presented in (Fumo & Biswas, 2015), considering hourly and daily data. The time interval is shown to be a relevant factor affecting model quality. In addition, MLR models using outdoor temperature and solar radiation improve the coefficient of determination. In (Baltputnis, Petrichenko, & Sobolevsky, 2018), a polynomial regression model is applied and grouping residuals by hour-of-day significantly reduces forecast error.
- Support Vector Machines (SVMs): short-term multi-step heat load prediction models for district heating consumers were developed using SVMs with Firefly Algorithm (FFA) optimization in (Al-Shammari, et al., 2016). Support Vector Regression (SVR) was applied in (Wang, Srinivasan, & Shi, Artificial Intelligent Models for Improved Prediction of Residential Space Heating, 2016) to

predict hourly residential heating electricity use for a typical single-family house. SVR outperformed other AI models in this application.

- Artificial Neural Networks (ANNs): a methodology was proposed in (Deb, Eang, Yang, & Santamouris, 2016) to forecast diurnal energy consumption of institutional buildings, using two years of data with high variability due to university schedules and vacation periods. The ANN trained on five previous days of data was able to forecast next-day energy use with good accuracy. In (Pino-Mejías, Pérez-Fargallo, Rubio-Bellido, & Pulido.Arcas, 2017), ANN and regression model results were compared: linear regression models performed better when predictor variables were transformed, whereas MLPs trained on untransformed variables provided greater accuracy in predicting heating and cooling demand. In (Jovanovic, Sretenovic, & Zivkovic, 2015) several ANN configurations were tested, including feed-forward backpropagation neural networks (FFNN), radial basis function networks (RBFN) and adaptive neuro-fuzzy inference systems (ANFIS). To further improve prediction accuracy, ensembles of ANNs were examined, with three different output combination strategies. All proposed architectures predicted heating consumption with high accuracy, with ensembles achieving the best results.

- Ensemble and hybrid models: in the hybrid-modelling approach, several combinations of physics-based models' parameters with data-driven methods are considered. The hybrid models provide solutions for several limitations of data-driven approaches. To be more specific, the input parameters of the physical-based models can be enhanced by the usage of data-driven methods. In (Siddharth, Ramakrishna, Geetha, & Sivasubramaniam, 2011), the input parameters of the analytical model were pre-processed through the usage of genetic algorithms, providing interesting findings. Another example of a hybrid forecasting model follows a Gaussian algorithmic process for the correction and adjustment of errors in the analytical model (Massa Gray & Schmidt, 2018). Moreover, the usage of ML algorithms as parts of the analytical models is reported in several cases (Banihashemi, Ding, & Wang, 2016), such as the Energy Plus (National Renewable Energy Laboratory (NREL), 2017) open-source environment. The Article by (El Khantach, Hamlich, & Belbounaguia, 2019) proposes an energy demand model applicable to short-term forecasting. This approach utilizes the periodic variation of historical load data, dismantled per hour of the day in time-series format and used as the input of ML methods.

2.3. State of the art in PV and ST modelling techniques

The literature shows extensive research on short-term forecasting techniques for PV and ST generation. These methods are often extensions of those developed for solar irradiance prediction. Broadly, three categories can be distinguished: physical models, statistical models and AI-based approaches.

- Physical models describe the system as a function of independent variables (e.g., PV cell characteristics, meteorological parameters).

- Statistical models include time series and regression-based methods, which rely on explicit mathematical relationships between inputs and outputs.

- AI methods (such as ANNs or SVMs) learn complex nonlinear relationships directly from data without requiring explicit models. While they can deliver very good results, their performance depends strongly on parameter tuning, which is often achieved through trial-and-error.

The following subsections present a literature review of the most common approaches for ST, PV and hybrid Photovoltaic-Thermal (PVT) systems.

2.3.1. ST modelling techniques

Analytical models for ST systems require significant computational effort to achieve accurate solutions. As an alternative, data-driven and ML techniques have gained relevance, providing useful information for performance estimation and fault detection. To mitigate the strong dependence on data quality and availability, ensemble learning has also been explored, as it generally outperforms individual models and provides better generalization.

(Ahmad, Reynolds, & Rezgui, 2018) demonstrated that tree-based ensemble methods, including Random Forest (RF), ETs, SVR and Decision Trees (DT), can improve accuracy and reduce computational cost. Despite this, most research has focused on ANNs, as shown by (Bonilla, Carballo, Fernández-Reche, & Valenzuela, 2019).

2.3.2. State of the art in PV power forecasting

Different approaches have been investigated to address PV power forecasting, which can be grouped into three main categories: time series statistical methods, physical models and ensemble methods. The latter have the advantage of mitigating the weaknesses of individual techniques by combining their strengths. By combining several of these approaches, there are two different hybrid perspectives:

- Hybrid-statistical, combining two or more statistical techniques.
- Hybrid-physical, combining statistical techniques with PV performance models (Antonanzas, et al., 2016).

AI-based methods are increasingly popular due to their ability to capture nonlinear and complex structures in the data. In this report, the focus is placed on statistical and AI approaches, including regression, Autoregressive Integrated Moving Average (ARIMA) models, ANNs, SVMs and DL.

- Regression: (Li, He, Su, & Shu, 2016) applied Multivariate Adaptive Regression Splines (MARS) to forecast solar energy production, which does not require any assumptions about the power–predictor relationship. It retains the simplicity of classical MLR and handles nonlinearities effectively.
- ARIMA models: (Lee & Jung, 2017) developed a short-term 24-hour forecasting model, considering weather factors affecting PV generation at a South Korean site. The best single model for each time horizon was selected to minimize forecast error. (Yang, A., & Xie, 2015) proposed an autoregressive model with exogenous variables, integrating local and geographically distributed PV production data.
- ANNs: (Leva, Dolara, Grimaccia, Mussetta, & Ogliari, 2017) proposed an ANN with backpropagation for PV plant forecasting under different weather conditions (sunny, partially cloudy and cloudy). Accuracy was strongly linked to pre-processing of historical data. (Izgi, Yerli, Kaymak, & Şahin, 2012) developed an ANN for a 750 W PV panel, achieving best results for horizons of 5 and 35 minutes.
- SVMs: SVMs have shown strong performance compared to ANNs in many studies (Sobri, Koohi-Kamali, & Rahim, 2018). (Das, et al., 2018) reviewed forecasting techniques and concluded that ANN and SVM are the most common approaches. (Wolff, Kühnert, Lorenz, Kramer, & Heinemann, 2016) compared physical modelling and SVR, showing that SVR achieved promising results. (Shi, Lee, Liu, Yang, & Wang, 2011) proposed a one-day-ahead SVM-based forecast, classifying days by weather conditions and using 15-minute historical production data combined with next-day weather forecasts (maximum, minimal and average temperature).

- DL: DL techniques constitute a part of the ML methods which automatically learns data representations based on deploying a multi-layered NNs. (Kazem, Yousif, Chaichan, & Al-Waeli, 2019) applied DL methods such as Recurrent Neural Networks (RNNs) and Time-Lagged Recurrent Networks (TLRNs) to predict PV performance from temperature and irradiance data, achieving high accuracy.

2.3.3. The PVT problem

Hybrid PVT systems combine ST and PV technologies to maximize overall energy production (Shahsavar, Moayedi, Al-Waeli, Sopian, & Chelvanathan, 2020). Compared to standalone PV panels and ST collectors, PVT systems can deliver higher efficiencies. PV cells integrated in PVT units typically achieve 4–12% higher efficiency than equivalent standalone PV cells, since they are cooled by heat extraction. Moreover, part of the spectrum otherwise lost as heat in PV panels can be recovered for Domestic Hot Water (DHW) or space heating (Shahsavar, Moayedi, Al-Waeli, Sopian, & Chelvanathan, 2020). This integration improves global efficiency while simplifying system design (Barbu & Darie, 2019).

PVT is sometimes used for the primary purpose of improving the electrical performance of the PV component. Other times, the main goal is to improve thermal performance of the ST component; in both scenarios, the objective is to maximize energy gains and minimize losses (Shahsavar, Moayedi, Al-Waeli, Sopian, & Chelvanathan, 2020).

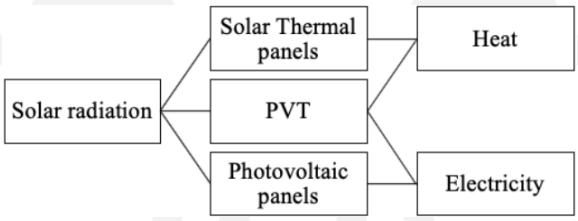


Figure 1: Principle of ST, PV and PVT systems (Barbu & Darie, 2019)

(Barbu & Darie, 2019) simulated PVT panel performance with TRNSYS, using daily DHW profiles based on parameters such as household size, appliances, ambient conditions, seasonal variations and yearly demand. Three meteorological parameters most strongly affect PVT output: ambient temperature, solar irradiation and wind speed. Solar irradiation is the dominant factor, as it drives electrical output and influences material temperatures.

AI-based soft computing approaches are increasingly explored to enhance PVT efficiency and support grid integration, avoiding costly and time-consuming experimental studies (Ahmadi, 2020). ANN models have shown strong performance (Shahsavar, Moayedi, Al-Waeli, Sopian, & Chelvanathan, 2020), while other promising techniques include Adaptive-Network-Based Fuzzy Inference Systems (ANFIS), SVMs, Hidden Markov Models (HMMs) and Extreme Learning Machines (ELMs). (Mojumder, Ong, Chong, Izadyar, & Shamshirband, 2017) estimated electrical and thermal efficiency of a PVT system using ELM, comparing results with Genetic Programming (GP) and ANN, with ELM outperforming both. (Diwania, Agrawal, Siddiqui, & Singh, 2020) also reported strong results for ELM. (Ahmadi, 2020) compared multiple approaches (MLP-ANN, RBF-ANN, ANFIS, LSSVM), finding that Least Squares SVM (LSSVM) achieved the best performance when modelling inputs such as temperature, flow rate, solar radiation and heat.

2.4. Algorithms for time series forecasting

A time series is an ordered sequence of values recorded at equal time intervals. Modelling such data typically involves two stages: (i) identifying the underlying structure of the series and (ii) fitting a model to generate future predictions. This section briefly introduces the most frequently used time-series models in the literature for forecasting energy demand, ST and PV generation. These models take historical time series as input and produce future values with comparable temporal resolution.

- Multiple Linear Regression (MLR): MLR models the relationship between a dependent variable and several independent variables. It is widely applied to long-term building-level prediction due to its simplicity and ease of use. However, it cannot adequately capture nonlinear relationships (Wang & Srinivasan, A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models, 2016).

- Support Vector Regression (SVR): this technique appears with the perceptron theory (Rosenblatt, 1962). It is based on the concept of decision hyperplanes that distribute the data in two groups, so that when it is found, the largest margin is found between the two groups (maximum width of the plane parallel to the hyperplane that does not contain interior points). In addition to classification, this technique can be used to develop a regression model. The function has to be as simple as possible: an overly complex one will consider all variations in the training set, generating a small error and not generalizing as expected when using invisible data points (data overfitting situation). The original problem is reformulated through the use of a kernel function, which allows mapping lower-dimensional data to a higher one and the implications of using it is that the optimization problem consists of determining the flattest possible function in feature space (Deb, Zhang, Lee, & Shah, 2017). Commonly used kernels include polynomial, sigmoid and radial basis functions (RBF).

- ANNs: ANNs provide a powerful framework for addressing nonlinear forecasting problems. Model development generally involves three phases:
  - Design, defining the network architecture (type of ANN, number of neurons, layers and input/output parameters).
  - Training, where network weights are adjusted to minimize error; this is often the most critical phase.
  - Validation, assessing performance with previously unseen datasets.

- ARIMA models: ARIMA models describe a value as a linear function of previous observations and random errors, while also incorporating cyclical or seasonal components. The general ARIMA (p, d, q) model expresses a time series differentiated d times to address non-stationarity, combining autoregressive (p) and moving average (q) components. Its formula (Eq. 1) encompasses the models previously described, so that if any of the components p, d, q is equal to zero, its terms are removed from the general formula.

$$\left(1-\phi_1 \mathrm{B}-\phi_2 \mathrm{B}^2-\cdots-\phi_p \mathrm{B}^p\right)\mathrm{X}_t=\left(1-\left(1-\nu_1 \mathrm{B}-\nu_2 \mathrm{B}^2-\cdots-\nu_q \mathrm{B}^q\right)a_t \quad \text{Eq. 1}$$

Its seasonal extension, ARIMA (p, d, q) (P, D, Q), captures both long-term trends (regular part) and seasonal cycles (cyclical part). Simplified variants include autoregressive (AR), moving average (MA), or combined ARMA models.

2.5. Most common metrics for model evaluation using ML methods

Several metrics are commonly used to validate predictions and quantify the error of ML models:

- R-squared (R^2): this coefficient measures the proportion of variance in the dependent variable explained by the independent variables. Values range from 0 (no improvement over the average model) to 1 (perfect fit). R^2 is widely used due to its availability in most statistical packages.

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{n}\left(y_t - \tilde{y}_t\right)^2} \quad \text{Eq. 2}$$

- ME (Mean Error): ME represents the average of the errors, where an error is defined as the difference between the measured ($\mathbf{y_t}$) and the predicted value ($\tilde{y}_t$) (Eq. 3):

$$ME = \frac{1}{n}\sum_{t=1}^{n}(y_t - \tilde{y}_t) \quad \text{Eq. 3}$$

  However, its usefulness is limited, as positive and negative errors may cancel each other out.

- RMSE (Root Mean Squared Error): RMSE is the square root of the average squared differences between observed and predicted values (Eq. 4):

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \tilde{y}_t)^2} \quad \text{Eq. 4}$$

  Lower RMSE values indicate better fit. RMSE is sensitive to large errors, making it suitable to detect significant deviations, but some studies highlight its limited reliability for generalization (Amstrong & Collopy, 1992).

- MAE (Mean Absolute Error): MAE represents the average of the absolute differences between observed and predicted values (Eq. 5):

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t - \tilde{y}_t| \quad \text{Eq. 5}$$

  where $\mathbf{y_t}$ is the measured value and $\tilde{\mathbf{y}}_\mathbf{t}$ the predicted value. It is less sensitive to outliers than RMSE and offers a straightforward interpretation as the average prediction error in the original units of the variable.

- MAPE (Mean Absolute Percentage Error): MAPE measures accuracy as a percentage (Eq. 6):

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \tilde{y}_t}{y_t}\right| \quad \text{Eq. 6}$$

  This unit-free metric is easy to interpret and widely applied. Lewis proposed classification ranges to assess prediction quality (Lewis, 1982).

- SMAPE (Symmetric Mean Absolute Percentage Error): SMAPE is a common variant of MAPE, and is designed to address some of its limitations, particularly the division-by-zero

issue when the actual value is zero and the fact that MAPE produces asymmetrical penalties. SMAPE bounds the error between 0% and 200%, providing a more balanced and robust measure of forecast accuracy, especially for data with intermittent or zero values Eq. 7:

$$SMAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{\tilde{y}_t - y_t}{|y_t| + |\tilde{y}_t|}\right| \quad \text{Eq. 7}$$

2.6. Factors influencing the demand

Several factors significantly affect demand variation:

- Meteorological conditions: external weather parameters are fundamental drivers of energy demand. Beyond general ambient temperature and solar radiation, the forecasting models integrate specific meteorological data such as average, maximum and minimum air temperature, solar radiation, air humidity, wind speed, wind direction and derived indicators such as Heating Degree Days (HDD) or Cooling Degree Days (CDD). The inclusion of outdoor temperature and solar radiation has been shown to improve the quality of demand prediction models.
- Internal environmental conditions: information gathered from parameters monitored within the building provides crucial insights into the internal comfort and operational state. For instance, data related to the system's operational mode (such as heating or cooling), the desired temperature setpoints, the actual measured temperature inside the building, CO_2 concentration, indicators of user interaction with controls or the fan speed help to understand the thermal comfort levels and system responses. This detailed internal feedback is vital for accurate demand profiling.
- Occupation and periodicity: demand patterns are heavily influenced by human activity and building occupancy, in particular in residential contexts, such as those of MiniStor demo sites. This factor accounts for variations depending on the day of the week (e.g., weekdays or weekends) or the period of the year (e.g., seasonal periods). To capture these influences, seasonality data such as the month, day of the year, day of the week and hour are explicitly extracted from timestamps and utilized as additional parameters in the forecasting models. This enables the models to learn and predict demand based on recurring patterns.
- Other events: beyond routine factors, unforeseen or external events can significantly impact demand. Where possible, considering other factors such as socio-economic circumstances or specific real-world scenarios helps to explain and account for anomalous demand situations. These "other events" represent external influences that might deviate from typical patterns.

2.7. Factors influencing the PV and ST production

Atmospheric and environmental conditions strongly affect PV and ST system performance (Fouad, Shihata, & Morgan, 2017) (Hosenuzzaman, Rahim, Selvaraj, & Hasanuzzaman):

- Solar irradiance: directly drives energy generation, with continuous variations due to weather and cloud cover.
- Solar angle: changes throughout the day affect incident irradiance on panels, which in turn affect the energy production.
- Module temperature: increases in cell temperature reduce open-circuit voltage, lowering PV efficiency.

- Shading: shadows from trees, buildings and other elements surrounding solar systems reduce output significantly. Losses depend on the area and material affected by the shadows; a fully shaded cell can lead to major energy losses.
- Dust accumulation: dust blocks sunlight and reduces generation, although this can be mitigated by cleaning or high irradiance levels.

2.8. Weather forecast services

Weather information is a critical component for accurately forecasting both energy demand and renewable energy generation. Data-driven forecasting approaches rely on two main types of data: historical information and future predictions.

Weather information is essential for forecasting both demand and renewable energy generation. Data-driven approaches require two types of information: historical measurements and forecast data.

- Historical data: obtained from meteorological stations installed at pilot sites, typically recording 4–12 samples per hour depending on station characteristics.
- Forecast data: harder to access, as most services limit the temporal resolution or require paid licenses.

The correlation between weather and energy demand is well-established. For instance, fluctuations in ambient temperature directly influence the use of heating and cooling systems, which are significant contributors to overall energy consumption. Similarly, factors like humidity, wind speed and even cloud cover can impact energy usage patterns in residential and commercial buildings. On the generation side, solar irradiance, cloud dynamics, temperature and wind conditions are decisive drivers of renewable production, particularly for PV, PVT and ST systems, where even short-term variations can cause substantial fluctuations in output. Therefore, access to detailed and accurate weather forecasts is essential for building robust demand and generation prediction models.

Several services offer weather forecasting data with varying capabilities that can be leveraged for both demand-side management and renewable generation forecasting:

- WeatherBit[1]: global coverage, hourly forecasts up to 48h (licensed), daily forecasts up to 16 days (free) and minutely forecasts for precipitation/snowfall. Provides temperature, humidity, cloud cover and solar irradiance (DHI, DNI, GHI). Data are available through an open API (Application Programming Interface) in JSON format (JavaScript Object Notation).
- SoDa project[2] provides historical solar radiation and meteorological data via web services.
- MERRA-2[3]: global reanalysis dataset including temperature, humidity, wind, rainfall and snow variables since 1980, updated monthly, with time steps in historical data which can be set from 1 min up to 1 month.

[1] Weatherbit.io, "Weather API Documentation", 2024 [Online]. Available: https://www.weatherbit.io/

[2] SoDa, “Solar Radiation Data”, 2024 [Online]. Available: https://www.soda-pro.com/

[3] Global Modeling and Assimilation Office (GMAO) (2015), “Modern-Era Retrospective analysis for Research and Applications (MERRA-2)”, 2015 [Online]. Available: https://gmao.gsfc.nasa.gov/gmao-products/merra-2/

- CAMS (Copernicus Atmosphere Monitoring Service)[4]: offers global, direct and diffuse irradiance time series.
- Visual Crossing[5]: this platform provides both historical and forecast data for up to 15 days. Key metrics for demand forecasting include temperature, humidity and wind speed. The ability to receive data in formats like CSV can simplify integration with various modelling and analysis tools.
- Open-Meteo[6]: free and open-source option, Open-Meteo provides up to 16 days of forecasted weather data and extensive historical data. For demand forecasting, the availability of granular data on temperature, humidity and other relevant parameters makes it a strong contender. The inclusion of specialized data, while often used for renewable energy forecasting, can also be used to refine demand models.

For analysis and modelling within the project, data from local weather stations have been used for model training. However, access to forecast data such as that provided by WeatherBit is necessary to generate predictions.

[4] ECMWF (EU), "CAMS (Copernicus Atmosphere Monitoring Service)", 2025 [Online]. Available: https://atmosphere.copernicus.eu/

[5] Visual Crossing, "Global Weather Data & API for Every Application", 2025 [Online]. Available: https://www.visualcrossing.com/

[6] Open-Meteo, "Free Weather API", 2025 [Online]. Available: https://open-meteo.com/

3. PVT energy forecasting

This section describes the approach developed for DER forecasting using ANNs, with meteorological data as input variables; in particular, hybrid PVT solar energy. The methodology is based on MLP models organized in layers that sequentially estimate irradiance, thermal power and electrical power. Additionally, aspects related to model validation across different geographical contexts are included, together with the main challenges encountered in the data and the lessons learned during the implementation process.

3.1. Data handling

This subsection outlines the strategic framework adopted for handling data in the development of forecasting models within the MiniStor project. The workflow comprises several stages, from the collection and preparation of raw data to the evaluation of forecasting performance.

3.1.1. Data description

3.1.1.1. Ministor pilots - data streams

Table 1 summarizes the relevant data streams collected from the MiniStor pilots for production forecasting.

Table 1: Data streams from MiniStor pilots regarding production forecasting

Pilot site	Source/ Device	Main monitored parameters	Units	Data resolution
Thessaloniki	Weatherbit API (WA), Weather Station (WS), Ministor System (MS)	GHI (WA), DNI (WA), TEMP (WA), WIND_SP (WA), CLOUDS (WA), AirTemperature (WS), SolarRadiation (WS), averageWindSpeed (WS), Effective_elevation (WS), Deltasol_Pump_1 (MS), Deltasol_Temperature_sensor_1 (MS), Deltasol_Temperature_sensor_10 (MS), Deltasol_Temperature_sensor_7 (MS), Deltasol_Temperature_sensor_8 (MS), Deltasol_Flow_rate_sensor_13_H (MS), froniusgen_module1DCW (MS),	W/m^2 W/m^2 °C m/s % °C W/m^2 m/s ° l/h °C °C °C °C °C W	WA: 1h WS: 15min MS: 15seg (All resampled to 1 h)
Sopron	Weatherbit API, Weather Station (DeltaOHM), MS	GHI (WA), DNI (WA), TEMP (WA), WIND_SP (WA), CLOUDS (WA), AirTemperature (WS), SolarRadiation (WS), averageWindSpeed (WS),	W/m^2 W/m^2 °C m/s % °C W/m^2 m/s	WA: 1h WS: 15min MS: 15seg (All resampled to 1 h)

		Effective_elevation (WS), Deltasol_Pump_1 (MS), Deltasol_Temperature_sensor_1 (MS), Deltasol_Temperature_sensor_10 (MS), Deltasol_Temperature_sensor_7 (MS), Deltasol_Temperature_sensor_8 (MS), Deltasol_Flow_rate_sensor_13_H (MS),	° l/h °C °C °C °C °C	
Kimmeria	Weatherbit API, Weather Station (DeltaOHM), MS	GHI (WA), DNI (WA), TEMP (WA), WIND_SP (WA), CLOUDS (WA), AirTemperature (WS), SolarRadiation (WS), averageWindSpeed (WS), Effective_elevation (WS),	W/m^2 W/m^2 °C m/s % °C W/m^2 m/s °	WA: 1h WS: 15min MS: 15seg (All resampled to 1 h)
Santiago de Compostela	Weatherbit API, Weather Station (Davis), MS	GHI (WA), DNI (WA), TEMP (WA), WIND_SP (WA), CLOUDS (WA), AirTemperature (WS), SolarRadiation (WS), averageWindSpeed (WS), Effective_elevation (WS), Deltasol_Pump_1 (MS), Deltasol_Temperature_sensor_1 (MS), Deltasol_Temperature_sensor_10 (MS), Deltasol_Temperature_sensor_7 (MS), Deltasol_Temperature_sensor_8 (MS), Deltasol_Flow_rate_sensor_13_H (MS), froniusgen_module1DCW (MS),	W/m^2 W/m^2 °C m/s % °C W/m^2 m/s ° l/h °C °C °C °C °C W	WA: 1h WS: 15min MS: 15seg (All resampled to 1 h)
Cork	Weatherbit API, Weather Station (Davis), MS	GHI (WA), DNI (WA), TEMP (WA), WIND_SP (WA), CLOUDS (WA), AirTemperature (WS), SolarRadiation (WS), averageWindSpeed (WS), Effective_elevation (WS), Deltasol_Pump_1 (MS), Deltasol_Temperature_sensor_1 (MS), Deltasol_Temperature_sensor_10 (MS), Deltasol_Temperature_sensor_7 (MS), Deltasol_Temperature_sensor_8 (MS), Deltasol_Flow_rate_sensor_13_H (MS), froniusgen_module1DCW (MS)	W/m^2 W/m^2 °C m/s % °C W/m^2 m/s ° l/h °C °C °C °C °C W	WA: 1h WS: 15min MS: 15seg (All resampled to 1 h)

3.1.1.2. Forecasting targets

The main goal of the forecasting models is to predict the energy production of PVT assets, focusing on irradiance, electrical and thermal power. Table 2 summarizes the forecasting targets by pilot site.

Table 2: Forecasting targets for MiniStor pilots

Pilot site	Device name	Target category	Forecast purpose	Forecast horizon
Thessaloniki	PVT (PV+ST)	Irradiance, electrical power, thermal power	High level optimization and control solar subsystems	48h
Kimmeria	-	Irradiance	High level optimization and control solar subsystems	48h
Sopron	ST	Irradiance, thermal power	High level optimization and control solar subsystems	48h
Santiago de Compostela	PVT (PV+ST)	Irradiance, electrical power, thermal power	High level optimization and control solar subsystems	48h
Cork	PVT (PV+ST)	Irradiance, electrical power, thermal power	High level optimization and control solar subsystems	48h

3.1.2. Data preparation

3.1.2.1. Pre-processing

Prior to training and prediction, data consistency was ensured through several pre-processing steps:

- Data alignment: all sources aligned to a common UTC-based index.
- Cleaning & outlier handling: invalid or extreme values (e.g., negative power, unrealistic spikes) removed or capped via quantile-based rules.
- Missing data imputation: gaps filled by linear interpolation. For gaps >1h, available data from other days were used to maintain consistency.
- Time resolution: all data resampled to a 1-hour interval using methods tailored to each variable:
    - Accumulated variables: summed within the interval.
    - Status/state variables: last observed value retained (e.g., operation mode).
    - Continuous variables: median applied to smooth variability and mitigate outliers.

3.1.2.2. Feature engineering

The pre-processed datasets were further refined through feature engineering:

- Interpolation & standardization: all signals, from multiple heterogeneous sources, were initially retrieved and interpolated and resampled to a common 1h granularity. This

resampling step ensures temporal consistency across variables and facilitates downstream alignment.

- Temporal synchronization: inner joins on shared timestamps ensured only periods with valid entries across all sources were retained. This strategy guarantees the availability of multi-dimensional inputs at each time step, mitigating the impact of missing segments in individual sources and preserving data integrity for model consumption.

- Quality control: the merged dataset underwent rigorous data quality control, where outliers and anomalous values (e.g., sensor faults, implausible spikes) were identified and corrected using domain thresholds and statistical rules (e.g., quantile clipping). This step ensures numerical stability and robustness during training.

- Final dataset: the final output was a cleaned, gap-free and aligned multivariate time series, formatted as a 2D input tensor (samples × features), suitable for training FFNN such as MLP-based training.

The dataset was then split into training, validation and testing subsets to support robust model development and evaluation.

3.2. Training strategy and architecture

The methodology used for PV forecasting is based on ML using ANNs, specifically Multi-Layer Perceptron (MLP) models with dense architectures. This approach captures the nonlinear relationships between meteorological variables and energy production, leveraging the ability of ANNs to approximate complex functions without requiring explicit physical models of the system.

The process begins with the acquisition of 48-hour hourly weather forecasts from the Weatherbit API, including GHI, cloud cover, air temperature, Direct Normal Irradiance (DNI) and wind speed. These variables are selected for their demonstrated impact on solar energy generation. All input data are normalized using pre-trained scalers to ensure consistency, improve stability and enhance prediction accuracy.

The models employed have been pre-trained for each demonstrator and are organized into three independent ANNs, each optimized for a specific task: irradiance prediction, PV electrical power prediction and ST power prediction. Each network features a three-layer structure: an input layer corresponding to the number of predictor variables, an intermediate hidden layer with 64 neurons and ReLU (Rectified Linear Unit) activation and a second hidden layer with 32 ReLU-activated neurons. Dropout regularization is applied to prevent overfitting, thereby improving generalization to unseen data.

After inference, the normalized outputs of each model are rescaled using the inverse of the original scalers to obtain values in their real units. The predicted hourly irradiance is used as a common input for the power models, assuming that the relationship between irradiance and energy production may dynamically vary depending on ambient temperature and other local conditions. Finally, the results can be stored in a SQL Server database or sent to an external API for integration into control or visualization systems.

This ANN-based approach (conceptually summarized in Figure 2) offers accurate, flexible and automated predictions of hybrid PVT behaviour under diverse meteorological conditions. Its adoption is supported by its demonstrated effectiveness over other techniques such as SVR or MLR, especially in contexts with high variability and nonlinear interactions among predictor variables.

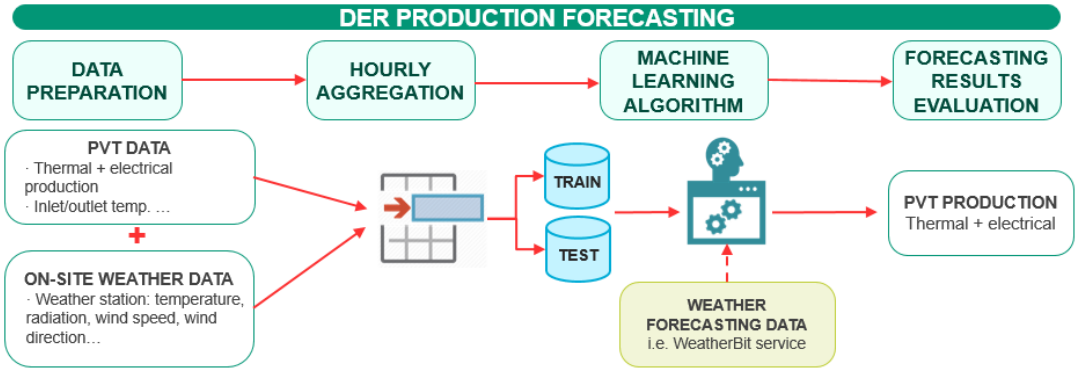


Figure 2: Process for renewable energy generation forecasting (schematic workflow)

IRRADIANCE PREDICTION MODEL

The irradiance model follows the principles established in solar forecasting literature, applying a fully data-driven approach with DL techniques. A feedforward MLP ANN trained through supervised learning is employed, justified by its proven ability to address nonlinear and multivariate forecasting problems.

The methodology applied in the development of the neural model for irradiance prediction is based on principles established in the solar generation literature, combining a purely data-driven approach with DL techniques. The goal is to accurately anticipate the hourly behaviour of real GHI, based on available meteorological variables, in order to enhance the operation and planning of renewable energy systems.

In this case, a feedforward ANN is selected, trained through supervised learning. This choice is justified by its proven ability to address nonlinear and multivariate problems, such as short-term solar irradiance forecasting. The network uses a dense architecture with two intermediate hidden layers, placing it within the framework of MLP models, which are widely recognized in the state-of-the-art for their robustness and versatility in energy prediction tasks.

To feed the model, a set of independent variables derived from hourly weather forecasts is selected, such as predicted GHI, weighted cloud cover, air temperature, DNI and wind speed. This selection is based on previous studies identifying these parameters as the most influential in solar generation, particularly due to their direct relationship with energy capture in PV modules and PVT hybrid systems. The model training process is carried out on filtered and validated historical data, including steps to clean outliers or missing values, followed by normalization using statistical standardization techniques. To assess the model's ability to generalize, a chronological validation strategy is employed that simulates real operating scenarios, reserving a set of specific dates for blind testing outside the training set.

A unique feature of the developed model is the application of custom weights to the input variables, exploring combinations that allow tuning the model's sensitivity to the relative influence of each variable. Although this practice is uncommon in conventional models, it is comparable to regularization techniques and hyperparameter optimization methods commonly found in the specialized literature, particularly when aiming to optimize the architecture without resorting to overly complex models.

Training is conducted using loss functions based on mean squared error (MSE) and evaluated using standard metrics such as MAE and R^2, as recommended in the technical literature. This deep ANN-based approach aligns with the most promising methods in irradiance and solar power forecasting according to the reviewed studies, especially when compared to linear or physics-based methods, as it offers greater capacity to model the nonlinear dynamics typical of complex solar environments.

THERMAL POWER PREDICTION MODEL

This second layer continues the methodological logic previously described, employing MLP ANNs for the prediction of energy-related variables. In this specific case, the ANN is trained to estimate the instantaneous thermal power output of a hybrid solar system, based on meteorological variables and the output of the previous irradiance model.

The strategy also falls within the supervised learning paradigm, using a dense architecture with three hidden layers, ReLU activation functions and dropout regularization. This configuration aligns with approaches reviewed in the state of the art, where MLP models are shown to be particularly effective for multivariable energy forecasting tasks when reliable meteorological and historical measurement data are available.

Unlike the previous model, the dataset used here includes only two input variables: GHI, previously processed through another ANN and the forecasted air temperature. This selection is driven by the need to reduce dimensionality and focus on the variables with the most significant physical impact on thermal generation in solar systems, as widely demonstrated in the literature on PVT systems. Specifically, irradiance directly determines energy capture, while ambient temperature affects the useful thermal gradient and, therefore, the overall efficiency of the thermal collector.

The evaluation metrics used for the model are standard in regression problems within the energy domain: MAE, coefficient of determination (R^2) and MAPE. As highlighted in the literature, these metrics provide a comprehensive view of the model's performance, both in terms of overall accuracy and its ability to explain the variability of real-world data.

In summary, this second layer represents a logical extension of the previous one, complementing the irradiance prediction with an accurate estimation of thermal output, while applying a methodological approach consistent with best practices recognized in the literature on ST systems prediction supported by AI.

ELECTRICAL POWER PREDICTION MODEL

This third layer follows exactly the same methodology used in the previous case, both in pre-processing and in the architecture and training. The only difference lies in the objective of the model, which in this case is to estimate the electrical power generated by the PV system. Although structurally identical, this layer is adjusted to learn the relationship between meteorological conditions and actual electrical output, instead of thermal output.

3.3. Evaluation of generation prediction performance

This section presents the validation results of the ANNs developed to forecast key energy variables in hybrid solar systems. Specifically, it evaluates a single ANN, common to all demonstrators, designed to predict GHI from meteorological variables, as well as two additional networks trained individually for each demonstrator: one to estimate electrical power and another one for thermal power.

Using a single globally trained ANN for multiple demonstrators is a justifiable strategy. Although local climatic differences exist among sites, the main meteorological variables (GHI, cloud cover and temperature) share structural patterns that can be captured by a jointly trained architecture. This approach reduces system complexity, facilitates maintenance and enhances robustness in the presence of data variability. Moreover, it mitigates overfitting to site-specific conditions, promoting a more general and flexible model capable of adapting to different climatic contexts.

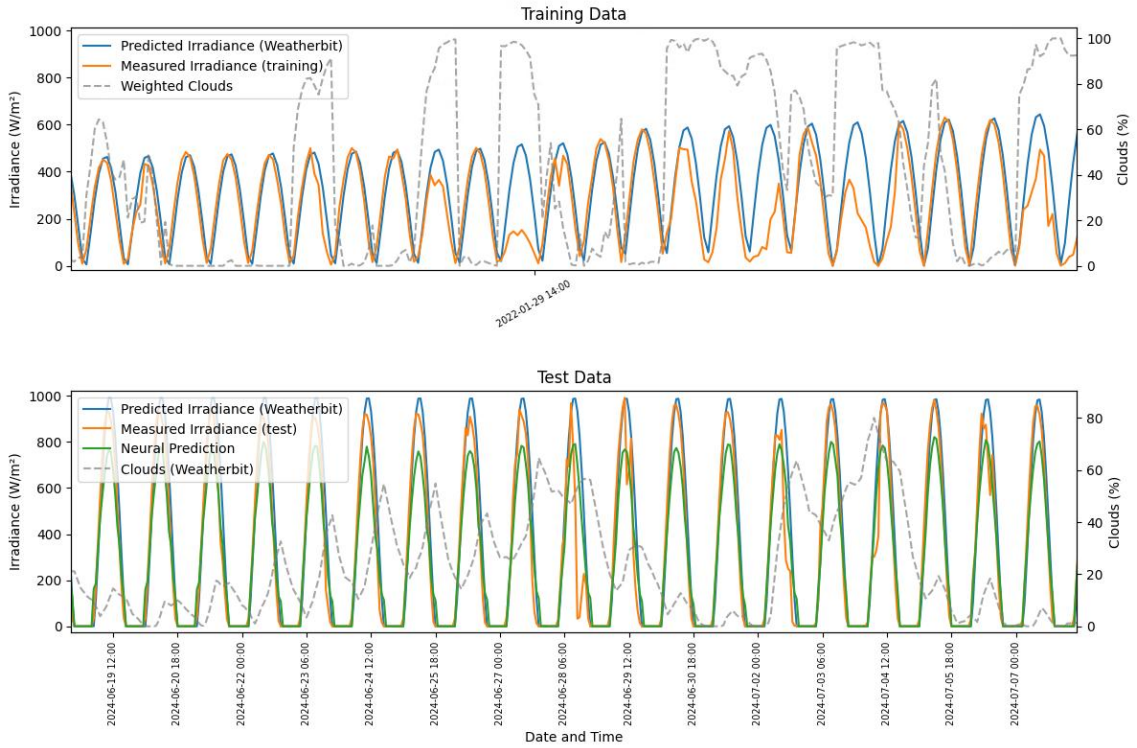


Figure 3: Performance of the ANN irradiance model: training and test results compared with meteorological forecasts and real measurements

The ANN designed for irradiance prediction shows a notably positive performance compared to raw meteorological forecasts. The analysed graphs demonstrate how the network's estimates closely follow the measured irradiance values, particularly on clear or fully overcast days, where the model effectively captures the trends (Figure 3). However, under variable cloud cover or intermittent shading, larger discrepancies appear. These deviations are mainly due to the complexity of atmospheric phenomena such as clouds, whose local and temporal variability is challenging to capture using conventional forecasting sources.

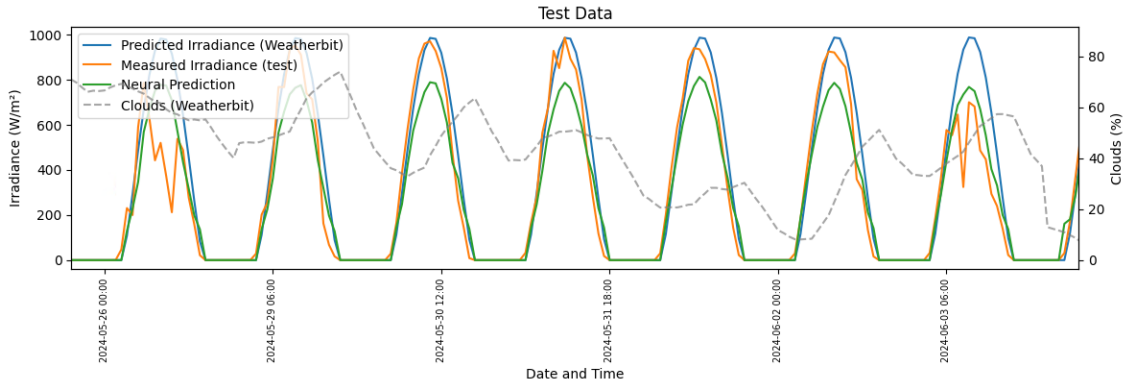


Figure 4: Detailed view of the ANN irradiance model performance on test data

Clouds, due to their highly dynamic and localized behaviour, represent one of the main challenges in irradiance prediction. Although services such as Weatherbit provide hourly forecasts with key variables like global irradiance, temperature and cloud cover, their spatial and temporal resolution is limited. In real-world environments, small cloud formations not detected by satellites or rapid

changes in cloud cover can lead to sudden drops in irradiance not reflected in the model's input data. Consequently, even a well-trained ANN model, being dependent on these inputs, may smooth out or fail to anticipate such events (see Figure 4).

Nevertheless, the use of ANNs, and specifically MLP models trained with meteorological variables such as horizontal irradiance, weighted cloud cover, temperature, and wind, provides a significant improvement over relying solely on raw forecasts. This approach reduces the mean absolute error (MAE, MAPE; see Table 3) and improves the coefficient of determination (R^2; see Table 3), demonstrating its ability to adapt to nonlinear contexts and capture hidden structures among variables. The model is particularly effective in mitigating systematic biases present in meteorological forecasts, enabling a more accurate adjustment of the hourly irradiance profile.

Table 3: Performance metrics of the irradiance ANN

MAE	R^2	MAPE
53.2 W	0.912	12.5 %

The results reflect a satisfactory performance for an irradiance prediction model in real-world contexts, especially considering the inherent limitations of the meteorological input data. This error level is reasonable within the domain of hourly solar forecasting and confirms that the chosen architecture; i.e., an MLP ANN, has been an appropriate choice. The model achieves a good balance between complexity and generalization capacity, capturing the nonlinear relationships between meteorological variables without overfitting the data. Techniques such as input normalization and dropout regularization further contributed to achieving high accuracy while maintaining stability on unseen data.

3.3.1. Thessaloniki

The ANN trained to estimate the electrical power output of PV panels shows behaviour consistent with recent literature on energy forecasting using AI. As highlighted in previous studies, MLP networks have proven particularly effective when well-selected input variables, such as irradiance and temperature, are available together with a sufficient volume of data to capture the system's dynamics. The graph (Figure 5) indicates that the model not only accurately reproduces the daily shape of the production curve, but also maintains strong stability in its predictions across different days of the test period.

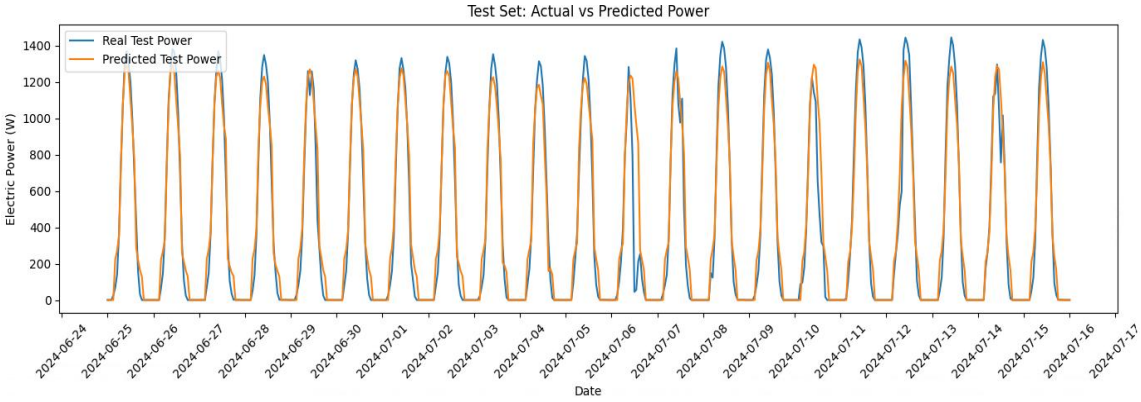


Figure 5: Performance of the ANN electric power model in Thessaloniki: comparison of real and predicted data

The quantitative results (see Table 4) are in line with, and even exceed, the reference values reported in numerous studies using similar models for PV energy forecasting. As noted in the state

of the art, one of the main challenges in these models is achieving a balance between accuracy and generalization capability, which has been accomplished here thanks to an architecture that is sufficiently deep without being overly complex. Moreover, since the network was trained individually for each demonstrator, it was able to capture the specific characteristics of each system in greater detail, which explains its strong performance even under less stable weather conditions.

Table 4: Performance metrics of the electric power ANN in Thessaloniki

MAE	R^2	MAPE
65.9 W	0.94	11.7 %

The ANN trained to predict the thermal power generated by the hybrid system delivers results that, although somewhat more modest than in the electrical case due to the variability, remain consistent with findings in the state of the art on solar energy forecasting. As noted in the literature, estimating thermal production is more complex due to the high sensitivity of this type of generation to cloudiness, which constitute one of the greatest challenges to predict locally as previously mentioned, as well as to other aspects such as variations in ambient temperature, thermal fluid flow rate, and heat losses; factors that are not always monitored. In this regard, the graph (Figure 6) shows that the network is able to follow the general trend and correctly anticipate the start and end of production periods, but it tends to underestimate thermal generation peaks, particularly on days with higher actual performance.

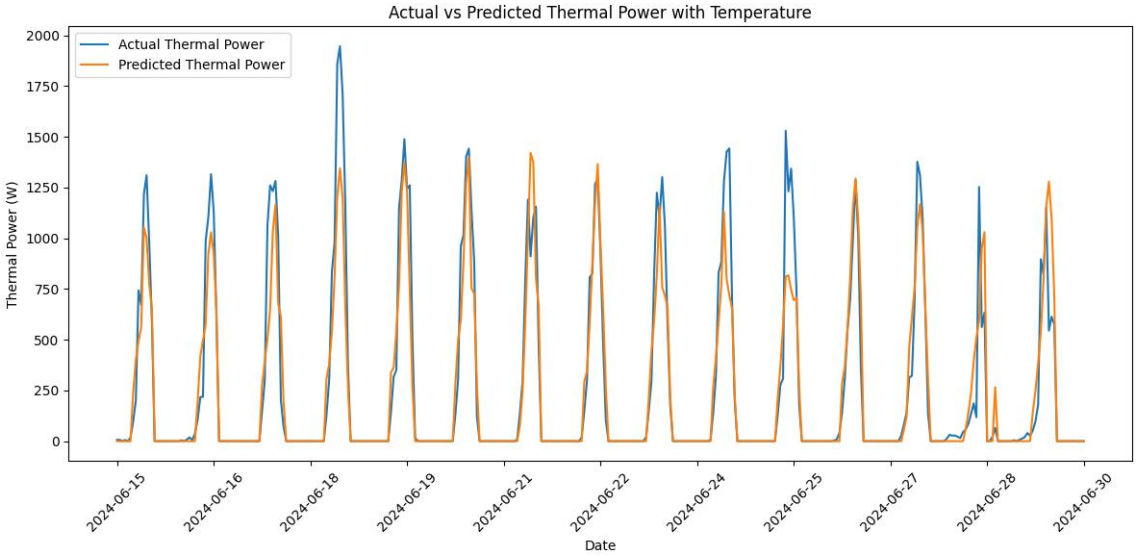


Figure 6: Performance of the ANN thermal power model in Thessaloniki: comparison of real and predicted data

The performance indicators (Table 5) support this interpretation. Although the accuracy is lower than in the cases of irradiance or electrical power, the results are reasonable and align with other studies that use similar methods to predict thermal variables. As several authors have pointed out, one of the key aspects of such models lies in the selection of input variables. In this case, the decision to limit inputs to irradiance and temperature, while appropriate from a simplification standpoint, may restrict the model's ability to capture the full thermal dynamics of the system. Even so, the chosen architecture (an individual network per demonstrator with moderately sized hidden layers and regularization) makes it possible to capture useful patterns and provide valid predictions for most days. These results support the use of MLP networks for thermal estimation, provided it is understood that their performance will largely depend on the richness and quality of the available data.

Table 5: Performance metrics of the thermal power ANN in Thessaloniki

MAE	R^2	MAPE
108.4 W	0.83	16.8 %

3.3.2. Sopron

In the case of the demonstrator located in Sopron, no ANN was developed or validated for estimating electrical power generation, as the system operates through a proprietary inverter (pre-existing, prior to the installation of MiniStor system) that does not allow direct or programmatic access to production data. This situation is not uncommon in commercial installations, where devices often incorporate closed protocols or non-standardized interfaces, limiting the ability to automatically extract data for ML-based prediction models.

In this context, and given that the objective of this work is to develop general approaches transferable across demonstrators, the training and evaluation of electrical power models were restricted to sites where direct and reliable measurements of this variable were available. This decision ensures methodological consistency in cross-validation and avoids indirect estimations that could compromise the comparability of results.

As highlighted in the state of the art, predicting thermal power is more complex than predicting electrical power, as it does not depend solely on irradiance. As previously mentioned, it is also highly influenced by factors such as cloudiness (which in turn affects the irradiance prediction), ambient temperature, the effective use of heat, the dynamics of the thermal fluid and thermal losses, many of which are not totally accurate or unavailable as input variables in the models. The graph (Figure 7) confirms this difficulty: the model successfully identifies the system activation periods and the overall trend on sunny days but clearly underestimates peak values and slightly overestimates output on days with low actual production. This behaviour suggests a smoothing effect typical of MLP networks when dealing with noisy or weakly representative patterns.

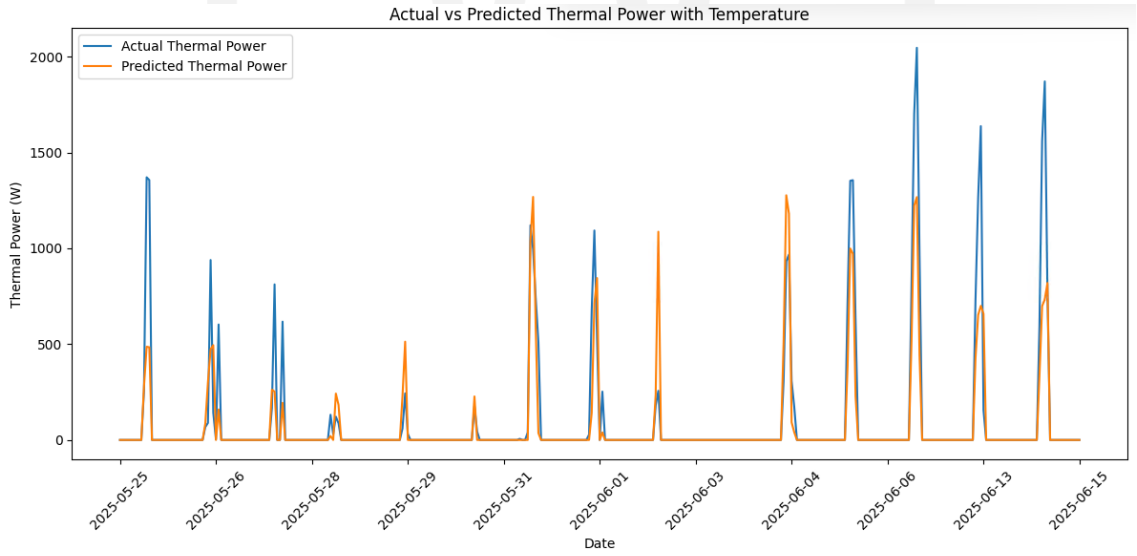


Figure 7: Performance of the ANN thermal power model in Sopron: comparison of real and predicted data

The ANN trained for thermal power prediction at the Sopron demonstrator shows significantly more limited performance than that observed in Thessaloniki (see Table 6), which aligns with the literature on the sensitivity of ML models to local conditions. In this case, the R^2 value drops to 0.58, meaning the model explains only about 58% of the variance in the real data, despite maintaining a moderate MAE of 85.1 W and a MAPE of 13.5%. This behaviour is expected, considering that Sopron, located

in Central Europe, experiences greater meteorological variability, lower irradiance levels, and more unstable temperatures compared to the Mediterranean region of Thessaloniki.

Table 6: Performance metrics of the thermal power ANN in Sopron

MAE	R^2	MAPE
85.1 W	0.58	13.5 %

3.3.3. Kimmeria

At the Kimmeria demonstrator, it was not possible to develop prediction models for either electrical or thermal power from the solar circuit due to the specific architecture of the installed energy system. In this case, electrical production is linked to a solar thermal hybrid installation owned by the local partner, which operates under restrictions and data access limitations that prevent the availability of stable and complete time series on actual generation.

Moreover, the system's operation is governed by an independent control logic in which heat support does not directly respond to irradiance. Instead, it is routinely activated during daytime hours, regardless of whether solar production is actually required. This condition introduces a discontinuity between meteorological variables and the observable thermal response, making it unfeasible to apply supervised ML models such as those used in other demonstrators.

Given these circumstances, and in order to maintain methodological consistency and rigor, Kimmeria was excluded from the training and validation of ANNs for both electrical and thermal power prediction, since the system conditions do not allow for establishing a reliable causal relationship between meteorological inputs and energy response.

3.3.4. Santiago de Compostela

The ANN trained to predict the electrical power output of the PV system at the Santiago de Compostela demonstrator shows performance that, while generally adequate, reflects some of the typical limitations encountered when modelling environments with irregular irradiance and Atlantic climates. With a MAE of 227.7 W, an R^2 of 0.81, and a relative error of 19.8% (see Table 7), the model maintains a reasonable ability to reproduce the daily production profile, although with greater deviations than in other locations such as Thessaloniki. The graph (Figure 8) illustrates how the model correctly tracks the overall shape of the generation cycles but tends to underestimate peaks on sunny days and to smooth the output under partially cloudy conditions.

This type of deviation is common in regions where variable cloud cover and intermittent shading exert a nonlinear influence on electrical production. As discussed in the state of the art, MLP ANNs can deliver strong results when input conditions are well represented in the training dataset. However, in climates like Galicia region (where Santiago de Compostela is located), with high intraday variability and short-duration meteorological events, the PV system's response to irradiance becomes more difficult to capture, particularly in the absence of complementary measured variables such as DNI or high-resolution cloud cover data.

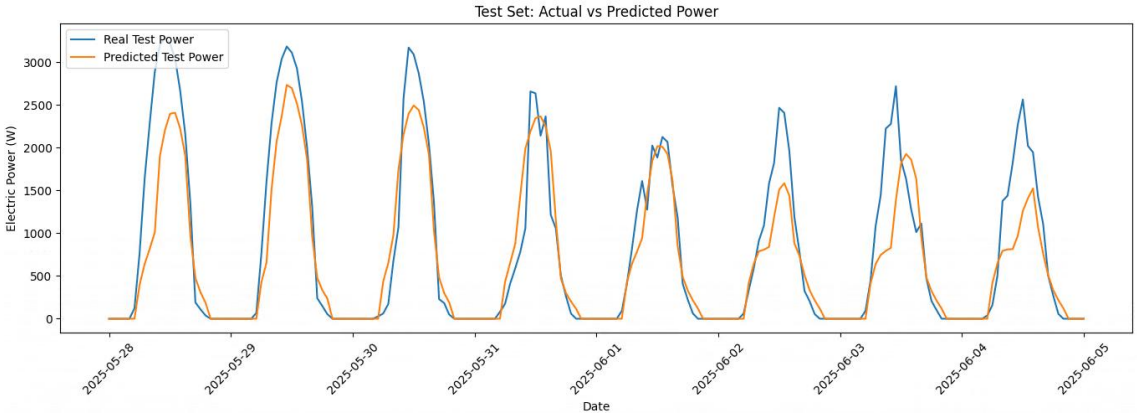


Figure 8: Performance of the ANN electric power model in Santiago de Compostela: comparison of real and predicted data

Compared to other demonstrators, the results in Santiago highlight a challenge that is less pronounced in Mediterranean or continental contexts with more stable solar conditions. While in Thessaloniki errors below 12% and R^2 values above 0.9 were observed, the model in Santiago exhibits more modest performance, though still within acceptable thresholds for practical energy applications. This difference does not necessarily indicate modelling flaws, but rather the physical complexity of the training environment, an issue that aligns with conclusions from recent studies emphasizing the need to tailor predictive approaches to the specific characteristics of each site.

Table 7: Performance metrics of the electric power ANN in Santiago de Compostela

MAE	R^2	MAPE
227.7 W	0.81	19.8 %

The ANN trained for thermal power prediction at this demonstrator shows significantly more limited performance compared to other cases (see Table 8), with a very low R^2 value (0.18), indicating that the network barely manages to explain the variance observed in the actual data. Although the MAE (106.1 W) and relative error (15.5%) are not excessively high in absolute terms, the combined analysis of the temporal graph and 3D projections reveals a highly disrupted behavioural pattern, which hinders effective model learning.

Table 8: Performance metrics of the thermal power ANN in Santiago de Compostela

MAE	R^2	MAPE
106.1 W	0.18	15.5 %

As shown in the left 3D scatter plot (Figure 9), a large proportion of the thermal power (PTERM) values cluster at zero, reflecting a low rate of active production throughout the analysis period[7]. Additionally, high dispersion is observed in the instances where thermal power is actually generated, resulting in a pattern difficult to model with an MLP network. These networks rely on identifying regularities between input variables (in this case, GHI and temperature) and the thermal response, which becomes challenging when such regularities are sparse or inconsistent.

[7] In Figure 9, unlike the left 3D plot which includes the complete dataset, the right 3D plot excludes zero values in order to provide a clearer visualization of non-zero production patterns.

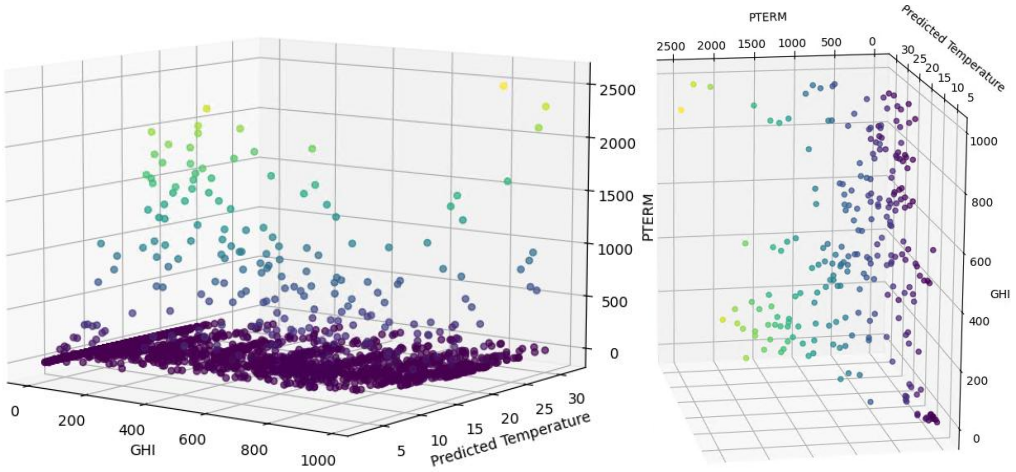


Figure 9: Multivariable dispersion plots of thermal generation data in Santiago de Compostela: GHI, predicted temperature, and thermal power (PTERM)

This situation is consistent with the state of the art regarding the challenges of modelling thermal production when auxiliary devices such as heat pumps are involved. In this demonstrator, the complementary use of a heat pump within the solar circuit introduces a significant alteration to the expected behaviour: the measured thermal output does not depend solely on meteorological conditions, but also on an additional control logic that activates or inhibits ST generation based on internal energy management criteria or specific thermal demands.

From a modelling perspective, this introduces a source of unobservable noise into the training data, breaking the assumption of a direct causal relationship between meteorological inputs and thermal output. As a result, even though the model is properly trained on the available data, it lacks sufficient information to distinguish whether the thermal power originates from the solar system or from an external contribution, nor can it anticipate hidden operational decisions. This explains why the network tends to predict smoothed values and loses its ability to react to actual thermal peaks (Figure 10).

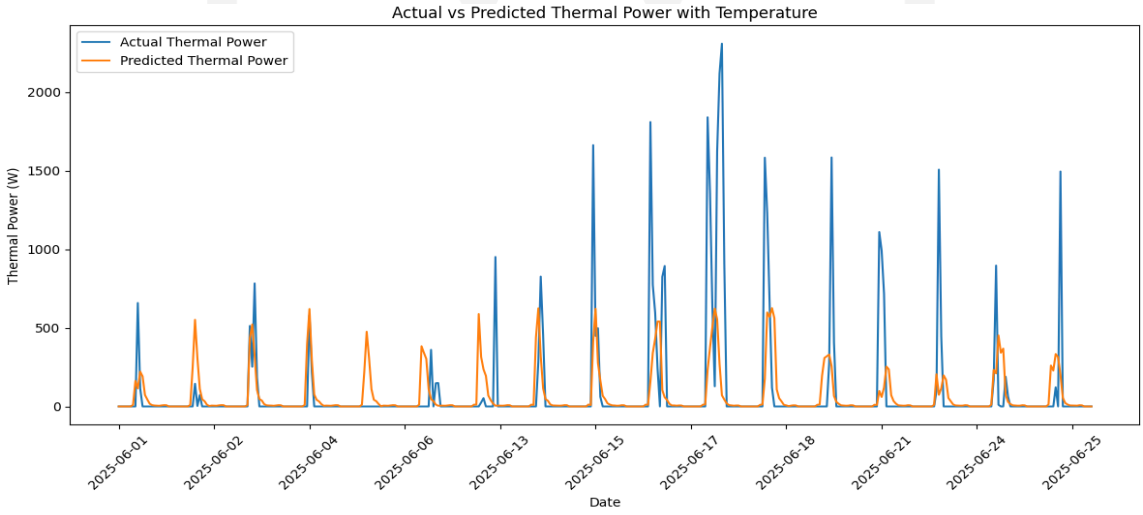


Figure 10: Performance of the ANN thermal power model in Santiago de Compostela: comparison of real and predicted data

Compared to other demonstrators, where thermal production more clearly followed the evolution of GHI, this case highlights a structural limitation of the purely supervised approach, which requires

functional coherence between inputs and outputs. Although the model retains some utility in estimating the general thermal envelope, the results indicate that in contexts like this, where multiple generation sources and complex activation logics coexist, it would be necessary to expand the set of input variables to include system state signals or adopt more advanced approaches that integrate control information.

3.3.5. Cork

The ANN trained to predict electrical power at the Cork demonstrator delivers moderately satisfactory results (see Table 9 and Figure 11), although they differ noticeably from those obtained at other European sites. With an R^2 of 0.6, a MAE of 35.9 W, and a relative error of 20.2%, the model is able to capture the general shape of daily production but shows a clear tendency to overestimate generation on partially cloudy days. This difference in model fit aligns with challenges highlighted in the state of the art regarding energy system modelling in oceanic climates, where high humidity, variable cloud cover, and the predominance of diffuse irradiance create more complex forecasting conditions than in other environments.

Table 9: Performance metrics of the electric power ANN in Cork

MAE	R^2	MAPE
35.9 W	0.6	20.2 %

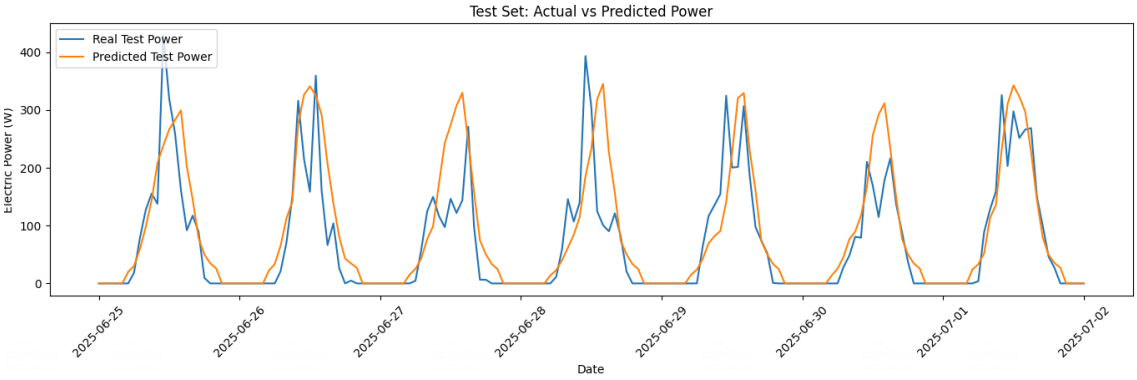


Figure 11: Performance of the ANN electric power model in Cork: comparison of real and predicted data

Unlike the models trained in locations such as Thessaloniki or Santiago, where the PV system's response followed more direct irradiance patterns and peak values were more consistent, the electrical prediction in Cork faces the challenge of discriminating between direct and diffuse irradiance. This issue has been described in the literature as a limitation of models based solely on GHI. As noted in (Kazem, Yousif, Chaichan, & Al-Waeli, 2019), neural models may lose accuracy in complex climates if specific atmospheric regime variables such as the diffuse fraction or spectral composition are not included, factors that were unavailable in this case.

Additionally, this demonstrator features the lowest scale of electrical production among all those analysed, with peak values barely exceeding 400 W. In such scenarios, even small absolute deviations can result in high relative errors, since minor discrepancies represent a significant portion of the total output. This effect is also documented in the literature, which highlights that metrics such as MAPE can become less representative in low-power systems and should therefore be interpreted in light of the site's energy profile; see (Izgi, Yerli, Kaymak, & Şahin, 2012) and (Das, et al., 2018).

Overall, the results in Cork for the electrical power forecasting show that the model retains operational usefulness and remains capable of predicting the general production envelope, but they clearly highlight the structural limitations of using MLP networks in regions with high atmospheric variability and without auxiliary variables. This case emphasizes the generalization capacity of the model, being functional even in adverse climatic environments, but also underlines, as many recent studies point out, that improving performance in such contexts requires more refined input selection and more intensive local tuning when richer datasets are available. Nevertheless, the validation in Cork was successfully completed with an operational model, contributing to the methodological consistency across all demonstrators.

The predictive model for thermal power in the Cork demonstrator did not achieve the expected performance levels, presumably due to the limited availability of data with the necessary granularity and structure to adequately capture the complex dynamics of thermal production. The nonlinear and abrupt nature of certain thermal transitions in the system requires a more extensive and representative dataset to enable the model to identify and generalize the underlying patterns with greater accuracy.

3.4. Data challenges and lessons learned

During the development of energy forecasting models based on ANNs for hybrid solar systems, several limitations were identified in data availability and quality, many of which align with challenges widely reported in the scientific literature on energy systems. This section provides a critical synthesis of the main issues encountered, enriched with contextual lessons derived from the demonstrators.

First, one of the most frequently cited problems in recent studies is the absence of key meteorological variables, particularly the separation between DNI and DHI. Many advanced PV forecasting models require this distinction to refine the calculation of useful incident energy, especially in climates with a high proportion of diffuse radiation such as Atlantic or oceanic regions. As noted in (Shi, Lee, Liu, Yang, & Wang, 2011) and (Kazem, Yousif, Chaichan, & Al-Waeli, 2019), the lack of separated components reduces model sensitivity to rapid changes in cloud cover—a phenomenon particularly difficult to capture with dense MLP networks when this information is missing.

Second, the incomplete or temporally limited recording of historical time series emerged as a major barrier to training robust models. Some demonstrators provided only a few weeks or months of data, making it impossible to capture seasonal patterns or train models that generalize beyond the training range. According to (Deb, Eang, Yang, & Santamouris, 2016) and (Leva, Dolara, Grimaccia, Mussetta, & Ogliari, 2017), the lack of sufficient historical coverage directly affects the predictive capability of ANNs, especially for thermal power forecasting, which is more sensitive to small variations in temperature, fluid flow, or internal thermal losses that are not always monitored. This constraint was evident in Cork's thermal case, where limited granularity prevented the ANN from learning the abrupt nonlinear dynamics of the system.

Third, local climatic variability strongly shaped the effectiveness of the models. In environments with high intraday atmospheric variability, such as Santiago de Compostela or Sopron (i.e., Northern Spain or Central Europe), ANNs tended to excessively smooth the response, underestimating production peaks and overestimating low-output periods. This behaviour, documented in (Sobri, Koohi-Kamali, & Rahim, 2018) , (Das, et al., 2018) and (Ahmadi, 2020), results from the underrepresentation of extreme events in the training data. The literature emphasizes that in such contexts, models should be complemented with auxiliary variables (cloud fraction, visibility, relative humidity) or with hybrid physical-statistical approaches that incorporate system-level physics.

Fourth, operational and architectural constraints played a key role. In Sopron, the impossibility to access electrical data (PV generation) due to proprietary pre-existing inverters constituted a challenge. In Kimmeria and Santiago, the influence of auxiliary devices (e.g., heat pumps, independent control logics) introduced unobservable noise into the thermal signals, invalidating the assumption of a direct causal link between meteorology and thermal output. These cases demonstrate that data challenges are not only meteorological but also systemic: reliable forecasting requires access not just to environmental variables but also to operational metadata (pump states, inverter signals, control logic triggers). Without this, purely supervised approaches face intrinsic structural limitations.

Regarding the meteorological services used, such as Weatherbit, the available hourly forecasts proved sufficient to feed short-term ANNs, but their temporal and spatial resolution was insufficient to capture rapid variations or localized microclimatic effects, such as clouds. As underlined in (Kazem, Yousif, Chaichan, & Al-Waeli, 2019) and (Barbu & Darie, 2019), this shortcoming remains one of the main sources of residual error in models based solely on numerical weather predictions. Recent research suggests that integrating satellite-derived irradiance data, sky cameras, or nowcasting techniques can mitigate this gap and should be considered in future developments.

Finally, the experience gained during this work underscores the importance of balancing local specialization and generalization across sites. The use of a common ANN for irradiance prediction,

complemented by site-specific power models, represents a hybrid approach that mirrors strategies reported in (Jovanovic, Sretenovic, & Zivkovic, 2015) and (Yang, A., & Xie, 2015). The global model captures structural patterns across locations, while the site-specific layers account for local idiosyncrasies. This dual-level approach proved essential for maintaining methodological consistency while ensuring adaptability to diverse climates and system architectures.

In summary, the challenges encountered were not only technical but also methodological and logistical, reflecting the inherent complexity of forecasting in hybrid PVT systems. The lessons learned can be summarized as follows:

- Forecast reliability depends on both meteorological resolution and system observability (including operational signals).
- Short datasets and missing variables limit model generalization and call for hybrid strategies combining data-driven and physics-based approaches.
- Climatic variability amplifies smoothing effects in ANNs, underscoring the need for auxiliary atmospheric descriptors.
- General models can be effective for shared variables such as irradiance, but thermal forecasting requires richer and more localized datasets.

These findings converge with the literature: effective forecasting in real-world hybrid solar systems requires the co-evolution of models and data. Progress depends not only on refining ANN architectures but also on improving data availability, quality, and contextual richness to bridge the gap between theoretical potential and operational deployment of AI in renewable energy systems.

4. Demand energy forecasting

4.1. Data handling

This section outlines the high-view methodology for the data management to develop demand forecasting models within the Ministor project. The entire process can be broken down into several key stages, from initial data handling to the final evaluation of the forecasts.

4.1.1. Data description

4.1.1.1. Ministor pilots - data streams

Table 10 provides a summary of the data streams from the Ministor pilots that are relevant to demand forecasting. For example, regarding the Thessaloniki pilot site, the primary data sources are the Fan Coil Meters, which provide detailed operational data.

Table 10: Summarized view of data streams from Ministor pilots regarding demand forecasting

Pilot site	Source/ Device	Main monitored parameters	Units	Data resolution
Thessaloniki	Fan Coil Meter 1,2	operationMode, swing, accumulatedPower, setTemp, userControl, fanSpeed, tempAct, status,	Mode: Heating, Cooling, Auto; Power: kW; Temperature: °C; FanSpeed: 1-3; Status: ON/OFF	15 min (Resampled to 1h)
Sopron	-	-	-	-
Kimmeria	-	-	-	-
Santiago de Compostela	Energy, Thermal Meter (Apartment)	thermalEnergy, KWh_Tot (electricEnergy)	KWh	Varying resolution (1-5mins) Resampled to 1h
Cork	-	-	-	-

Regarding weather data, both historical weather data that are used for the training of the forecasting algorithms and weather forecasts, used for the actual prediction, are being fetched by local available weather station (e.g. Table 11) or from an external API (e.g. Visual Crossing, Open-Meteo, as analysed in section 2.8)

Table 11: Data stream of local weather station (Thessaloniki)

Source	Device name	Monitored Parameters	Units	Data Resolution
Weather Station	CERTH SmartHome Weather Station	gustWindSpeed, averageWindSpeed, airTemperature, solarRadiation, airHumidity, windDirection, timestamp	Wind Speed: m/s; Temperature: °C; Solar Radiation: W/m^2; Humidity: %; Wind Direction: deg	15 min

4.1.1.2. Forecasting targets

The primary goal of the demand forecasting is to predict the electrical power consumption of the assets for heating and cooling purposes (Table 12).

Table 12: Summarized view of demand energy forecasting targets

Pilot site	Device name	Target category	Forecast purpose	Forecast horizon
Thessaloniki	Fan Coil Meter (1,2)	Elec. power consumption	Heating/cooling demand forecasting	24h
Kimmeria	-	-	-	-
Sopron	-	-	-	-
Santiago de Compostela	Energy, Thermal Meter (Apartment)	Electric & thermal energy consumption	Energy & heating demand forecasting	24h
Cork	-	-	-	-

4.1.2. Data preparation

4.1.2.1. Pre-processing

Before any procedure related to training and prediction, consistency of the data has to be assured and therefore some pre-processing steps are necessary.

- Data alignment: all sources are time-aligned to a common UTC-based index.
- Cleaning and outlier handling: invalid or extreme values (e.g., negative power, unrealistic spikes) are removed or capped using quantile-based rules. . Specifically, any value above the 99.9th percentile is capped at the 99.9th percentile value to mitigate the influence of rare, extreme consumption spikes.
- Missing data imputation: gaps in data are filled via linear interpolation, preserving temporal continuity.
- The time resolution of the data is not fixed but is dependent on the pilot site and the way energy data are measured for each of the pilot's assets. The resolution can be 15 minutes, 30 minutes or 1 hour.
- Seasonality data (i.e., month, day of the year, day of week and hour) are extracted from the said timestamps and are utilized as extra parameters. Consequently, the final dataset is a combination of energy data and weather and seasonality parameters, all merged under their unique timestamps per time interval.

Depending on the pilot site characteristics a resampling strategy was implemented. As a typical example timeseries are resampled to a common interval (e.g., 1 hour), applying variable-specific aggregation functions:

- Cumulative variables: summation over the interval.
- State/status variables (e.g., device status): last observed value.
- Continuous variables: median value to reduce sensitivity to short-term fluctuations.

4.1.2.2. Feature engineering

Sliding Window: historical lags and multi-step-ahead target features are generated to support direct forecasting. This typically involves a 24-, 48- or 96-step ahead shifting of the data, according to the respective time resolution of the asset, so that every row of the dataset is enhanced with the information of the previous day. When necessary, the normal two-dimensional Sliding Window is replaced by a 3D version of it, for example in the case of an LSTM Recurrent ANN-based model.

Moreover, **seasonality data** are extracted from the said timestamps and are utilized as extra parameters. Cyclic time variables (hour, day, month, etc.) are encoded using sine/cosine functions. **Cleaning & Outlier Handling:** Invalid or extreme values (e.g., negative power, unrealistic spikes) are removed or capped using quantile-based rules.

Weather parameters: if available, weather-related variables (e.g., temperature, wind speed, solar radiation) are integrated into the feature set. These can be sourced from local weather stations or external APIs to enrich the forecasting models with relevant exogenous factors.

Consequently, the final dataset is a combination of energy data and weather and seasonality parameters, all merged under their unique timestamps per time interval.

4.2. Training strategy and architecture

4.2.1. Forecasting design

The methodology used for PV forecasting was based on the utilization ML and DL algorithms (Figure 12). This strategy addresses the need to capture the nonlinear relationships between meteorological variables and power/energy demands, leveraging the ability of such models to model complex functions without explicitly formulating a physical model of the system.

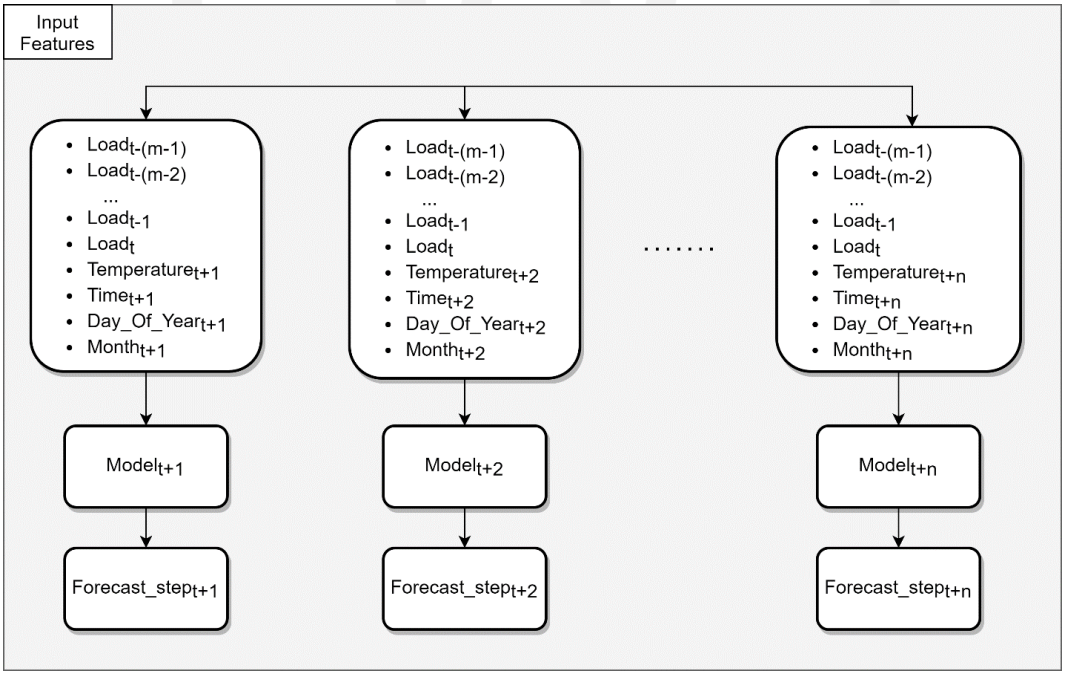


Figure 12: Day ahead load forecast (direct strategy) flowchart

4.2.2. Algorithms and methods

1 Tree-based regressors and ensemble methods

Among the selected algorithms and methods, ensemble methods, particularly tree-based regressors and their combinations, were a primary choice to experiment with. Ensemble learning is a favourable choice because such solutions generally demonstrate better performance and enhanced generalization capabilities compared to individual models. These methods combine multiple individual models to produce a more robust and accurate prediction. Specific algorithms that have been explored and evaluated for demand prediction, as indicated by the analysis of the Thessaloniki pilot site, include:

- Voting Regressor (VotRegr)
- Stacked Regressor (StackRegr)
- Zero inflation voting regressor (ZIVR)
- ETs
- Light Gradient Boosting Machine (LGBM)
- CatBoost (CB)
- XGBoost (XGB)
- Hist Gradient Boosting Regressor (HGBR)
- DT regressor

2 ANNs

- RNN (LSTM): for DL, LSTM is an artificial RNN architecture (Ansari, Stella, Turkmen, Zhang, & Mercado, 2024) Standard feed-forward ANNs lack feedback connections, whereas LSTM does. Both single data points (like pictures) and complete data sequences (like time series, speech, or video) can be processed using it. According to a number of studies, LSTMs are capable of handling sequences of different lengths (huge timeseries, text, etc.) and are particularly good at modelling intricate temporal patterns and long-term dependencies (Siddharth, Ramakrishna, Geetha, & Sivasubramaniam, 2011). They can also be quite expandable and flexible with sophisticated designs.
    - Sequence to sequence (seq2seq) LSTM
- Transformer-based models: the ability of transformer-based models to capture long-range dependencies through self-attention mechanisms has led to a significant increase in interest in the field of time series forecasting (Banihashemi, Ding, & Wang, 2016). Transformers have better scalability than RNN-based models like LSTMs since they handle incoming sequential information in parallel, especially for big datasets (Energy, 2019). In a variety of energy and traffic forecasting benchmarks, their adaptability has enabled many modifications in multivariate and multi-horizon forecasting tasks, demonstrating competitive or state-of-the-art results (El Khantach, Hamlich, & Belbounaguia, 2019)
    - seq2seq transformer
- Foundational models: generally pre-trained on large and diverse time series datasets to learn generalizable representations. This pre-training allows them to be adapted to specific downstream tasks, like forecasting or classification, often with minimal task-specific data (few-shot learning) or even no additional training (zero-shot learning). Typically, such pre-trained models are based on transformer architectures. This emerging field of AI aims to provide powerful, "plug and play" solutions that can be easily deployed without requiring extensive computational resources for training.

- Chronos: a time series forecasting model that tokenizes data through scaling and quantization. It utilizes T5-family encoder-decoder models to produce probabilistic forecasts by sampling from a distribution of token bins. Exogenous variables can be incorporated by training a separate regressor and then having Chronos forecast the residuals.

4.2.3. Implementation per pilot site

4.2.3.1. Thessaloniki

For the Thessaloniki pilot, three datasets were utilised:

- Two identical fan–coil accumulated power readings
- Ambient weather forecast from the ITI smart home weather station

Next, each fan–coil power series is joined on its timestamp to the full suite of meteorological predictors (air temperature, relative humidity, wind speed). This produces two harmonized tables capturing the interplay between equipment load and ambient conditions. These fan–coil–weather tables are concatenated into a single dataset. This augmentation strategy increases the diversity of operating scenarios and supports the development of a more robust, generalizable forecasting model. A single model trained on two Fan Coil power consumption datasets. The multi-step forecasting strategy was employed to train one model per horizon step (1 to 24 h ahead) using a sliding-window approach.

Post-processing, operation-mode propagation and stream merging

- Non-negativity enforcement: Any forecasted power-difference below 1 Watt was truncated to zero to ensure physical plausibility and filter out numerical noise.
- Operation-mode propagation: The most recent observed mode (e.g., heating, cooling, standby) is carried forward and applied uniformly across all forecasted timesteps to maintain coherent system status.
- Combined forecast: Forecasts for Fan Coil 1 and Fan Coil 2 are generated independently, aligned by timestamp and summed to produce the final combined power forecast

4.2.3.2. Sopron

Boilers 1F, attic 1 and attic 2 aggregated forecasting implementation:

The Sopron pilot site's Boilers 1F, attic 1 and attic 2 aggregated energy usage was modelled through a 96-step ahead multi-output forecasting task using a range of ML and DL models. Sensor inputs included total consumption, weather measurements and derived time features. All models shared a consistent sliding window setup, with 96-step input and output horizons (equivalent to 24 hours).

The dataset comprised the following:

- Target variable: KWh_Tot_diff, representing the incremental energy consumption of the Boilers system.
- Extra features: weather variables (temperature, humidity, solar radiation, wind direction and speed), internal environmental conditions and temporal features such as hour, weekday, day of year and month.

The pre-processing pipeline included:

- Timestamp alignment and sorting

- Drop of irrelevant energy variables
- Normalization of all features and target using StandardScaler
- Feature engineering with time features (hour, day, month, weekday)
- Sequence creation via a sliding window

A comparison of four multi output forecasting models was conducted:

- ET regressor
- DT regressor
- LSTM network
- Transformer Encoder-Decoder (TF)

Regarding the training setup:

- An 80%/20% train/test split was used.
- The ETs model used 80 estimators and trained on multivariate inputs.
- For neural models, 30 epochs for the LSTM model and 50 epochs were used with the Adam optimizer and MSE loss.

For the evaluation, the metrics used for each of the 96 steps included NRMSE and R^2, with the performance was assessed post inverse transformation to the original scale.

Lab forecasting implementation:

The Sopron pilot site's Lab forecasting framework was developed using a 96-step ahead multi-output modeling approach to predict energy consumption based on environmental and internal variables. The target variable Lab KWh_Tot_diff reflects differential usage over time, requiring robust models to capture subtle changes.

Preprocessing steps included:

- Sorting and aligning timestamps
- Removal of non-relevant consumption fields (e.g., boilers, HVAC)
- Normalization of features and target using StandardScaler
- Generation of month, hour, weekday and day-of-year features
- Sequence windowing (96 input → 96 output horizon)

Modelling comparisons included:

- DT regressor
- ET regressor
- LSTM network
- Transformer Encoder-Decoder (TF)

The training pipeline includes:

- Lag features (e.g., Boilers_KWh_Tot_diff_lag_1, Boilers_KWh_Tot_diff_lag_2, Boilers_KWh_Tot_diff_lag_3) to capture past consumption trends.
- Sliding window approach for sequence generation: 96-timestep input → 96-timestep output.
- StandardScaler normalization for both inputs and targets.
- Train/test split with 85% training data.
- Comparison by training a single multi output models like ET regressor, DT regressor, sequence to sequence LSTM and seq2seq Transformer to output the full 96-step vector

HVAC forecasting implementation:

For the Sopron pilot site and HVAC (Heating, Ventilation and Air Conditioning) energy consumption forecasting, a 96-step ahead multivariate forecasting pipeline was implemented using a multi-output model comparison strategy. The models evaluated included:

- DT regressor
  ET regressor
- LSTM network
- Transformer-based Encoder-Decoder (TF)

The dataset comprised the following:

- Target variable: hvac_KWh_Tot_diff, representing the incremental energy consumption of the HVAC system.
- Extra features: Weather variables (temperature, humidity, solar radiation, wind direction and speed), internal environmental conditions and temporal features such as hour, weekday, day of year and month.

For the preprocessing pipeline, the following steps were conducted:

- All unnecessary energy columns were excluded to focus on incremental load.
- Applied Z-score normalization with StandardScaler for both inputs and target.
- Created input/output sequences using a sliding window approach with a 96-step input and 96-step output horizon.
- Inputs were flattened for tree-based models and maintained in 3D format for DL models.

Regarding the training setup:

- An 85%/15% train/test split was used.
- The ET model used 100 estimators and trained on multivariate inputs.
- For neural models, 50 epochs were used with the Adam optimizer and MSE loss.

The final selected model for deployment was the Transformer, demonstrating robustness and better generalization for longer horizons.

4.2.3.3. Kimmeria

For the Kimmeria pilot site, a demand forecasting model was implemented using a multi-step ahead forecasting strategy with a single multi output model comparison. The approach leverages a sequence-to-vector direct forecasting technique to predict 96 future steps of energy consumption.

The dataset includes:

- Energy consumption data (Total_Energy_kWh), processed to represent energy differences (kwh_diff) over time.
- Internal sensor data (e.g., room temperature, humidity, wind speed) from various building zones.
- Time-based features such as month, hour, weekday, day of the year, week number, quarter and year.

Weather and internal sensor data were interpolated and aligned, while outliers were handled through percentile capping (e.g. 99%). Data was resampled to a consistent 1h resolution.

The training pipeline includes:

- Lag features (e.g., kwh_diff_lag_1, kwh_diff_lag_2, kwh_diff_lag_3) to capture past consumption trends.
- Sliding window approach for sequence generation: 96-timestep input → 96-timestep output.
- MinMaxScaler normalization for both inputs and targets.
- Train/test split with 85% training data.
- Comparison by training a single multi output models like ET regressor, DT regressor, sequence to sequence (seq2seq) LSTM and seq2seq Transformer to output the full 96-step vector

Predictions are evaluated per horizon step and metrics such as R^2 and NRMSE are computed and stored.

4.2.3.4. Santiago de Compostela

For the Santiago de Compostela pilot site, the focus was on two targets:

1. Electric energy load (apartment -kWh)
2. Thermal energy load (apartment – kWh)

A unified data preprocessing methodology is applied to both electric and thermal energy data to prepare it for 24-hour ahead forecasting models. This ensures consistency and reliability for the project's energy management system. The core steps of the process are:

- Deriving Interval energy demand: Raw cumulative energy readings are converted into interval-based consumption by calculating the difference between consecutive measurements. This standardized approach is used for both thermal and electric loads.
- Hourly Resampling and Aggregation: The data is resampled into one-hour intervals. During this step, the consumption values for each interval are summed to produce a total hourly energy consumption figure. This provides the necessary granularity for the forecasting models.
- Feature Engineering: The feature set was enhanced by creating historical lags using a sliding window approach and by adding temporally encoded features (e.g., sine/cosine transformations for the hour of the day and day of the year) to better capture cyclical patterns.

4.2.3.5. Cork

For the Cork pilot site and “electric house” target (total electric energy usage), a 96-step ahead energy consumption forecasting pipeline was developed using a multi-output model comparison between DTs, LSTMs, Transformers ET regressor, using the last as final model trained on multivariate sensor and time features.

The dataset included:

- Target variable: electric_house, representing total electric energy usage.
- Environmental and internal variables: gas flow, solar radiation (UV, GHI), wind (speed and direction), relative humidity, pressure, various temperatures (e.g., dew point, wet bulb) and internal room measurements.
- Temporal features: month, hour, weekday, day of the year.

For the preprocessing part, the pipeline included the following steps:

- Removal of irrelevant fields (e.g., immersion energy).

- Alignment and interpolation of missing values.
- Z-score normalization using sklearn's StandardScaler for all inputs and target values.
- Sequence generation via a sliding window approach with 96-time steps as both input and prediction horizon.
- Flattening multivariate inputs into 1D arrays per sample (96 × N features) (or 3D for the DL architectures).

Regarding the modelling workflow, a Train/test split (85%/15%) ratio was used on the final model training using a 100-estimator ExtraTreesRegressor. As for the inference, a setup was conducted followed by inverse transformation to the original scale and step-wise evaluation of predictions over the 96 horizons using NRMSE and R^2 metrics (for all models)

For the Cork pilot site and gas flow target, a similar multi-step ahead forecasting framework comparison between multi output models was developed. Similar to the previous implementations, the data consisted of timestamped sensor readings, where the target variable gas_flow displayed highly skewed behaviour (Figure 13).

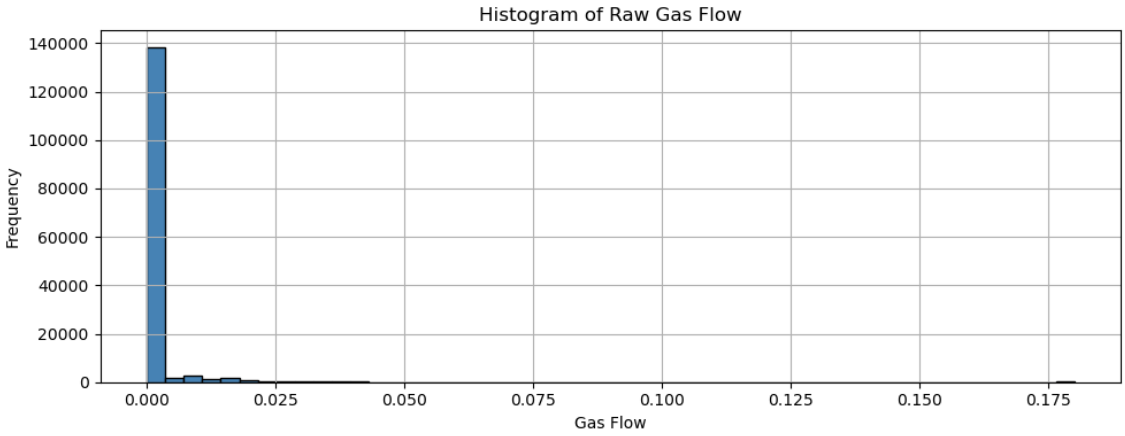


Figure 13: Original histogram of Cork Gas Flow

To stabilize variance and improve learning performance, a logarithmic transformation was applied (Figure 14), followed by normalization via StandardScaler.

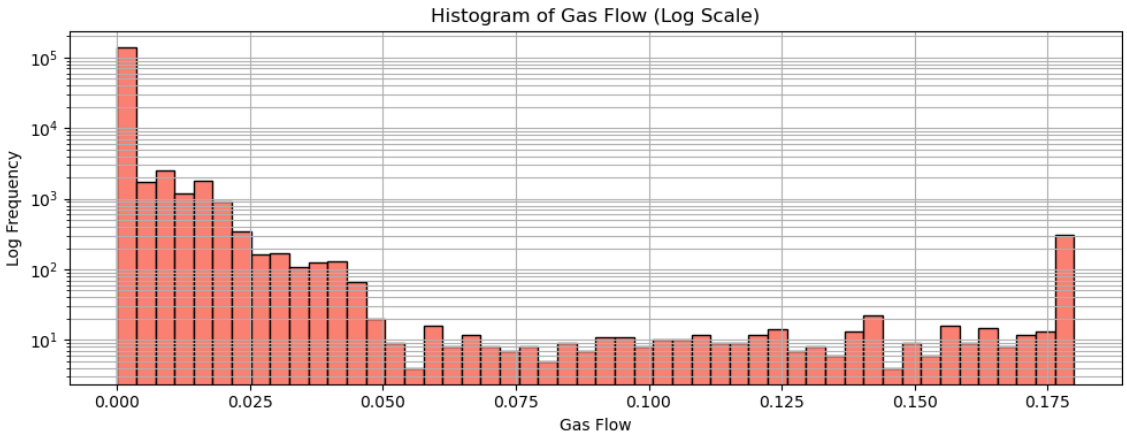


Figure 14: Histogram of Gas Flow with Log Scale

Besides the target variable the data also include

- Environmental and internal variables: gas flow, solar radiation (UV, GHI), wind (speed and direction), relative humidity, pressure, various temperatures (e.g., dew point, wet bulb) and internal room measurements.
- Temporal features: month, hour, weekday, day of the year,

while for the preprocessing part, the pipeline included:

- Removal of irrelevant fields (e.g., immersion energy).
- Alignment and interpolation of missing values.
- Z-score normalization using sklearn's StandardScaler for all inputs and target values.
- Sequence generation via a sliding window approach with 96-time steps as both input and prediction horizon.
- Flattening multivariate inputs into 1D arrays per sample (96 × N features) (or 3D for the DL architectures).

The training task was formulated using a direct multi-output strategy, predicting the next 96-time steps (i.e., 24 hours assuming 15-minute resolution) from a single past observation window. Overall, four models were implemented and evaluated:

- DT regressor
- ET regressor
- LSTM network
- Transformer-based Encoder-Decoder (TF)

With the TF model being used on the final pipeline.

4.3. Evaluation of demand prediction performance

Each pilot is examined and therefore evaluated separately. For the validation and the evaluation of the performance of the different models, the aforementioned error metrics are adopted.

4.3.1. Thessaloniki

Table 13 provides a comprehensive comparison of the evaluation metrics for the various forecasting models tested on the Thessaloniki pilot site.

Table 13: Thessaloniki pilot 1-step ahead (1h) forecast metrics (FanCoil1,2 combined power load)

Model	MAE	RMSE	R^2	NRMSE
VotRegr	0.407	0.883	0.822	0.042
StackRegr	0.485	0.99	0.786	0.0472
ZIVR	0.403	1.006	0.768	0.0479
ET	0.449	0.896	0.816	0.0427
LGBM	0.46	0.93	0.802	0.0443
CB	0.462	0.939	0.798	0.0447
XGB	0.47	0.947	0.789	0.0451
LSTM	0.48	0.949	0.795	0.0455
HGBR	0.481	0.955	0.798	0.0455

While an ANN (LSTM) was tested, the results indicate that tree-based models and ensemble methods, particularly the VotRegr provided superior performance with lower error rates. The evaluation results are depicted in Figure 15 , which provides a visual comparison of the day-ahead demand forecast against the actual measured power consumption for the Thessaloniki pilot site. This graph illustrates the performance of the selected model (VotRegr) over a 24-hour period.

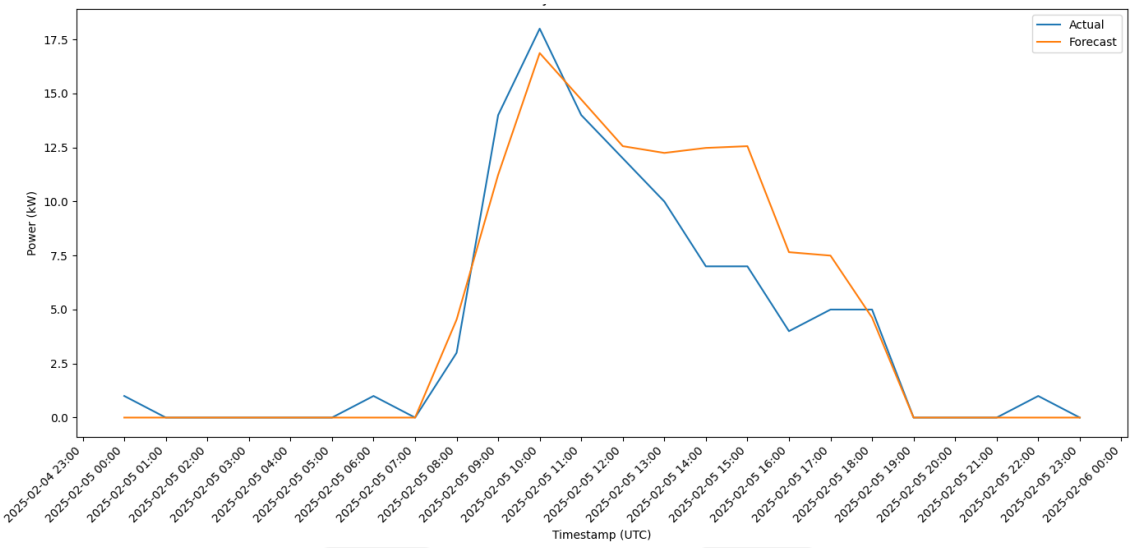


Figure 15: Day-ahead forecasting for Thessaloniki pilot (FanCoil1,2 combined power load)

4.3.2. Sopron

The Sopron pilot site focused on a 96-step ahead forecasting task targeting Boiler 1, Boiler 2 and HVAC energy usage, based on multivariate weather and sensor readings. A comprehensive evaluation of multi-output models was conducted to assess short- and long-horizon prediction capabilities. Figures below summarize performance across the full forecast window, using R^2 and NRMSE as primary evaluation metrics.

<u>For Boilers 1F, attic 1 and attic 2 aggregated implementation:</u>

Among the models, the LSTM achieved the most stable forecasting performance, consistently outperforming tree-based and transformer models. Its R^2 remained close to 0.55-0.40 (Figure 16) even at long horizons, while its NRMSE stabilized around 0.14–0.26 (Figure 17) throughout the 96-step window.

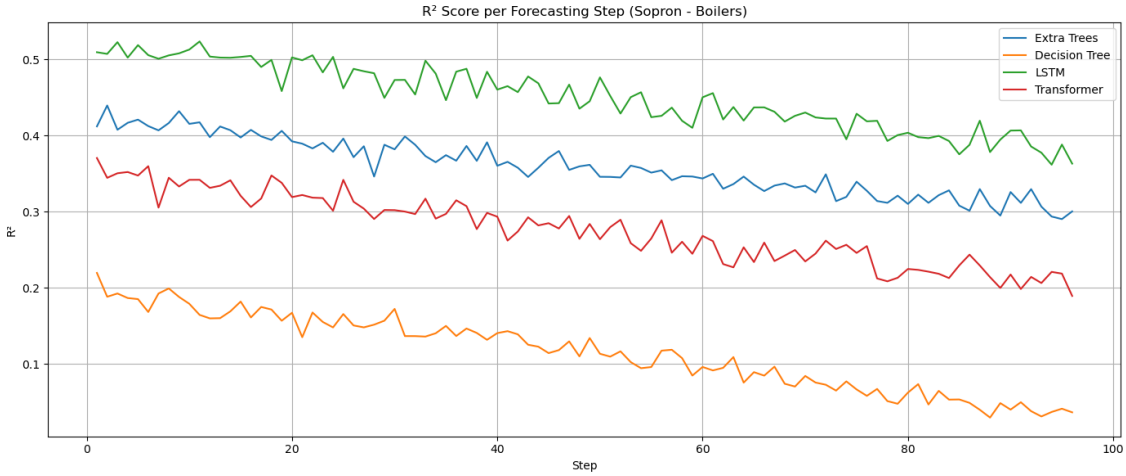


Figure 16: R^2 score across all 96 forecasting steps for Boilers aggregated target

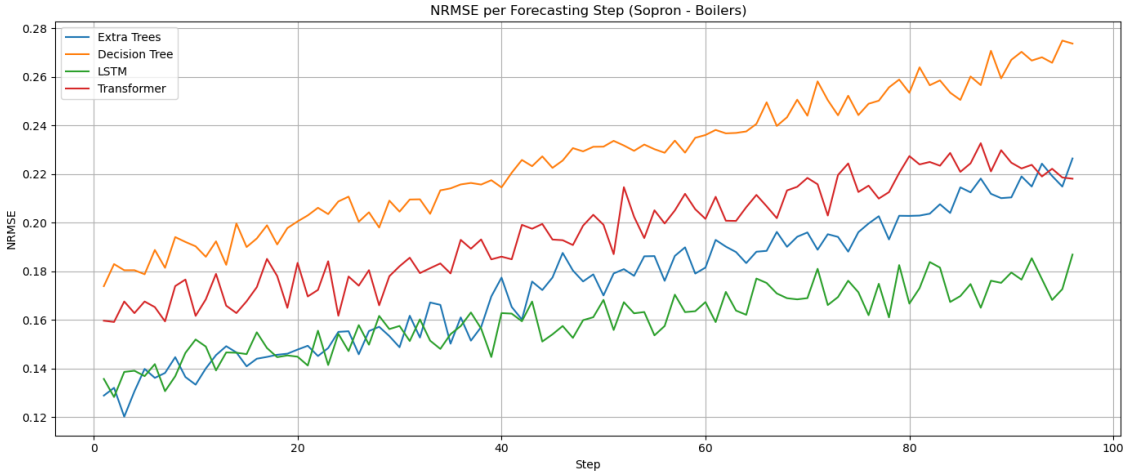


Figure 17: NRMSE across all 96 forecasting steps for Boilers aggregated target

ET models showed lower short-term error but deteriorated faster beyond 40 steps. The Transformer, although performing well in Cork and HVAC cases, showed suboptimal behaviour for this particular target, possibly due to lack of sufficient structural trends in the input sequences.

For the Lab forecasting implementation:

For this task, the ET regressor yielded the best performance, maintaining R^2 around 0.5 on average (Figure 18) and NRMSE below 0.12 (Figure 19) even at deeper forecast steps. Its ensemble nature captured multivariate dependencies effectively, making it suitable for structured consumption signals like Boiler 2.

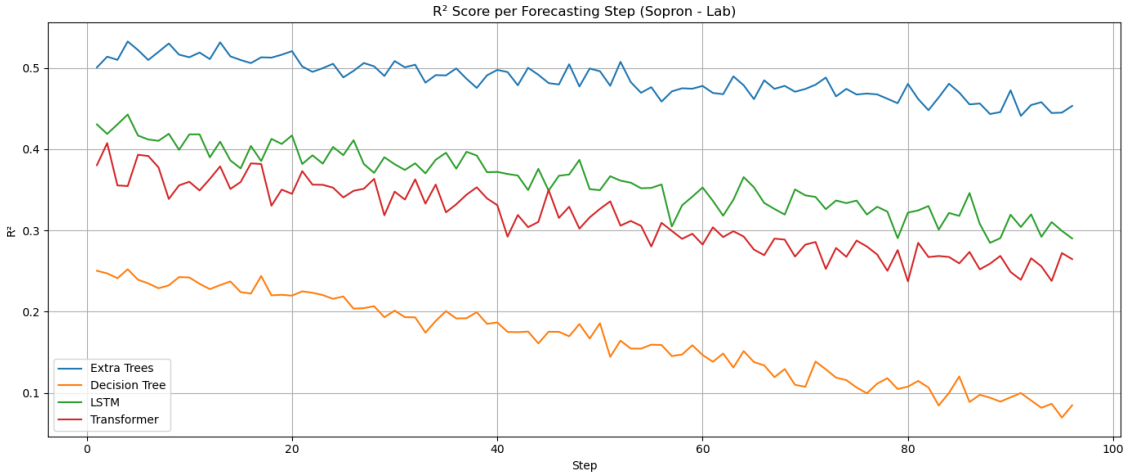


Figure 18: R^2 score evolution across 96 forecasting steps (Lab target)

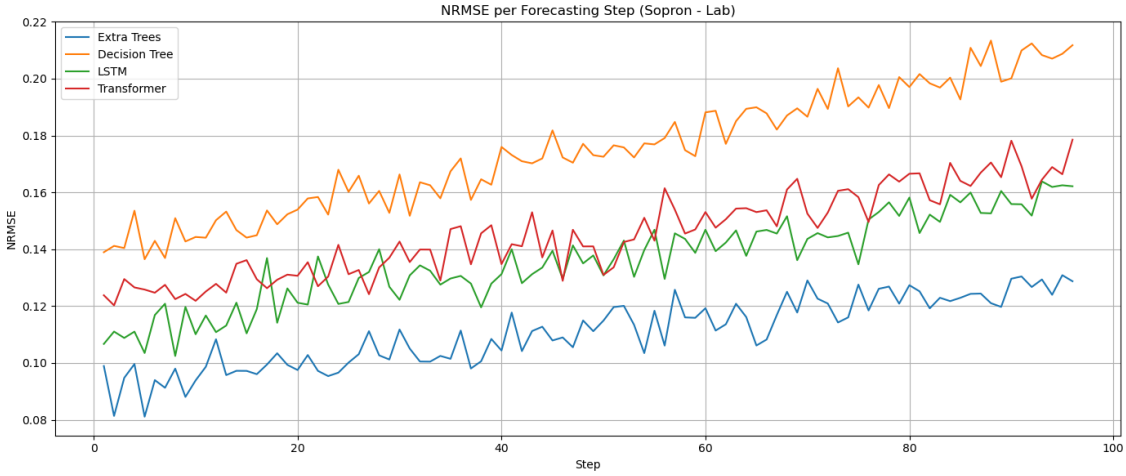


Figure 19: NRMSE values across the 96 forecasting steps (Lab target)

The LSTM model also performed reasonably, although with slightly worse stability. The Transformer and DT models showed significant variability and underperformance in long horizons, indicating limited generalization for this specific energy signature.

Regarding the HVAC implementation:

The Transformer-based model demonstrated the most consistent accuracy, with R^2 scores beginning near 0.55 and sustaining above 0.42 across all 96 steps (Figure 20). It also achieved the lowest average NRMSE (Figure 21), maintaining values between 0.09 and 0.16, showcasing its ability to retain predictive power at later horizons.

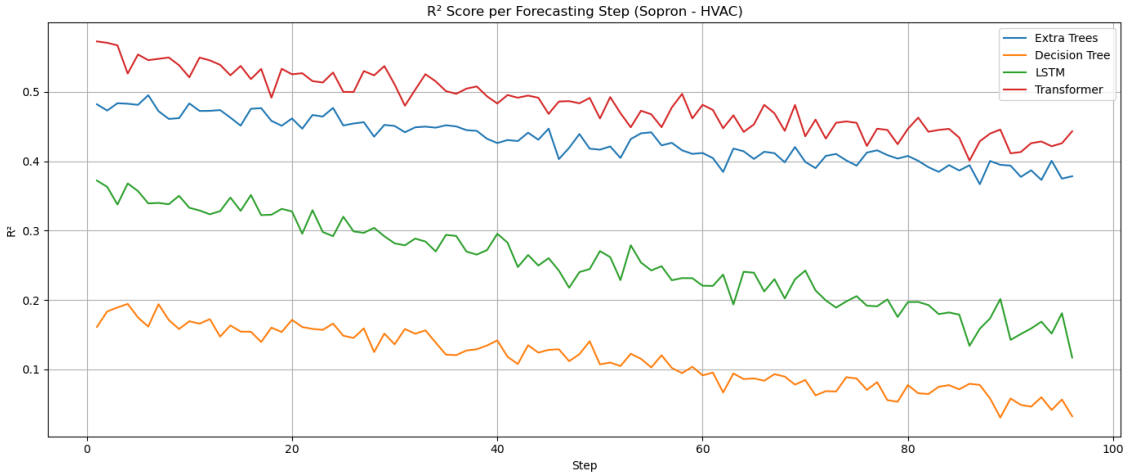


Figure 20: R^2 results for Sopron pilot site (HVAC target)

The ET regressor followed closely, with R^2 scores ranging from 0.48 to 0.38 and NRMSE values between 0.10 and 0.22. Its ensemble nature allowed for relatively stable performance but exhibited minor degradation over time.

DT models, while computationally efficient, suffered from greater prediction variability and weaker generalization, with R^2 dipping below 0.1 after step 60 and NRMSE exceeding 0.24 at later steps.

The LSTM model showed reasonable mid-range performance but degraded more noticeably toward the forecast horizon's end (R^2 near 0.15, NRMSE > 0.2), suggesting difficulty in long-term temporal pattern retention under the current data setup.

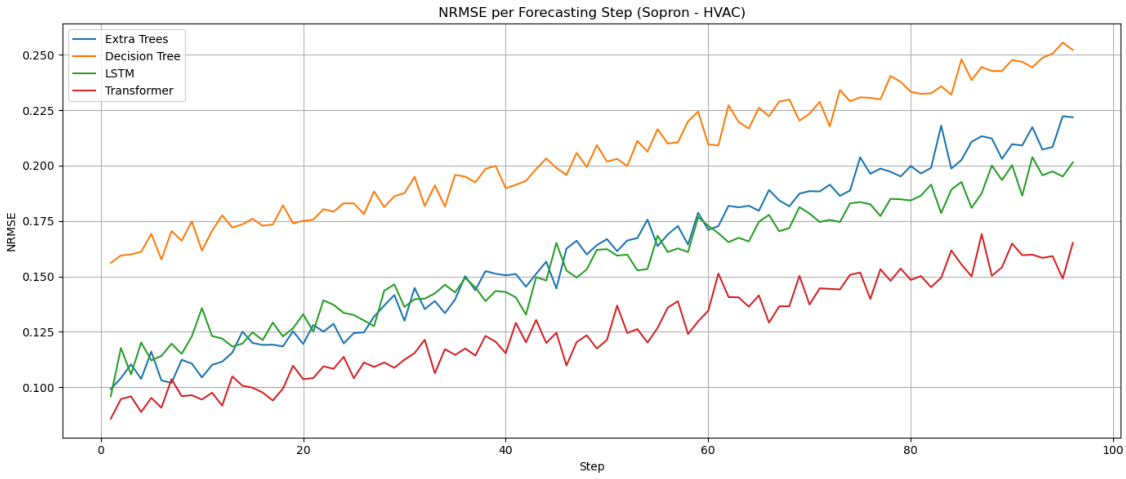


Figure 21: NRMSE results for Sopron pilot site (HVAC target)

Overall, the comparison reinforces the effectiveness of Transformer and ensemble tree methods in complex multi-step forecasting tasks with multivariate time series inputs. Further tuning of ANN architectures could potentially enhance their long-term accuracy.

4.3.3. Kimmeria

The Kimmeria pilot site involved a 96-step ahead electrical energy forecasting task based on weather and historical load inputs. Multiple multi-output models were evaluated to capture temporal dynamics across the full horizon. Results are summarized using R^2 (Figure 22) and NRMSE (Figure 22) as primary metrics.

The ET regressor achieved the best performance overall, maintaining an average R^2 around 0.91 and NRMSE values below 0.03, even in later steps. Its robustness and ability to capture non-linear dependencies made it the most stable model throughout the forecast window.

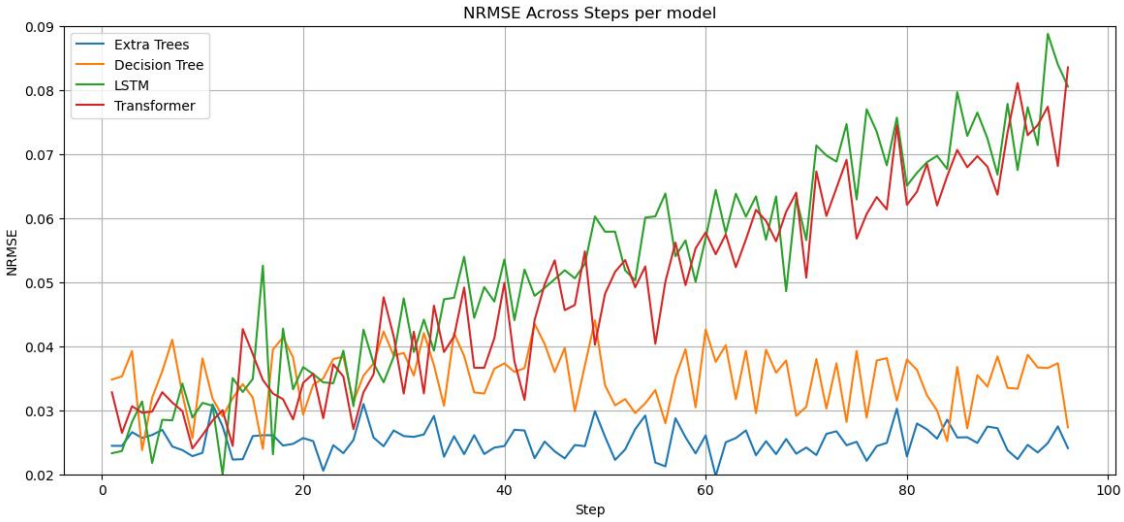


Figure 22: NRMSE results for Kimmeria pilot site

The DT regressor provided decent performance ($R^2 \approx 0.87$, NRMSE ≈ 0.035–0.04) but suffered from higher variability and reduced accuracy in longer horizons.

DL models such as LSTM and Transformer showed more noticeable degradation over time. The LSTM's R^2 dropped gradually to 0.70 by step 96, while its NRMSE rose beyond 0.07, indicating reduced reliability in long-range forecasts.

The Transformer architecture followed a similar trend, with R^2 falling toward 0.75 and NRMSE rising past 0.06 near the end of the forecast horizon

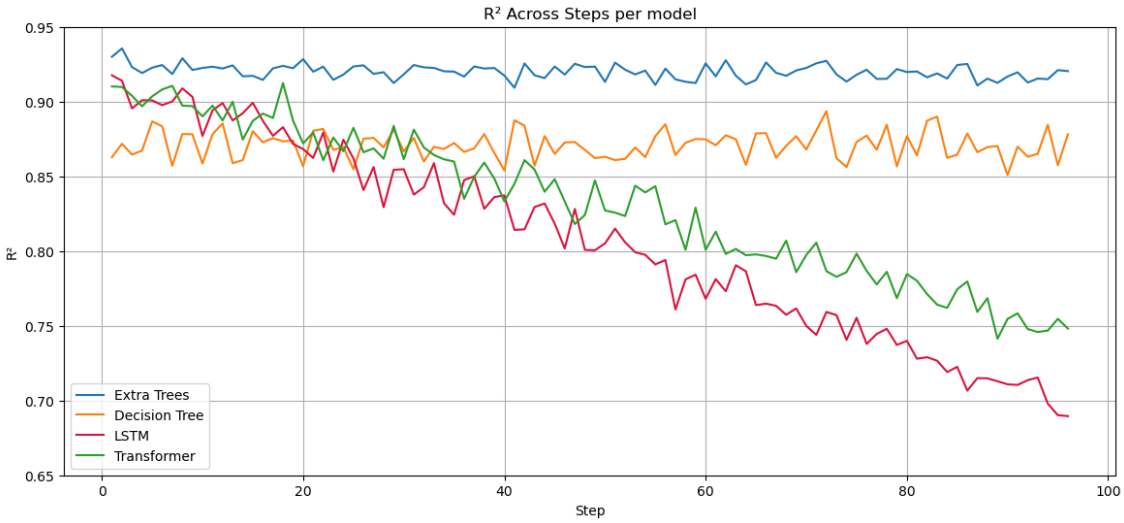


Figure 23: R^2 results for Kimmeria pilot site

Results comparison indicate the effectiveness of tree-based ensemble methods at this pilot, especially ETs, for mid-to-long range multi-output forecasting under limited data conditions. While neural models offer promise, they require further tuning or architectural enhancements to achieve comparable generalization.

4.3.4. Santiago de Compostela

Table 14 provides a comprehensive comparison of the evaluation metrics for the various forecasting models tested on the Santiago de Compostela pilot. This involves two targets hourly electric energy and hourly thermal energy demand, on an apartment level.

Table 14: Santiago de Compostela pilot 1-step ahead (1h) forecast metrics (Electric energy demand)

Model	R^2	NRMSE
Chronos	0.668	0.116
LGBM	0.661	0.119
CB	0.649	0.1225
ET	0.635	0.126
XGB	0.62	0.13
LSTM	0.598	0.136
HGBR	0.575	0.1415

As shown in Table 14, for the electric energy demand forecasting, the Chronos model demonstrated superior performance across all key metrics. It achieved the lowest MAE of 0.039, RMSE of 0.092 and Normalized Root Mean Squared Error (NRMSE) of 0.116, along with the highest R^2 of 0.668.

Following Chronos, traditional tree-based models like LightGBM (LGBM), CatBoost (CB), ETs and XGBoost (XGB) also delivered robust results, though with a slight decrease in accuracy. The performance of the LSTM and HGBR models was comparatively weaker, suggesting that for this particular forecasting task, the foundational Chronos model and standard ensemble methods were the most effective. It is important to note that as the forecast horizon extends further into the future

(i.e., more hours ahead), the R^2 value decreases while the NRMSE increases, indicating a reduction in model accuracy over longer prediction periods.

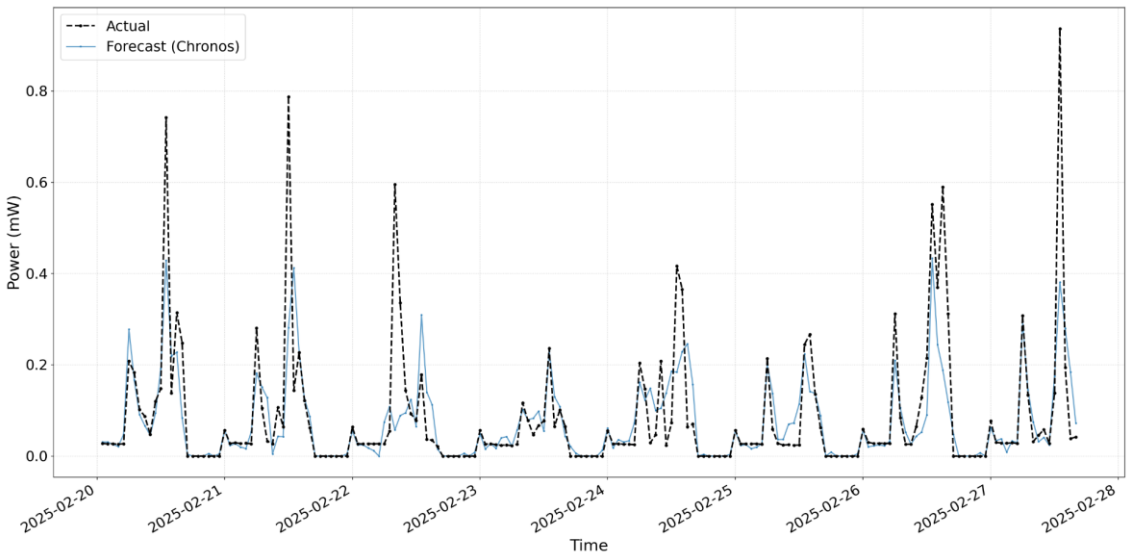


Figure 24: Day-ahead forecasting for Santiago de Compostela pilot-apartment (Electric energy -kWh)

Figure 24 shows a consecutive 24-hour ahead forecast of electric energy demand for the Santiago de Compostela pilot. A typical daily energy usage pattern is evident, with low consumption during the night, a sharp rise in the morning, a peak during the day and a decline in the evening. The forecast successfully captures these key patterns of the actual consumption profile, including the morning ramp-up and evening decline, indicating a robust prediction. For the daytime peak, the model correctly identifies the timing of the peak consumption; however, it tends to underestimate the full magnitude of the energy spike, predicting a more smoothed or sustained level of consumption rather than capturing the sharp maximum.

Table 15: Santiago de Compostela pilot 1-step ahead (1h) forecast metrics (Thermal energy demand)

Model	R^2	NRMSE
Chronos	0.739	0.1182
CB	0.729	0.1232
XGB	0.719	0.1282
LGBM	0.708	0.1337
ET	0.696	0.1397
LSTM	0.685	0.1452
HGBR	0.674	0.1507

In the case of thermal energy demand, the evaluation results are highlighted in Table 15. Chronos slightly outperformed the other models, achieving the highest R^2 of 0.739. It is important to note that these results are for the non-summer months (November to May), due to the thermal energy demand being typically close to zero in the other months (more details provided in Section 4.4.4). As with the electric energy forecast, the performance metrics are expected to degrade as the forecast moves further into the future, with the R^2 decreasing and the NRMSE increasing for each subsequent hour in the prediction horizon.

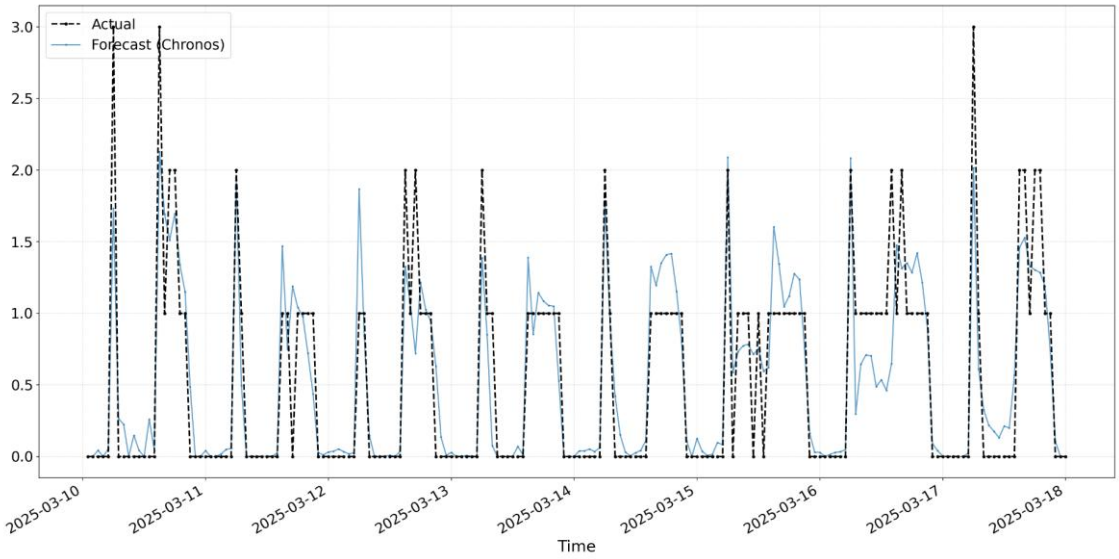


Figure 25: Day-ahead forecasting for Santiago de Compostela pilot-apartment (Thermal energy -kWh)

Figure 25, shows a consecutive 24-hour ahead forecast of thermal energy demand for the Santiago de Compostela pilot. Compared to the electric energy graph, the thermal energy consumption shows a more regularized daily pattern in the non-summer months (November to May), which explains the increased accuracy of the models. The forecast effectively captures the general trends and timings of consumption. The model's ability to closely track the consumption profile, including the sharp peaks, is consistent with the higher R^2 and lower error metrics reported for the thermal energy models.

4.3.5. Cork

For the Cork electric house target, the ETs model outperformed all other approaches, maintaining consistently high R^2 (Figure 26) values (around 0.82) and low NRMSE (Figure 27) values (starting near 0.03). This indicates its robustness in capturing the signal in multistep electric load forecasting.

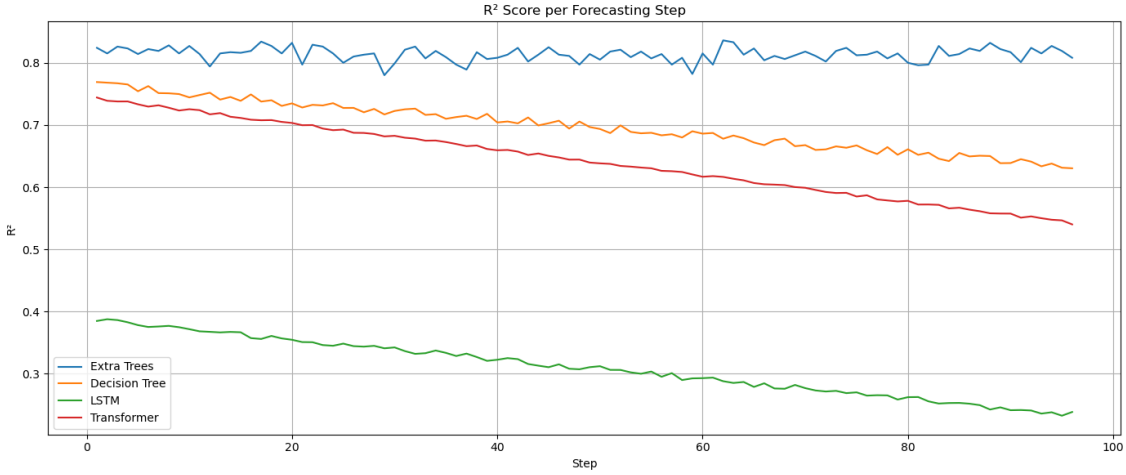


Figure 26: R^2 per step model comparison for Cork electric house

The DT model followed a moderately declining trajectory in performance, with R^2 gradually decreasing from 0.78 to 0.64 and NRMSE increasing up to 0.083. Despite some variability, it showed limited generalization capacity compared to ETs.

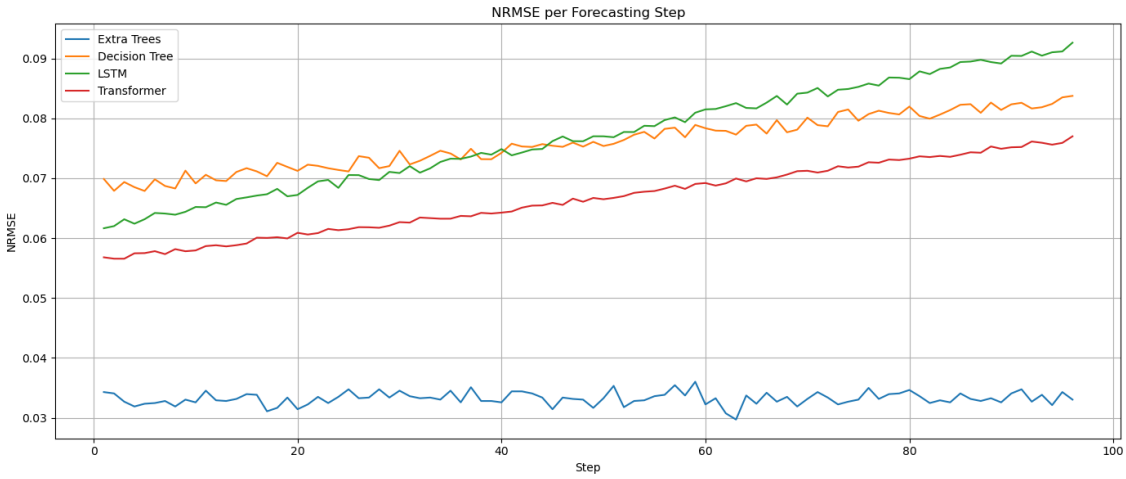


Figure 27: NRMSE per step model comparison for Cork electric house

Transformer and LSTM models demonstrated the weakest performance. The Transformer started with R^2 around 0.70 and decreased steadily toward 0.55, while LSTM began as low as 0.40, deteriorating further across the forecast horizon. Their NRMSE values also increased significantly, indicating difficulty in modelling long-range dependencies in this setting without additional tuning.

For the Cork gas flow, all models were trained on a normalized and log-transformed target, and predictions were evaluated using inverse transformations to compute real-scale metrics including R^2 and NRMSE.

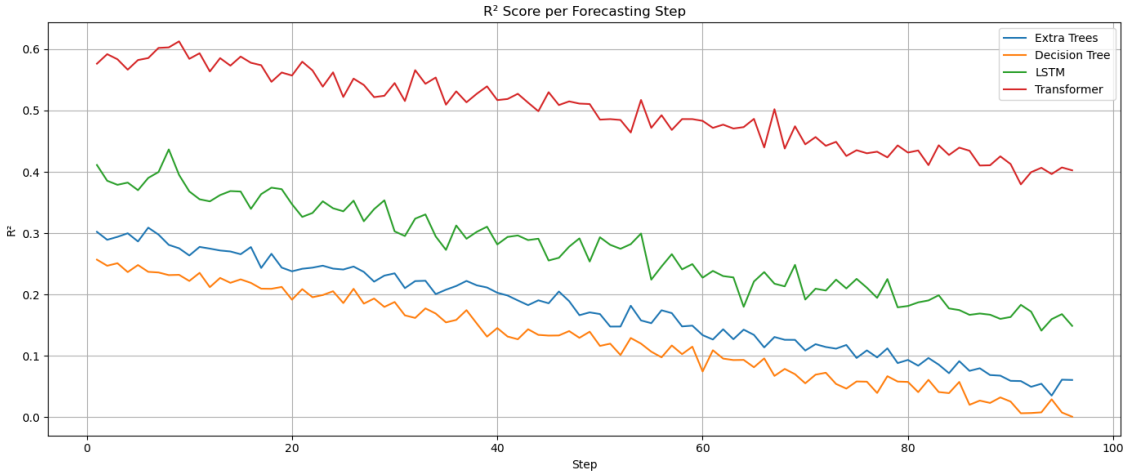


Figure 28: R^2 per step model comparison for Cork gas flow

The Transformer model outperformed the others, achieving R^2 (Figure 28) scores starting above 0.6 and gradually declining toward 0.4 across the forecast horizon. In contrast, tree-based models (DT and ET) started lower (around 0.25–0.3) and deteriorated faster. The LSTM model provided intermediate performance, starting around 0.4 R^2 and declining toward 0.15. NRMSE (Figure 29) followed an inverse trend, with the Transformer yielding the lowest error levels throughout.

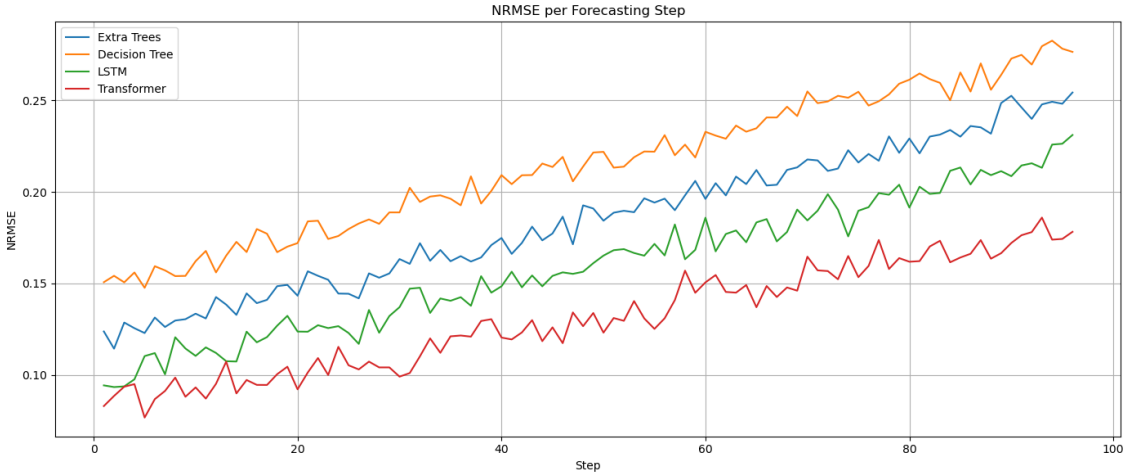


Figure 29: NRMSE per step model comparison for Cork gas flow

These findings highlight the Transformer's ability to capture temporal dynamics more effectively in long-horizon gas flow forecasting tasks, particularly under log-transformed conditions.

4.4. Data challenges and lessons learned

Unlike section 3.4, where data-related challenges and lessons learned on DER forecasting were synthesised across all demonstrators to highlight cross-cutting issues, in this case, focusing on demand forecasting, it was considered more appropriate to structure the discussion by pilot site. This approach allows capturing the specificities of each building and context, which strongly condition data quality, signal behaviour, and ultimately model performance.

4.4.1. Thessaloniki

The implementation of forecasting models for the Thessaloniki pilot site encountered specific data challenges, primarily related to data quality and availability.

- Due to **limited data availability**, a strategy was chosen to concatenate data from the two fan coil meters. This approach involved joining each fan-coil power series with the suite of meteorological predictors (air temperature, relative humidity, wind speed and direction, solar radiation) to create harmonized tables. Concatenating these fan-coil-weather tables into a single dataset increased the diversity of operating scenarios, supporting the development of a more robust and generalizable forecasting model. A single model was then trained on this combined dataset, employing a multi-step forecasting strategy where one model was trained per horizon step (1 to 24 hours ahead)

- **Minimum load handling:** the primary challenge here is physical implausibility and numerical noise in raw forecast outputs. Forecasting models can sometimes produce negative power difference values or values extremely close to zero due to inherent model limitations or minor numerical fluctuations. From a physical standpoint, power consumption cannot be negative. Furthermore, for the Thessaloniki pilot site, the minimum combined load for both fan coils cannot be under 1 kW. Truncating values below 1 Watt to zero ensures that the forecasts adhere to real-world physical constraints and helps in filtering out insignificant numerical noise that doesn't represent actual consumption.

- Operation-mode propagation: the challenge addressed by this mechanism is ensuring coherent and consistent system status across all forecasted timesteps. Without propagating the most recently observed operation mode (e.g., heating, cooling, standby), there's a risk that the forecasting model might generate power consumption predictions that are inconsistent with the actual or intended operational state of the system over the prediction horizon. This propagation maintains a continuous and logical representation of the system's status, crucial for accurate and actionable forecasts.

4.4.2. Sopron

The forecasting task in the Sopron pilot revolved around predicting differential energy consumption patterns for thermal loads such as boiler and HVAC systems. Several key challenges were encountered during model development and evaluation:

- Multivariate sensor integration: the dataset included readings from multiple internal and external sensors (e.g., temperatures, humidity, solar radiation, wind), combined with energy-related targets. The fusion of these heterogeneous inputs required synchronized time alignment and conversion to a uniform 15-minute resolution. Ensuring consistency across sources was essential, especially during daylight-saving transitions and occasional logging delays.

- Target granularity and signal weakness: the target variables for the boilers were formulated as kWh differences over time, which often exhibited low magnitude and variability. This made learning meaningful patterns more challenging, especially for DL models prone to overfitting on small signal fluctuations.

- Limited correlation and redundancy: in contrast to strongly seasonal or reactive targets (e.g., electric heating), the boiler signals were only weakly correlated with environmental conditions. This was particularly evident in feature importance analyses and was reflected in the modest R^2 scores obtained, rarely exceeding 0.55 across most modelling strategies.

- Modelling trade-offs: while the ETs regressor achieved the most consistent results for boiler 2, other models like LSTM and Transformer demonstrated reduced generalization and sensitivity to hyperparameter tuning. DL methods particularly struggled to capture long-term dependencies, possibly due to the weak predictive signal and small data scale.

These lessons highlight the importance of hybrid and ensemble approaches when dealing with sparse or low-resolution targets. Further improvements may be possible with domain-specific feature engineering or inclusion of external context (e.g., user activity schedules, occupancy signals).

4.4.3. Kimmeria

The forecasting task in Kimmeria focused on predicting power load, using a variety of environmental and operational features. Several key challenges emerged during model development:

- Heterogeneous sensor inputs: the dataset contained measurements from diverse sources (e.g., weather, energy and environmental sensors) with varying sampling rates. Harmonizing these sources into a unified time series required careful resampling and alignment, particularly ensuring consistency in temporal resolution and handling time zone shifts or daylight-saving adjustments.

- Missing data and imputation: sensor outages and irregularities led to gaps in both features and target variables. A combination of forward-fill and interpolation techniques was applied,

with additional attention paid to not biasing the model with artificial smoothness, especially for weather variables like solar radiation.

- Target volatility: the power consumption signal showed sharp transitions and irregularities, likely driven by operational shifts and user behaviour. To stabilize training, smoothing strategies and sequence-to-sequence forecasting approaches were applied. However, even advanced architectures struggled with long-term accuracy beyond 6–12 steps ahead, prompting a shift to hybrid modelling using both tree-based (ETs) and neural models.

4.4.4. Santiago de Compostela

For the Santiago de Compostela pilot site, the project focused on developing 24-hour ahead forecasting models for the apartment's electric and thermal energy loads. However, several data-specific challenges necessitated methodological adjustments:

- DHW forecasting: an additional target, DHW energy consumption, was part of the initial experimentation. However, the DHW data was characterized by high sparsity and intermittent demand patterns. This resulted in poor forecasting accuracy from all tested models for short-term horizons (e.g., one hour ahead). Consequently, the development of a reliable short-term DHW forecasting model was deemed unfeasible. Although models for long-term horizons (e.g., weekly or monthly) were considered a possibility, such forecasts lack the granularity required for the project's real-time control and monitoring platform and were therefore determined to be outside the project's scope.

- Thermal energy data skewness: the thermal energy data was found to be heavily skewed. During the summer months (approximately June to September), consumption was consistently near-zero. To address this, a specific strategy was developed to set the forecast to zero for this period, improving model accuracy and avoiding unnecessary computations during low-demand seasons.

- Older data irregular patterns: a key challenge was the high degree of data irregularities observed during the initial years of the available data (mainly 2022-2023). This was characterized by increased inconsistent patterns and anomalous readings. In contrast, data from 2024 and 2025 showed a more normalized and stable behaviour, which significantly improved the reliability of the experimented dataset, utilized to benchmark and developed the forecasting models.

4.4.5. Cork

The Cork pilot site dealt with the forecasting of electric house and gas flow using historical energy and environmental data. The following data-specific issues were encountered:

- Target distribution skewness: the gas_flow target variable exhibited heavy skewness and long-tail behaviour. A log transformation was applied to stabilize variance and make the data more suitable for modelling. Following this, standard normalization was used to ensure balanced model training.

- Feature redundancy: certain features (e.g., immersion_energy) were found to be strongly correlated or irrelevant for the forecasting objective and were removed to reduce noise and overfitting risks.

- One-to-many forecasting architecture: a multi output forecasting strategy was implemented using multi-output regressors, where each model predicted 96 steps (24

hours) ahead from a single timestamped input. Models including ETs, DTs, LSTM and Transformer architectures were evaluated.

- Log-space metrics handling: special care was taken during evaluation to apply inverse transformations (including exponentiation) before computing real-world metrics like R^2 and NRMSE, to maintain physical interpretability of forecasts.
- Model performance divergence: while tree-based models showed relatively low performance (R^2 ~0.3), the Transformer-based model consistently outperformed others, maintaining R^2 above 0.8 in early steps and around 0.5 near the 96th horizon. This confirmed the suitability of attention-based mechanisms for complex temporal dependencies in gas dynamics.

5. Conclusions

This deliverable has presented the development and evaluation of forecasting models for both renewable energy generation and energy demand within the MiniStor system. The work has shown that forecasting approaches based on ANNs and advanced ML methods provide valuable predictive capabilities, with performance outcomes closely linked to the characteristics and quality of the available datasets.

In the case of DER generation forecasting, MLP ANNs were trained to predict GHI, electrical power and thermal power at different demonstration sites. The results demonstrated clear improvements compared to raw meteorological forecasts, as the networks were able to reduce systematic biases and capture nonlinear relationships between variables such as irradiance, temperature and cloud cover. At the same time, the experiments underlined the importance of context-specific conditions: in sites where auxiliary devices such as heat pumps or control logics influenced the system response, the causal link between meteorological inputs and observed outputs became less direct. This confirms that, in such complex environments, richer datasets including additional meteorological or operational variables further enhance model accuracy and generalization potential.

For demand forecasting, models were successfully developed and validated. The results confirmed that no single algorithm is universally superior, and the most suitable approach varies depending on the dataset. In several cases, tree-based and ensemble methods such as ETs and VotRegrs delivered robust and reliable performance, sometimes exceeding that of more complex DL models such as LSTM networks. Transformer-based models also showed potential, although their performance over longer forecasting horizons highlighted the need for additional tuning or architectural refinements to improve generalization.

Data quality and availability proved to be decisive factors in shaping the final methodologies. The challenge of limited or heterogeneous data was mitigated by combining multiple sources into unified and more robust datasets. Quality issues, such as irregularities and anomalous readings in older datasets, were addressed by prioritizing recent and more stable measurements, which substantially improved reliability. Furthermore, normalization and scaling techniques such as Z-score normalization and Min-Max scaling were successfully applied, helping to stabilize variance and ensure balanced training across diverse conditions.

The forecasting models consistently captured daily energy demand patterns. While some smoothing was observed during rapid consumption peaks and the accuracy naturally declined with longer forecasting horizons, these results reflect the inherent complexity of user-driven demand behaviour rather than modelling deficiencies. Importantly, the developed models proved effective in providing actionable forecasts within the operational context of the MiniStor system.

Overall, the results of both generation and demand forecasting confirm the operational relevance of predictive models in hybrid energy systems. The outputs developed here constitute a solid foundation for the optimization and control strategies defined in related tasks, particularly Task 5.1, Task 5.3 and Task 5.4, and their associated deliverables D5.2 and D5.5. By integrating forecasting with the SHEMS and the IoT platform, the MiniStor system can better align renewable generation with demand profiles, optimize operation, and contribute directly to the project's overarching objectives of maximizing renewable self-consumption, reducing costs and preserving user comfort.

References

Ahmad, M. W., Reynolds, J., & Rezgui, Y. (2018). Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees. *Journal of Cleaner Production*(203), 810-821.

Ahmadi, M. H. (2020). Evaluation of electrical efficiency of photovoltaic thermal solar collector,. *Engineering Applications of Computational Fluid Mechanics*, *14*(1), 545-565.

Al-Shammari, E. T., Keivani, A., Shamshirband, S., Mostafaeipour, A., Yee, P. L., Petkovic, D., & Ch, S. (2016). Prediction of heat load in district heating systems by Support Vector Machine with Firefly searching algorithm. *Energy*, *95*, 266-273.

Amstrong, J. S., & Collopy, F. (1992). Error measures for generalizing about forecasting methods. Empirical comparisons. *International Journal of forecasting*(8), 69-80.

Ansari, A. F. (2024). *Chronos* (Vol. 10).

Ansari, A. F., Stella, L., Turkmen, C., Zhang, X., & Mercado, P. (2024). *Chronos: Learning the language of time series* (Vol. 10). Transactions on Machine Learning Research.

Antonanzas, J., Osorio, N., Escobar, R., Urraca, R., Martinez-de-Pison, F., & Antonanzas-Torres, F. (2016). Review of photovoltaic power forecasting. *Solar Energy*(136), 78-111.

Baltputnis, K., Petrichenko, R., & Sobolevsky, D. (2018). Heating Demand Forecasting with Multiple Regression: Model Setup and Case Study. *IEEE 6th Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE).* Vilnius, Lithuania.

Banihashemi, S., Ding, G., & Wang, J. (2016). Developing a hybrid model of prediction and classification algorithms for building energy consumption. *1st International Conference on Energy and Power, ICEP2016* (pp. 14-16). Melbourne, Australia: RMIT University.

Barbu, M., & Darie, G. S. (2019). Analysis of a residential photovoltaic-thermal (PVT) system in two similar climate conditions. *Energies*, *12*(3595).

Bonilla, J., Carballo, J., Fernández-Reche, J., & Valenzuela, L. (2019). Machine learning perspectives in concentrating solar thermal technology. *EUROSIM 2019*. Logroño, Spain.

Das, U., Tey, K., Seyedmahmoudian, M., Mekhilef, S., Idris, M., Deventer, W., . . . Stojcevski, A. (2018). Forecasting of photovoltaic power generation and model optimization: A review. *Renewable and Sustainable Energy Reviews*, *81*(1), 912-928.

Deb, C., Eang, L. S., Yang, J., & Santamouris, M. (2016). Forecasting diurnal cooling energy load for institutional buildings using Artificial Neural Networks. *Energy and Buildings*, *121*(1), 284-297.

Deb, C., Zhang, F., Lee, S. E., & Shah, K. W. (2017). A review on time series forecasting techniques for building energy consumption. *Renewable and Sustainable Energy Reviews*, *74*, 902-924.

Diwania, S., Agrawal, S., Siddiqui, A. S., & Singh, S. (2020). Photovoltaic–thermal (PV/T) technology: a comprehensive review on applications and its advancement. *International Journal of Energy and Environmental Engineering*(11), 33-54.

El Khantach, A., Hamlich, M., & Belbounaguia, N. E. (2019). Short-term load forecasting using machine learning and periodicity decomposition. *AIMS Energy*, *7*(3), 382-394.

Energy, U. D. (2019). EnergyPlus. *version 9.2.0*. Washington, D.C. Retrieved from https://energyplus.net

Fouad, M. M., Shihata, L. A., & Morgan, E. I. (2017). An integrated review of factors influencing the perfomance of photovoltaic panels. *Renewable and Sustainable Energy Reviews*(80), 1499-1511.

Fumo, N., & Biswas, M. R. (2015). Regression analysis for prediction of residential energy consumption. *Renewable and Sustainable energy reviews*, *47*, 332-343.

Hosenuzzaman, M., Rahim, N. A., Selvaraj, J., & Hasanuzzaman, M. (n.d.). Factors affecting the PV based power generation. *3rd IET International Conference on Clean Energy and Technology (CEAT) 2014*. Kuching, Malaysia.

Izgi, E. Ö., Yerli, B., Kaymak, M., & Şahin, A. (2012). Short.mid-term solar power prediction by using artificial neural networks. *solar Energy*, *725-733*(86).

Jovanovic, R. Z., Sretenovic, A. A., & Zivkovic, B. D. (2015). Ensemble of various neural networks for prediction of heating energy consumption. *Energy and Buildings*, *94*(1), 189-199.

Kazem, H. A., Yousif, J., Chaichan, M. T., & Al-Waeli, A. H. (2019). Experimental and deep learning artificial neural network approach for evaluation grid-connected photovoltaic systems. *Int J Energy Res.*, 1-20.

Lee, M., & Jung, J. (2017). 24-Hour photovoltaic generation forecasting using combined very-short-term and short-term multivariate time series model. *IEEE Power & Energy Society General Meeting*. Chicago, IL, USA.

Leva, S., Dolara, A., Grimaccia, F., Mussetta, M., & Ogliari, E. (2017). Analysis and validation of 24 hours ahead neural network forecasting of photovoltaic output power. *Mathematics and Computers in Simulation*, *131*, 88-100.

Lewis, D. (1982). Industrial and business forecasting methods. *London: Butterworths*, 40.

Li, Y., He, Y., Su, Y., & Shu, L. (2016). Forecasting the daily power output of a grid-connected photovoltaic system based on multivariate adaptive regression splines. *Applied Energy*, *180*, 392-401.

Massa Gray, F., & Schmidt, M. (2018). A hybrid approach to thermal building modelling using a combination of Gaussian processes and grey-box models. *Energy and Buildings*, *165*, 56-63.

Mojumder, J., Ong, H. C., Chong, W. T., Izadyar, N., & Shamshirband, S. (2017). The intelligent forecasting of the performcances in PV/T collectors based on soft computing method. *Renewable and sustainable energy reviews*, *72*.

National Renewable Energy Laboratory (NREL). (2017). EnergyPlus™.

Pino-Mejías, R., Pérez-Fargallo, A., Rubio-Bellido, C., & Pulido.Arcas, J. A. (2017). Comparison of linear regression and artificial neural networks models to predict heating and cooling energy demand, energy consumption and CO2 emissions. *Energy*, *118*(1), 24-36.

Rosenblatt, F. (1962). Perceptrons and the Theory of Brain Mechanisms. *Principles of Neurodynamics*.

Shahsavar, A., Moayedi, H., Al-Waeli, A., Sopian, K., & Chelvanathan, P. (2020). Machine learning predictive models for optimal design of building-integrated photovoltaic-thermal collectors. *International Journal of Energy Research*.

Shi, J., Lee, W., Liu, Y., Yang, Y., & Wang, P. (2011). Forecasting power output of photovoltaic systems based on weather classification and support vector machines. *IEEE Industry Applications Society Annual Meeting.* Orlando, US.

Siddharth, V., Ramakrishna, P. V., Geetha, T., & Sivasubramaniam, A. (2011). Automatic generation of energy conservation measures in buildings using genetic algorithms. *Energy and Buildings*, *43*(10), 2718–2726.

Sobri, S., Koohi-Kamali, S., & Rahim, N. A. (2018). Solar photovoltaic generation forecasting methods: A review. *Energy Conversion and Management*(156), 459-497.

Wang, Z., & Srinivasan, R. S. (2016). A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models. *Renewable and Sustainable Energy Reviews*, *75*, 796-808.

Wang, Z., Srinivasan, R. S., & Shi, J. (2016). Artificial Intelligent Models for Improved Prediction of Residential Space Heating. *Journal of Energy Engineering*, *142*.

Wolff, B., Kühnert, J., Lorenz, E., Kramer, O., & Heinemann, D. (2016). Comparing support vector regression for PV power forecasting to aphysical modeling approach using measurement, numerical weatherprediction, and cloud motion dat. *Solar Energy*(135), 197-208.

Yang, C., A., T., & Xie, L. (2015). Multitime-scale data-driven spatio-temporal forecast of photovoltaic generation. *IEEE Transactions on Sustainable Energy*.