

Evaluating the Role of Input Window Size and Horizon-Specific Models in Multi-Horizon Solar Irradiance Forecasting

Ifran Rahman Nijhum^{a,b}, David Lillis^{b,c}, and Soumyabrata Dev^{*, a,b,d}

^aThe ADAPT Research Centre, Dublin, Ireland

^bSchool of Computer Science, University College Dublin, Dublin, Ireland

^cBeijing Dublin International College, Beijing, China

^dSchool of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland

*Corresponding author: devs@tcd.ie

ABSTRACT

Solar irradiance forecasting is key to reliable smart grid operation under high photovoltaic penetration. However, the interaction between historical input window length, forecasting horizon, and learning strategy remains insufficiently understood. In this study, we investigate the coupling between temporal window size and multi-horizon forecasting at different lead times using lightweight recurrent models. We investigate the performance of single-task models on predicting individual horizons with multi-task variants of similar models to forecast multiple horizons at once. Observations show that gated recurrent units perform best in single-task settings regardless of the window size and forecasting horizon. A simple multi-task RNN obtains comparable RMSE with reduced model redundancy. The results provide practical guidance for engineering efficient and scalable short-term solar forecasting systems.

Keywords: Solar irradiance forecasting, multi-horizon prediction, time-series forecasting, recurrent neural networks, lightweight models

1. INTRODUCTION

Solar energy is one of the largest sources of renewable energy. However, solar energy is highly intermittent and strongly dependent on atmospheric conditions. Hence, it is challenging to reliably integrate it into power grids [1, 2]. Accurate solar estimation and forecasting are crucial for efficient energy management and grid operation [3–6]. Particularly, short-term forecasts spanning minutes to hours ahead are valuable for real-time decision making and grid operation [7, 8]. Beyond solar energy applications, understanding the impact of temporal design choices is relevant to a broader class of time-series forecasting problems.

In recent years, researchers have increasingly explored data-driven approaches to forecasting. In particular, solar forecasting using time-series models from meteorological data is becoming popular. With the advancement of machine learning and deep learning, modern architectures such as RNNs, LSTMs, and GRUs have been widely adopted. This notably improved the predictive performance [9–11].

Despite these advances, key design choices in multi-horizon forecasting remain insufficiently understood. Window size refers to the amount or length of historical data used for forecasting. On the contrary, lead time defines how far into the future the model can predict. Forecasting at multiple lead times (e.g., 10, 30, and 60 minutes) can be addressed either by training separate models trained for each lead time (horizon-specific models) or by using a single model that jointly predicts all horizons. Similarly, longer input windows are often assumed to improve forecasting accuracy by providing richer temporal context. Different studies have used different window sizes. However, these design choices are rarely systematically validated. Longer windows may improve prediction accuracy yet increase computational cost. Similarly, horizon-specific models may perform better as these are specifically trained for particular lead times, yet require maintaining multiple predictors, which complicates deployment. It therefore remains unclear to what extent these design choices influence forecasting performance in multi-horizon forecasting.

Many studies have used different deep learning models for solar irradiance forecasting, including recurrent neural networks (RNNs), transformer-based models, and MLP-Mixer architectures [12–15]. These models have shown good forecasting performance. However, different studies use different experimental settings, such as input window size and forecasting horizon, which makes comparison difficult. In earlier studies, separate models were often trained for each forecasting

horizon [12, 13]. More recent works have explored using a single model to predict multiple future horizons through multi-step or multi-task forecasting approaches [14, 15]. Although these approaches have shown promising results, some questions still remain. For example, the effect of input window size is not always clear. Some studies use short windows while others use longer windows, but the choice is often based on empirical settings. It is therefore difficult to know whether increasing the amount of historical data consistently improves forecasting performance. In addition, while multi-task models can reduce the number of models that need to be deployed, it is not clear whether they can achieve similar accuracy to horizon-specific models. These issues motivate the work presented in this paper.

This work¹ addresses these gaps by investigating two fundamental questions: (i) how input window size affects forecasting performance, and (ii) whether separate horizon-specific models are required. To this end, we conduct a controlled comparison of single-task and multi-task learning strategies using lightweight recurrent architectures (RNN, LSTM, and GRU). All models are trained under identical data splits, normalization, and optimization settings to ensure a fair evaluation of the learning strategies.

The main contributions of this work are summarized as follows:

- We perform a systematic empirical investigation into the effect of historical input window length on single and multi-horizon solar irradiance forecasting, using three different window sizes of 10, 30, and 60 minutes.
- We conduct a comparative evaluation of single-task and multi-task learning strategies across three recurrent architectures, RNN, LSTM, and GRU at forecasting horizons of 10, 30, and 60 minutes.
- We provide an analysis of the efficiency-accuracy trade-off between single-task and multi-task recurrent architectures for multi-horizon solar irradiance forecasting.

2. METHODOLOGY

2.1 Problem Formulation

We address short-term multi-horizon solar irradiance forecasting. Given a historical window of length W ,

$$X_t = [x_{t-W+1}, \dots, x_t]$$

the objective is to predict future irradiance at horizons $H = \{10, 30, 60\}$ minutes. Two learning settings are considered:

- **Single-task (ST):** one model predicts a single horizon, i.e., $f_h(X_t) \rightarrow y_{t+h}$.
- **Multi-task (MT):** one model jointly predicts all horizons, i.e., $f(X_t) \rightarrow [y_{t+10}, y_{t+30}, y_{t+60}]$.

2.2 Data Preparation

We use a subset of the Girisol dataset [16], containing synchronized irradiance and meteorological measurements sampled every 15 seconds. The dataset comprises 71,750 observations and seven variables: temperature, dew point, atmospheric pressure, wind direction, wind velocity, humidity, and irradiance, with irradiance used as the prediction target. Data from 8 December 2017 to 7 January 2018 (46,527 observations) are used for model development, while the subsequent 15 days (25,223 observations) are reserved for testing. An 80:20 chronological split is applied to the development set, resulting in 37,221 training samples and 9,306 validation samples. Samples are constructed using a sliding-window formulation, where the previous W observations form the input and future irradiance values at the specified horizons form the targets. In the ST setting, each sample has a single target, whereas the MT setting jointly predicts all three horizons. All features are normalized using Min–Max scaling fitted on the training set only.

2.3 Model Architectures

Three recurrent architectures are evaluated: RNN, LSTM, and GRU. All models use a single recurrent layer with a hidden size of 4 to ensure comparable model capacity across experiments. The core architecture remains identical for all settings, with the only difference in the output layer depending on the learning strategy. In the single-task (ST) setting, the model has a single output neuron corresponding to one forecasting horizon. In contrast, in the multi-task (MT) setting, the model uses three output neurons to simultaneously predict all horizons (10, 30, and 60 minutes). In the ST formulation, a separate

¹ In the spirit of reproducible research, the code to reproduce the results of this paper can be found in <https://github.com/ifran-rahman/SolarForecasting-DesignTradeoffs>

model is trained for each forecasting horizon, with each model optimized independently using a scalar regression target. This results in multiple models, each specialized for a specific horizon. In the MT formulation, a single model jointly predicts all horizons using a shared representation. The model outputs a three-dimensional vector, and training is performed using mean squared error applied across all predicted horizons.

Table 1: Experimental setup

Setting	Models	Windows	Horizons
ST	RNN, LSTM, GRU	10, 30, 60 min	10, 30, 60 min
MT	RNN, LSTM, GRU	10, 30, 60 min	Joint (10/30/60)

2.4 Experimental Design

The experiments are designed to study the effect of input window size and to compare horizon-specific and multitask forecasting models. Three input window sizes are considered: 10, 30, and 60 minutes. These correspond to 40-, 120-, and 240-time steps. Forecasts are generated at 10-, 30-, and 60-minute horizons using RNN, LSTM, and GRU models. In the single-task setting, a separate model is trained for each forecasting horizon; in the multi-task setting, a single model predicts all horizons simultaneously. This results in 27 and 9 experiments, respectively.

2.5 Training and Evaluation

All training codes are implemented in PyTorch. The models are trained using the Adam optimizer with a learning rate of 10^{-3} . The batch size is set to 1024, and training is performed for a maximum of 500 epochs. Early stopping with a patience of 5 epochs is used to prevent overfitting.

For all experiments, the hidden size is fixed at 4. This value was selected based on initial experiments and then kept the same across RNN, LSTM, and GRU models to keep the comparison fair. Apart from the learning strategy and input window size, all other settings, such as preprocessing, data splits, and training configuration are kept unchanged.

Model performance is evaluated using MSE and RMSE. In the single-task (ST) setting, a separate model is trained for each forecasting horizon and the MSE loss is used during training. In the multi-task (MT) setting, a single model predicts the 10-, 30-, and 60-minute horizons together. The training loss is calculated as the average MSE over the three outputs:

$$\mathcal{L}_{MT} = \frac{1}{3N} \sum_{j=1}^3 \sum_{i=1}^N \left(y_i^{(j)} - \hat{y}_i^{(j)} \right)^2 \quad (1)$$

Results are reported separately for the 10-, 30-, and 60-minute forecasting horizons. Average values across the three horizons are also used when comparing different models and settings.

3. EXPERIMENTS AND RESULTS

3.1 Model Efficiency

Table 2 shows the parameter counts and model sizes of the architectures. When forecasting multiple horizons with single-task (ST) models, a separate model is required for each forecasting horizon (10, 30, and 60 min). Consequently, the three single-task models together require 207 (3×69), 783 (3×261), and 591 (3×197) parameters for the RNN, LSTM, and GRU architectures, respectively, with total model sizes of 0.9 KB, 3.0 KB, and 2.4 KB. On the contrary, a multi-task model can predict all three horizons at once. Hence, this requires only 79, 271, and 207 parameters for RNN, LSTM, and GRU (0.3 KB, 1.0 KB, and 0.8 KB, respectively).

Table 2: Model Efficiency Comparison

Model	Single-Task		Multi-Task	
	Param	KB	Param	KB
RNN	69	0.3	79	0.3
LSTM	261	1.0	271	1.0
GRU	197	0.8	207	0.8

3.2 Impact of Temporal Design Choices

The effect of window size and learning strategy is shown in Fig. 1a, Fig. 1b, and Tables 3 and 4. In general, the prediction error increases as the forecasting horizon becomes longer for all models. However, the impact of window size is not very clear and depends on the model. Some models show small improvements when more historical data is used, while others do not. For example, in the single-task (ST) setting, the GRU model improves slightly when the window size increases, with the average RMSE decreasing from 0.136 to 0.132. On the contrary, the LSTM models do not show a consistent pattern in results. A similar trend can be observed for the multi-task (MT) models, where the differences between window sizes are generally small. Overall, increasing the input window from 30 to 60 minutes did not lead to significant improvements in our experiments.

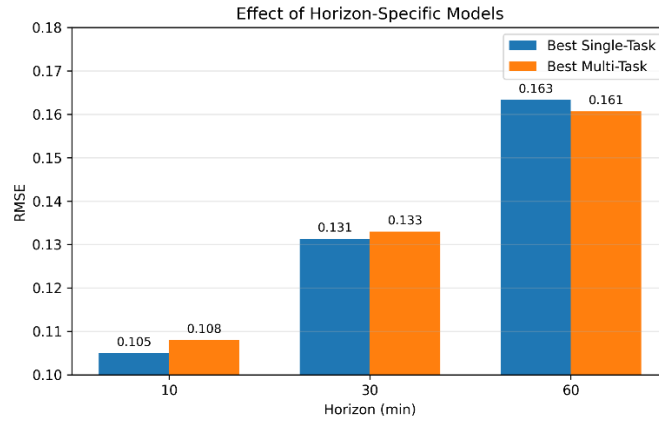
Table 3: Comparison of **single-task (ST)** models grouped by window size

Window	Horizon	RNN		LSTM		GRU	
		RMSE	MSE	RMSE	MSE	RMSE	MSE
10	10	0.102	0.0104	0.109	0.0119	0.113	0.0128
	30	0.144	0.0207	0.158	0.0250	0.131	0.0172
	60	0.165	0.0272	0.163	0.0266	0.163	0.0266
	Avg	0.137	0.019	0.143	0.021	0.136	0.019
30	10	0.108	0.0117	0.105	0.0110	0.107	0.0114
	30	0.137	0.0188	0.138	0.0190	0.129	0.0166
	60	0.178	0.0317	0.171	0.0292	0.166	0.0276
	Avg	0.141	0.021	0.138	0.020	0.134	0.019
60	10	0.105	0.0110	0.126	0.0159	0.102	0.0104
	30	0.133	0.0177	0.135	0.0182	0.134	0.0180
	60	0.183	0.0335	0.167	0.0279	0.161	0.0259
	Avg	0.140	0.021	0.143	0.021	0.132	0.018

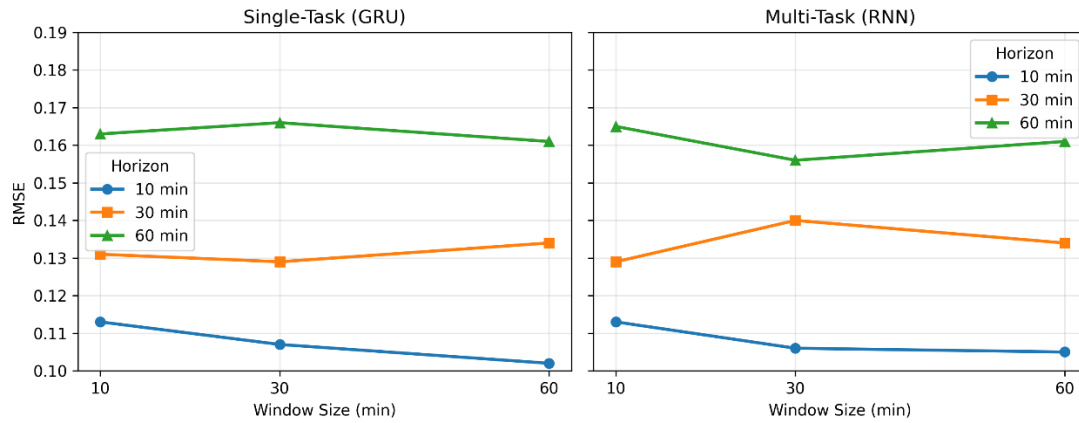
Table 4: Comparison of **multi-task (MT)** models grouped by window size

Window	Horizon	RNN		LSTM		GRU	
		RMSE	MSE	RMSE	MSE	RMSE	MSE
10	10	0.113	0.0128	0.123	0.0151	0.111	0.0123
	30	0.129	0.0166	0.156	0.0243	0.136	0.0185
	60	0.165	0.0272	0.160	0.0256	0.172	0.0296
	Avg	0.136	0.019	0.146	0.022	0.140	0.020
30	10	0.106	0.0112	0.146	0.0213	0.114	0.0130
	30	0.140	0.0196	0.140	0.0196	0.131	0.0172
	60	0.156	0.0243	0.163	0.0266	0.178	0.0317
	Avg	0.134	0.0180	0.150	0.023	0.141	0.021
60	10	0.105	0.0110	0.109	0.0119	0.106	0.0112
	30	0.134	0.0180	0.134	0.0180	0.132	0.0174
	60	0.161	0.0259	0.161	0.0259	0.168	0.0282
	Avg	0.133	0.018	0.135	0.019	0.135	0.019

To evaluate the benefit of horizon-specific models, Fig. 1a compares the best-performing recurrent architecture in the ST and MT settings. For each forecasting horizon, the reported RMSE is averaged across the three input window sizes (10, 30, and 60 min). At a shorter horizon, the ST model slightly outperforms the MT model (e.g., at the 30-minute horizon, the difference between ST and MT is negligible (0.131 vs. 0.133)). At the 60-minute horizon, however, the best MT model slightly outperforms the best ST model (0.161 vs. 0.163). Thus, it can be said that overall, the performance gap between these models is small across all horizons.



(a) Comparison of the best-performing recurrent architecture in the ST and MT settings. RMSE values are averaged across the three input window sizes for each forecasting horizon.



(b) Effect of input window size on RMSE across forecasting horizons.

Figure 1: Impact of temporal design choices. (a) Single-task and multi-task learning achieve comparable performance across forecasting horizons. (b) Increasing the input window size provides limited and inconsistent performance improvements across models and horizons.

4. DISCUSSION AND CONCLUSION

In this work, we provide a systematic analysis of temporal design choices in multi-horizon solar forecasting using lightweight recurrent models. The results indicate that increasing the input window does not lead to consistent performance improvements. While certain configurations, such as GRU in the single-task setting, show minor gains with longer windows, this trend is not general. In most cases, extending the window beyond 30 minutes yields only marginal benefits, suggesting that a moderate historical context is sufficient for capturing the relevant temporal dynamics. A similar observation emerges when comparing single-task and multi-task learning. Although single-task models achieve slightly better performance at the shortest horizon, the difference remains small. Multi-task models maintain competitive accuracy across all horizons while significantly reducing model redundancy, as a single model can produce multiple forecasts. Overall, increasing the input window beyond 30 minutes and using dedicated horizon-specific models yield only limited performance improvement. As a result, moderate window sizes seem sufficient for the evaluated forecasting task. Future work will be extended to multiple datasets and different environmental conditions to assess generalizability. In addition, multi-modal inputs will also be investigated utilizing both imaging and environmental sensor data. In the future, advanced architectures such as CNN- and Transformer-based models will also be explored.

ACKNOWLEDGEMENTS

This research was conducted with the financial support of Research Ireland under Grant Agreement No. 13/RC/2106_P2 at the ADAPT Centre at University College Dublin. ADAPT, the Research Ireland Centre for AI-Driven Digital Content Technology, is funded through the Research Ireland Centres Programme. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. The authors used ChatGPT (OpenAI) during manuscript preparation for language editing and text organization. All AI-assisted content was reviewed, revised, and verified by the authors, who take full responsibility for the accuracy and integrity of the manuscript.

REFERENCES

- [1] Kumar, D. S., Yagli, G. M., Kashyap, M., and Srinivasan, D., “Solar irradiance resource and forecasting: a comprehensive review,” *IET Renewable Power Generation* 14(10), 1641–1656 (2020).
- [2] Ajith, M. and Martínez-Ramón, M., “Deep learning algorithms for very short term solar irradiance forecasting: A survey,” *Renewable and Sustainable Energy Reviews* 182, 113362 (2023).
- [3] National Renewable Energy Laboratory, “Solar and wind forecasting,” (2024). Accessed: 2024-06-08.
- [4] International Renewable Energy Agency (IRENA), “World energy transitions outlook 2022,” (2022). Accessed: 2024-06-08.
- [5] Yang, D., Wang, W., Gueymard, C. A., Hong, T., Kleissl, J., Huang, J., Perez, M. J., Perez, R., Bright, J. M., Xia, X., et al., “A review of solar forecasting, its dependence on atmospheric sciences and implications for grid integration: Towards carbon neutrality,” *Renewable and Sustainable Energy Reviews* 161, 112348 (2022).
- [6] Kostylev, V., Pavlovski, A., et al., “Solar power forecasting performance—towards industry standards,” in 1st international workshop on the integration of solar power into power systems, Aarhus, Denmark, 1–8, Energynautics GmbH Mühlenstraße Langen, Germany (2011).
- [7] Terrén-Serrano, G. and Martínez-Ramón, M., “Deep learning for intra-hour solar forecasting with fusion of features extracted from infrared sky images,” *Information Fusion* 95, 42–61 (2023).
- [8] Al-Ali, E. M., Hajji, Y., Said, Y., Hleili, M., Alanzi, A. M., Laatar, A. H., and Atri, M., “Solar Energy Production Forecasting Based on a Hybrid CNN-LSTM-Transformer Model,” *Mathematics* 11(3), 676 (2023).
- [9] Nijhum, I. R., Hossain, M. Y., Rakib, M. M. H., and Dev, S., “On improving solar forecasts with enhanced IR imaging and CNNs,” in *IET Conference Proceedings CP887*, 2024(10), 95–102, IET (2024).
- [10] Nijhum, I. R., Kenny, P., and Dev, S., “Intra-Hour Solar Irradiance Estimation Using Infrared Sky Images and MobileNetV2-Based CNN Regression,” in *2023 IEEE 7th Conference on Energy Internet and Energy System Integration (EI2)*, 3534–3539 (2023).
- [11] Nijhum, I. R., Hossain, M. Y., Akinola, T., Rakib, M. M. H., and Dev, S., “A Hybrid CNN-LSTM Framework and Infrared Image Processing for Solar Irradiance Forecasting,” in *17th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI 2024)*, 26–28 (October 2024).
- [12] Yu, Y., Cao, J., and Zhu, J., “An LSTM Short-Term Solar Irradiance Forecasting Under Complicated Weather Conditions,” *IEEE Access* 7, 145651–145666 (2019).
- [13] Perera, M., De Hoog, J., Bandara, K., and Halgamuge, S., “Multi-resolution, multi-horizon distributed solar PV power forecasting with forecast combinations,” *Expert Systems with Applications* 205, 117690 (2022).
- [14] Boussif, O., Boukachab, G., Assouline, D., Massaroli, S., Yuan, T., Benabbou, L., and Bengio, Y., “Improving day-ahead Solar Irradiance Time Series Forecasting by Leveraging Spatio-Temporal Context,” *Advances in Neural Information Processing Systems* 36, 2342–2367 (2023).
- [15] Zhang, Z., Wang, J., Xia, Y., Wei, D., and Niu, Y., “Solar-Mixer: An Efficient End-to-End Model for Long-Sequence Photovoltaic Power Generation Time Series Forecasting,” *IEEE Transactions on Sustainable Energy* 14(4), 1979–1991 (2023).
- [16] Terrén-Serrano, G., Bashir, A., Estrada, T., and Martínez-Ramón, M., “Girasol, a sky imaging and global solar irradiance dataset.” *Dryad* (2021).