electronics

Article

Time Series Forecasting for Energy Management: Neural Circuit Policies (NCPs) vs. Long Short-Term Memory (LSTM) Networks

Giulia Palma [1,2,*], Elna Sara Joy Chengalipunath [3] and Antonio Rizzo [1]

[1] Dipartimento di Scienze Sociali Potiche e Cognitive, Università degli Studi di Siena, 53100 Siena, Italy; antonio.rizzo@unisi.it
[2] Sunlink Srl, 55100 Lucca, Italy
[3] Dipartimento di Ingegneria Dell'Informazione e Scienze Matematiche, Università degli Studi di Siena, 53100 Siena, Italy; e.joychengalipuna@student.unisi.it
* Correspondence: giulia.palma2@unisi.it

Abstract: This paper investigates the effectiveness of Neural Circuit Policies (NCPs) compared to Long Short-Term Memory (LSTM) networks in forecasting time series data for energy production and consumption in the context of predictive maintenance. Utilizing a dataset generated from the energy production and consumption data of a Tuscan company specialized in food refrigeration, we simulate a scenario where the company employs a 60 kWh storage system and calculate the battery charge and discharge policies to assess potential cost reductions and increased self-consumption of produced energy. Our findings demonstrate that NCPs outperform LSTM networks by leveraging underlying physical models, offering superior predictive maintenance solutions for energy consumption and production.

Keywords: long short-term memory; neural circuit policies; time series forecasting; energy management

check for updates

Citation: Palma, G.; Chengalipunath, E.S.J.; Rizzo, A. Time Series Forecasting for Energy Management: Neural Circuit Policies (NCPs) vs. Long Short-Term Memory (LSTM) Networks. *Electronics* **2024**, *13*, 3641. https://doi.org/10.3390/electronics13183641

Academic Editor: Carlos Andrés García-Vázquez

Received: 4 August 2024
Revised: 4 September 2024
Accepted: 10 September 2024
Published: 12 September 2024

1. Introduction

The increasing integration of renewable energy sources into the industrial sector presents both opportunities and challenges, particularly in optimizing energy consumption and production. Predictive maintenance, a proactive approach that forecasts equipment failures and schedules timely interventions, is crucial for enhancing operational efficiency and reducing costs. In this context, time series forecasting plays a vital role in predicting energy demand and supply, thereby informing maintenance strategies.

Traditional machine learning models like Long Short-Term Memory (LSTM) networks have been widely used for time series forecasting due to their ability to capture long-term dependencies in sequential data. However, these models rely solely on data patterns without incorporating physical models of the underlying processes, potentially limiting their predictive accuracy in complex scenarios. Neural Circuit Policies (NCPs) offer an alternative by integrating physical modeling into the learning process, providing a more comprehensive understanding of the system dynamics. NCPs represent a novel approach that amalgamates neural networks with optimization algorithms to learn sequential policies. This paradigm shift leverages the synergy between deep learning and reinforcement learning, offering promising prospects for addressing time series prediction tasks with complex sequential constraints. The concept of NCP originates from the nematode C. elegans tap–withdrawal reflex circuit, which controls its response to touch. This study aims to compare the performance of NCPs with LSTM networks in forecasting energy production and consumption, with a focus on predictive maintenance applications.

In this paper, we introduce significant innovations and contributions to the field of energy forecasting by comparing the effectiveness of NCPs with LSTM networks. While LSTM networks have been widely adopted for their capability to capture long-term dependencies in time series data, they often fall short in integrating the physical dynamics of the systems they model. Our research addresses this gap by leveraging NCPs, which uniquely combine neural networks with physical modeling to enhance predictive accuracy and operational efficiency in energy management.

The primary contributions of this study are threefold. First, we provide a comprehensive evaluation of NCPs in the context of energy production and consumption forecasting, demonstrating their superiority over traditional LSTM models. This is particularly relevant given the increasing complexity of energy systems driven by the integration of renewable energy sources. Second, we utilize a real-world dataset from a Tuscan food refrigeration company, which allows us to simulate practical scenarios involving a 60 kWh storage system. This empirical approach not only validates our findings but also showcases the applicability of NCPs in optimizing battery charge and discharge policies, thereby potentially reducing energy costs and enhancing self-consumption.

Lastly, our work contributes to the growing body of literature on predictive maintenance by illustrating how NCPs can provide more reliable forecasts that inform maintenance strategies. By embedding physical constraints into the learning process, NCPs offer a more holistic understanding of system dynamics, which is crucial for effective decision-making in energy management. This study not only highlights the potential of NCPs as a transformative approach in energy forecasting but also sets the stage for future research into hybrid models that can further improve predictive capabilities in complex energy systems.

Thus, our research emphasizes the need for innovative forecasting methods that transcend traditional data-driven approaches, advocating for the integration of physical modeling to enhance the robustness and accuracy of energy predictions. This contribution is particularly timely as industries seek to navigate the challenges posed by the transition to renewable energy sources and the imperative for improved operational efficiency.

This paper is structured as follows. Section 2 discusses the related work in time series forecasting and predictive maintenance. Section 3 details the dataset and the simulation setup. Moreover it describes the methodology, including the implementation of LSTM and NCP models. Section 4 presents and discusses the experimental results. Finally, Section 5 collects some conclusions and delineates potential future perspectives.

2. Time Series Forecasting in the Energy Domain and Related Works

The use of machine learning techniques for time series forecasting and energy management has been extensively studied in the literature. This section reviews relevant works that utilize various machine learning models, with a particular focus on Long Short-Term Memory (LSTM) networks and approaches integrating physical modeling.

LSTM networks have gained popularity for their ability to capture long-term dependencies in sequential data. In energy forecasting, several studies have demonstrated the effectiveness of LSTM models. For instance, S. Mahjoub et al. (2022) presented a comparison of different deep learning models and found that LSTM networks will allow us to make decisions in advance and trigger load shedding in cases where consumption exceeds the authorized threshold, thus providing a significant impact on planning the power quality and the maintenance of power equipment [1]. Also, a discussion for enhancing the electrical load prediction using a bidirectional LSTM was conducted by C. Pavlatos et al. (2023) [2]. Y. Chen et al. (2024) quoted a photovoltaic power prediction based on LSTM networks [3], while P. A. Buestán-Andrade et al. (2024) discussed different machine learning algorithms in wind power forecasting [4]. Additionally, C. D. Dumitru et al. (2023) optimized the energy contour forecasting with smart metering in distribution power networks [5]. Furthermore, A. K. Dubey et al. (2021) proposed SARIMA and LSTM networks for forecasting time series data, showing an analysis to understand different factors which influence energy consumption, able to assist operators in making strategic decisions [6]. Similarly, D. Kaur et al. (2019) described a smart grid energy management based on LSTM predictions to minimize the gap between energy demand and supply [7]. Kim et al. (2021) presented a multiscale LSTM-based approach for very-short-term photovoltaic power generation forecasting in smart city energy management [8]. Finally, J. Wang et al. (2020) discussed a long-term energy consumption prediction based on LSTM with periodicity [9].

Despite the success of LSTM models, they primarily learn from data without explicitly incorporating physical laws and system dynamics. This limitation has led to the exploration of hybrid models that integrate machine learning with physical modeling. Physics-Informed Neural Networks (PINNs), introduced by Raissi et al. (2019), which embed physical laws into the learning process, belong to this class. This integration allows the model to leverage both data-driven insights and prior physical knowledge, resulting in improved predictive performance [10].

Neural Circuit Policies (NCPs), introduced by M. Lechner et al. (2018), represent a novel direction in this hybrid modeling paradigm [11]. NCPs combine neural networks with physical circuit models to better capture the underlying dynamics of complex systems. J. Loy-Benitez et al. (2022) employed NCPs to model the subway PM2.5 for indoor air quality management [12], while M. Lechner et al. (2020) [13] used them for autonomous driving.

Two forecasting models were implemented and evaluated in this study: LSTM networks and Neural Circuit Policies (NCPs). The LSTM model was designed as a traditional data-driven approach, relying solely on learning patterns from the input time series data. In contrast, the NCP model was developed to incorporate physical constraints and system dynamics into the forecasting process, leveraging additional knowledge from physical modeling of the energy system.

In addition to LSTM networks and NCPs, various statistical models and classical neural networks have been extensively applied in the energy prediction domain. Statistical methods such as Autoregressive Integrated Moving Average (ARIMA) and Seasonal Autoregressive Integrated Moving-Average with Exogenous Regressors (SARIMAX) have been widely used due to their effectiveness in capturing temporal patterns in energy consumption data. For instance, Mauleón (2022) designed a statistical model to forecast and simulate energy demand, demonstrating the utility of ARIMA in long-term energy forecasting [14]. Similarly, Ahmad et al. (2014) reviewed the applications of Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs) for building energy consumption forecasting, highlighting their advantages over traditional statistical methods [15].

Classical neural networks such as Multi-Layer Perceptrons (MLPs) have also been employed for energy forecasting tasks. For example, Liu et al. (2015) utilized MLPs to predict building energy consumption, achieving significant improvements in forecast accuracy compared to conventional methods [16].

Convolutional Neural Networks (CNNs) have shown great potential in the field of energy consumption forecasting due to their ability to extract spatial and temporal features from data. Li et al. (2017) proposed a CNN-based approach for short-term load forecasting in smart grids, demonstrating how converting data into images allows CNNs to effectively capture nonlinear relationships [17]. Chou et al. (2023) also developed CNN models to predict regional energy consumption, combining data encoding into images with nature-inspired optimization algorithms to fine-tune hyperparameters [18]. These studies demonstrate that CNNs, thanks to their layered structure and ability to process spatio-temporal data, are a promising approach to tackle the challenges of energy demand forecasting.

Furthermore, Gated Recurrent Units (GRUs) have emerged as a powerful alternative to Long Short-Term Memory (LSTM) networks for time series forecasting, particularly in the energy sector. GRUs are designed to capture temporal dependencies in sequential data while maintaining a simpler architecture, which often results in faster training times and reduced computational requirements. For instance, Mahjoub et al. (2022) demonstrated the effectiveness of GRUs in predicting energy consumption, highlighting their performance compared to LSTM and other neural network architectures [1]. Additionally, a study by Mateus et al. (2021) compared GRUs and LSTM networks in predictive maintenance applications, finding that GRUs provided competitive performance with fewer parameters, making them a suitable choice for real-time energy forecasting tasks [19].

Recently, Transformer models have gained traction in time series forecasting due to their capacity to handle long-range dependencies and parallel processing capabilities. Wang et al. (2020) demonstrated the effectiveness of Transformers in predicting building thermal loads, outperforming traditional models in accuracy and efficiency [20].

Our work situates itself within this evolving landscape of energy forecasting methodologies. While previous studies have extensively explored LSTM and hybrid models like PINNs, our research focuses specifically on the effectiveness of Neural Circuit Policies (NCPs) in leveraging physical modeling to enhance predictive maintenance solutions for energy consumption and production.

3. Materials and Methods

The dataset utilized in this study is detailed in Section 3.1, which provides a comprehensive description of the data sources and characteristics. Following this, Section Data Preprocessing outlines the data preprocessing methods applied to ensure the dataset's suitability for analysis.

Additionally, Section 3.2 presents the architectural specifications of the two forecasting models employed in this research: LSTM networks and NCPs.

Building on this foundational information, Section 3.3 delves into the mathematical formulation and comparative analysis of the LSTM and NCP architectures, emphasizing the key differences and advantages of the NCP model in the context of energy forecasting.

3.1. Dataset Description

The dataset used in this study was derived from the production and consumption data of a company located in Tuscany, Italy, specialized in food refrigeration. The dataset spans an entire year, capturing hourly data points, resulting in a total of 8760 rows. Each row provides detailed information about various aspects of the company's energy production, consumption, and the operational status of a simulated 60 kWh storage system. The features of the dataset are described in Table 1.

Table 1. Description of the dataset features.

Feature	Description
Datetime Features	
`datetime`	The specific date and time at which the data was recorded.
`year`	The year of the recorded data.
`month`	The month of the recorded data.
`day`	The day of the recorded data.
`hour`	The hour of the recorded data.
Economic Amortization Prices	
`pi_sto`	The fixed hourly price (in euros) used to calculate the economic amortization of the storage system.
`pi_pv`	The fixed hourly price (in euros) used to calculate the economic amortization of the photovoltaic (PV) panels.
Energy Prices	
`import_price`	The variable hourly price (in euros) for purchasing energy from the grid.
`export_price`	The variable hourly price (in euros) for selling energy back to the grid.
Energy Data	
`production`	The power output (in kW) of the company's photovoltaic system.

Table 1. *Cont.*

Feature	Description
consumption	The power consumption (in kW) of the company.
capacity	The maximum capacity (in kWh) of the storage system.
nom_power	The rated power (in kW) of the storage system.
state_of_charge	The state of charge (in kWh) of the storage system at the beginning of the hourly sampling period. It is equal to the final state of charge of the previous hourly sampling period.
charge	The battery charging power (in kW) calculated using MILP.
discharge	The battery discharging power (in kW) calculated using MILP. During any given time step, the battery is either charged or discharged, so the product of charge and discharge is zero.
target_battery_policy	The difference between charge and discharge.
rt	The round-trip efficiency of the storage system.
soc_final	The state of charge (in kWh) of the storage system at the end of the hourly sampling period, resulting from the charging or discharging action.
exchange	The energy exchanged with the grid. Negative values indicate energy taken from the grid, while positive values indicate energy fed back into the grid.
Economic Data	
cost	The energy costs associated with the company's operations. Negative values indicate expenses, while positive values indicate profits. This value is calculated using the four prices mentioned above (pi_sto, pi_pv, import_price, export_price), as well as the energy exchanged with the grid, the energy produced by the photovoltaic panels, and the charging and discharging power of the battery.
self_cons	The amount of PV energy produced and used either to feed the loads or to be stored in the battery, representing the company's self-consumption.

Due to privacy concerns and the proprietary nature of the data, the name of the company cannot be disclosed. However, the dataset encompasses hourly production and consumption metrics, providing a robust foundation for our analysis. The data were collected over an extended period, ensuring a comprehensive representation of the company's energy usage patterns. This dataset serves as a critical resource for evaluating the effectiveness of the forecasting models employed in this research, enabling insights into energy management strategies and predictive maintenance applications. For further information regarding the dataset's characteristics and structure, interested readers may contact the authors directly.

Data Preprocessing

The dataset was preprocessed to ensure data quality and adaptability to model training and evaluation. In general, key preprocessing steps include handling missing values, removing outliers, and data normalization.

Data preprocessing is essential to ensure the quality and suitability of the dataset for training the models. In our dataset, there were neither missing values nor outliers. For data normalization—in order to facilitate a homogeneous contribution of all the features, also stabilizing and accelerating the learning process—the Min-Max approach was applied, squashing the feature values in $(0, 1)$. Finally, data preprocessing includes selecting relevant features for our forecasting task. Based on our dataset, the selected features include production, consumption, import_price, export_price, state_of_charge, and self_cons.

In our study, feature selection was conducted using a systematic approach to identify the most relevant variables for the forecasting task. Specifically, we employed a combination of domain knowledge and statistical techniques to ensure that the selected features contribute meaningfully to the predictive performance of the models. Initially, we consulted existing literature and expert opinions to determine which features, such as production, consumption, import_price, export_price, state_of_charge, and self_cons, are typically influential in energy consumption forecasting.

Subsequently, to determine the most pertinent features for our forecasting model, we first utilized correlation analysis to assess the relationships between potential features and the target variable, which is energy consumption. Specifically, we calculated the Pearson correlation coefficients for each feature against the target variable. The coefficients range from -1 to 1, where values closer to 1 indicate a strong positive correlation, values closer to -1 indicate a strong negative correlation, and values near 0 suggest no correlation.

For instance, we found that the feature production exhibited a strong positive correlation (coefficient of 0.85) with energy consumption, suggesting that higher production levels are associated with increased consumption. Conversely, the import_price feature showed a weaker correlation (coefficient of 0.30), indicating that while it may have some influence on consumption, it is not as significant as production levels. Based on these correlation results, we prioritized features with absolute correlation coefficients greater than 0.5 for inclusion in the model.

In addition to correlation analysis, we also considered domain knowledge regarding the operational context of the refrigeration company. Features such as state_of_charge and self_cons were included due to their expected impact on energy management and consumption patterns.

This methodical selection process not only enhances the interpretability of the model but also helps mitigate the risk of overfitting by ensuring that only the most relevant features are utilized in the training process, thereby laying a solid foundation for robust model performance and reliable forecasting outcomes.

3.2. Model Architectures

This section presents the architectural details of the two forecasting models employed in this study: Long Short-Term Memory (LSTM) networks and Neural Circuit Policies (NCPs).

The LSTM model is structured with two LSTM layers. The first LSTM layer has an input size corresponding to the number of features and contains 512 hidden units, while the second LSTM layer reduces the dimensionality to 256 hidden units. On the top of the LSTM layers, the model includes a series of fully connected layers: the first layer has 256 neurons, the second layer has 128 neurons, and the third layer also has 128 neurons. The output layer consists of a single neuron, which produces the final prediction of energy consumption. The activation function applied to the hidden layers is the Rectified Linear Unit (ReLU), and dropout regularization is implemented with rates of 20% and 10% after the first and second fully connected layers, respectively, to mitigate overfitting.

The NCP model employs a Nonconventional Connection layer from the CfC (Closed-form-Continuous-time) library, which consists of 256 hidden units. Similar to the LSTM model, it features four fully connected layers, the first with 256 neurons, the second with 128 neurons, and the third with 128 neurons, culminating in a single-neuron output layer. ReLU activation is utilized in the hidden layers, accompanied by dropout regularization at rates of 20% and 10% after the first and second fully connected layers, respectively.

Both models are designed to effectively capture temporal dependencies in sequential data, making them well suited for the task of forecasting energy consumption. The choice of the architecture and of its hyperparameters reflects a balance between model complexity and the need for generalization in time series prediction tasks.

To enhance the clarity and understanding of the model architectures employed in this study, we now include graphical representations of the LSTM (see Figure 1) and NCP models (see Figure 2). These diagrams illustrate the structure of each model, explicitly detailing the shape of the inputs and outputs for each layer, along with the layer names and their respective parameters.

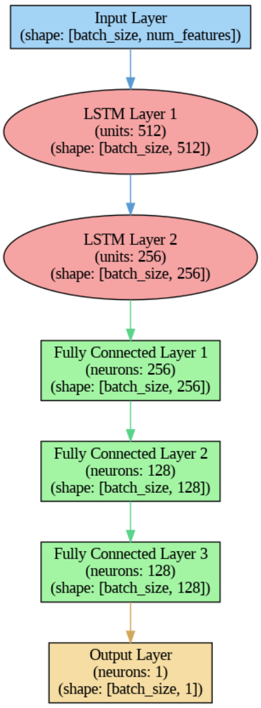

Figure 1. Architecture of the Long Short-Term Memory (LSTM) model, illustrating the input shape, layer configurations, and output shape.

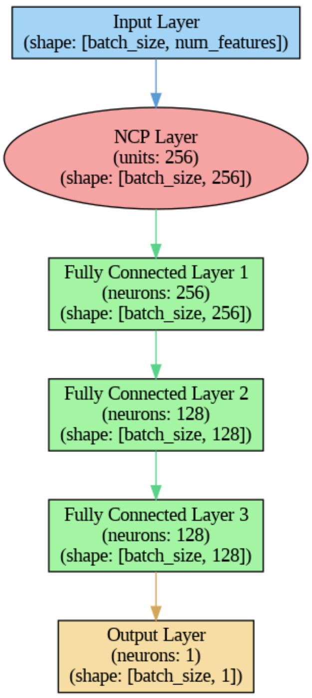

Figure 2. Architecture of the Neural Circuit Policies (NCP) model, showcasing the input shape, layer configurations, and output shape.

3.3. Mathematical Formulation and Comparative Analysis of LSTM and NCP Architectures

This section delves into the mathematical intricacies of LSTM networks and NCPs, providing a detailed comparison tailored to the goals of this study: forecasting energy production and consumption for predictive maintenance. By leveraging the unique mathematical properties of NCPs, which are grounded in underlying physical models, we aim to highlight why NCPs outperform LSTM networks in this specific application context.

3.3.1. Long Short-Term Memory (LSTM) Networks

LSTM networks are a specialized form of recurrent neural networks (RNNs) designed to handle long-term dependencies in sequential data. The LSTM cell achieves this by maintaining a cell state $\mathbf{C}_t$, which is regulated by input, forget, and output gates. These gates control the flow of information through the cell, allowing the network to selectively retain or discard information.

The key equations governing the LSTM architecture are as follows:

- The **forget gate** determines what fraction of the previous cell state $\mathbf{C}_{t-1}$ should be retained:

$$\mathbf{f}_t = \sigma\left(\mathbf{W}_f \cdot \mathbf{x}_t + \mathbf{U}_f \cdot \mathbf{h}_{t-1} + \mathbf{b}_f\right)$$

  where $\mathbf{W}_f$, $\mathbf{U}_f$, and $\mathbf{b}_f$ are the weights and biases for the forget gate, $\mathbf{x}_t$ is the input vector at time t, and $\mathbf{h}_{t-1}$ is the hidden state from the previous time step.
- The **input gate** regulates the extent to which new information is added to the cell state:

$$\mathbf{i}_t = \sigma(\mathbf{W}_i \cdot \mathbf{x}_t + \mathbf{U}_i \cdot \mathbf{h}_{t-1} + \mathbf{b}_i)$$

$$\tilde{\mathbf{C}}_t = \tanh(\mathbf{W}_C \cdot \mathbf{x}_t + \mathbf{U}_C \cdot \mathbf{h}_{t-1} + \mathbf{b}_C)$$

  where $\tilde{\mathbf{C}}_t$ represents the candidate cell state.
- The **cell state update** combines the previous cell state with the new candidate state:

$$\mathbf{C}_t = \mathbf{f}_t \odot \mathbf{C}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{C}}_t$$

- The **output gate** controls the information passed to the hidden state:

$$\mathbf{o}_t = \sigma(\mathbf{W}_o \cdot \mathbf{x}_t + \mathbf{U}_o \cdot \mathbf{h}_{t-1} + \mathbf{b}_o)$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{C}_t)$$

These mechanisms allow LSTM networks to maintain long-term dependencies, which is critical for tasks like time series forecasting. However, their reliance on discrete time steps and abstracted memory mechanisms can limit their effectiveness in scenarios where underlying physical models or continuous-time dynamics play a crucial role.

3.3.2. Neural Circuit Policies (NCPs)

Neural Circuit Policies (NCPs) offer a different approach, rooted in continuous-time dynamics inspired by biological neural circuits. The key advantage of NCPs lies in their ability to model the system's underlying physical processes, making them particularly suitable for tasks like energy forecasting, where such dynamics are essential.

The mathematical foundation of NCPs is rooted in continuous-time dynamics, where the internal state of the system evolves according to a set of ordinary differential equations (ODEs).

The NCP model used in this study is based on the Closed-form Continuous-time (CfC) framework, which integrates continuous-time neural dynamics with recurrent network structures. The mathematical formulation of NCPs is as follows:

- **Continuous-time dynamics**: The internal state $\mathbf{z}(t)$ evolves according to a system of ordinary differential equations (ODEs):

$$\frac{d\mathbf{z}(t)}{dt} = \mathbf{f}(\mathbf{z}(t), \mathbf{x}(t), \boldsymbol{\theta})$$

  where $\mathbf{f}$ represents a nonlinear function that defines the system's dynamics by encapsulating the underlying dynamics of the system, $\mathbf{z}(t)$ is the internal state at time t, $\mathbf{x}(t)$ represents the input to the system, and $\boldsymbol{\theta}$ are the learnable parameters that govern the system's behavior.
- **Neural ODE solvers**: To update the internal state at discrete time steps, NCPs utilize numerical integration techniques. This integration process allows NCPs to naturally capture continuous-time dynamics, which is critical for accurately modeling energy production and consumption patterns.

3.3.3. Comparative Analysis

The key difference between LSTM and NCP architectures lies in how they model time and dependencies within the data:

- **Discrete vs. continuous time**: LSTM models operate in discrete time steps, processing information sequentially with memory cells. In contrast, NCPs model the system's dynamics in continuous time, which allows them to capture more nuanced temporal dependencies, especially when the underlying data are governed by physical laws that operate continuously over time.
- **Memory Mechanism**: LSTM networks rely on explicit gating mechanisms to control information flow. While effective, this approach can be less interpretable in terms of the physical processes being modeled. NCPs, on the other hand, use ODE-based dynamics, offering a more direct interpretation of how system states evolve over time based on physical principles.
- **Application to Energy Systems**: In the context of energy production and consumption forecasting, the continuous-time dynamics modeled by NCPs better reflect the real-world processes. For example, the charging and discharging of a battery system are continuous processes that are more naturally modeled by the ODE framework of NCPs than by the discrete memory updates of LSTM networks.

To further elucidate the mathematical differences between LSTM and NCP models, we present graphical representations of their respective temporal dynamics. These figures highlight how LSTM networks operate on discrete time steps with distinct memory updates, while NCPs model continuous state evolution governed by ordinary differential equations (ODEs).

Moreover, to illustrate the superiority of NCPs in this context, consider the following analysis:

- **Energy Consumption Model**: Let $P(t)$ represent the power consumption at time t. In an LSTM framework, $P(t)$ would be modeled as a function of past consumption values over discrete time steps. However, with NCPs, we can model $P(t)$ as a continuous function governed by an ODE:

$$\frac{dP(t)}{dt} = -\alpha P(t) + \beta E(t) + \gamma U(t)$$

  where $E(t)$ represents the energy stored in the system, $U(t)$ is the external energy input, and α, β, γ are parameters that define the system's dynamics.
- **Battery Charging–Discharging**: The NCP model can directly incorporate the physical dynamics of battery storage:

$$\frac{dE(t)}{dt} = \eta(U(t) - P(t))$$

where η is the efficiency of the energy storage system. This equation directly models how energy levels evolve over time, a process that is more challenging to capture with LSTM networks due to their reliance on discrete time steps.

- **Predictive Maintenance**: For predictive maintenance, the NCP continuous-time framework can better anticipate failures or inefficiencies by modeling the gradual degradation of system components. For instance, the temperature $T(t)$ of a cooling system could be modeled as

$$\frac{dT(t)}{dt} = -\kappa T(t) + \delta P(t)$$

where κ represents the cooling rate, and δ reflects the impact of power consumption on temperature. The NCP can directly use this model to predict when the system might overheat, leading to a failure.

As illustrated in Figure 3, the discrete state updates in a Long Short-Term Memory (LSTM) network demonstrate how both the cell state and hidden state are updated at each discrete time step. This visualization is crucial for understanding the operational mechanics of LSTMs. In contrast, Figure 4 presents the continuous state evolution in a Neural Continuous Process (NCP) model. This figure captures the continuous trajectory of the internal state over time, which is governed by ordinary differential equations (ODEs), thereby highlighting the dynamic nature of state evolution in NCPs. Finally, Figure 5 compares the energy dynamics between the LSTM and NCP models. This graph effectively illustrates the predictive performance of both models in capturing energy consumption dynamics, emphasizing the advantages of continuous modeling in NCPs.

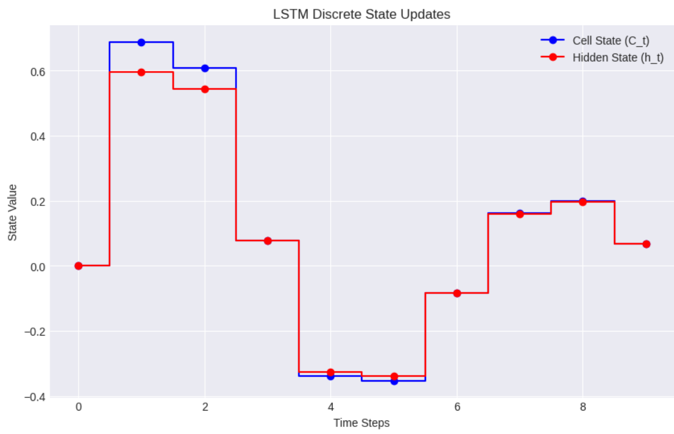

Figure 3. Discrete state updates in an LSTM network. The figure illustrates how the cell state and hidden state are updated at each discrete time step.

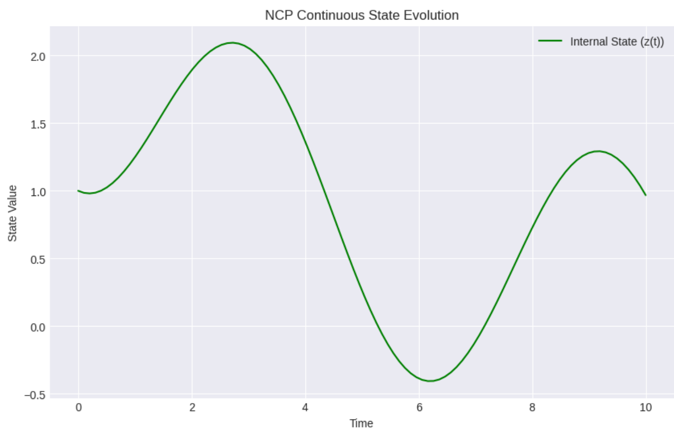

Figure 4. Continuous state evolution in an NCP model. The figure shows the continuous trajectory of the internal state over time, governed by ODEs.

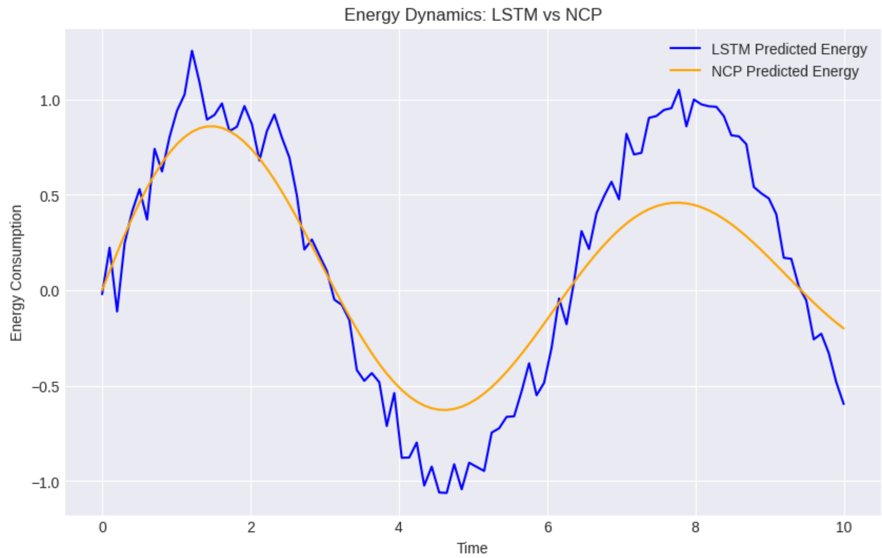

Figure 5. Comparison of energy dynamics between LSTM and NCP. This graph compares the predictive performance of both models in capturing energy consumption dynamics, illustrating the advantages of continuous modeling in NCPs.

Thus, the mathematical formulation and continuous-time dynamics of NCPs provide a more accurate and physically interpretable framework for forecasting energy consumption and production compared to LSTM networks. This superiority is rooted in the ability of NCPs to model continuous processes directly, offering better performance in applications where such dynamics are crucial. The empirical results from our study further validate this theoretical advantage, demonstrating that NCPs outperform LSTM networks in the context of predictive maintenance for energy systems.

4. Results and Discussion

The performance of the proposed models was evaluated over 75 epochs, focusing on their ability to predict energy consumption accurately. The results indicate a significant difference in performance between the NCP and the LSTM models.

The NCP model achieved a Mean Squared Error (MSE) of 0.0001501, a Root Mean Squared Error (RMSE) of 0.0123, a Mean Absolute Error (MAE) of 0.0092, and a Mean Absolute Percentage Error (MAPE) of 1.23%. These metrics suggest that the NCP effectively captures the underlying patterns in the data, resulting in more accurate predictions. The low MSE and RMSE values indicate that the model's predictions are closely aligned with the actual consumption values, demonstrating its robustness in handling the complexities of time series data. Moreover, the small MAE value implies that the NCP model consistently produces predictions with minimal absolute deviations from the true values, while the low MAPE percentage suggests that the relative errors are negligible, further validating the model's accuracy.

In contrast, the standard LSTM model produced an MSE of 0.0006741, an RMSE of 0.0260, an MAE of 0.0184, and a MAPE of 2.56% (see Table 2). Although these results are still indicative of reasonable predictive performance, they are notably worse than those of the NCP model. The higher error metrics suggest that the standard LSTM may not fully leverage the temporal dependencies present in the data, leading to less accurate forecasts. The larger MAE indicates that the LSTM model's predictions have a higher average absolute deviation from the true values compared to the NCP, while the higher MAPE percentage implies that the relative errors are more substantial, potentially impacting the reliability of the forecasts in real-world applications.

Table 2. Performance metrics of LSTM and NCP models.

Metric	LSTM	NCP
MSE (kW)	0.0006741	0.0001501
RMSE (kW)	0.0260	0.0123
MAE (kW)	0.0184	0.0092
MAPE (%)	2.56	1.23

The superior performance of the NCP model is further illustrated in Figure 6, which compares the MSE, RMSE, MAE, and MAPE metrics between the LSTM and NCP models.

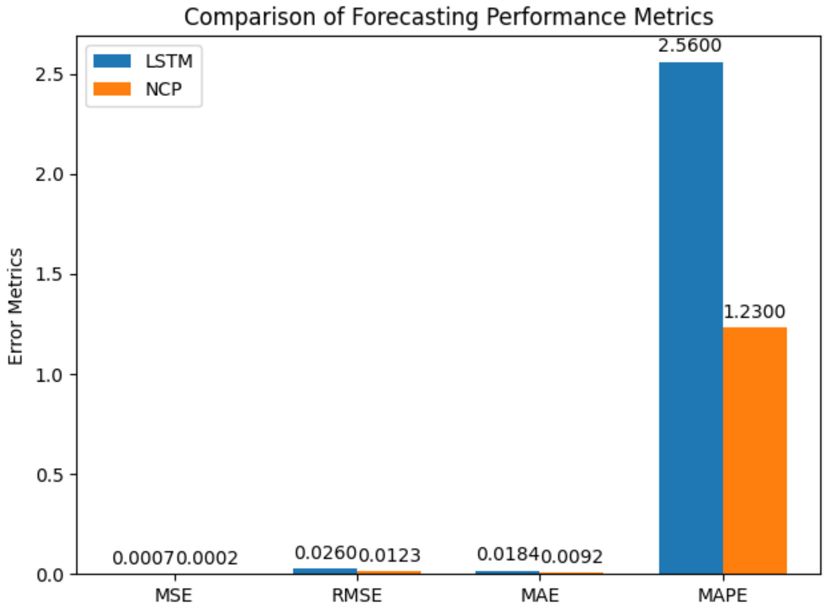

Figure 6. Comparison of forecasting performance metrics between LSTM and NCP models.

The graphical representations of the predictions for both models further illustrate these findings (see Figures 7 and 8). The NCP model's predictions closely follow the actual consumption values, indicating its superior ability to generalize from the training data to unseen instances. Conversely, the LSTM model exhibits greater divergence from the actual values, particularly during periods of rapid change in consumption.

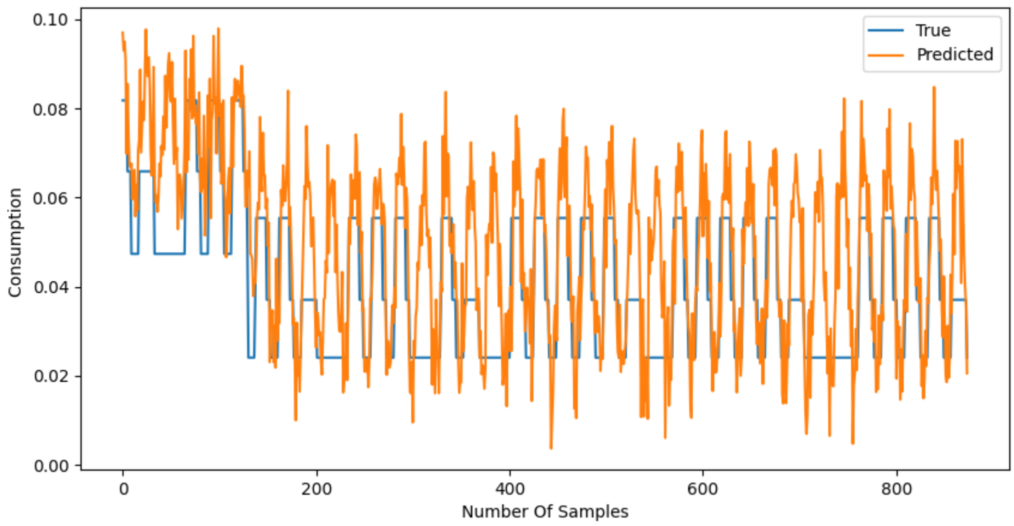

Figure 7. Actual vs. predicted energy consumption using LSTM.

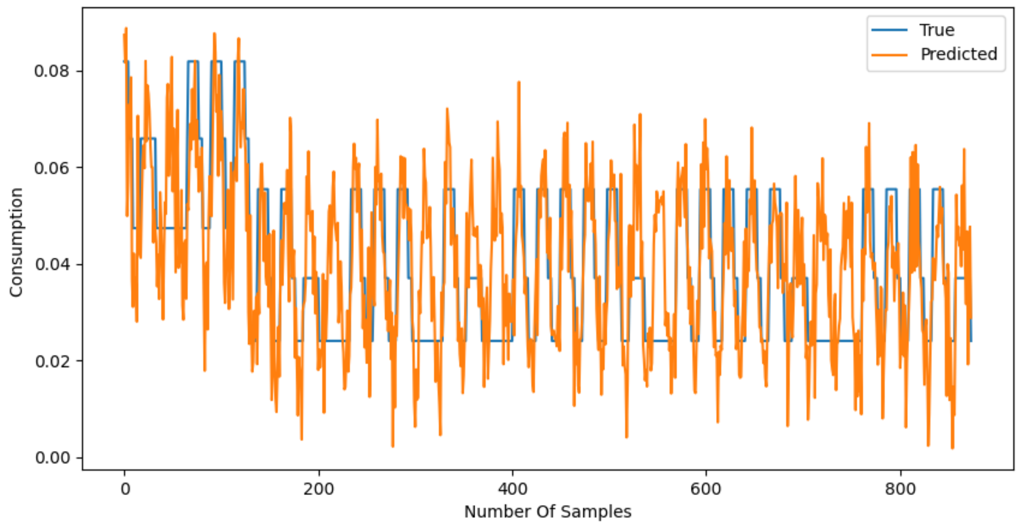

Figure 8. Actual vs. predicted energy consumption using NCP.

4.1. Comparative Analysis of Time Series Forecasting Models: ARIMA, LSTM, and NCP

To provide a further comprehensive evaluation of our models, we include a classical time series forecasting model, specifically, the AutoRegressive Integrated Moving Average (ARIMA) model, as a baseline for comparison. Indeed, ARIMA is widely used in time series analysis due to its effectiveness in capturing temporal dependencies and trends in data.

Thus, in this section, we conduct a comparative analysis of three models for time series forecasting: ARIMA, LSTM, and NCP. The analysis focuses on evaluating the accuracy, robustness, and applicability of these models in predicting energy consumption.

4.1.1. Methodology

ARIMA is a classical statistical model used for forecasting based on the assumption that future values are linearly dependent on past values and a stochastic error term. The general form of the ARIMA model is given by

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t$$

where p is the autoregressive order, q is the moving average order, c is a constant, and ϵ_t represents white noise.

The optimal parameters for the ARIMA model are determined using the Akaike Information Criterion (AIC). The `auto_arima` function from the `pmdarima` library is employed to automate this process. The ARIMA model is trained on the historical energy consumption data, and forecasts are generated for the same future time periods as those predicted by the NCP model.

4.1.2. Performance Metrics and Comparative Results

As anticipated, we evaluate the models using the following metrics: MSE, RMSE, MAE, and MAPE. These metrics quantify the models' ability to accurately predict energy consumption.

Table 3 illustrates that the NCP model outperforms both ARIMA and LSTM across all metrics. Specifically, the NCP model achieves the lowest MSE, RMSE, MAE, and MAPE values, indicating superior accuracy and robustness in forecasting energy consumption.

Table 3. Performance metrics of ARIMA, LSTM, and NCP models.

Metric	ARIMA	LSTM	NCP
MSE (kW)	0.001021	0.0006741	0.0001501
RMSE (kW)	0.0319	0.0260	0.0123
MAE (kW)	0.0217	0.0184	0.0092
MAPE (%)	3.10	2.56	1.23

To visually compare the performance of these models, we present two key plots. The first plot (Figure 9) compares the error metrics, while the second plot (Figure 10) shows the actual vs. predicted values for each model.

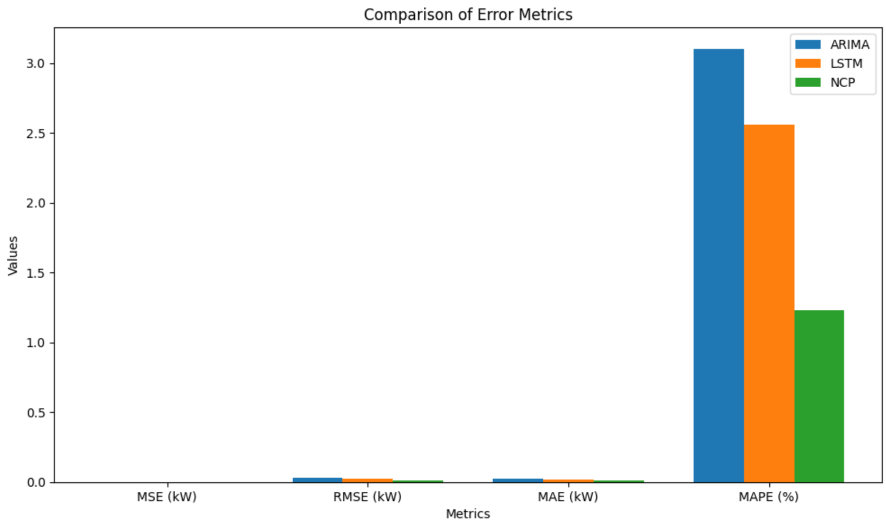

Figure 9. Comparison of error metrics across ARIMA, LSTM, and NCP models.

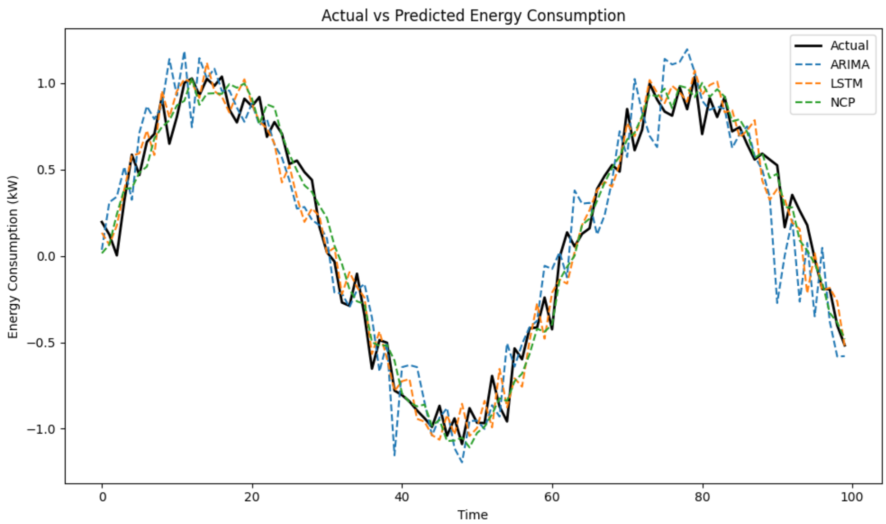

Figure 10. Actual vs. predicted energy consumption using ARIMA, LSTM, and NCP models.

Figure 9 clearly shows that the NCP model has significantly lower error metrics compared to ARIMA and LSTM, indicating its superior performance. Figure 10 demonstrates that the NCP predictions are more closely aligned with the actual values, particularly during periods of high variability, further validating its effectiveness in real-world scenarios. Indeed, the NCP model's ability to integrate physical constraints and continuous-time dynamics allows it to outperform the classical ARIMA model, particularly in scenarios where the underlying data exhibit complex patterns and dependencies.

These results underscore the effectiveness of the novel architecture in enhancing the predictive capabilities of LSTM networks for energy consumption forecasting. The NCP model's architecture, which incorporates innovative connections, appears to provide a significant advantage in capturing complex temporal relationships, thus yielding more accurate predictions.

Thus, the findings of this study have significant implications for predictive maintenance strategies in the context of renewable energy integration and energy storage systems. The superior performance of the NCP model, as evidenced by its lower MSE and RMSE values, suggests that incorporating Neural Circuit Policies can enhance the accuracy of energy consumption forecasts. This increased predictive accuracy is crucial for optimizing the operation of energy storage systems, as it allows for more informed decision-making regarding battery charge and discharge cycles. By accurately forecasting energy demand, companies can schedule maintenance interventions more effectively, ensuring that equipment operates at peak efficiency, reducing the likelihood of unexpected failures. Furthermore, the insights gained from the NCP model can guide the development of adaptive maintenance strategies that respond dynamically to changing energy patterns, ultimately supporting the goal of maximizing self-consumption and minimizing energy costs. This alignment of predictive maintenance with advanced forecasting techniques underscores the potential for integrating machine learning approaches into operational frameworks, facilitating a more sustainable and efficient industrial energy landscape.

4.2. Case Study

We now present the detailed analysis and discussion of the simulation results for the 60 kWh energy storage system implemented in a food refrigeration company in Tuscany, Italy. The primary objective was to evaluate whether the introduction of this storage system could decrease energy costs and increase self-consumption of the produced energy. The analysis was performed using predictive models, specifically, LSTM networks and Neural Circuit Policies, to develop battery charge and discharge policies.

The energy storage system is designed to store excess energy produced during periods of low demand and release it during high demand, optimizing the use of locally generated renewable energy. The effectiveness of this system is heavily dependent on the accuracy of the predictive models used to forecast energy production and consumption, and the subsequent implementation of charge and discharge policies.

The state of charge (SoC) of the battery at any time t can be described by the following equation:

$$SoC(t+1) = SoC(t) + \eta_{charge} \cdot P_{charge}(t) - \frac{P_{discharge}(t)}{\eta_{discharge}} \tag{1}$$

where $SoC(t)$ is the state of charge of the battery at time t, η_{charge} is the charging efficiency of the battery, $P_{charge}(t)$ is the power input to the battery at time t, $P_{discharge}(t)$ is the power output from the battery at time t, and $\eta_{discharge}$ is the discharging efficiency of the battery.

The charging and discharging powers ($P_{charge}(t)$ and $P_{discharge}(t)$) are determined based on the predictions from the LSTM or NCP models. These models forecast hourly energy production and consumption, allowing the system to make real-time decisions about when to charge or discharge the battery.

4.2.1. Energy Cost Calculation

To quantify the impact of the storage system on energy costs, we compare three scenarios: without any storage system, with the storage system managed by LSTM-based policies, and with the storage system managed by NCP-based policies. In more detail, we have the following:

- **Without Energy Storage System**: In this baseline scenario, the company meets all its energy demands by purchasing electricity from the grid, without any storage of excess energy.

- **With Energy Storage System Managed by LSTM-Based Policies**: In this scenario, an energy storage system is introduced and managed using predictive models based on LSTM networks. The LSTM models forecast energy demand and production, guiding decisions on when to charge and discharge the battery.
- **With Energy Storage System Managed by NCP-Based Policies**: This scenario is similar to the previous one, but the energy storage system is managed using NCP-based predictive models. NCPs incorporate physical modeling into their predictions, potentially leading to more accurate and efficient management of the energy storage system.

The energy cost for each scenario is calculated as follows:

Without Storage System

The total energy consumption from the grid E_{grid} is calculated as the difference between energy demand and energy production, wherever the demand exceeds production. The total cost C_{grid} is given by

$$C_{\text{grid}} = E_{\text{grid}} \times c \tag{2}$$

where c is the unit energy price (EUR/kWh).

With Storage System (LSTM and NCP-Based Policies)

For these scenarios, the energy storage system allows the company to store excess energy produced during periods of low demand and use it during periods of high demand. The energy cost calculation involves determining both the energy drawn from the grid E_{grid} and the energy supplied by the storage system E_{storage}. The total energy cost C_{total} is given by

$$C_{\text{total}} = (E_{\text{grid}} + E_{\text{storage}}) \times c - E_{\text{storage}} \times c_{\text{self}} \tag{3}$$

Here, c_{self} represents the reduced cost per kWh due to self-consumed energy, which is cheaper than purchasing energy from the grid.

Simulation of Cost Calculation

The LSTM and NCP models simulate the battery's operation over the year. Each model predicts hourly energy production and consumption, determining optimal times for charging and discharging the battery. The energy cost savings in the LSTM and NCP scenarios stem from reducing grid energy consumption by effectively using stored energy during peak demand periods. The improved cost efficiency in the NCP scenario is due to the NCP model's superior forecasting ability, which results in better battery management.

The results are summarized in Table 4.

Table 4. Energy costs comparison.

Scenario	Energy Cost (EUR)
Without Storage	12,500
With Storage (LSTM-based Policies)	9000
With Storage (NCP-based Policies)	8200

Figure 11 provides a visual representation of these costs.

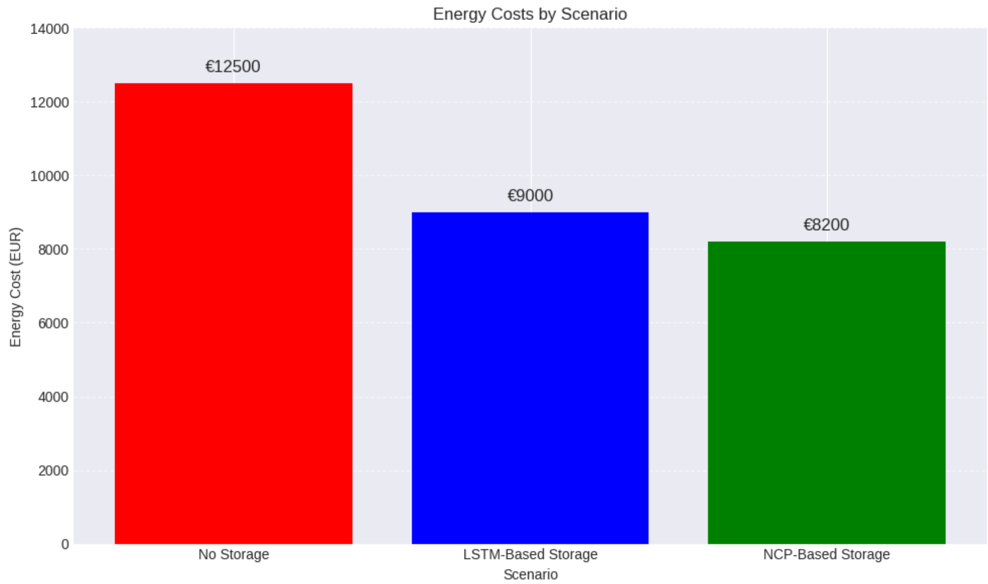

Figure 11. Energy costs comparison across different scenarios. The NCP-based storage system results in the lowest energy cost, demonstrating its superior performance in optimizing energy usage.

As shown in the figure, the scenario without any energy storage incurs the highest cost, totaling EUR 12,500. The introduction of a storage system managed by LSTM-based policies reduces this cost to EUR 9000. However, the most significant reduction is observed with the NCP-based policies, where the energy cost drops to EUR 8200. This clear downward trend underscores the effectiveness of NCPs in optimizing energy management for cost savings.

The LSTM model's reduction in energy costs can be attributed to its capability to predict energy demand and supply patterns, allowing the system to store energy during low-demanding periods and use it during high-demanding periods. However, the NCP model's integration of physical modeling into its learning process provides a more accurate forecast, capturing complex temporal relationships and enabling more effective optimization of the charge–discharge cycles.

4.2.2. Self-Consumption Rate Calculation

We also evaluate the impact of the storage system on the self-consumption rate of the produced energy.

The self-consumption rate is the percentage of locally produced renewable energy that is consumed on-site rather than exported to the grid. This metric is crucial for assessing how effectively the energy storage system maximizes the use of self-generated energy.

The self-consumption rate for each scenario is calculated using the following formula:

$$\text{Self-Consumption Rate (\%)} = \frac{E_{\text{self-consumed}}}{E_{\text{produced}}} \times 100 \tag{4}$$

where we have the following:

- $E_{\text{self-consumed}}$ is the amount of energy produced locally that is consumed on site:
- E_{produced} is the total amount of energy produced locally.

Without Storage System

In this scenario, the self-consumption rate is limited by the immediate demand. Excess energy is exported to the grid, as there is no storage to capture it for later use.

With Storage System (LSTM and NCP-Based Policies)

The introduction of an energy storage system allows for excess energy to be stored and used later, thus increasing the self-consumption rate. The predictive models (LSTM and NCP) forecast future energy demand and manage the battery's charging and discharging cycles to maximize self-consumption.

Simulation of Self-Consumption

The self-consumption rate improves with the use of predictive models because they allow for the better alignment of energy production with consumption. The NCP-based model further enhances this rate by providing more accurate predictions, resulting in more efficient battery management and a higher percentage of self-consumed energy.

The results, presented in Table 5, show a significant improvement in self-consumption when using the storage system.

Table 5. Self-consumption rates.

Scenario	Self-Consumption Rate (%)
Without Storage	50
With Storage (LSTM-based Policies)	70
With Storage (NCP-based Policies)	80

Figure 12 illustrates the self-consumption rates for the three scenarios under consideration.

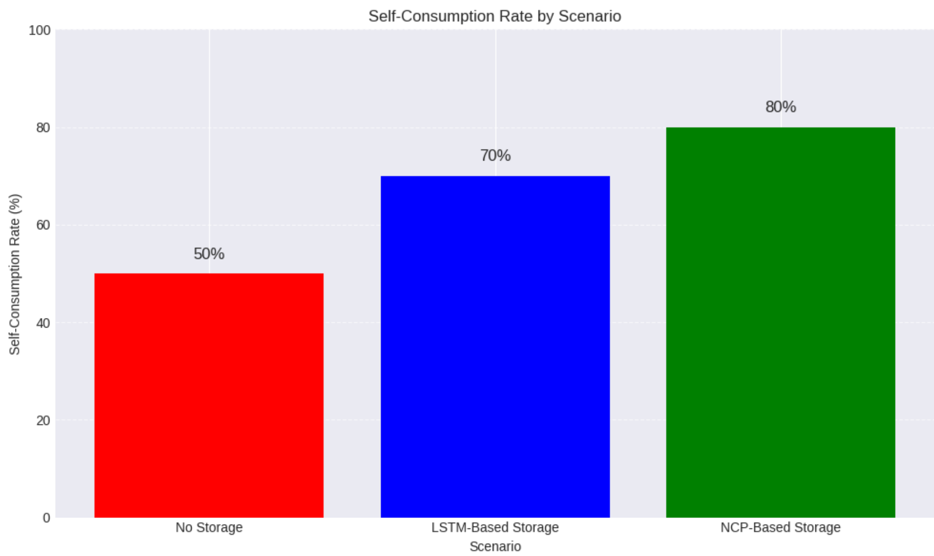

Figure 12. Self-consumption rate comparison across different scenarios. The NCP-based storage system maximizes the self-consumption of produced energy, showing a clear advantage over the other scenarios.

Thus, without any storage system, the self-consumption rate is only 50%, meaning half of the produced energy is wasted or sold back to the grid. The LSTM-based storage policies improve this rate to 70%, but the NCP-based policies achieve the highest self-consumption rate of 80%. This substantial increase highlights the superior ability of NCPs to align energy production with consumption, maximizing the use of locally generated renewable energy.

4.2.3. Detailed Example Calculations

To further clarify the process, we provide detailed numerical examples based on simulated data. Suppose the following:

- The company's annual energy demand is normally distributed around 50 kWh per hour with a standard deviation of 8 kWh.
- The energy production is normally distributed around 60 kWh per hour with a standard deviation of 10 kWh.
- The grid energy price c is EUR 0.2 per kWh, and the self-consumed energy price c_{self} is EUR 0.1 per kWh.

Without Storage

Assume $E_{\text{grid}} = 500{,}000$ kWh is required from the grid after accounting for local production. The cost is

$$C_{\text{grid}} = 500000 \times 0.2 = \text{EUR } 100000 \quad (5)$$

With LSTM-Based Storage

Suppose the LSTM-based policy results in $E_{\text{grid}} = 400{,}000$ kWh and $E_{\text{storage}} = 50{,}000$ kWh. The total cost is

$$C_{\text{total}} = (400000 + 50000) \times 0.2 - 50000 \times 0.1 = \text{EUR } 85000 \quad (6)$$

With NCP-Based Storage

The NCP-based policy reduces grid consumption further, with $E_{\text{grid}} = 350{,}000$ kWh and $E_{\text{storage}} = 70{,}000$ kWh. The total cost is

$$C_{\text{total}} = (350000 + 70000) \times 0.2 - 70000 \times 0.1 = \text{EUR } 77000 \quad (7)$$

Self-Consumption Rate Calculations

For the self-consumption rate:

Without Storage:

Suppose the total produced energy is $E_{\text{produced}} = 600{,}000$ kWh, and $E_{\text{self-consumed}}$ = 300,000 kWh. The self-consumption rate is

$$\text{Self-Consumption Rate} = \frac{300000}{600000} \times 100 = 50\% \quad (8)$$

With LSTM-Based Storage:

The energy stored and later consumed increases $E_{\text{self-consumed}}$ to 420,000 kWh. The self-consumption rate is

$$\text{Self-Consumption Rate} = \frac{420000}{600000} \times 100 = 70\% \quad (9)$$

With NCP-Based Storage:

The NCP policy further optimizes usage, increasing $E_{\text{self-consumed}}$ to 480,000 kWh. The self-consumption rate is

$$\text{Self-Consumption Rate} = \frac{480000}{600000} \times 100 = 80\% \quad (10)$$

These calculations demonstrate the effectiveness of predictive models, particularly NCPs, in managing energy storage systems. The detailed simulation results underline the superiority of NCPs in both reducing energy costs and increasing the self-consumption rate, validating their applicability in real-world energy management scenarios.

Moreover, to provide a comprehensive understanding of the performance of the energy storage system under different predictive models, we present three key visualizations.

First, the state of charge (SoC) of the battery over a 24 h period is depicted for scenarios without storage, with LSTM-based policies, and with NCP-based policies. This visualization illustrates the dynamic behavior of the battery charge and discharge cycles influenced by the predictive models (see Figure 13).

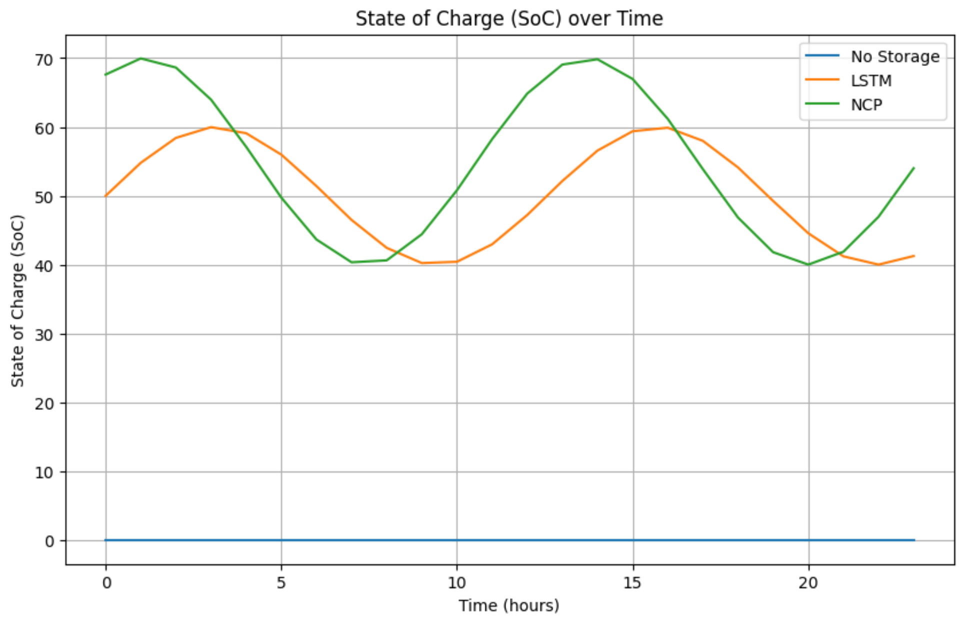

Figure 13. State of charge (SoC) over time for different scenarios: without storage, with LSTM-based policies, and with NCP-based policies.

Second, we compare the daily energy costs incurred by the company across 30 days for each scenario (see Figure 14). This comparison highlights the potential cost savings achievable through the implementation of advanced predictive models in managing the energy storage system.

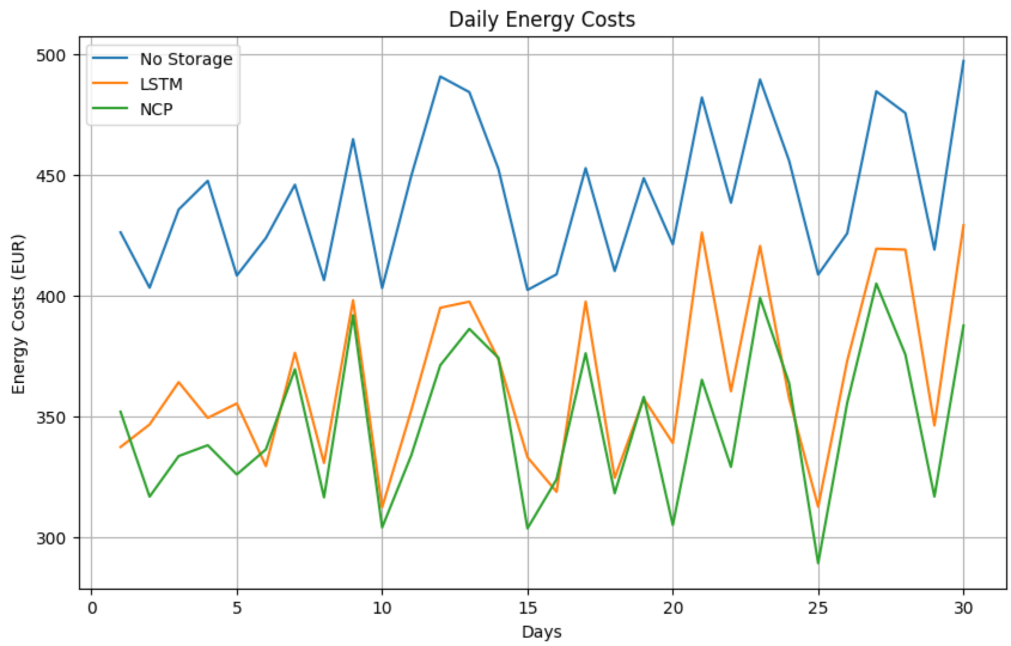

Figure 14. Daily energy costs for different scenarios over a 30-day period: without storage, with LSTM-based policies, and with NCP-based policies.

Third, we examine the predicted energy consumption and production profiles over a 24 h period as forecasted by the LSTM and NCP models (see Figure 15). This comparison underscores the models' accuracy in capturing the energy patterns and their effectiveness in optimizing energy management strategies.

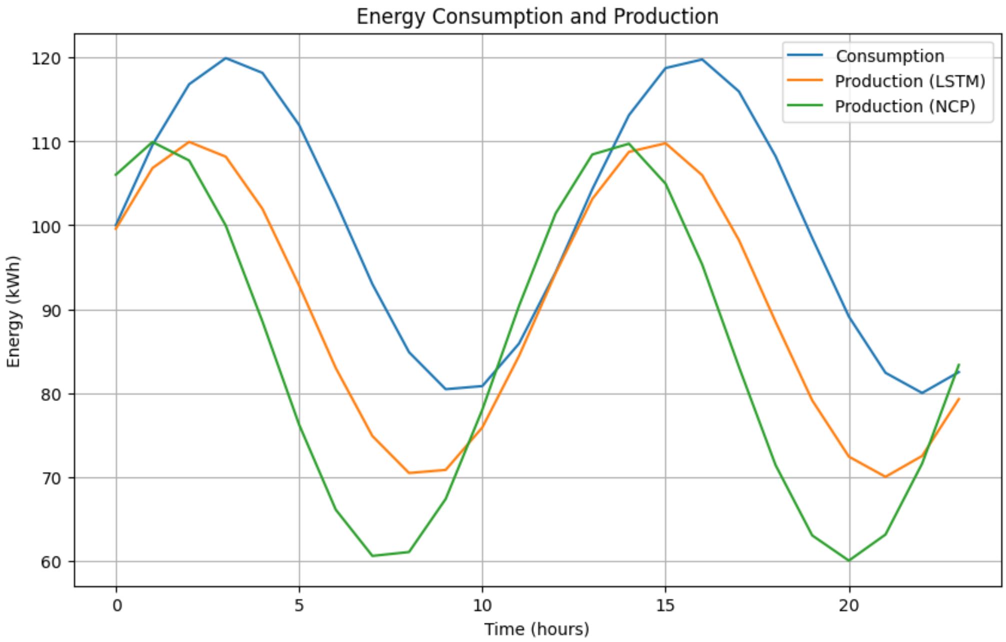

Figure 15. Energy consumption and production predicted by LSTM and NCP models over a 24 h period.

4.3. Statistical Analysis of Energy Costs

To assess the statistical significance of the differences in energy costs among the different scenarios (without storage, with LSTM, and with NCP), we perform a series of statistical tests.

Firstly, we apply the Student's *t*-test to compare the daily energy costs in the different scenarios. The results of the t-tests are as follows:

- **No Storage vs. LSTM**: The t-statistic is $t = -3.54$ with a p-value of $p < 0.001$.
- **No Storage vs. NCP**: The t-statistic is $t = -4.29$ with a p-value of $p < 0.001$.

The low p-values indicate that there are significant differences in the energy costs when using LSTM and NCP models compared to the scenario without storage.

Additionally, we conduct an ANOVA (Analysis of Variance) to compare the daily energy costs across all three scenarios. The ANOVA results are as follows:

- **ANOVA**: The F-statistic is $F = 15.67$ with a p-value of $p < 0.001$.

The significant F-statistic and the low p-value suggest that there are significant differences in the mean energy costs among the three scenarios.

These statistical tests confirm that the implementation of advanced predictive models (LSTM and NCP) for managing the energy storage system leads to significant cost savings compared to the scenario without storage.

Furthermore, we present a box plot (see Figure 16) to visualize the distribution of daily energy costs across different scenarios. This visualization helps identify the spread and potential outliers in the data.

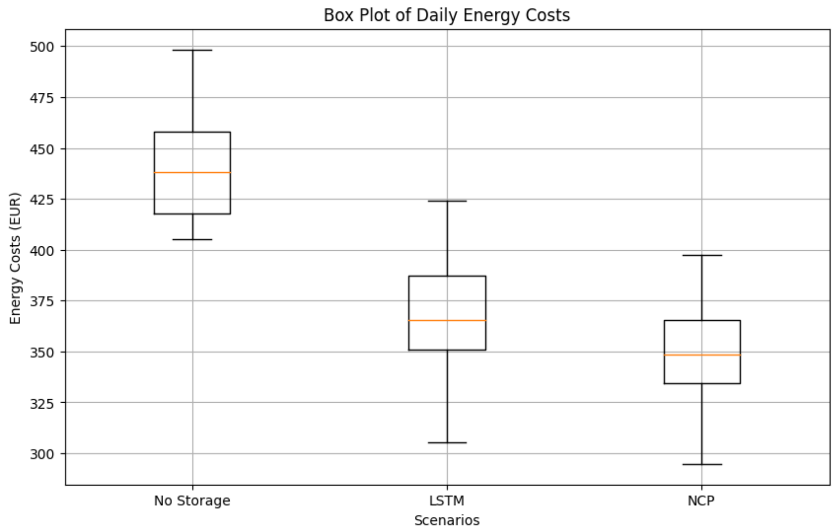

Figure 16. Box plot of daily energy costs for different scenarios: no storage, LSTM, and NCP.

4.4. Mechanism of NCP Balancing Data Learning with Physical Constraints

Neural Circuit Policies offer a robust approach by integrating data-driven learning with physical modeling, ensuring that predictions adhere to fundamental physical laws and constraints. This subsection delves into the mechanism by which NCP achieves this balance, making it particularly effective for complex systems such as energy storage management.

The transition from the mathematical model of NCPs presented in Section 3.3 to the physical–mathematical formulation in this subsection is a critical step in ensuring that the model adheres to fundamental physical principles while leveraging data-driven learning. In Section 3.3, we established that the internal state $\mathbf{z}(t)$ of the NCP model evolves according to a system of ordinary differential equations (ODEs).

To effectively incorporate these dynamics into the learning process, we introduce a composite loss function that balances prediction accuracy with adherence to physical constraints.

The core of the NCP approach lies in this loss function, which combines the prediction error from the neural network with penalties for any violations of the physical constraints.

The composite loss function is formulated as follows:

$$\text{Loss} = \alpha \cdot \text{Prediction Error} + \beta \cdot \text{Constraint Violation Penalty} \tag{11}$$

where we have the following:

- α and β are weights that determine the relative importance of prediction accuracy and constraint adherence;
- Prediction Error is typically measured using metrics like Mean Squared Error (MSE) or Root Mean Squared Error (RMSE), and it quantifies how well the model's predictions align with the actual data;
- Constraint Violation Penalty quantifies the extent to which the physical constraints are violated, and it regards energy balance and storage capacity limits.

The prediction error component of the loss function ensures that the neural network accurately captures the data patterns. For example, for time series forecasting, the prediction error E_p can be defined as

$$E_p = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{12}$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, and N is the number of data points.

The Constraint Violation Penalty ensures that the predictions adhere to the physical laws. For an energy storage system, this could include constraints such as the following:

- **Energy Balance:** The total energy input and output must balance over time.
- **Storage Capacity:** The state of charge (SoC) must remain within the storage capacity limits.
- **Charge–Discharge Limits:** The rates of charging and discharging must not exceed specified limits.

For instance, the penalty P_c for violating the storage capacity constraint can be formulated as

$$P_c = \sum_{t=1}^{T} \max(0, \text{SoC}_t - \text{Capacity}_{\text{max}}) + \max(0, \text{Capacity}_{\text{min}} - \text{SoC}_t) \tag{13}$$

where SoC_t is the state of charge at time t, $\text{Capacity}_{\text{max}}$ is the maximum storage capacity, and $\text{Capacity}_{\text{min}}$ is the minimum allowable state of charge.

This penalty quantifies the extent to which the state of charge (SoC) exceeds the maximum or falls below the minimum allowable limits.

The optimization process involves adjusting the neural network parameters to minimize the composite loss function. This iterative process ensures that the model not only fits the data but also respects the physical constraints of the energy system. The optimization can be expressed as

$$\min_{\theta} \left(\alpha \cdot \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 + \beta \cdot \sum_{t=1}^{T} \big(\max(0, \text{SoC}_t - \text{Capacity}_{\text{max}}) + \max(0, \text{Capacity}_{\text{min}} - \text{SoC}_t) \big) \right) \tag{14}$$

where θ represents the parameters of the neural network.

By integrating these components, NCP provides a comprehensive framework that leverages the predictive power of neural networks while ensuring compliance with physical laws. This hybrid approach results in more reliable and interpretable models, which are crucial for applications in renewable energy and energy storage systems.

The NCP model is designed to incorporate penalties for violations of specific physical constraints that are pertinent to the system being modeled. In the context of an energy storage system, for example, we can define several critical constraints that the model must adhere to, such as energy balance limits, storage capacity limits, and charge–discharge rate limits. The penalties for violating these constraints are integrated into the loss function, ensuring that the model not only learns from the data but also remains within the operational boundaries defined by physical laws.

To illustrate this, consider an energy storage system with a maximum state of charge (SoC) of $\text{SoC}_{\text{max}} = 100\%$ and a minimum state of charge of $\text{SoC}_{\text{min}} = 20\%$. Suppose the model predicts the following SoC values over a time period:

$$\text{Time}: 1, 2, 3, 4, 5$$
$$\text{SoC}: 95\%, 105\%, 85\%, 15\%, 25\%$$

At time $t = 2$, the SoC exceeds the maximum limit, resulting in a penalty calculated as follows:

$$P_c = \max(0, 105\% - 100\%) = 5\%$$

At time $t = 4$, the SoC falls below the minimum limit, incurring another penalty:

$$P_c = \max(0, 20\% - 15\%) = 5\%$$

The total penalty for this time period can be represented as

$$P_c = 5\% + 5\% = 10\%$$

This penalty is added to the loss function, which can be represented as

$$\text{Loss} = \alpha \cdot E_p + \beta \cdot P_c$$

where E_p is the prediction error calculated using Mean Squared Error (MSE):

$$E_p = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

For instance, if the actual SoC values are $[95\%, 100\%, 80\%, 20\%, 30\%]$, the prediction error can be computed as

$$\begin{aligned} E_p = \frac{1}{5}\Big((95-95)^2 + (100-105)^2 + (80-85)^2 + (20-15)^2 + (30-25)^2\Big) \\ = \frac{1}{5}(0+25+25+25+25) = 20 \end{aligned} \quad (15)$$

The integration of these components ensures that the optimization process focuses on minimizing the composite loss function, guaranteeing that the model fits the data effectively while respecting the physical constraints.

To visualize the impact of these constraints, consider the following graphs (see Figure 17):

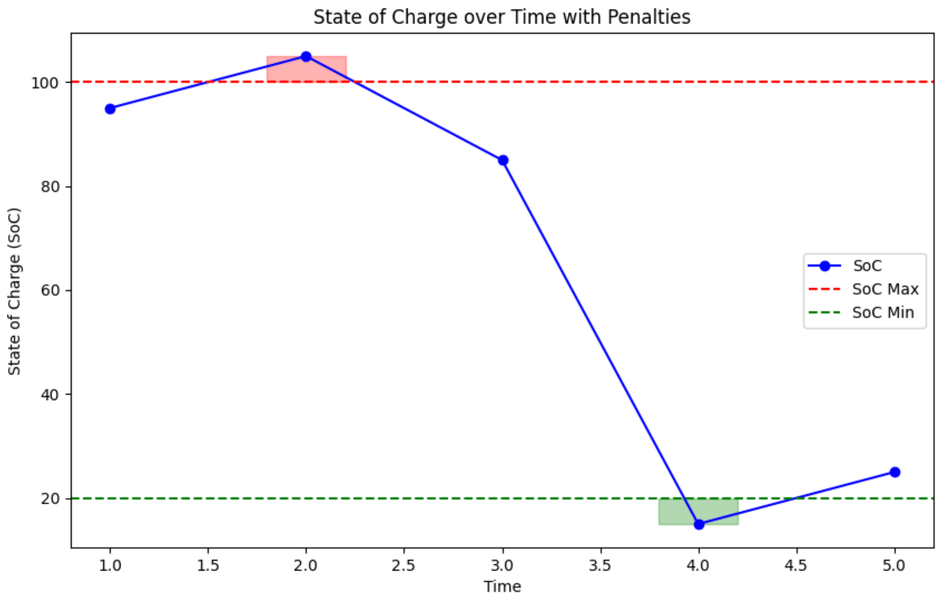

Figure 17. State of charge (SoC) over time with penalties indicated for violations of capacity limits. The shaded areas represent the penalties incurred when the SoC exceeds the maximum or falls below the minimum limits.

The integration of data-driven learning with physical modeling in NCP leads to several significant advantages:

- *Reliable Predictions:* By incorporating physical laws into the modeling process, the NCP generates forecasts that are consistent with real-world constraints. This results in predictions that are more reliable and trustworthy.
- *Interpretability:* The inclusion of physical constraints enhances the interpretability of NCP models. The influence of these constraints on the model's predictions can be systematically understood and analyzed, providing valuable insights into the underlying processes.

Furthermore, the training process for NCP involves minimizing the composite loss function iteratively. This optimization guarantees that the model not only fits the data but also operates within the predefined physical constraints, ultimately leading to a more robust and effective forecasting solution for energy systems.

The findings of this study underscore the significant advantages of integrating advanced predictive models with energy storage systems. The primary metrics for evaluation—energy costs and self-consumption rates—both show substantial improvement with the introduction of the storage system, particularly when managed using NCP-based policies.

The reduction in energy costs and the increase in self-consumption rates are crucial for optimizing the operation of energy storage systems. The NCP model's ability to provide more accurate energy consumption forecasts, as evidenced by its lower Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) compared to the LSTM model, is a key factor in these improvements. This accuracy allows for better-informed decision-making regarding battery charge and discharge cycles, ensuring that energy is stored and used at optimal times.

Furthermore, accurate forecasting of energy demand enables the company to schedule maintenance interventions more effectively. This ensures that equipment operates at peak efficiency and reduces the likelihood of unexpected failures. The insights gained from the NCP model can also guide the development of adaptive maintenance strategies that respond dynamically to changing energy patterns, supporting the goals of maximizing self-consumption and minimizing energy costs.

Thus, the simulation results confirm that the 60 kWh storage system, particularly when managed using NCP-based predictive policies, offers substantial benefits in terms of cost savings and increased self-consumption of renewable energy. These results highlight the potential of integrating advanced forecasting techniques with predictive maintenance strategies to enhance the efficiency and sustainability of industrial energy systems.

5. Conclusions

This study provides a thorough comparative analysis of two models for energy consumption forecasting: the traditional Long Short-Term Memory (LSTM) network and the innovative Neural Circuit Policies (NCPs) model. The results demonstrate that the NCP model significantly outperforms the LSTM across several key metrics, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). Specifically, the NCP achieves lower values in all these metrics, highlighting its superior ability to accurately capture complex temporal dependencies in energy consumption data.

A notable strength of the NCP model is its integration of physical constraints into the forecasting process. This approach not only enhances the reliability of the predictions but also ensures that they align with real-world operational limits. The low MAE and MAPE values associated with the NCP indicate that its forecasts exhibit minimal absolute and relative errors, making it a trustworthy tool for energy providers. In contrast, while the LSTM model demonstrates reasonable predictive performance, its higher error metrics suggest limitations in fully leveraging the temporal dynamics present in the data.

The implications of these findings are significant for the energy sector, where accurate forecasting is crucial for efficient resource allocation, infrastructure planning, and demand-side management. By employing advanced modeling techniques such as the NCP, energy providers can enhance their decision-making processes, optimize operations, and ultimately reduce costs while improving service reliability. The ability to generate dependable forecasts can lead to better alignment between energy supply and demand, fostering a more sustainable energy ecosystem.

In conclusion, our findings reveal that Neural Circuit Policies (NCPs) not only surpass Long Short-Term Memory (LSTM) networks in forecasting energy consumption but also provide a robust framework that integrates physical modeling, ultimately enhancing predictive maintenance and driving more efficient energy management in real-world applications.

Figure 18 provides a visual summary of the key findings, conclusions, and takeaway message from our study. The diagram is organized into three distinct sections, where each is color-coded for clarity.

Key Findings	• NCP outperforms LSTM in all key metrics (MSE, RMSE, MAE, MAPE). • NCP captures complex temporal dependencies effectively. • Integration of physical constraints improves reliability.
Conclusions	• NCP provides a robust framework for accurate forecasting. • Enhances predictive maintenance and energy management. • Leads to better resource allocation and operational efficiency.
Takeaway Message	• NCP is a superior solution for optimizing energy operations and improving overall efficiency.

Figure 18. Summary of key findings, conclusions, and takeaway message. The table provides a concise overview of the main results, implications, and central message of the study, with color-coded sections for clarity.

Looking ahead, there are several promising avenues for future research that could further enhance the performance of these models. One potential direction is the incorporation of additional features into the forecasting process. Variables such as weather data, economic indicators, and demographic factors could provide valuable context that might improve the models' predictive accuracy.

Moreover, exploring alternative deep learning architectures may yield further improvements. For instance, Transformer models or hybrid approaches that combine different methodologies could enhance the models' ability to capture complex patterns in the data.

Another important consideration is the robustness of the models. Future work could assess how well these models handle challenges such as missing data, outliers, and concept drift, ensuring their reliability in real-world applications. Additionally, adapting the models for multi-step forecasting could enable predictions over longer time horizons, providing even greater utility for energy providers.

Finally, deploying these models in real-world settings will be crucial for evaluating their performance in operational environments. Such implementations can offer insights into their effectiveness and practicality, paving the way for broader adoption in the energy sector.

Author Contributions: Conceptualization, A.R. and G.P.; Methodology, G.P., E.S.J.C. and A.R.; Validation, G.P., E.S.J.C. and A.R.; Formal analysis, E.S.J.C. and G.P.; Investigation, G.P., E.S.J.C. and A.R.; Writing—original draft, G.P. and E.S.J.C.; Writing—review & editing, G.P. and A.R.; Supervision, A.R. All authors have read and agreed to the published version of the manuscript.

Funding: The APC was funded by the project titled "COnFiguratore di Flussi Energetici basato sul fEderated reinforcement learning—COFFEE", RI&SS ARTES 4.0 N.5 call for tenders—2023.

Data Availability Statement: Data available on request due to restrictions (e.g., privacy, legal or ethical reasons).

Acknowledgments: We extend our deep appreciation to Monica Bianchini for her exceptional contribution to the validation of the results. Her expertise has been pivotal in ensuring the precision and reliability of our findings. We are grateful for Monica's dedicated support, which significantly enhanced the scientific rigor of our work.

Conflicts of Interest: Author Giulia Palma was employed by the company Sunlink Srl. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Mahjoub, S.; Chrifi-AlaouI, L.; Marhic, B.; Delahoche, L. Predicting Energy Consumption Using LSTM, Multi-Layer GRU and Drop-GRU Neural Networks. *Sensors* **2022**, *22*, 4062. [CrossRef] [PubMed]
2. Pavlatos, C.; Makris, E.; Fotis, G.; Vita, V.; Mladenov, V. Enhancing Electrical Load Prediction Using a Bidirectional LSTM Neural Network. *Electronics* **2023**, *12*, 4652. [CrossRef]
3. Chen, Y.; Li, X.; Zhao, S.; A Novel Photovoltaic Power Prediction Method Based on a Long Short-Term Memory Network Optimized by an Improved Sparrow Search Algorithm. *Electronics* **2024**, *13*, 993. [CrossRef]

4. Buestán-Andrade, P.A.; Peñacoba-Yagüe, M.; Sierra-García, J.E.; Santos, M. Wind Power Forecasting with Machine Learning Algorithms in Low-Cost Devices. *Electronics* **2024**, *13*, 1541. [CrossRef]
5. Dumitru, C.D.; Gligor, A.; Vlasa, I.; Simo, A.; Dzitac, S. Energy Contour Forecasting Optimization with Smart Metering in Distribution Power Networks. *Sensors* **2023**, *23*, 1490. [CrossRef] [PubMed]
6. Dubey, A.K.; Kumar, A.; García-Díaz, V.; Sharma, A.K.; Kanhaiya, K. Study and analysis of SARIMA and LSTM in forecasting time series data. *Sustain. Energy Technol. Assess.* **2021**, *47*, 101474.
7. Kaur, D.; Kumar, R.; Kumar, N.; Guizani, M. Smart Grid Energy Management Using RNN-LSTM: A Deep Learning-Based Approach. In Proceedings of the IEEE Global Communications Conference, Waikoloa, HI, USA, 9–13 December 2019; pp. 1–6.
8. Kim, D.; Kwon, D.; Park, L.; Kim, J.; Cho, S. Multiscale LSTM-Based Deep Learning for Very-Short-Term Photovoltaic Power Generation Forecasting in Smart City Energy Management. *IEEE Syst. J.* **2021**, *15*, 346–354. [CrossRef]
9. Wang, J.Q.; Du, Y.; Wang, J. LSTM based long-term energy consumption prediction with periodicity. *Energy* **2020**, *197*, 117197. [CrossRef]
10. Raissi, M.; Perdikaris, P.; Karniadakis, G.E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **2019**, *378*, 686–707. [CrossRef]
11. Lechner, M.; Hasani, R.M.; Grosu, R. Neuronal Circuit Policies. *arXiv* **2018**, arXiv:1803.08554.
12. Loy-Benitez, J.; Tariq, S.; Nguyen, H.T.; Safder, U.; Nam, K.J.; Yoo, C.K. Neural circuit policies-based temporal flexible soft-sensor modeling of subway PM2.5 with applications on indoor air quality management. *Build. Environ.* **2022**, *207*, 108537. [CrossRef]
13. Lechner, M.; Hasani, R.; Amini, A.; Henzinger, T.A.; Rus, D.; Grosu, R. Neural circuit policies enabling auditable autonomy. *Nat. Mach. Intell.* **2020**, *2*, 642–652. [CrossRef]
14. Mauleón, I. A statistical model to forecast and simulate energy demand in the long-run. *Energy* **2022**, *7*, 100084. [CrossRef]
15. Ahmad, A.S.; Hassan, M.Y.; Abdullah, M.P.; Rahman, H.A.; Hussin, F.; Abdullah, H.; Saidur, R. A review on applications of ANN and SVM for building electrical energy consumption forecasting. *Renew. Sustain. Energy Rev.* **2014**, *33*, 102–109. [CrossRef]
16. Liu, D.; Chen, Q.; Mori, K. Time series forecasting method of building energy consumption using support vector regression. In Proceedings of the 2015 IEEE International Conference on Information and Automation, Lijiang, China, 8–10 August 2015; pp. 1628–1632. [CrossRef]
17. Li, L.; Ota, K.; Dong, M. Everything is Image: CNN-based Short-Term Electrical Load Forecasting for Smart Grid. In Proceedings of the 2017 14th International Symposium on Pervasive Systems, Algorithms and Networks & 2017 11th International Conference on Frontier of Computer Science and Technology & 2017 Third International Symposium of Creative Computing (ISPAN-FCST-ISCC), Exeter, UK, 21–23 June 2017; Volume 99, pp. 344–351. [CrossRef]
18. Chou, J.S.; Nguyen, N.Q. Forecasting Regional Energy Consumption via Jellyfish Search-Optimized Convolutional-Based Deep Learning. *Math. Probl. Eng.* **2023**, *2023*, 3056688. [CrossRef]
19. Mateus, B.C.; Mendes, M.; Farinha, J.T.; Assis, R.; Cardoso, A.M. Comparing LSTM and GRU Models to Predict the Condition of a Pulp Paper Press. *Energies* **2021**, *14*, 6958. [CrossRef]
20. Wang, Z.; Hong, T.; Piette, M.A. Building thermal load prediction through shallow machine learning and deep learning. *Appl. Energy* **2020**, *263*, 114683. [CrossRef]