# Uncertainty Quantification for Global Horizontal Irradiance Prediction

Florian Comte (346006), Oscar de Francesca (388476), David Gauch (394014)
*Department of Computer Science, EPFL, Switzerland*

***Abstract*—Uncertainty quantification in short-term global horizontal irradiance (GHI) prediction is essential for optimizing renewable energy systems and managing power grids. This work proposes two complementary approaches: quantile regression and Bayesian neural networks (BNNs), leveraging webcam imagery and meteorological features to enhance predictive accuracy and robustness. Experimental results show that quantile regression provides sharper, well-calibrated prediction intervals with high computational efficiency, particularly when meteorological data is incorporated. BNNs, on their side, yield a broader view of uncertainty producing wider intervals and requiring bigger computational resources.**

## I. Introduction

Accurate short-term global horizontal irradiance (GHI) forecasting is beneficial for optimizing renewable energy systems, balancing power grids, and ensuring the efficient integration of solar energy into modern energy networks. However, traditional deterministic models only produce single-point predictions, failing to account for the uncertainty inherent in GHI forecasts. This lack of uncertainty quantification can result in unreliable predictions, which, in turn, compromise grid stability, resource allocation, and decision-making processes.

To address this limitation, we build upon the work of Roy Sarkis and the Laboratory of Applied Photonics Devices (LAPD) at EPFL [1], introducing uncertainty quantification into short-term GHI forecasting. By explicitly modeling uncertainty, forecasts can better account for variability in environmental conditions, leading to more informed and actionable predictions for stakeholders in renewable energy.

The primary objectives of this project are:

1) To transition from deterministic single-point predictions to prediction intervals with a 95% confidence level, quantifying forecast uncertainty.
2) To incorporate meteorological features alongside webcam image data, improving both accuracy and robustness of the predictions.

To achieve these goals, we investigate two separate approaches for uncertainty quantification:

- **Bayesian Neural Networks (BNNs):** A probabilistic method that models neural network weights as distributions rather than fixed values, producing confidence intervals through Monte Carlo sampling.
- **Quantile Regression:** A direct approach for predicting specific quantiles (e.g., 2.5%, 50%, and 97.5%), enabling the estimation of prediction intervals while maintaining computational efficiency.

## II. Dataset and Preprocessing

### A. Dataset Description

The project utilizes two complementary datasets: webcam images and meteorological measurements, both collected on the EPFL campus. The webcam images are captured from two distinct angles, providing diverse visual perspectives of atmospheric conditions. The meteorological dataset includes key features influencing global horizontal irradiance (GHI), such as:

- Air Temperature,
- Cloud Opacity,
- Precipitable Water,
- Relative Humidity, and
- Zenith.

These variables are widely recognized for their relevance in solar energy modeling. Both image datasets as well as the meteorological dataset were partitioned into training, validation, and test subsets to enable robust model evaluation and ensure generalization to unseen data. Additionally, we made use of label files containing GHI values that were also divided into training, validation, and subset files. Specifically, for each image collected at a given time, the GHI value for 2 hours later is provided, ensuring a structured temporal alignment between inputs and targets.

### B. Preprocessing

A systematic preprocessing pipeline was implemented to clean, normalize, and prepare the data for training. The primary steps included:

*Image Preprocessing:* Webcam images were resized to a standardized resolution of $224 \times 224$ pixels, ensuring compatibility with convolutional neural network (CNN) architectures while retaining sufficient spatial information. Pixel values were normalized to the range $[0, 1]$ and further standardized using the mean and standard deviation computed over the training set:

$$\text{Normalized Pixel} = \frac{\text{Pixel Value} - \text{Mean}}{\text{Standard Deviation}}.$$

*Meteorological Feature Processing:* Meteorological features were standardized to ensure uniform scaling and alignment with the optimization requirements of machine learning models. Specifically, each feature was centered and scaled using the training set statistics:

$$\text{Standardized Value} = \frac{\text{Feature Value} - \text{Mean}}{\text{Standard Deviation}}.$$

GHI values, serving as the target variable, were normalized using min-max scaling to the $[0, 1]$ range to ensure numerical stability during model training.

*Handling Missing Data:* Missing values in the meteorological dataset, though minimal, were removed entirely to maintain data integrity and avoid introducing biases through imputation.

*Batch Processing:* To efficiently handle the dataset size, all preprocessing operations, including normalization and resizing, were performed in batches of 32 samples. This approach optimized computational performance and memory usage.

*Feature Selection:* Domain knowledge guided the selection of meteorological features, focusing on those with established influence on GHI predictions. Iterative testing of feature combinations further refined the selected subset, balancing predictive performance with computational efficiency.

## III. Methods and Models

### A. Quantile Regression

Quantile regression was employed to estimate prediction intervals, enabling the model to capture the range of possible GHI values. This approach transitions from single-point predictions to interval-based predictions, offering a more informative and robust output [2].

*Model Adaptation:* The original deterministic CNN-LSTM architecture was adapted to combine features from images and meteorological data effectively. The proposed neural network architecture integrates CNN, LSTM, and MLP components for quantile regression, designed to process paired input images and optional meteorological data. The CNN module comprises four convolutional layers with a kernel size of $3 \times 3$ and ReLU activations: two layers with 16 channels, followed by two with 32 channels. Batch normalization (`affine=False`) and max-pooling (kernel size 2) are applied after each convolutional block, with dropout rates of 0.3 and 0.1 for regularization. The meteorological data, when provided, is processed through an MLP with two fully connected layers (32 neurons each, ReLU activation, and 0.1 dropout). The combined features are passed through two stacked LSTM layers with 128 and 64 hidden units, respectively. A final MLP, comprising a hidden layer (64 neurons, ReLU) and an output layer, maps the features to the desired number of quantiles.

*Loss Function:* The quantile-specific pinball loss was implemented to optimize predictions at each quantile[3]. This asymmetric loss penalizes under- and over-predictions differently, encouraging accurate quantile estimation:

$$\mathcal{L}_q(y, \hat{y}) = \begin{cases} q \cdot (y - \hat{y}), & \text{if } y \geq \hat{y}, \\ (1 - q) \cdot (\hat{y} - y), & \text{if } y < \hat{y}, \end{cases}$$

where $q$ is the quantile, $y$ is the ground truth, and $\hat{y}$ is the predicted value.

*Non-Negativity Constraint:* To ensure physically realistic predictions (i.e., GHI values cannot be negative), predictions were clipped to zero during inference. Unlike ReLU, which restricts outputs during training and can hinder learning of lower quantiles, clipping allows the model to learn freely, including slight negative predictions that help optimize the loss. The non-negativity constraint is applied only at inference, striking a balance between flexibility during training and domain-specific constraints at test time. This approach proved computationally efficient and yielded superior results.

### B. Bayesian Neural Network (BNN)

The second model we tried is a BNN where the weights are treated as probability distributions rather than fixed values. This allows the model to propagate uncertainty through its layers, providing a principled estimate of predictive variability. We used the same model as the Quantile Regression except for the last layer for which we only predict one value and not the 3 quantiles.

*Probabilistic Framework:* The IntelLabs Bayesian Torch library [4] was used to replace deterministic weight layers with Bayesian equivalents. During inference, Monte Carlo (MC) sampling was performed by running multiple stochastic forward passes for the same input. The resulting outputs were aggregated to compute the mean prediction and standard deviation, forming the basis for prediction intervals.

*Training Procedure:* Training the BNN involved minimizing a combination of the standard regression loss and a Kullback-Leibler (KL) divergence term. The KL loss regularizes the learned posterior distributions, ensuring they remain close to the prior while fitting the training data. The overall loss is expressed as:

$$\mathcal{L} = \mathcal{L}_{\text{data}} + \beta \cdot D_{\text{KL}}(q(w)||p(w)),$$

where $q(w)$ is the posterior distribution of the weights, $p(w)$ is the prior, and $\beta$ is a scaling factor to balance the two terms.

*Prediction Intervals:* By aggregating multiple MC samples, the BNN provides a mean prediction and confidence intervals derived from the predictive uncertainty.

### C. Training and Evaluation Techniques

Given the computational demands, hyperparameter tuning was performed manually. This process involved iteratively adjusting learning rates, batch sizes, and the inclusion of meteorological features, with results monitored through observed loss and qualitative predictions. Although manual tuning restricted the search space, it allowed for focused improvements within the constraints of long training times and resource availability.

To expedite the training process, an early stopping mechanism was implemented, halting training when no significant improvements in loss were observed across epochs. This approach mitigated overfitting and reduced unnecessary computations, particularly for the BNN.

The models were trained using the Adam optimizer with a learning rate of 0.003 for both models and a weight decay of 0.005 for quantile regression and 0 for bayesian neural network. Training was performed for 200 epochs with a batch size of 32, and early stopping was applied with a patience of 30 epochs and a minimum decrease in validation loss of 0.3% for both models. For the BNN, we used 3 samples for Monte Carlo sampling during training and 50 during the testing.

*Evaluation Process:* Model evaluation focused on visual analysis of prediction intervals on plots and quantitative evaluation of the model's performance focusing on 3 metrics:

1. Percentage within bounds: This metric calculates how often the true GHI falls within the predicted lower and upper bounds. Mathematically, it is defined as:

$$\text{Percentage within bounds} = \frac{1}{N}\sum_{i=1}^{N} I\left(\hat{y}_i \in [L_i, U_i]\right) \times 100$$

where:

- $N$ is the total number of samples,
- $y_i$ is the true value for sample $i$,
- $L_i$ and $U_i$ are the lower and upper bounds for sample $i$,
- $I(\cdot)$ is the indicator function, which equals 1 if the true value is within the bounds, and 0 otherwise.

2. Average certainty: This metric quantifies the model's confidence in its predictions, with narrower intervals reflecting higher certainty. It is computed as the average of the normalized width of the prediction intervals, defined as:

$$\text{Average certainty} = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{1 + (U_i - L_i)}$$

where:

- $U_i - L_i$ is the width of the prediction interval for sample $i$,
- The normalization factor $(1 + (U_i - L_i))$ ensures that narrower intervals result in higher certainty.

3. Mean Squared Loss: This metric measures the average squared difference between the model's predictions and the true values. It quantifies the overall accuracy of the model's predictions, where smaller values indicate better performance. Mathematically, MSE is defined as:

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N} \left(\hat{y}_i - y_i\right)^2$$

where:

- $N$ is the total number of samples,
- $\hat{y}_i$ is the predicted value for sample $i$,
- $y_i$ is the true value for sample $i$.

These metrics provide a comprehensive assessment of both the model's **accuracy** (via the percentage within bounds and MSE) and its **confidence** in the predictions (via the average certainty).

## IV. Results

### A. Model Performance

The quantile regression and Bayesian neural network (BNN) approaches were evaluated for their ability to capture uncertainty and produce realistic prediction intervals.

*Quantile Regression:* Quantile regression demonstrated reliable performance with well-calibrated prediction intervals and accurate median predictions. We analyzed its behavior with and without meteorological features, as shown in Figures 1 and 2.

- **With Meteorological Data (Figure 1):** Incorporating meteorological features such as cloud opacity, direct normal irradiance (DNI), diffuse horizontal irradiance (DHI), precipitable water, and solar zenith angle resulted in a clear improvement in prediction quality. The intervals tightly captured the variability in true GHI values with minimal deviations in median predictions.
- **Without Meteorological Data (Figure 2):** Omitting meteorological inputs significantly degraded performance. The prediction intervals became wider, and the median predictions deviated more from the true values, underscoring the importance of meteorological features in reducing uncertainty.

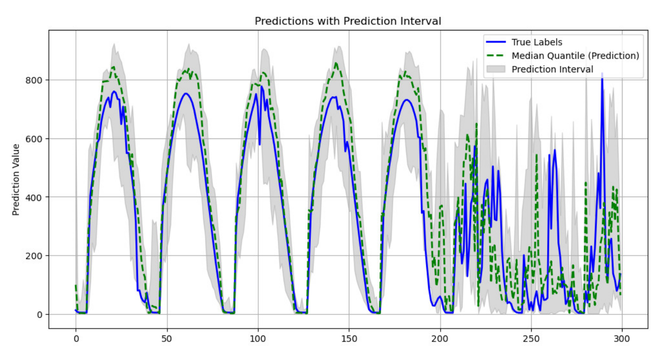


Figure 1. Quantile Regression with meteorological data: True values, prediction intervals, and median predictions over 300 samples.

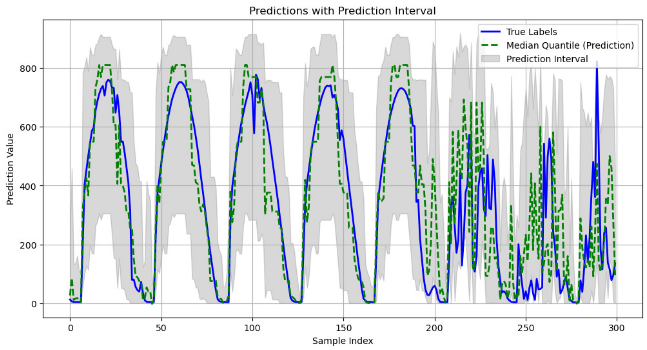


Figure 2. Quantile Regression without meteorological data: True values, prediction intervals, and median predictions over 300 samples.

*Bayesian Neural Network:* The BNN approach provided accurate predictions but exhibited broader prediction intervals compared to quantile regression, as shown in Figure 3.

- The BNN successfully captured the uncertainty inherent in the data; however, the intervals were often excessively wide. This suggests overestimation of uncertainty, likely stemming from over-regularization of the weight distributions.
- Computationally, BNNs proved more demanding due to the repeated Monte Carlo sampling required during inference.

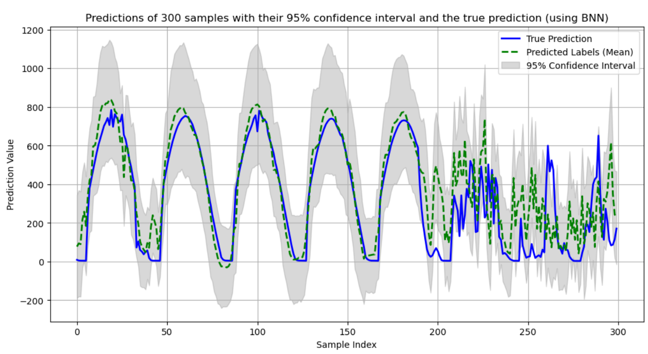


Figure 3. Bayesian Neural Network Predictions with meteorological data: True values, prediction intervals, and predicted means over 300 samples (3 samples used for Monte Carlo sampling during training and 50 during the testing).

The table I summarizes the quantitative results.

| Model | Percentage Within Bounds (%) | Average Certainty | MSE |
|---|---|---|---|
| Quantile Regression (With Meteorological Data) | 95.13% | 0.7474 | 0.0214 |
| Quantile Regression (Without Meteorological Data) | 94.93% | 0.6533 | 0.0329 |
| Bayesian Neural Network (With Meteorological Data) | 92.22% | 0.6624 | 0.0244 |
| Bayesian Neural Network (Without Meteorological Data) | 96.55% | 0.5432 | 0.0200 |

Table I
PERFORMANCE COMPARISON OF THE MODELS

## V. Discussion and Conclusion

This project investigated uncertainty quantification for short-term global horizontal irradiance (GHI) forecasting using two approaches: quantile regression and Bayesian neural networks (BNNs). Both methods demonstrated their strengths and limitations.

Quantile regression provided sharp, well-calibrated prediction intervals with high computational efficiency, particularly when incorporating meteorological data. Its ability to produce interpretable and reliable intervals makes it ideal for real-time applications where accuracy and efficiency are paramount. In contrast, the Bayesian neural network offered a more principled probabilistic framework for uncertainty quantification but produced significantly broader intervals, likely due to weight over-regularization. While the BNN captured a more comprehensive view of uncertainty, its higher computational cost and longer training times limit its practical deployment.

### A. Future Work

Future work should focus on integrating the strengths of both methods. A hybrid approach combining quantile regression's sharp intervals with Bayesian techniques' uncertainty estimates could enhance robustness. Additionally, we observed promising preliminary results with a pretrained ResNet18 model (average certainty of 0.8197, MSE of 0.0223 and 79.04% of true GHI within the bounds with only 14 epochs), which warrants further exploration. Given more time and computational power, we could optimize hyperparameters and explore ensemble techniques to improve accuracy and reliability. Expanding the dataset to include richer meteorological features or alternative probabilistic models could further advance GHI forecasting performance.

## VI. Ethical Risks

A key ethical concern in this project is data privacy. Since webcam images are used for GHI prediction, there is a potential risk that identifiable information, such as people or vehicles, could be captured, raising privacy concerns. Individuals inadvertently recorded in these images may face risks related to unauthorized data usage or exposure of private information. Given the increasing importance of data protection laws like GDPR, this risk is significant, especially when working with publicly captured visual data.

The Laboratory of Applied Photonics Devices (LAPD) at EPFL has already addressed this concern by implementing a privacy-preserving method. They used OpenCV's pre-trained YOLO v3 model for object detection, which identifies individuals in the raw images and applies a Gaussian blur to anonymize them. This approach ensures privacy protection without compromising model performance. Due to time constraints and the significant training time required, we decided not to implement it in our current project. However, given its proven effectiveness, it could easily be incorporated into our system to provide similar privacy safeguards.

Additionally, the LAPD team explored using cropped images focused solely on the sky to further protect privacy. However, this approach significantly impacted model accuracy due to the loss of critical information like shadows and reflections. Given these findings, the blurring method remains the preferred solution for balancing privacy and prediction quality, and it could be implemented in our project in the future.

By incorporating such privacy-preserving measures, as demonstrated by the LAPD, our project could ensure compliance with data protection regulations and mitigate the risk of privacy breaches.

## Acknowledgments

We extend our deepest gratitude to Roy Sarkis and Yanan Niu from the Laboratory of Applied Photonics Devices (LAPD) at EPFL for their invaluable support, insights, and resources throughout this project. This work was carried out as part of the EPFL Machine Learning course, and we are especially grateful to Professors Nicolas Flammarion and Martin Jaggi for their guidance and expertise. We also wish to thank the teaching assistants and student assistants for their dedicated feedback and assistance.

## References

[1] R. Sarkis, I. Oguz, D. Psaltis, M. Paolone, C. Moser, and L. Lambertini, "Intraday solar irradiance forecasting using public cameras," *Solar Energy*, vol. 275, p. 112600, 2024. [Online]. Available: https://doi.org/10.1016/j.solener.2024.112600

[2] S. Matsumoto, "Quantile Loss in Neural Networks. Implementation, results tendency, and…" [Online]. Available: https://shrmtmt.medium.com/quantile-loss-in-neural-networks-6ea215fcee99

[3] C. Molnar, "How I made peace with quantile regression," Jan. 2024. [Online]. Available: https://mindfulmodeler.substack.com/p/how-i-made-peace-with-quantile-regression

[4] Intel Labs, "Bayesian torch: Bayesian neural networks for pytorch," https://github.com/IntelLabs/bayesian-torch/tree/main, 2024, accessed: 2024-12-10.