

Article

Enhancing Peak Shaving Efficiency in Small Hydro Power Plants Through Machine Learning-Based Predictive Control

Francesca Mangili * , Marco Derboni , Lorenzo Zambon , Vincenzo Giuffrida  and Matteo Salani 

Istituto Dalle Molle di Studi sull'Intelligenza Artificiale (IDSIA), University of Lugano (USI), University of Applied Sciences and Arts of Southern Switzerland (SUPSI), 6962 Lugano, Switzerland; marco.derboni@supsi.ch (M.D.); lorenzo.zambon@supsi.ch (L.Z.); vincenzo.giuffrida@supsi.ch (V.G.); matteo.salani@supsi.ch (M.S.)

* Correspondence: francesca.mangili@supsi.ch

Abstract

Small hydropower plants (HPPs) equipped with water storage play an important role in managing fluctuating energy demand. This article presents a real-world case study in which model predictive control (MPC), driven by energy-demand and water-inflow forecasts produced using the Light Gradient Boosting Machine (LGBM), is applied to optimize the operation of a small hydropower plant for peak shaving. A comparative analysis is conducted between the current non-predictive control strategy, which relies on operator decisions for peak shaving, and a fully automatic controller that optimally schedules the utilization of available water resources based on ML predictions. Results show that the MPC can outperform the operator-based scheduling and that this has the potential to improve the peak shaving capabilities of small HPPs. Unlike previous studies that predominantly focus on large and complex hydropower systems or introduce new control formulations evaluated under idealized assumptions, this work offers a pragmatic solution to the underexplored context of peak shaving for small HPPs operated with limited data and resources, that small utilities can adopt with minimal effort using their own data. We show that even these small-scale hydropower operations have room for improvement through optimal scheduling.

Keywords: model predictive control; LGBM; energy demand forecast; peak shaving; hydropower plant



Academic Editors: Sebastijan Seme and Klemen Sredenšek

Received: 10 January 2026

Revised: 7 February 2026

Accepted: 9 February 2026

Published: 13 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Managing the variability of energy demand throughout the day poses a significant challenge for energy providers, potentially leading to increased electricity costs. As a result, utilities are increasingly concerned with developing effective storage and power generation strategies to address peak loads [1]. Recently, several techniques have been explored to reduce such undesired behavior [2]. In particular, the optimal management of hydropower plants (HPPs) is considered effective for peak shaving purposes [3,4].

Villeneuve et al. [5] review the current state of hydropower scheduling, optimization, prediction, and production forecasting, providing a comprehensive overview of diverse case studies and objectives related to short-term scheduling as summarized in Tables 1 and 2 of [5]. Some of these works address the problem of smoothing system load (peak shaving). Wu et al. [6] propose a composite model including multi-objective optimal peak shaving, inter-grid power distribution and load fluctuation balance for the short-term

management of a massive system of 92 HPPs totaling 41 GW nominal power. Liao et al. [3] formulated an improved daily peak shaving operation model for one-reservoir and multi-cascade HPPs by developing a method to represent the water delay time and addressing non-linear constraints for cascade HPPs. The main results demonstrate that the proposed mixed-integer quadratic program (MIQP), integrating the delay time, effectively reduces peak-valley differences by 25% in the dry and 33% in the wet season. Zhao et al. [7] propose a decomposition and iterative search strategy (DISS) to address the challenges in the short-term peak-shaving operation of complex, very large-scale hydropower installations with multiple reservoirs and cascades, which are characterized by modeling difficulties and high computational cost. The DISS method involves dividing the system into independent subsystems, using mixed-integer linear programming (MILP) to find an initial feasible power output, and then iteratively refining the solution by adjusting peak-valley differences of the residual load. Wang et al. [8] address the challenge of coordinating the power dispatch of a cascade HPP with pumping facilities to compensate for the increasing randomness of PV power. The authors propose a day-ahead dispatch strategy based on multi-objective stochastic optimization. Across all these works, the primary goal is to introduce a new optimization framework and demonstrate its theoretical effectiveness within highly complex, large-scale installations that coordinate the operation of multiple plants and interconnected systems. Their experiments typically assume perfect knowledge of future demand and inflow, and when robustness is desired, they rely on stochastic formulations that generate scenario variations around these known historical profiles rather than forecasting them. The operational problem of optimally scheduling energy production in small local HPPs is instead ignored by the existing literature, and these studies seldom evaluate real-time operational performance against truly unknown future demand, as their focus is not on real-time deployment or operational validation. Our work contributes to filling this gap by addressing the need for pragmatic, cost-effective solutions tailored to small utilities operating storage-based HPPs (i.e., plants equipped with a reservoir that allows shifting water use over time), primarily for peak-shaving applications. Small HPPs offer a dual advantage of environmental sustainability and operational flexibility, both of which are essential for maintaining grid stability [9]. These decentralized systems have the potential to contribute significantly to local peak load management. However, during dry periods, generation is often constrained because limited inflow and finite reservoir storage reduce the available hydraulic energy; therefore, careful management of the stored water is critical. Based on stakeholder workshops conducted with local utilities in the Canton of Ticino (Switzerland), we found that these operators often manage their HPPs autonomously, using non-optimized strategies, which limits their impact on grid flexibility. Similar situations have been observed in other areas [10]. Operators play a central role in ensuring efficient resource utilization and maintaining a stable energy supply, drawing on their knowledge of system behavior and operational constraints. Nevertheless, especially for less experienced personnel, decision-making can be improved by decision-support tools such as automated optimal scheduling systems, which generate production plans aligned with the utility's objectives. Even when operators remain directly in charge of the final decision, their operational effectiveness can be enhanced through accurate short-term forecasts of energy demand and water availability [11].

To support operators in the scheduling task and in light of the current trend toward semi- or fully-automated control in small HPPs, this study explores how accurate energy demand and water inflow forecasts coupled with the automatic optimal scheduling of HPP production can meaningfully improve peak shaving efficiency. Achieving an improvement in this context would also deliver economic benefits for utilities and end users in those regions where power transfer prices paid to the national energy providers increase

with peak demand. While these pricing policies are meant to incentivize peak reduction, adoption by small utilities is feasible only if solutions are low cost, affordable, and easy to implement using existing historical data. Therefore, this work addresses two practical research questions: (i) how current robust and computationally efficient machine learning (ML) models and model predictive control (MPC) can be leveraged to smooth load profiles through the optimal use of water storage during peak-demand periods; and (ii) how effective such an approach is in real-world operational settings when compared with current operator-driven practices.

We demonstrate the proposed approach in the context of a small HPP managed by a local utility, the Azienda Elettrica di Massagno (AEM), in Ticino, Switzerland, and compare the operator decisions to those of the MPC fed with different inputs. Historical data provided by the utility are used to train the energy demand and water flow forecasters. We design experiments simulating the MPC decision based on data from periods characterized by different water availability. These experiments allow us to evaluate the impact of the MPC scheduler and ML forecasters on the subsequent decisions made by the MPC and to compare these decisions with those made by operators. These experiments highlight the importance of feeding the MPC with accurate and reliable forecasts and show the potential of the MPC to support operators in their daily decision-making processes, thus enhancing energy storage management.

The main contributions of this work can be summarized as follows:

- Unlike most studies that investigate large and complex hydropower systems (e.g., 92 plants and 41 GW systems with multi-grid HVDC interactions [3,6]), this work examines peak-shaving scheduling for small HPPs—a scale largely absent in academic research—showing that even single plant scheduling problems have room for improvement through optimal scheduling.
- Because the prior literature focuses on large installations, it can adopt complex multi-objective models, and requires high-performance computing [3,6–8]. In contrast, we propose a pragmatic and computationally lightweight solution that can be easily customized to other small HPPs. Our approach, based on LGBM forecasting and deterministic MPC, runs in real time on standard hardware (solution times < 3 s), providing actionable guidance for small utilities with limited IT and engineering resources.
- Consistent with their theoretical nature, most existing studies evaluate their models in an ideal setting. We instead test the proposed approach in a real-world operational setting where future load and inflow must be predicted, and provide a direct comparison between MPC-generated schedules and human operator decisions. This constitutes an empirical validation rarely found in the literature.
- Finally, we report concrete metrics on peak reduction, load fluctuations, and estimated monthly savings. Such micro-scale energetic and economic evaluations quantify the real operational benefits achievable in a small HPP.

2. Methodological Background: Optimization and Forecasting Approaches

While the Introduction provides application-oriented related work to contextualize the relevance and motivation of this study, the present section focuses specifically on the methodological background. Here, we review prior work on optimization frameworks for energy assets and on short-term energy demand forecasting methods, as these constitute the key methodological components underpinning our approach.

Approaches for optimal management of energy assets include control algorithms based on Convex Optimization (CO) [12] and multi-objective MILP for energy arbitrage in an energy community [13,14]. A simple rule-based controller (RBC) is presented in [15],

which can mitigate peak loads in the distribution grid of a Swiss utility. Zhang et al. [16] and Hannan et al. [17] propose more sophisticated approaches based on the Genetic Algorithm and Binary Particle Swarm Optimization. Other techniques consider ML approaches such as [18], where the authors propose a Decision Tree (DT)-based approach for energy scheduling. Training the DT involves creating learning sets via Monte Carlo simulations, with short-term and long-term firm power capacity provisions considered. Multi-agent reinforcement learning (RL) is considered in [19], where the RL controller employs soft actor–critic agents to manage storage charging/discharging decisions. Zhang et al. [20] compare MPC and RL, focusing on cost and carbon emissions. The results slightly favor MPC. Finally, Rocchetta et al. [21] investigate RBCs and RL agents, combining MaskPPO and a deep neural network.

Although advanced regression-based models can provide very good forecasting accuracy, optimization strategies still suffer from residual errors and uncertainties in estimating future energy demand and water availability. To address this challenge, several studies have explored advanced control techniques—most notably, stochastic model predictive control (SMPC) and RL, which are explicitly designed to handle uncertainty. For instance, Sahu et al. [22] develop an SMPC framework for adaptive real-time reservoir operation under uncertain inflow forecasts. Likewise, Kumar et al. [23] compare deterministic and stochastic MPC for stationary battery systems, highlighting the benefits of explicitly anticipating system disturbances. Hybrid approaches have also been proposed: Arroyo et al. [24] introduce an RL-MPC algorithm, showing that combining data-driven learning with model-based control can achieve performance comparable to conventional MPC. More recently, Ge et al. [25] apply multi-task deep reinforcement learning to optimize real-time scheduling in hybrid renewable energy systems that include hydropower resources.

Within this landscape, MPC emerges as a particularly suitable choice for our problem setting. Unlike heuristic rule-based controllers, MPC explicitly incorporates the physical and operational constraints, ensuring the feasibility of all scheduling decisions, while, in contrast to meta-heuristic and RL approaches, it offers full interpretability, predictable real-time solvability, and requires no expensive offline training. Although SMPC and RL-based methods have shown promising performance in handling uncertainty, their practical deployment in operational hydropower systems remains challenging: SMPC is based on reliable probabilistic inflow forecasts and scenario generation, which are rarely available in day-to-day plant operation, while RL methods demand large historical datasets and exhibit closed-loop behaviors that are difficult to interpret and certify in safety-critical water management applications. Deterministic MPC, by contrast, is inherently designed to exploit short-term forecasts—the type of information most readily available to small hydropower plants—while providing transparent constraint handling, modest computational requirements, and straightforward integration into existing supervisory control architectures. Empirical evidence further supports its effectiveness in practice: for example, Cao et al. [26] show that real-time deterministic MPC can successfully enhance peak-shaving performance in interprovincial hydropower systems even under varying inflow and demand conditions.

A conceptually similar MPC approach was previously proposed by Derboni et al. [27] to flexibly schedule residential loads in a self-consumption community, with the goal of reducing peak electricity demand. The controller operates on adjustable loads according to their availability for curtailment or shifting. In both [27] and the present work, the MPC bases its decisions on the estimates of future energy demand and the availability of controllable resources. In [27], the prediction task is performed with a forecasting model combining a seasonal method and a simple classification model. This approach, which is used as a comparison baseline in this work, belongs to classical time-series forecasting

methodologies (moving average, seasonal method with error correction, autoregressive models, etc.) that base their predictions on the most recent observations and assume specific patterns within a time series (trends and seasonal fluctuations) [28]. These approaches work well for short-term forecasting, while for medium to medium-term forecasting, it is important to include a set of explanatory variables or features, such as calendar effects, weather variations, and so on. Therefore, medium and long-term forecasting are more effectively treated as a regression problem [29].

In the last decade, several ML regression frameworks have been applied to time-series forecasting with covariates. Among them, LGBM is a boosted tree-based algorithm that has proven very effective and widely used. Thanks to its versatility, performance, and relatively simple usage, it enables the development of models with competitive accuracy while requiring limited development times and computational resources [29–31]. For instance, boosted trees algorithms were among the most used frameworks adopted by teams in the BigDEAL Challenge 2022, which focused on energy load and peak forecasting [32]. These models have also been successfully used in water availability forecasting [33]. Although studies on foundation models for energy forecasting are beginning to emerge and have shown promising results [34], they remain in an early stage, and their complexity and need for domain-specific adaptation place them beyond the scope of this work.

In this work, we focus on predicting energy demand and water availability over a two-day-ahead horizon with a 5 min and 1 h resolution, respectively. This requires generating tens or hundreds of outputs (the target variable at each time step of the two days ahead), making it a complex multi-output regression task. The forecasters are trained using LGBM. However, LGBM does not natively provide tools for multi-output regression.

Deep learning models naturally support it, making them particularly well suited for multi-step-ahead time-series forecasting [35]. However, designing and tuning deep neural models typically requires substantial expertise and extensive computational resources, and, more importantly, they generally rely on large amounts of training data to achieve good generalization performance [36]. For these reasons, deep learning approaches do not align with the lightweight and off-the-shelf methodology targeted in this work. Consequently, we opt to adapt LGBM to the multi-output setting instead. Traditional approaches to multi-output predictions are based on training multiple independent or chained models [37]. Some recent works, e.g., [38,39], have explored extensions of the gradient boosted trees framework to handle multiple outputs. We address the lack of native multivariate regression support by training a set of independent models for the water flow forecaster, while for the energy demand forecasts, due to the higher resolution of this time series, we design a more efficient strategy that exploits the LGBM capability of handling missing data to learn a global model that is sufficiently accurate across the entire considered time horizon.

In summary, among the wide variety of ML-based methods applied to demand forecasting, deep learning and gradient boosting models have proven highly effective for medium- and long-term forecasting, with the former being natively well-suited for multi-step-ahead forecasts. However, deep-learning forecasters, although powerful, typically require large training datasets, extensive hyperparameter tuning, and significant computational resources, and do not necessarily outperform gradient boosting [31]. In contrast, our use of LGBM provides a lightweight, data-efficient, and easily deployable forecasting solution that can be quite easily adapted to multi-step-ahead predictions, making it particularly suitable for small utilities with limited historical data.

3. Materials and Methods

3.1. Description of the Small-Scale HPP Under Study

We describe a real-world application of MPC for the optimal management of water resources in the small-scale HPP managed by AEM, a utility distributing energy primarily to the municipalities of Massagno, Capriasca, and Isonne in Ticino, Switzerland. The price of the electrical energy paid by AEM to the cantonal electricity supplier is based on both the energy volume and the power transfer price that is proportional to the peak demand reached each month. AEM exploits its HPP primarily to reduce peak demand and thus contain the energy price.

The plant collects water from two rivers and accumulates it in a reservoir from which it can draw during peak demand periods as described in Figure 1. The maximum production of the power plant is 4 MW, achieved at a water flow rate of $2.11 \text{ m}^3/\text{s}$. The reservoir can store about $18,500 \text{ m}^3$ of usable water, based on the specified minimum and maximum reservoir water levels, corresponding to approximately 9.75 MWh of stored energy.

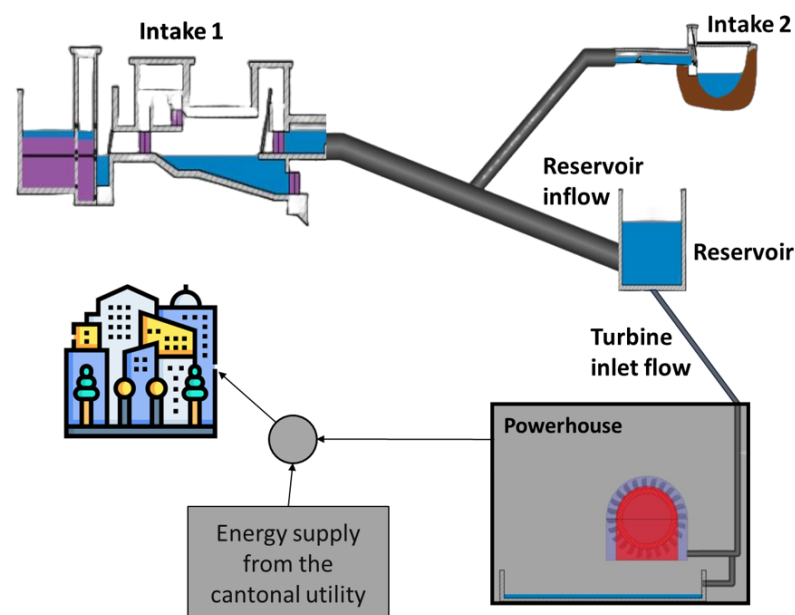


Figure 1. Simplified scheme of the AEM HPP.

The production scheduling of the HPP is a combination of real-time corrective control (that adjusts system operations in response to observed variations in demand) and operator decisions. The primary objective of the operators is to minimize the size of the monthly demand peak, which is usually observed during evenings. When the reservoir is full, the plant is set to reduce the monthly peak as much as possible. If the storage reservoir is not full, despite the goal of reducing peaks, the operators must also ensure that a sufficient water volume will be available for the potential future consumption spikes. The operators determine the target value of the daily peak based on their expertise. They consider the current and short-term water availability, weather conditions, and forecasts that influence both the demand and the future availability of water, the size of the peak already reached during the current month, and the expected seasonal trends of the energy demand. A real-time corrective controller then schedules the production based on the difference between the most recently observed demand and the target peak value. Outside peak hours, production is limited by the need to fill the reservoir. If there is no need to fill the reservoir, the plant produces electricity based on the available water flow: the water entering the system is immediately used by the HPP to generate electricity.

3.2. Power Production Scheduling

The HPP we model within the MPC framework is a plant without pumping facilities, meaning that its reservoir is filled by the inflow from rivers only. This type of storage system has an uncontrollable inflow rate, which can fill up the reservoir up to a maximum capacity (any excess water spills out when the storage is full). We now describe our modeling choices that are driven by requirement aspects: (i) the physical characteristics of the AEM HPP, (ii) operational practices currently used by HPP operators, and (iii) the need for a computationally lightweight formulation suitable for small utilities and widely accessible computational resources.

The production of the HPP is managed using an MPC control strategy that models the system's behavior of interest over a finite time horizon, optimizes decisions for the immediate time step, and simulates future steps to anticipate the system's trajectory. We define the duration and granularity of the horizon by an interval of time T , consisting of N steps, each of duration Δt . In our experiments, T covers 48 h starting from the current scheduling time t_0 , with granularity $\Delta t = 5$ min, for a total of $N = 576$ steps. The expected inflow rate $\hat{in}_t$ at each future step $t \in T$, expressed in m^3/s , as well as the energy demand $\hat{p}_t$ are predicted at t_0 by the forecasters described in the following Sections 3.3 and 3.4. The storage status is represented by the variable SOC_t , expressed in m^3 , which denotes the estimated volume of water available at the end of step $t \in T$. Given these characteristics, the discharging power of the HPP at each timeslot is controlled via the continuous variable pd_t as defined by the following set of equations:

$$z = \min wp \quad (1)$$

$$wp \geq \hat{p}_t - th - pd_t \quad \forall t \in T \quad (2)$$

$$pd_{min} \leq pd_t \leq pd_{max} \quad \forall t \in T \quad (3)$$

$$|pd_t - pd_{t-1}| \leq \Delta P \quad \forall t \in T \setminus \{0\} \quad (4)$$

$$SOC_{min} \leq SOC_t \leq SOC_{max} \quad \forall t \in T \quad (5)$$

$$SOC_t = SOC_{t-1} - f_d(SOC_{t-1}, pd_t, \Delta t) + \hat{in}_t \Delta t \quad \forall t \in T \setminus \{0\} \quad (6)$$

where wp represents the amount by which the power requested from the grid exceeds the threshold th , minimized over the 48 h control horizon. th represents the peak power recorded during the previous days of the same month, since lowering the peak below that point would not improve the power transfer price. Constraints in Equation (2) ensure that wp captures the maximum amount by which the net value is expected to exceed th at any time within T . Operational constraints in Equations (3) and (4) set the HPP to operate between given boundaries $[pd_{min}, pd_{max}]$, and they impose a maximum permissible power variation, denoted by ΔP . Constraints in Equation (5) ensure that the HPP reservoir operates within a given interval. Finally, constraints in Equation (6) determine how the reservoir changes according to the selected power and the expected inflows of the rivers. Function f_d is the discharge rate function that maps the power output pd_t and the reservoir status SOC_{t-1} to the amount of water consumed during a time slot of duration Δt .

For the AEM HPP:

- All powers ($\hat{p}_t$, th and pd_t) are expressed in MW.
- $[pd_{min}, pd_{max}]$ is set to $[0, 4.0]$ MW.
- ΔP is set to 4 MW; these are the current operational constraints of the equipment, which is capable of reaching the maximum production in less than 30 s.
- $[SOC_{min}, SOC_{max}]$ is set to $[4622, 23145]$ m^3 , corresponding to the usable storage volume of the reservoir. The lower bound enforces the minimum operational water level adopted by AEM to mitigate debris ingestion and protect the turbine, while the upper bound represents the maximum allowable storage capacity.

- The water discharge in m^3 is modeled using the linear function $f_d(pd_t, \Delta t) = 0.509 \cdot pd_t \cdot \Delta t$, where $0.509 \text{ m}^3/\text{s}$ represents the discharge rate at 1 MW. Practical experiments performed at multiple reservoir water levels confirmed that the relationship between power output and discharge can be reasonably approximated as linear within the normal operating range of the plant, making this model adequate for the proposed MPC formulation.

Post Optimization Settings

Based on input from HPP control experts, we implement simple rule-based post-optimization decisions. When the estimated demand is “low” (i.e., below a parametrized threshold of the observed past peak) and the reservoir is “nearly full” (i.e., it exceeds a predefined minimum storage threshold, denoted as sc_{min}), the HPP plant switches to self-consumption mode, preventing spillover while still generating some useful energy. In this mode, the plant adjusts its output to the river inflow rate, in a way that still allows for a slightly positive storage balance so that the reservoir can gradually reach SOC_{max} . The power output is computed using a convex combination:

$$r = (SOC_t - sc_{min}) / (1.0 - sc_{min})$$

$$pd_t = r \cdot \hat{in}_t / f_d(SOC_{t-1}, 1.0, \Delta t) + (1 - r) \cdot pd_{min}$$

where r is a weight between minimum power output and full inflow consumption. When the reservoir level is only slightly above the threshold sc_{min} , the plant operates close to its minimum power output (0 in our settings), thereby minimizing power generation to maximize inflow storage. As the reservoir level increases, the operating point transitions smoothly toward a self-consumption regime, eventually reaching a power output at which water consumption matches the river inflow.

3.3. Energy Demand Forecasts

The goal of the energy demand predictive model is to forecast the utility’s energy demand over a 48 h control horizon, with a resolution of 5 min. At the current time step t_0 , the model generates forecasts $\hat{p}_{t_h}$ of the energy demand p_{t_h} for each prediction time step $t_h = t_0 + h\Delta t$, with $h = 1, \dots, 576$. Model inputs x_{t_h} include calendar features, past energy demand features and weather forecasts.

- Calendar features: they are both categorical (namely, day of the week, month, and a binary feature indicating holidays) and numerical (the sine and cosine of the day of the year and of the time of the day, to model circularity). These variables allow us to capture seasonal patterns.
- Past energy demand features: these include short-term lags (from 5 to 30 min behind with a step of 5 min) and medium-term lags of 1 day, 2 days, and 1 week. For medium-term lags, also the average, minimum, and maximum values over two time windows of 10 and 30 min are included.
- Weather forecasts: these include temperature, precipitation, and the average global horizontal irradiance (GHI), computed over 1 h.

The features identified in this work are traditional in the context of energy forecasting [40]. We include a relatively large set of features covering a time horizon that is meaningful for this application. We do not apply explicit feature selection methods because LGBM inherently performs feature selection during training. While different or more extensive feature sets could be explored, determining an optimal set is not the goal of this study, given the dataset’s specificity and the expectation that any improvement in predictive accuracy would be limited, this being a well-explored domain, and, thus, not necessarily relevant for the control objective.

Given that LGBM does not inherently support multi-output regression and since for this task we would need 576 outputs, we opt against constructing so many separate models. We adopt a strategy that learns only two single-output models. The short-term model is trained specifically for $h = 1$, as this forecast directly determines the only MPC decision that is actually implemented—subsequent decisions are revised as new information becomes available. The medium-term model is used to provide forecasts for $h = 2, \dots, 576$. Of the aforementioned predictor features x_{t_h} , not all are available at t_0 for every value of h . For instance, short-term lags are not available for $h > 6$. The values of these unavailable inputs are left missing, exploiting the capability of the LGBM algorithm to deal with missing values. As for the weather forecasts, the most recent forecasts available at time t_0 are used.

In the training phase, to account for all possible control horizons, h , we adopt the following strategy. From each historical instance (x_t, p_t) relative to some time t , we derive multiple training instances by considering different values of $t_0 = t - h\Delta t$. For each of them, we set as missing all predictors that are not available at the sampled time t_0 . We also add h as an additional feature of the model. While in principle we could consider 576 different instances for each t , we choose to sample only 30 values of h , without replacement, for each historical instance, as keeping all of them would exceedingly increase the training dataset without adding much useful information to learn from. Sampling follows the non-uniform distribution $\mathbb{P}(h) \propto 1/h$, where h is the number of hours corresponding to h , rounded up to the nearest integer. During training, this approach gives more weight to shorter time horizons, for which more accurate predictions are both feasible and more valuable.

As a baseline for comparison with the LGBM-based demand forecaster, we adopt the statistical model introduced in [27], hereafter referred to as *Stat*. This model assumes daily periodicity in power demand and estimated values as the average demand observed at the same time of day over a sliding window, including the past 10 days. For the first hour after t_0 , this 10-day average is combined with the demand p_{t_0} at the current time t_0 by a weighted average assigning to p_{t_0} a weight which decays with t as $1 - \frac{t-t_0}{1h}$. To account for the weather conditions that may affect the demand, observations are classified into five categories, based on the observed temperature and irradiation conditions. For predicting the demand at a given future time t , only observations in the same weather category are averaged.

3.4. Inflow Forecasts

The goal of the inflow forecaster is to predict the reservoir inflow rate in_{t_h} for the forecast horizons $t_h = t_0 + h\Delta t$, $h = 1, \dots, 48$, with $\Delta t = 1$ h. Since the output size is moderate, due to the lower resolution, we choose to use 48 different LGBM models, each of which produces the forecast $\hat{in}_{t_h}$ for h hours ahead. The forecaster is not directly trained to predict the future inflow rate, but the inflow rate variation $in_{t_h} - in_{t_0}$. The models use as inputs date-time features, past inflow features, and precipitation forecasts. Date-time features are the month and the sine and cosine of the day of the year and of the time of the day. Past inflow features include the last 6 observed values before time t_0 (i.e., in_t for $t = t_0 - \Delta t, \dots, t_0 - 6\Delta t$) and the average, minimum, and maximum values of in_t over a 6- and a 12 h window before time t_0 . Precipitation forecasts include the predicted rainfall at the reservoir location, as well as the average rainfall forecasted at different positions along the rivers feeding into the reservoir. We use both the precipitation forecasts relative to 1, 2, and 3 h before t_h and the cumulative forecasted rain in the 2, 4, 6, 12, 18, 24, and 36 h before t_h .

3.5. Experimental Setup

The analysis presented in the next Section is based on the data provided by the utility from December 2018 to August 2023. Due to technical problems with the utility data management system, data were lost for four long periods between December 2021 and March 2023, covering about 18 months. This affects the quality of the demand forecaster in that period. However, data from a total of 1184 days spread over 46 months are still available. We selected four testing periods that are removed from the training data of the LGBM demand and flow rate forecasters: period 04/2021 (from 27 March to 13 April 2021), period 08/2022 (from 1 to 28 August 2022), period 05/2023 (from 1 to 28 May 2023), and period 08/2023 (from 1 to 28 August 2023). The selection of test periods that are non-consecutive to training ones is thus motivated. We aim to evaluate the model across diverse conditions, not only in different seasons but also in different years. This choice would be fully justified if similar energy demand patterns could be assumed before and after the testing period. In reality, however, demand patterns may evolve over time due to technological changes, such as increased PV penetration. Such changes could reduce model validity in the long term and may not necessarily improve forecast performance in the tested periods; in fact, they could even degrade it. In both cases, we expect the results to be slightly biased. The test on the last period (August 2023), taken at the end of the dataset, represents a good proxy for a realistic application scenario.

Over the four selected periods, we compare:

1. The LGBM demand forecaster with the *Stat* model from [27];
2. The LGBM water flow, forecaster with the naive forecaster hereafter referred to as *const*, which assumes a flow rate constantly equal to the value observed at t_0 ;
3. The MPC power production decisions in three input configurations with those historically executed by the HPP operator. The three configurations considered are:
 - (a) The unrealistic configuration, hereafter referred to as *oracle*, where the MPC knows the actual water flow rate or energy demand, which serves as an upper bound for the MPC performance;
 - (b) The *LGBM* configuration, when the MPC inputs are the demand and the water flow rate predicted by the LGBM models;
 - (c) The *Stat* configuration when the demand is predicted by the *Stat* model and the flow rate by the naive *const* forecaster.

The data provided by the utility include the historical energy demand time series (p_t), the historical HPP power production (pd_t^*), the reservoir level, and its inflow rate. The water availability SOC_t at a given time t is obtained from a cubic relation with the reservoir level derived from reservoir recovery measures provided by the utility.

Weather forecasts were purchased from the Swiss weather forecast provider *meteoblue* (<https://content.meteoblue.com/en>, accessed: 6 February 2026) and recorded by the utility only for the area it supplies. Because historical *meteoblue* forecasts were not available for purchase for the catchment area feeding the reservoir, we complemented the dataset with the free and openly accessible forecasts provided by *Open-Meteo* (<https://open-meteo.com/>, accessed: 6 February 2026), which ensured adequate spatial coverage of the relevant basin. This choice does not affect modeling consistency, as the two meteorological data sources are used as inputs to different models: *meteoblue* forecasts feed the demand forecaster, whereas *Open-Meteo* forecasts are used for inflow prediction. Unfortunately, *Open-Meteo* does not provide a history of past forecasts for years 2021–2023 but only the most recent forecast available for a given time. As a result, our evaluation may overestimate the performance of both the LGBM predictors and MPC since we are partially relying on weather forecasts of higher quality than those that would have actually been available at any given time t_0 .

We observe inconsistencies between the measured inflow rate and the net inflow derived from the reservoir variation, $SOC_t - SOC_{t-\Delta t}$, and the water consumed to produce pd_t^* , that is $f_d(pd_t^*, \Delta t)$. Therefore, due to the higher reliability of power production and reservoir level records, in the following analyses, the measured inflow rate is replaced by its estimate in_t^* :

$$in_t^* = \frac{(SOC_t - SOC_{t-1}) + f_d(pd_t^*, \Delta t)}{\Delta t} \quad (7)$$

The high variability of the resulting signal (see Figure 2) is attributed to measurement noise and is smoothed for training using a moving average with a 4 h window.

4. Results

4.1. Forecasting Performance

In this section, we present the results of the assessments described in Section 3.5. Figure 3A shows examples of the LGBM and *Stat* forecasts, while Figure 3B,C compare the performance of the two approaches across each of the 576 time horizons considered, using as performance metrics the Root Mean Square Error (RMSE) and the symmetric Mean Absolute Percentage Error (sMAPE). We notice from Figure 3A that LGBM systematically overestimates demand valleys during the 2023 period. This mismatch is mainly due to the increased penetration of photovoltaic (PV) installations, which has altered the demand patterns compared to previous years. The LGBM forecaster used for periods 05/2023 and 08/2023 is trained by assigning an increased weight to recent observations, but this is insufficient to solve the PV valleys issue, both because 2023 data is limited, as many entries were not properly recorded by the utility, and because total installed PV capacity almost doubled in the Ticino region between 2021 and 2023 [41]. However, since this issue affects only demand valleys and not the peaks, it is not critical in the context of this paper. The metrics shown in Figure 3B,C are computed considering only peak hours. The black lines focus on daily peaks, identified as the time steps at which the demand exceeds the median demand for that day. In addition, to better account for monthly rather than daily peaks, the green line shows the same metrics computed on monthly peak samples. Specifically, for each of the four periods—each representing a calendar month—we identify peak events by calculating the 0.8 quantile of the demand distribution and retaining only the samples exceeding this threshold. Such comparisons highlight that, by accounting for the predicted weather conditions, LGBM achieves better accuracy starting already from a 20 min ahead forecast. LGBM outperforms *Stat* also in the 5 min ahead forecast since a short-term model specifically trained for this prediction horizon is used in this case. We also observe that in both cases, the error increases as the focus narrows to higher monthly peaks, and this effect is more pronounced for the *Stat* model. Further analyses of this model, including feature importance and an ablation study, are reported in Appendix A.

Figure 2A shows examples of the LGBM and constant (*Const*) forecasts, and Figure 2B,C the RMSE and sMAPE of the two approaches across all considered time horizons. We notice from Figure 2A that the predicted signal appears significantly smoother than the observed data, and that the LGBM flow rate forecaster consistently outperforms the constant flow baseline, except in very short-term forecasts, where their performance is comparable.

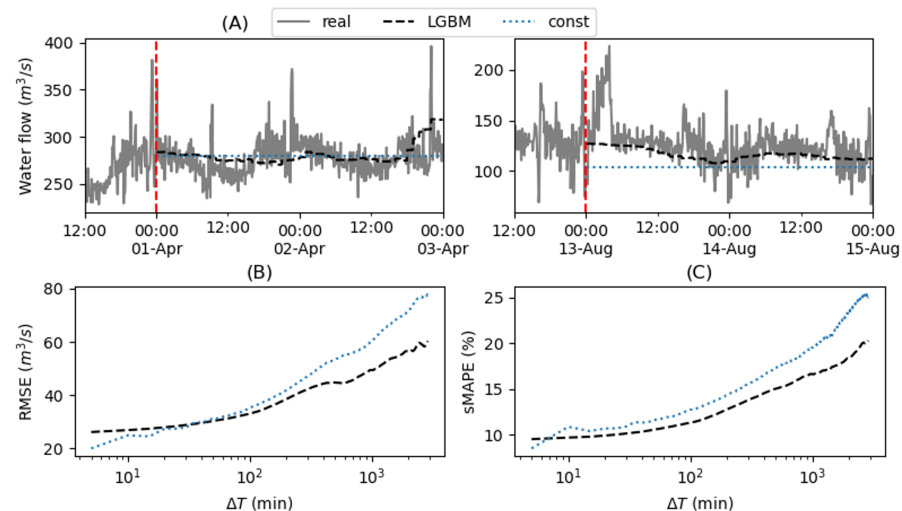


Figure 2. (A) Example of water flow forecasts. The vertical red line indicates the time t_0 at which the 48-h-ahead forecast is issued. (B) RMSE and (C) sMAPE, resulting from the comparison of the LGBM and constant-model inflow predictions against the inflow estimated from the observed reservoir-level variation (Equation (7)).

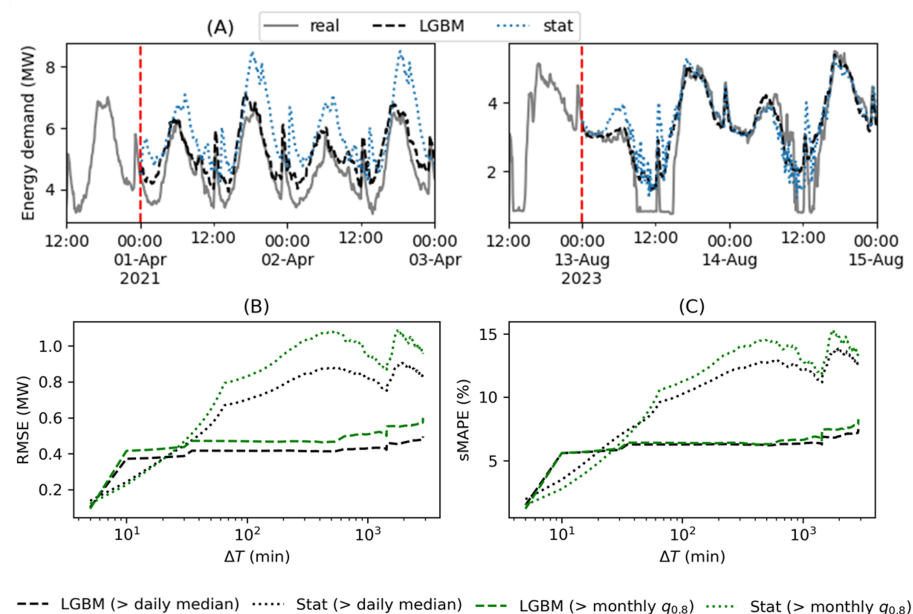


Figure 3. (A) Example of energy demand forecasts. The vertical red line indicates the time t_0 at which the 48-h-ahead forecast is issued. (B) RMSE and (C) sMAPE resulting from the comparison of the LGBM and Stat demand predictions against the observed demand. These metrics are computed on all observations above the daily median (black) and on the subset of observations where the demand exceeds the 0.8 quantile of the monthly demand.

4.2. Peak-Shaving Performance of MPC

Figure 4 and Table 1 analyze the MPC performance across the four simulation periods. For each day, the maximum net energy demand (daily peak) is calculated. Figure 4 compares the distribution of daily peaks, contrasting the operator's performance with that of the MPC under the three input configurations considered. For each period, we compute the mean and the maximum values of the daily peaks and perform a paired t -test for the mean difference in the peak values achieved by the MPC and the operator. Since each analyzed period corresponds to a specific month, the maximum daily peak value directly determines the power transfer price and, consequently, the cost of the energy purchased by

the utility. To further quantify the ability of the MPC to flatten the load curve, we compute the load fluctuation index (LFI) as in Equation (8):

$$LFI = \frac{1}{n_{\pi}} \sum_t \left| \frac{nd_t - \overline{nd}_{\pi}}{\overline{nd}_{\pi}} \right| \quad (8)$$

where nd_t is the net demand at time t , n_{π} is the number of time steps in period π , and $\overline{nd}_{\pi}$ is the average net demand over π . The LFI measures how much the net energy demand deviates from its average value over a given period: the higher the LFI, the more pronounced the peaks and valleys, whereas a lower LFI indicates a flatter and more stable demand profile. This metric is computed as the average relative deviation of each time step from the period's mean demand as shown in Equation (8).

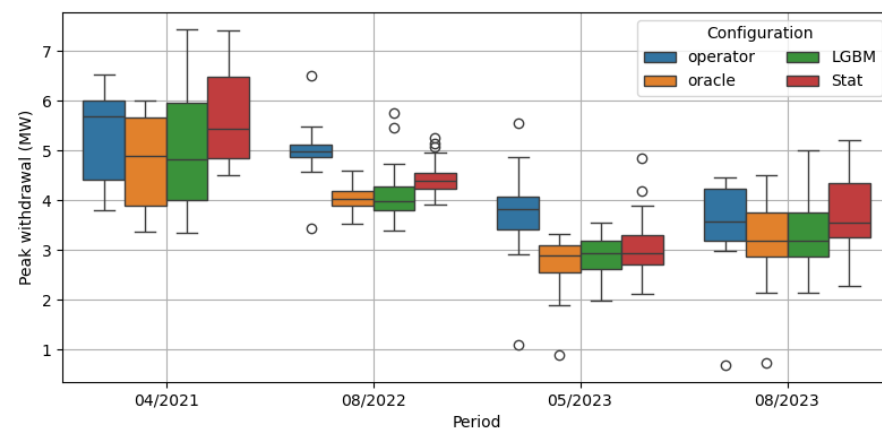


Figure 4. Boxplots of the daily peaks distribution in the different periods and under different settings.

Table 1. Peak-shaving performance achieved by the operator or the MPC in the *oracle* and LGBM and *Stat* configurations. *Mean* and *max* refer to the average and maximum values of daily peaks, respectively; *p-val* indicates the significance of the mean difference in peak values achieved by the operator and the MPC.

	LFI	Max	Mean	<i>p-Val</i>	LFI	Max	Mean	<i>p-Val</i>
04/2021					08/2022			
operator	0.257	6.53	5.30	-	0.207	6.50	5.03	-
oracle	0.203	6.00	4.79	4×10^{-3}	0.100	4.61	4.02	5×10^{-9}
LGBM	0.202	7.45	5.04	0.11	0.103	5.76	4.11	7×10^{-14}
Stat	0.222	7.41	5.70	0.898	0.121	5.26	4.45	1×10^{-6}
05/2023					08/2023			
operator	0.564	5.55	3.77	-	0.407	4.47	3.61	-
oracle	0.351	3.32	2.72	9×10^{-10}	0.267	4.51	3.22	3×10^{-6}
LGBM	0.394	3.55	2.83	2×10^{-12}	0.300	5.00	3.38	0.010
Stat	0.415	4.85	3.07	1×10^{-7}	0.315	5.21	3.71	0.863

5. Discussion

5.1. Discussion of Forecasting Performance

The analysis of the LGBM energy-demand forecasters in Figure 3A–C shows satisfactory performance around demand peaks: although errors increase when the evaluation is restricted to higher monthly peaks, they remain consistently—and often substantially—lower than those of the baseline. By contrast, the LGBM model systematically overestimates demand valleys in 2023, largely due to the rapid increase in PV penetration

in the region, where installed PV capacity nearly doubled between 2021 and 2023, significantly altering local demand patterns. In a similar case currently under study, we observe that including installed PV capacity as an input feature—when available—substantially improves the model’s ability to reproduce PV-induced fluctuations. However, this is deemed unnecessary in the present context since only peak prediction influences the scheduler’s decisions.

The ablation study in Appendix A further shows that all feature sets considered contribute positively to performance, consistent with the prior literature such as [31], highlighting the value of considering rich sets of covariates.

5.2. Discussion of MPC Peak-Shaving Performance

The following discussion focuses on the operational behavior of the MPC scheduler, assuming the forecasting performance characterized above. From the boxplot and from the average and maximum peak values, we observe that the MPC peaks are smaller, on average, than those achieved by the operator, and its ability to flatten the load curve, as measured by the LFI, is larger across all periods and in all input configurations, except occasionally in the *Stat* configuration which generally performs worse than the others. The *p*-values in Table 1 show that the mean MPC difference with the operator peaks is almost always significant (*p*-value < 0.01). However, the maximum peak of the period can sometimes be higher for the MPC. This is evident, for example, in period 04/2021 for both *LGBM* and *Stat*, and in period 08/2023 for all configurations. These higher peaks typically occur due to water shortages during peak times, caused by overestimated inflow or underestimated demand. Further analysis of the cases in which the MPC produces higher daily peaks than the operator is provided in Appendix B. This highlights how the operator’s more conservative approach can sometimes be preferable. However, in the extremely dry period of 08/2022, such a conservative approach resulted in the operator deciding to keep the facility at minimal output, thus giving up on mitigating the peaks. In contrast, the automatic MPC concentrated water use, though limited, during peak hours, significantly reducing them. Figure 5 shows an example of MPC simulation results from this period. The first two subfigures display the actual energy demand and water flow, along with the *LGBM* forecasts made at t_0 for the next 48 h. The third subfigure compares the power output decisions of the operator and MPC; the fourth shows the resulting net energy demand requested from the grid, obtained by subtracting the scheduled production from the actual demand (first plot, black line); and the fifth illustrates the water volume available in the reservoir. It can be noticed that the MPC exploits the available water mostly during peaks, thus managing to flatten them much more than the operator does, since, given the scarcity of water, the operator adopted a more conservative strategy. Even in this period, by blindly relying on forecasts, the deterministic MPC occasionally faces water shortages during peaks, leading to the two outliers in the *LGBM* boxplot of Figure 4. While there is still room for improvement, the MPC strategy achieves better performance than the operator’s conservative approach on average across the evaluation period as measured by reduced peak grid withdrawals and the associated cost-based performance indicators. In future work, this performance could be further improved by adopting stochastic optimization strategies that explicitly account for forecast uncertainty at the control level and yield more robust scheduling decisions.

The Swiss national transmission tariff (Swissgrid) consists primarily of a working tariff, charged in cents per kWh based on measured end-user consumption, and a power tariff, which is calculated from the average monthly peak load. The power tariff generates costs that increase linearly with the magnitude of the monthly peak, according to the relation

$$\text{cost} = (\text{power tariff}) \times (\text{monthly peak}).$$

According to the tariffs and remuneration rates for the transmission grid published by Swissgrid (<https://www.swissgrid.ch/en/home/customers/topics/tariffs.html>, accessed: 6 February 2026), the power tariff in kCHF/MW per year was 46.38 in 2024, 47.22 in 2025, and is 74.28 in 2026. The difference in peak load between the operator and the scheduler observed across the four test periods considered in this work ranged from -1 MW to $+2$ MW, with an average of $+0.3$ MW. Assuming this value as a reference for a rough estimate of average yearly savings, we would obtain approximately 14 kCHF/year in 2024 and 2025 and 22 kCHF/year in 2026. This amount should be interpreted in the context of the relatively small size of the plant. Beyond the transmission power tariff, the working tariff and additional components, such as hydropower production costs, also contribute to the total cost of electricity for the local utility. These components are not included in our calculation, as they are expected to have a comparatively minor impact on the incremental savings associated specifically with peak-load reduction.

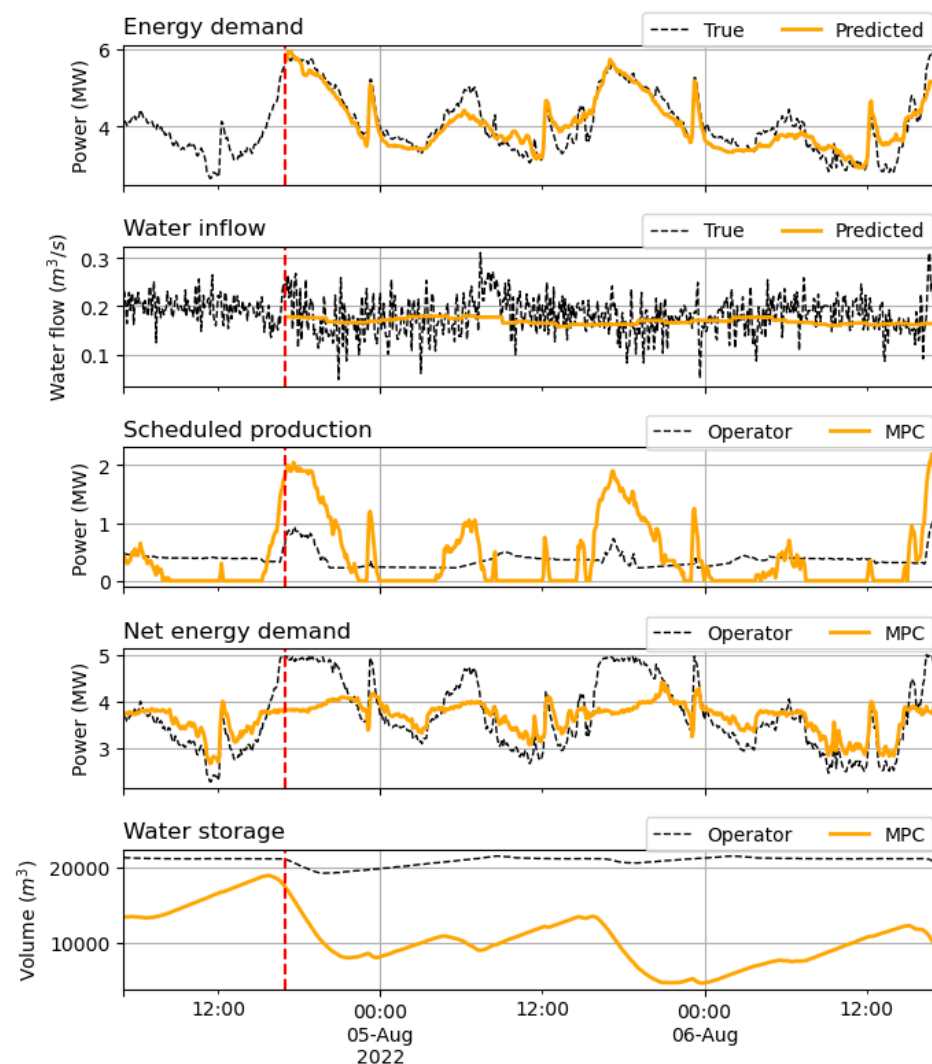


Figure 5. Dashboard with the MPC output. Black lines show real data—observed values and operator-actuated production decisions—while colored lines show model-generated outputs (forecasts, scheduler decisions, and their effects on net demand and reservoir level). The red vertical line indicates the time at which the forecasts in the first two subplots (top) are generated. The bottom subplot shows the results obtained with decisions refreshed every 5 min.

5.3. Sensitivity and Robustness

To assess robustness, we analyze the sensitivity of the MPC performance to variations in key parameters. Moderate changes in the ramping limit ΔP and in the storage bounds lead to only marginal variations in peak shaving performance. In contrast, the discharge coefficient proves to be more sensitive: its underestimation may result in excessive reservoir depletion over time, whereas overestimation can lead to overly conservative operation and underutilization of available water resources. The impact of wrong estimations is mitigated in an MPC framework, as real system measurements and updated forecasts are incorporated at each control step, reducing sensitivity to conservative or imperfect parameter settings in real-world operation. These results indicate that the controller's behavior is primarily driven by the accuracy of demand and inflow forecasts rather than by fine-tuning of most physical parameters. Demand and inflow forecasts are subject to uncertainty, but we still adopt a deterministic MPC formulation. The controller operates in a receding-horizon framework with frequent re-optimization, incorporating updated measurements of demand and reservoir state at each control step. This feedback mechanism substantially mitigates the impact of forecast errors. Moreover, the peak-shaving objective and the relatively slow dynamics of the water storage system make the controller tolerant to forecast inaccuracies. Stochastic or robust MPC approaches could be considered; their increased modeling and computational complexity would hinder practical adoption by small utilities.

5.4. Computational Performance

The MPC optimization problem is formulated as a linear program and solved using a commercial solver (Gurobi). All simulations are executed on a standard workstation equipped with an Intel i5 CPU, 2.00 GHz and 32 GB of RAM. For a 48 h horizon with 576 time steps, the average solution time per step is below 1 s, including model building, while worst-case solution times remain below 3 s, well within the 5 min control interval required for real-world operation. Preliminary experiments using non-commercial solvers (HiGHS) also demonstrate acceptable computational performance, with solution times always contained in 10 s. These results confirm that the proposed MPC formulation is computationally lightweight and suitable for real-time deployment on widely available hardware, requiring little to no investment in commercial optimization software.

6. Conclusions

This article explores the application of MPC to optimize production scheduling in small HPPs. The primary objective is to leverage the plant's flexibility to reduce peak electricity demand from the grid, thereby lowering energy costs. The findings demonstrate that automated scheduling can enhance operator decision-making, offering support that is particularly valuable during periods of water scarcity, when effective scheduling is most critical.

This work originated from a need observed across multiple local utilities, and likely shared by similar operators worldwide. We adopted a well-established approach to build predictive models using the relatively small datasets typically available to individual utilities, without aiming for state-of-the-art forecasting performance, the goal being to demonstrate the feasibility and effectiveness of such methods in combination with MPC in this real-world context. Results show clear advantages over the current reactive control and operator-based decisions; however, compared to an oracle scenario with perfect knowledge of demand and inflow, forecast inaccuracies can significantly increase daily and monthly peaks, and the controller may exhibit oversensitivity to prediction errors. Therefore, im-

proving prediction models and controller robustness remain two important directions for future work.

Building on these findings, future research should focus on improving forecast accuracy, e.g., by expanding training datasets through the aggregation of records from similar utilities and hydroelectric reservoirs and leveraging recent advances in foundational models for energy forecasting [34]; devising a model updating framework able to adapt models to reflect emerging demand patterns driven by technological and social changes, such as increased photovoltaic penetration; and upgrading the MPC to incorporate probabilistic forecasts to enable more robust decision-making by explicitly accounting for prediction uncertainty.

Author Contributions: Conceptualization, F.M. and M.S.; methodology, F.M., M.S. and L.Z.; software, M.D., L.Z. and V.G.; validation, L.Z., F.M., M.D. and V.G.; formal analysis, F.M., M.S. and L.Z.; investigation, F.M., L.Z. and M.D.; data curation, L.Z., F.M. and M.D.; writing—original draft preparation, F.M., M.S. and L.Z.; writing—review and editing, F.M., M.S. and L.Z.; visualization, F.M.; supervision, F.M. and M.S.; project administration, M.S.; funding acquisition, M.S. All authors have read and agreed to the published version of the manuscript.

Funding: This project has received funding in the framework of the joint programming initiative ERA-Net Smart Energy Systems' focus initiative Digital Transformation for the Energy Transition, with support from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 883973.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Additional Results on Energy Demand Predictive Model

Feature importances for short-term and long-term models

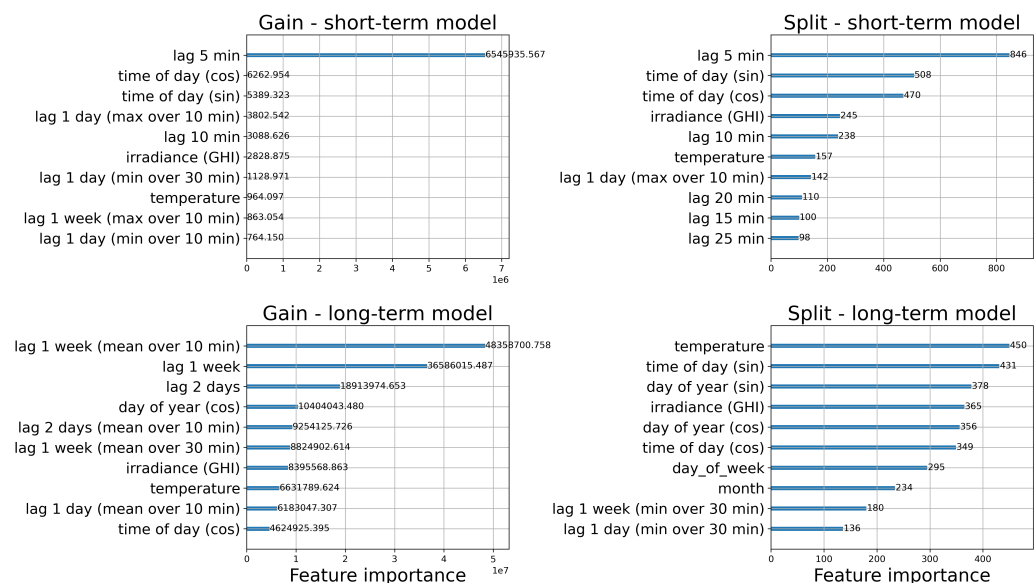


Figure A1. Feature importance comparison for the short-term (top) and medium-term (bottom) energy-demand models. Results are shown for both Gain (left) and Split (right) importance metrics. Feature labels correspond to the categories detailed in Section 3.3.

To assess the relative influence of different input variables on the energy demand forecasts, we analyze feature importance using *Gain*, which measures the cumulative

reduction in training loss, and *Split*, which counts how often a feature is used to partition data. The results, reported in Figure A1, show a clear shift in predictive strategy: the short-term model is dominated by the 5 min lag feature, indicating that short-term temporal dependency is the primary driver for 1-step ahead forecasts. In contrast, the medium-term model relies more heavily on weather variables, past demand lags for daily and weekly cycles, and calendar features for yearly trends.

These findings are further validated by the ablation study summarized in Tables A1 and A2. We systematically remove one feature category, described in Section 3.3, and re-evaluate the model, measuring performance for both the 1-step ahead (short-term model) and the full 48 h horizon (medium-term model) using RMSE and sMAPE. As in our main evaluation, these metrics are restricted to high-demand periods to test model robustness during critical regimes. Specifically, we compute the results for time steps where demand exceeds the daily median and for monthly peaks, defined by demand exceeding the 0.8 monthly quantile. For the 1-step ahead forecast, past demand features are crucial: their removal leads to a significant performance drop, as the model loses the immediate context of the preceding 5 min interval. In contrast, the 48 h horizon is most sensitive to the loss of weather data. Without temperature and irradiance inputs, the model struggles to capture the amplitude of peak events, which is reflected in the sharp RMSE increase during monthly peak periods. Finally, *calendar* features provide a modest but robust improvement for both the 5 min and 48 h ahead forecasts.

Table A1. Ablation study results across different feature groups. Values represent the Root Mean Squared Error (RMSE) and symmetric Mean Absolute Percentage Error (sMAPE) for the short-term model (5 min ahead forecast).

	RMSE		sMAPE	
	>Daily Median	>Monthly $q_{0.8}$	>Daily Median (%)	>Monthly $q_{0.8}$
All features	0.106	0.092	1.5%	1.2%
No past demand features	0.405	0.420	7.0%	6.1%
No calendar features	0.113	0.103	1.6%	1.3%
No weather features	0.105	0.094	1.5%	1.2%

Table A2. Ablation study results across different feature groups. Values represent the Root Mean Squared Error (RMSE) and symmetric Mean Absolute Percentage Error (sMAPE) for the medium-term model (48 h horizon).

	RMSE		sMAPE	
	>Daily Median	>Monthly $q_{0.8}$	>Daily Median (%)	>Monthly $q_{0.8}$
All features	0.443	0.474	7.0%	6.6%
No past demand features	0.454	0.498	7.7%	7.2%
No calendar features	0.550	0.585	8.3%	7.8%
No weather features	0.698	0.857	10.7%	12.3%

Appendix B. Analysis of Cases Where MPC Produces Higher Peak Loads

Table A3 reports a summary of all the days across the four testing periods in which the maximum daily peak produced by the MPC exceeds the corresponding peak scheduled by the operator. These failure cases are relatively infrequent, but they highlight important failure modes that help explain the limitations of the controller when forecast errors are significant. The table details, for each representative period, the number of occurrences

together with the average and maximum differences between the MPC peak and the operator's peak.

Table A3. Analysis of cases where a daily peak achieved by the MPC with LGBM forecasts is higher than the operator.

Metric	04/2021	08/2022	05/2023	08/2023
Total number of cases	16	28	28	28
Number of cases MPC peak is higher	6	0	1	2
Average difference	0.580	–	0.078	0.216
Maximum difference	1.023	–	0.078	0.259

Figures A2 and A3 show the MPC output dashboard over two periods, illustrating two representative examples of failure driven by different underlying causes.

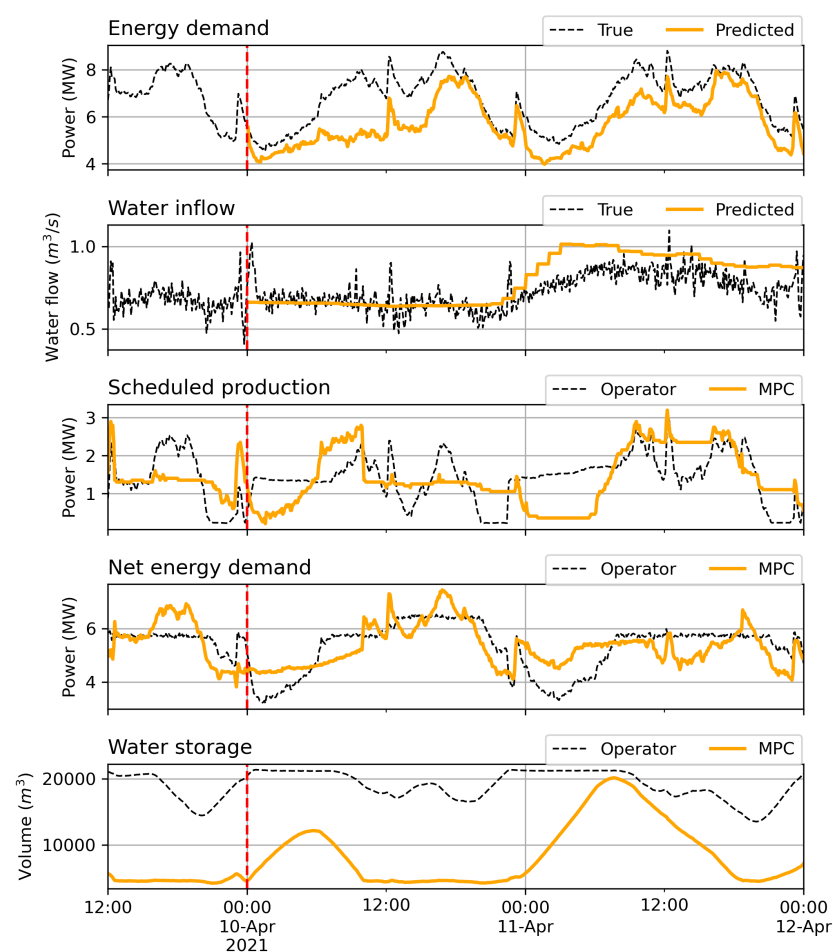


Figure A2. Predictions and scheduling decisions from 9 April 2021 at 12:00 to 11 April at 24:00. Between 12:00 and 20:00 on 10 April, the net energy-demand peaks reached by the MPC (solid orange line) exceed those of the operator (dashed black line).

Figure A2 illustrates a case in which the MPC produces a higher maximum peak because it initially underestimates the upcoming energy demand. As shown in the top panel, the predicted load remains significantly below the true demand for an extended period. Although the 5 min-ahead forecast is quite accurate, the scheduler only perceives short, isolated peaks and therefore assumes that the available water will be sufficient, dispatching production accordingly. As a result, the controller depletes the reservoir before

the unforeseen high peak later in the day is reached. This behavior reflects a failure of anticipatory scheduling directly caused by the underestimation of the energy demand.

Figure A3 illustrates a case in which the MPC behaves suboptimally because it overestimates the future water inflow. The inflow forecaster predicts substantially more water entering the reservoir than is actually observed, particularly between 9:00 and 15:00 on 28 August, when a sudden drop in inflow occurs. Owing to this overly optimistic forecast, the MPC schedules production under the assumption that reservoir water levels will be replenished more rapidly. As shown in the lower panels, this leads the controller to deplete the available water too early. When the expected inflow fails to materialize, the controller is left with insufficient water to counter the subsequent demand peak, ultimately producing a high peak. However, the operator also appears to have been unable to anticipate the inflow drop, and the MPC peak therefore only slightly exceeds the operator's peak.

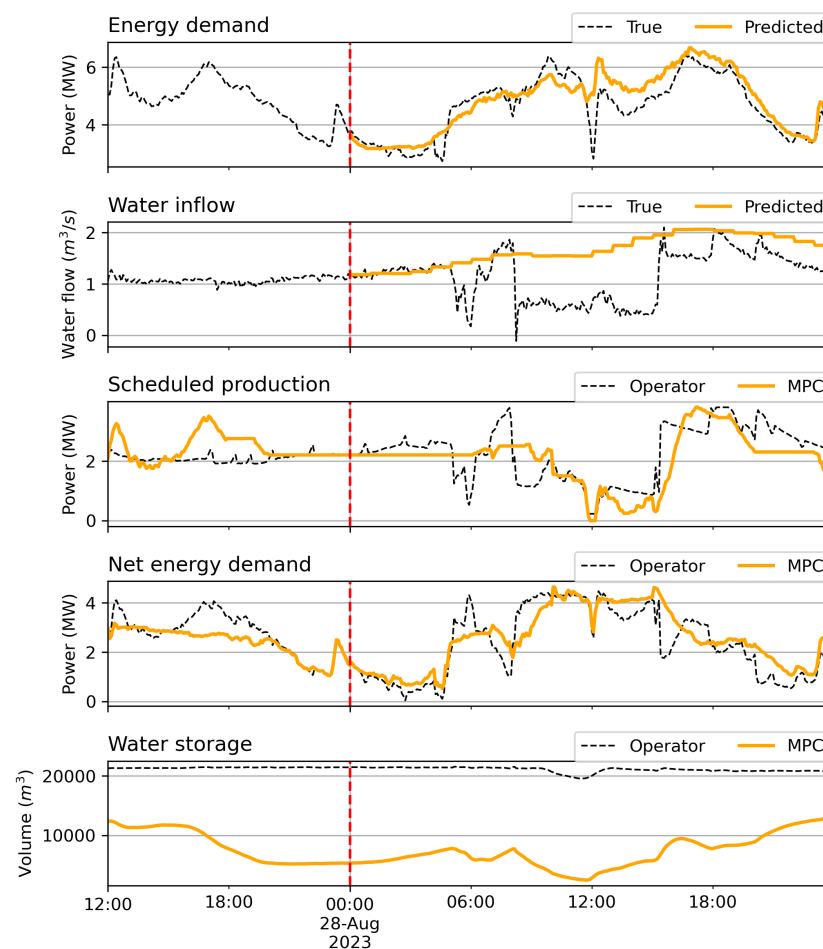


Figure A3. Predictions and scheduling decisions from 27 August 2023 at 12:00 to 28 August at 24:00. Between 9:00 and 15:00 on 28 August, the net energy-demand peaks reached by the MPC (solid orange line) exceed those of the operator (dashed black line).

These two examples highlight the critical role of accurate energy-demand and inflow forecasts in achieving consistently effective MPC scheduling. The underestimation of demand and overestimation of inflow can cause the controller to approach critical periods with inadequate water storage, which results in higher peaks. In light of this evidence, future work should aim to improve both the forecasting components and the MPC's robustness to forecast uncertainty.

References

1. Abdelfattah, A.I.; Shaaban, M.F.; Osman, A.H.; Ali, A. Optimal Management of Seasonal Pumped Hydro Storage System for Peak Shaving. *Sustainability* **2023**, *15*, 11973. [\[CrossRef\]](#)
2. Chua, K.H.; Bong, H.L.; Lim, Y.S.; Wong, J.; Wang, L. The state-of-the-arts of peak shaving technologies: A review. In Proceedings of the 2020 IEEE International Conference on Smart Grid and Clean Energy Technologies (ICSGCE), Sarawak, Malaysia 4–7 October 2020; pp. 162–166.
3. Liao, S.; Liu, Z.; Liu, B.; Cheng, C.; Wu, X.; Zhao, Z. Daily peak shaving operation of cascade hydropower stations with sensitive hydraulic connections considering water delay time. *Renew. Energy* **2021**, *169*, 970–981. [\[CrossRef\]](#)
4. Xu, Z.; Liu, C.; Liu, Q.; Xu, L.; Mo, L.; Zhang, Y. Optimal Peak-Shaving Dispatching of Hydropower Station in Response to Long-Term Load Demand of Power Grid. *Processes* **2025**, *13*, 489. [\[CrossRef\]](#)
5. Villeneuve, Y.; Séguin, S.; Chehri, A. Ai-based scheduling models, optimization, and prediction for hydropower generation: Opportunities, issues, and future directions. *Energies* **2023**, *16*, 3335. [\[CrossRef\]](#)
6. Wu, X.Y.; Cheng, C.T.; Shen, J.J.; Luo, B.; Liao, S.L.; Li, G. A multi-objective short term hydropower scheduling model for peak shaving. *Int. J. Electr. Power Energy Syst.* **2015**, *68*, 278–293. [\[CrossRef\]](#)
7. Zhao, H.; Liao, S.; Fang, Z.; Liu, B.; Ma, X.; Lu, J. Short-term peak-shaving operation of “N-reservoirs and multicascade” large-scale hydropower systems based on a decomposition-iteration strategy. *Energy* **2024**, *288*, 129834. [\[CrossRef\]](#)
8. Wang, Z.; Li, Y.; Wu, F.; Wu, J.; Shi, L.; Lin, K. Multi-objective day-ahead scheduling of cascade hydropower-photovoltaic complementary system with pumping installation. *Energy* **2024**, *290*, 130258. [\[CrossRef\]](#)
9. Sofi, M.S.; Bhat, S.U. Small hydropower nexus: Debunking myths and assessing real impacts. *Sustain. Nexus Forum* **2025**, *33*, 16. [\[CrossRef\]](#)
10. Grigoras, G.; Gârbea, R.; Neagu, B.C. Toward Smart SCADA Systems in the Hydropower Plants through Integrating Data Mining-Based Knowledge Discovery Modules. *Appl. Sci.* **2024**, *14*, 8228. [\[CrossRef\]](#)
11. Ruban, N.Y.; Suvorov, A.A.; Andreev, M.V.; Ufa, R.A.; Askarov, A.B.; Gusev, A.S.; Bhalja, B.R. Software and hardware decision support system for operators of electrical power systems. *IEEE Trans. Power Syst.* **2021**, *36*, 3840–3848. [\[CrossRef\]](#)
12. Babacan, O.; Ratnam, E.L.; Disfani, V.R.; Kleissl, J. Distributed energy storage system scheduling considering tariff structure, energy arbitrage and solar PV penetration. *Appl. Energy* **2017**, *205*, 1384–1393. [\[CrossRef\]](#)
13. Terlouw, T.; AlSkaif, T.; Bauer, C.; van Sark, W. Multi-objective optimization of energy arbitrage in community energy storage systems using different battery technologies. *Appl. Energy* **2019**, *239*, 356–372. [\[CrossRef\]](#)
14. Kriett, P.O.; Salani, M. Optimal control of a residential microgrid. *Energy* **2012**, *42*, 321–330. [\[CrossRef\]](#)
15. Efkarpidis, N.A.; Imoscopi, S.; Geidl, M.; Cini, A.; Lukovic, S.; Alippi, C.; Herbst, I. Peak shaving in distribution networks using stationary energy storage systems: A Swiss case study. *Sustain. Energy Grids Netw.* **2023**, *34*, 101018. [\[CrossRef\]](#)
16. Zhang, S.; Tang, Y. Optimal schedule of grid-connected residential PV generation systems with battery storages under time-of-use and step tariffs. *J. Energy Storage* **2019**, *23*, 175–182. [\[CrossRef\]](#)
17. Hannan, M.A.; Abdolrasol, M.G.M.; Faisal, M.; Ker, P.J.; Begum, R.A.; Hussain, A. Binary Particle Swarm Optimization for Scheduling MG Integrated Virtual Power Plant Toward Energy Saving. *IEEE Access* **2019**, *7*, 107937–107951. [\[CrossRef\]](#)
18. Moutis, P.; Hatziaargyriou, N.D. Decision trees aided scheduling for firm power capacity provision by virtual power plants. *Int. J. Electr. Power Energy Syst.* **2014**, *63*, 730–739. [\[CrossRef\]](#)
19. Nweye, K.; Sankaranarayanan, S.; Nagy, Z. MERLIN: Multi-agent offline and transfer learning for occupant-centric operation of grid-interactive communities. *Appl. Energy* **2023**, *346*, 121323. [\[CrossRef\]](#)
20. Zhan, S.; Lei, Y.; Chong, A. Comparing model predictive control and reinforcement learning for the optimal operation of building-PV-battery systems. *E3S Web Conf.* **2023**, *396*, 04018. [\[CrossRef\]](#)
21. Rocchetta, R.; Nespoli, L.; Medici, V.; Basso, S.; Derboni, M.; Salani, M. Rule-based deep reinforcement learning for optimal control of electrical batteries in an energy community. In Proceedings of the 33rd European Safety and Reliability Conference (ESREL 2023), Southampton, UK, 3–7 September 2023; pp. 639–646.
22. Sahu, R.K.; McLaughlin, D. An Ensemble Optimization Framework for Coupled Design of Hydropower Contracts and Real-Time Reservoir Operating Rules. *Water Resour. Res.* **2018**, *54*, 8401–8419. [\[CrossRef\]](#)
23. Kumar, R.; Jalving, J.; Wenzel, M.J.; Ellis, M.J.; ElBsai, M.N.; Drees, K.H.; Zavala, V.M. Benchmarking Stochastic and Deterministic MPC: A Case Study in Stationary Battery Systems. *Aiche J.* **2019**, *65*, e16551. [\[CrossRef\]](#)
24. Arroyo, J.; Spiessens, F.; Helsen, L. Comparison of Optimal Control Techniques for Building Energy Management. *Front. Built Environ.* **2022**, *319*, 114563. [\[CrossRef\]](#)
25. Ge, Y.; Xie, J.; Feng, S.; Chang, J.; Wang, Z. Multi-Task Deep Reinforcement Learning With Scenario Clustering for Real-Time Scheduling of Wind-Solar-Hydro Complementary Generation Systems. *IET Renew. Power Gener.* **2025**, *167*, 110635. [\[CrossRef\]](#)
26. Cao, R.; Shen, J.; Cheng, C.; Wang, J. Optimization Model for the Long-Term Operation of an Interprovincial Hydropower Plant Incorporating Peak Shaving Demands. *Energies* **2020**, *13*, 4804. [\[CrossRef\]](#)

27. Derboni, M.; Salani, M.; Rizzoli, A.E. Flexible Scheduling of Residential Energy Loads, the Optiflex concept. In *Proceedings of the 2nd FTAL Conference 2021 Sustainable Smart Cities and Regions, Lugano, Switzerland, 28–29 October 2021*; Mugellini, E., Carpanzano, E., Eds.; CEUR Workshop Proceedings: Aachen, Germany, 2021.
28. Bontempi, G.; Ben Taieb, S.; Le Borgne, Y.A. Machine learning strategies for time series forecasting. In *Business Intelligence: Second European Summer School, eBISS 2012, Brussels, Belgium, July 15–21, 2012, Tutorial Lectures*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 62–77.
29. Rubattu, N.; Maroni, G.; Corani, G. Electricity Load and Peak Forecasting: Feature Engineering, Probabilistic LightGBM and Temporal Hierarchies. In *Proceedings of the International Workshop on Advanced Analytics and Learning on Temporal Data, Turin, Italy, 18 September 2023*; Springer: Cham, Switzerland, 2023; pp. 276–292.
30. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In *Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017*; Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30.
31. Zhang, Y.; Ma, R.; Liu, J.; Liu, X.; Petrosian, O.; Krinkin, K. Comparison and explanation of forecasting algorithms for energy time series. *Mathematics* **2021**, *9*, 2794. [[CrossRef](#)]
32. Shukla, S.; Hong, T. BigDEAL Challenge 2022: Forecasting peak timing of electricity demand. *IET Smart Grid* **2024**, *7*, 442–459. [[CrossRef](#)]
33. Akbarian, M.; Saghafian, B.; Golian, S. Monthly streamflow forecasting by machine learning methods using dynamic weather prediction model outputs over Iran. *J. Hydrol.* **2023**, *620*, 129480. [[CrossRef](#)]
34. Ferdaus, M.M.; Dam, T.; Sarkar, M.R.; Uddin, M.; Anavatti, S.G. Foundation models for clean energy forecasting: A comprehensive review. *Renew. Sustain. Energy Rev.* **2026**, *226*, 116452. [[CrossRef](#)]
35. Chen, Z.; Ma, M.; Li, T.; Wang, H.; Li, C. Long sequence time-series forecasting with deep learning: A survey. *Inf. Fusion* **2023**, *97*, 101819. [[CrossRef](#)]
36. Goodfellow, I.; Bengio, Y.; Courville, A.; Bengio, Y. *Deep Learning*; MIT Press: Cambridge, UK, 2016; Volume 1.
37. Mastelini, S.M.; da Costa, V.G.T.; Santana, E.J.; Nakano, F.K.; Guido, R.C.; Cerri, R.; Barbon, S. Multi-output tree chaining: An interpretative modelling and lightweight multi-target approach. *J. Signal Process. Syst.* **2019**, *91*, 191–215. [[CrossRef](#)]
38. Nespoli, L.; Medici, V. Multivariate boosted trees and applications to forecasting and control. *J. Mach. Learn. Res.* **2022**, *23*, 1–47.
39. Zhang, Z.; Jung, C. GBDT-MO: Gradient-boosted decision trees for multiple outputs. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 3156–3167. [[CrossRef](#)]
40. Wen, Q.; Liu, Y. Feature engineering and selection for prosumer electricity consumption and production forecasting: A comprehensive framework. *Appl. Energy* **2025**, *381*, 125176. [[CrossRef](#)]
41. Soma, L.; Cereghetti, N.; Jardini, L.; Gut, L. Fotovoltaico in Ticino: Dati 2023. *Extra Dati* **2024**. Available online: https://www3.ti.ch/DFE/DR/USTAT/allegati/volume/ed_2024_14.pdf (accessed on 8 February 2026).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.