High-Resolution Solar Irradiance Prediction Using Variational Mode Decomposition and Entropy-Guided Multi-Expert Deep Learning Under Different Sky Conditions

Lamyaa Sabeeh Ashour
Department of Mathematics, College of Basic Education, University of Misan, Iraq

doi: https://doi.org/10.37745/ijncr.16/vol11030 Published September 28, 2025

Citation: Ashour L.S. (2026) High-Resolution Solar Irradiance Prediction Using Variational Mode Decomposition and Entropy-Guided Multi-Expert Deep Learning Under Different Sky Conditions, *International Journal of Network and Communication Research*, Vol.10, No.1, pp.10-30

Abstract: *Short-term solar irradiance prediction is investigated using regularly sampled measurements at one-minute or coarser resolutions, while decomposition-based learning on event-recorded, high-resolution measurements remains less explored. This study evaluates an entropy-guided multi-expert deep-learning framework combining variational mode decomposition (VMD), Sample Entropy, K-means clustering, and Neural Networks for high-resolution global horizontal irradiance (GHI) prediction. The framework uses the Natural Resources Canada high-resolution solar radiation dataset, comprising event-recorded measurements from two sensor networks in Quebec and Ontario. Eight observation days representing clear, overcast, variable, and very-variable conditions yield 164 sensor-day GHI series. Each series is decomposed into five modes using VMD; Sample Entropy is calculated for each mode, and K-means forms three complexity levels. Modes are assigned to multilayer perceptron (MLP), long short-term memory (LSTM), or one-dimensional convolutional neural network (1D-CNN) experts. Five mode-specific models are trained for each series under a fixed configuration, producing 820 trained model instances. Across the eight combinations, mean RMSE ranges from 0.57 to 6.40 W/m². Mean irradiance and relative error show a rank correlation of −0.933. The lowest-frequency mode (IMF1) was assigned to the MLP in all 164 series, while across all 820 IMF instances, 45.2% were assigned to the MLP, 43.0% to the LSTM, and 11.7% to the 1D-CNN. Because scaling and VMD precede chronological splitting and no baseline is included, results represent an empirical evaluation of the offline pipeline rather than strictly causal performance*.
Keywords: Solar irradiance prediction, Variational mode decomposition, Sample entropy, High-resolution measurements, Sensor Networks, Deep learning, K-means clustering, LSTM, 1D-CNN

INTRODUCTION

Solar photovoltaic (PV) generation is now the main driver of renewable capacity growth. The International Energy Agency expects new solar capacity to account for about 80% of the global growth in renewable power by 2030 [1]. As PV penetration rises, the variability of its output becomes an operational concern rather than a secondary detail. PV output follows the irradiance reaching the panels, and that irradiance can change sharply when clouds pass over a site. D. Yang et al. [2] reviewed how these fluctuations complicate grid integration and argued that forecasting is one of the main tools for managing them.

Much of this variability occurs at very short time scales. A moving cloud edge can reduce or enhance irradiance within seconds, and such changes are poorly represented by the hourly or even one-minute averages used in many solar studies. Chu et al. [3] linked the need for intra-hour irradiance forecasting to the operating challenges that intermittent solar generation creates for local and regional grids. The temporal resolution of the measurements themselves also matters. Using the Canadian sensor networks examined in this study, Gagné et al. [4] found that recording periods longer than about 400 ms can underestimate short-term variability. High-resolution irradiance data therefore carry information that coarser records lose.

Data-driven models have become the dominant approach to short-term irradiance prediction. Deep learning models such as long short-term memory (LSTM) networks and Convolutional Neural Networks (CNNs) are widely used for very short-term irradiance forecasting because they learn temporal dependencies directly from past observations [5,6]. Irradiance series are, however, nonstationary and combine slow diurnal trends with rapid cloud-induced fluctuations. A common response is to decompose the signal before learning. Variational mode decomposition (VMD) separates a signal into a fixed number of band-limited mode [7]. Hybrid models then predict each mode separately and sum the predictions to reconstruct the irradiance. Recent examples include VMD combined with temporal convolutional networks [8], with LSTM–Transformer models by Z. Liu and Zhao [9] and Sarango et al. [10], and with sequence-to-sequence models by Sibtain et al. [11].

Data-driven models have become the dominant approach to short-term irradiance prediction. Deep learning models such as LSTM and CNNs are Decomposition raises a further design question: should every mode be handled in the same way. The modes produced by VMD differ in frequency content and regularity, and several studies use entropy measures to describe these differences. Sample entropy quantifies the irregularity of a time series [12]. It and related measures have been used to merge similar modes, select informative components or build input features by Z. Liu and Zhao [9], Sibtain et al. [11], Wang et al. [8], and Zemouri et al. [13]. A smaller group of studies goes further and assigns different predictors to different kinds of modes. J. Liu et al. [14] predicted low-frequency VMD components with a CNN-LSTM-MLP model and high-frequency components with deep belief networks. Xue and Li [15] used sample entropy and K-means clustering to separate high- and low-frequency components of PV power before sending them to different learners. These studies suggest that matching model type to component complexity is a reasonable design principle.

The evidence behind this principle comes almost entirely from regularly sampled data at coarse resolution. The studies above work with hourly irradiance [9-14] or with PV and irradiance records at 5- to 60-minute resolution [10-15]. How decomposition and entropy-guided model assignment behave on sub-second, event-recorded irradiance is much less documented, even though this is where cloud-driven fluctuations are most pronounced. Most studies also report results for one or a few stations and do not examine how prediction error varies across many co-located sensors or across distinct sky regimes. Data availability adds a practical constraint. In a recent meta-review of 72 review articles on solar irradiance forecasting, only 14 of the benchmark datasets identified provided a direct access link [16], which limits the extent to which results can be checked or extended.

The high-resolution solar radiation datasets released by Natural Resources Canada (NRCan) are well suited to these questions. They were collected by CanmetENERGY with 41 synchronised photodiode pyranometer units: 17 at Varennes, Quebec, and 24 inside a solar farm at Alderville, Ontario. The data are publicly available for one clear-sky, one overcast, one variable and one very variable day at each site [17]. The logger samples irradiance every millisecond, averages over 10 ms, and saves a value whenever it changes by more than 5 W/m², and at least once per minute [17]. The resulting series are event-driven: dense during cloud transitions and sparse under stable skies. Despite these features, only one forecasting study that uses the dataset was identified. Sanchez-Lopez et al. [18] predicted irradiance 15 and 30 s ahead at Alderville using tree-based ensembles, with inputs selected from spatially and temporally correlated sensors. No previous study has applied decomposition-based deep learning to this dataset or evaluates the Varennes days for prediction.

This study addresses that gap with a framework that combines VMD, sample entropy and heterogeneous neural experts. For each sensor and day, the global horizontal irradiance (GHI) series is normalised and decomposed into five modes. The sample entropy of each mode is computed, and the entropy values are grouped with K-means into low-, intermediate- and high-complexity levels. Each mode is then assigned to a multilayer perceptron (MLP), LSTM or one-dimensional CNN expert according to its group. The mode predictions are summed and rescaled to obtain the GHI prediction. The framework is applied independently to all 164 sensor-day series of the NRCan dataset. Because the data are event-driven, the task is defined as one-step-ahead prediction in the ordered observation sequence from a window of 20 past observations, rather than as a forecast at a fixed horizon in seconds. The evaluation is offline: normalisation and decomposition are applied to each series before the chronological partition into training, validation and test sets. The results should therefore be read as evidence of how the integrated framework behaves on high-resolution data, not as the accuracy of a real-time forecasting system.

The primary objective is to evaluate this entropy-guided multi-expert framework for one-step-ahead prediction of high-resolution GHI across different sky conditions. Four questions follow from that objective: how the framework performs under clear, overcast, variable and very variable conditions; how performance differs between the Alderville and Varennes networks; how the entropy-guided assignment of modes to experts changes across sites and sky conditions; and how much prediction error varies between individual sensors within the same site and day.

The novelty of this study is not the introduction of a new neural network architecture. Rather, it lies in evaluating an entropy-guided assignment strategy that links the measured complexity of VMD modes to heterogeneous prediction experts, and in characterizing how this assignment behaves across high-resolution, event-recorded irradiance collected from two sensor networks and four labelled sky conditions. The study therefore focuses on empirical characterization rather than architectural superiority. This distinction is important because the objective is to determine how mode complexity, expert assignment, irradiance level and sensor-level variability interact in this high-resolution setting.

The study makes four contributions. First, it documents an integrated VMD sample entropy–multi-expert framework in which the choice of learner for each mode follows from its measured complexity. Second, it applies a decomposition-based deep learning framework to both NRCan high-resolution networks, which has not been reported in earlier work, covering 164 sensor-day series from two sites and four sky conditions. Third, it reports sensor-level errors and the resulting expert assignments rather than only site averages, which shows how prediction difficulty and component complexity change with site and sky regime. Fourth, it states the limits of the evaluation openly, including the offline decomposition, the event-driven observation step and the instability of percentage errors at low irradiance, so that later studies can build on the results with stricter protocols. The remainder of the paper reviews related work (Section 2), describes the data and methodology (Section 3) and the experimental setup (Section 4), presents the results (Section 5), discusses them (Section 6) and concludes (Section 7).

LITERATURE REVIEW

Solar radiation modelling and prediction

Very short-term irradiance prediction covers horizons from a few seconds to under an hour, and the methods used depend strongly on both the horizon and the data available at the site. Chu et al. [3] classified intra-hour techniques into regression, stochastic learning, deep learning, and evolutionary approaches. These methods can be applied through purely data-driven models, local sensing systems or hybrids of the two. At horizons of several minutes, sky images are a common source of information. For example, Hendrikx et al. [19] combined all-sky images taken every 15 s with on-site measurements and LSTM networks to predict GHI up to 20 min ahead. Ground sensor networks provide a different kind of local information because a cloud shadow reaches neighboring sensors at slightly different times. Sanchez-Lopez et al. [18] used this effect at Alderville, selecting input sensors through temporal and spatial correlation analysis before training tree-based ensembles for 15 s and 30 s horizons. Measurement resolution limits all of these approaches. Gagné et al. [4] using the same Canadian networks as in this study, showed that short-term variability is underestimated when the recording period exceeds about 400 ms. Many recent prediction studies nevertheless use one-minute or coarser records [6-19]. How prediction models behave on sub-second measurements has therefore received less attention.

Machine learning approaches for solar radiation prediction

Among data-driven approaches, deep learning is now the most common choice. Ajith and Martínez-Ramón [5] studied deep learning algorithms for very short-term irradiance forecasting, including recurrent and convolutional architectures. LSTM networks model temporal dependencies through gated memory cells. 1D-CNNs learn local patterns within a short input window. MLPs provide a simpler nonlinear mapping with no explicit temporal structure. Reported enhancements over baseline models at very short horizons are often modest. Barancsuk et al. [6] reported a forecast skill of about 10% relative to persistence for an LSTM-based model at horizons of 1–20 min. Hendrikx et al. [19] found that their LSTM models outperformed persistence , particularly beyond horizons of about 4–5 min. Together, these findings suggest that at the shortest horizons, the most recent observation already contains much of the predictable information. This makes the choice of baseline model and evaluation protocol central to how results are interpreted.

Decomposition-based hybrid models address the nonstationarity of irradiance by splitting the signal into simpler components before prediction. VMD decomposes a signal into a prescribed number of band-limited modes by solving a constrained variational problem, in which each mode is concentrated around an estimated centre frequency [7]. Because the number of modes and the bandwidth penalty are set in advance, VMD returns a fixed set of components, and one predictor can be built for each. VMD has been paired with a wide range of learners for solar prediction. Wang et al. [8] integrated an optimized VMD with a temporal convolutional network tuned by a metaheuristic algorithm and reported higher accuracy than nine comparison models. Sarango et al. [10] combined a five-mode VMD with LSTM and Transformer components for 1–12 h irradiance forecasts in mountainous terrain and compared the hybrid with persistence, climatology, and single-model baselines. Z. Liu and Zhao [9] applied VMD to hourly irradiance from four Chinese cities before an LSTM–Transformer model. Sibtain et al. [11] built VMD- and ICEEMDAN-based variants of a sequence-to-sequence model with spatio-temporal attention and reported the best results for the VMD variant. The studies differ in horizon, data resolution, the way the number of modes is chosen (fixed in some, optimized in others), and the learner used. They share the premise that separating slow and fast components makes each easier to learn.

Once a series has been decomposed into several modes, the next question is how each mode should be modelled. Entropy measures are the most common tool for describing the modes. Sample entropy estimates the irregularity of a series as the negative logarithm of the conditional probability that patterns similar over m points remain similar at m + 1 points, with self-matches excluded [12]. Low values indicate regular behavior, whereas high values irregular behaviour. In solar studies, entropy has mainly been used to reduce or reorganise the set of components. Wang et al. [8] and Z. Liu and Zhao [9] used fuzzy entropy to merge modes of similar complexity into fewer reconstructed components. Zemouri et al. [13] used sample entropy to retain the most informative component and discard redundant or noisy ones, and Sibtain et al. [11 incorporated sample entropy within their hybrid models.

A second line of work uses component characteristics to decide which learner handles which component. J. Liu et al. [14] predicted the low-frequency VMD components of hourly irradiance with a CNN-LSTM-MLP model, predicted the high-frequency components with deep belief networks, and added a stepwise error correction. Xue and Li [15] computed the sample entropy of the PV power modes obtained through a two-stage decomposition and grouped them into high- and low-frequency sets with K-means. They then predicted the two sets with an online kernel extreme learning machine and a CNN combined with an echo state network, respectively. The framework evaluated in this study follows the same principle but uses three complexity groups and three neural experts. In both earlier studies, the assignment rule is embedded within a larger pipeline that also includes secondary decomposition or error correction. Its separate effect is therefore difficult to isolate, a point that also applies to the present work.

Research Gap and Motivation

The evaluation of such models has received growing attention. D. Yang et al. [20] recommended comparing deterministic solar forecasts with a persistence-based reference and summarising them with a root mean square error (RMSE) skill score, so that results can be compared across sites and horizons. Error measures also behave differently at low irradiance. Chicco et al. [21] noted that the mean absolute percentage error (MAPE) is biased towards low forecasts and is defined only for strictly positive data. This becomes a practical problem when irradiance approaches zero near sunrise, sunset or under thick cloud. A further concern is specific to decomposition-based models. When the whole series is decomposed before it is split into training and test sets, the components of the training period are computed using information from the test period. X. Yang et al. [22] demonstrated this effect for empirical mode decomposition models and proposed sliding-window and repeated-decomposition schemes to avoid it. Decomposing before partitioning nonetheless remains common in solar studies, as reported by Sarango et al. [10]. These points do not invalidate earlier results. They mean that reported accuracy depends on the protocol, and that studies need to state whether scaling, decomposition and evaluation are causal. The present study states its protocol accordingly.

Research on the NRCan high-resolution data has focused mainly on variability rather than prediction. Gagné et al. [4] suggested the 41 units at Varennes and Alderville to quantify variability through the variability index, the daily clear-sky index and the variability score. They also related the averaged variability index to cloud speed and to the area covered by the sensors. The dataset documentation describes site differences relevant to modelling [17]. Some Varennes units stand close to trees and power lines that can shade them. The Alderville units are in a more open area, where shading events of this kind are less frequent. At both sites, birds or insects can briefly shade a sensor. The only prediction study identified was conducted by Sanchez-Lopez et al. [18], using four days of data from the 24 Alderville sensors. They inferred the wind direction from correlations between sensors, used it to select the most relevant sensors, and trained XGBoost, LightGBM and random forest models with training times short enough for near-real-time retraining. The dataset is also listed among country-specific solar data sources in a recent meta-review [16]. No previous study has applied decomposition-based models or deep learning experts to these data, or used the Varennes days for prediction.

The literature therefore supports three observations. First, decomposition-based hybrids, often guided by entropy measures, are an established way to handle nonstationary irradiance, and assigning learners according to component characteristics has been proposed in a small number of recent studies [14, 15]. Second, this evidence comes from regularly sampled data at resolutions between 5 min and 1 h. The fluctuations that motivate very short-term prediction, by contrast, occur at sub-minute scales that only high-resolution measurements capture [4]. Third, the NRCan dataset provides such measurements for two sensor networks and four labelled sky conditions, but it has been used for prediction only with tree-based models at Alderville [18]. What remains undocumented is how a VMD framework with entropy-guided assignment of heterogeneous neural experts behaves on event-recorded, sub-second irradiance. It is also unknown how its errors and expert assignments change between sky regimes and the two sites, and how much they vary among co-located sensors. The present study addresses these questions. In line with the evaluation concerns above, it reports the framework as an integrated pipeline under an offline protocol. It makes no claim of improvement over baseline models because baselines were not part of the experiments.

MATERIALS AND METHODS

Overview of the methodology

Figure 1 shows the overall framework of the proposed architecture. Each sensor-day series is treated independently and passes through the same sequence of stages. The measured global horizontal irradiance is first cleaned and scaled to a common range. Variational mode decomposition then separates the series into five band-limited modes, which isolates the slow diurnal trend from the rapid cloud-induced fluctuations. The sample entropy of each mode is computed as a measure of its irregularity, and the five entropy values are grouped with K-means into low, intermediate and high complexity levels. Each mode is assigned on that basis to one of three predictors: a multilayer perceptron for the most regular modes, a long short-term memory network for the intermediate ones and a one-dimensional convolutional network for the most irregular. Every expert predicts the next value of its own mode from a window of past values. The five predictions are then summed and converted back to the original units, and the reconstructed series is compared with the reference values using four error measures. The subsections that follow describe each stage in turn, together with the settings used.

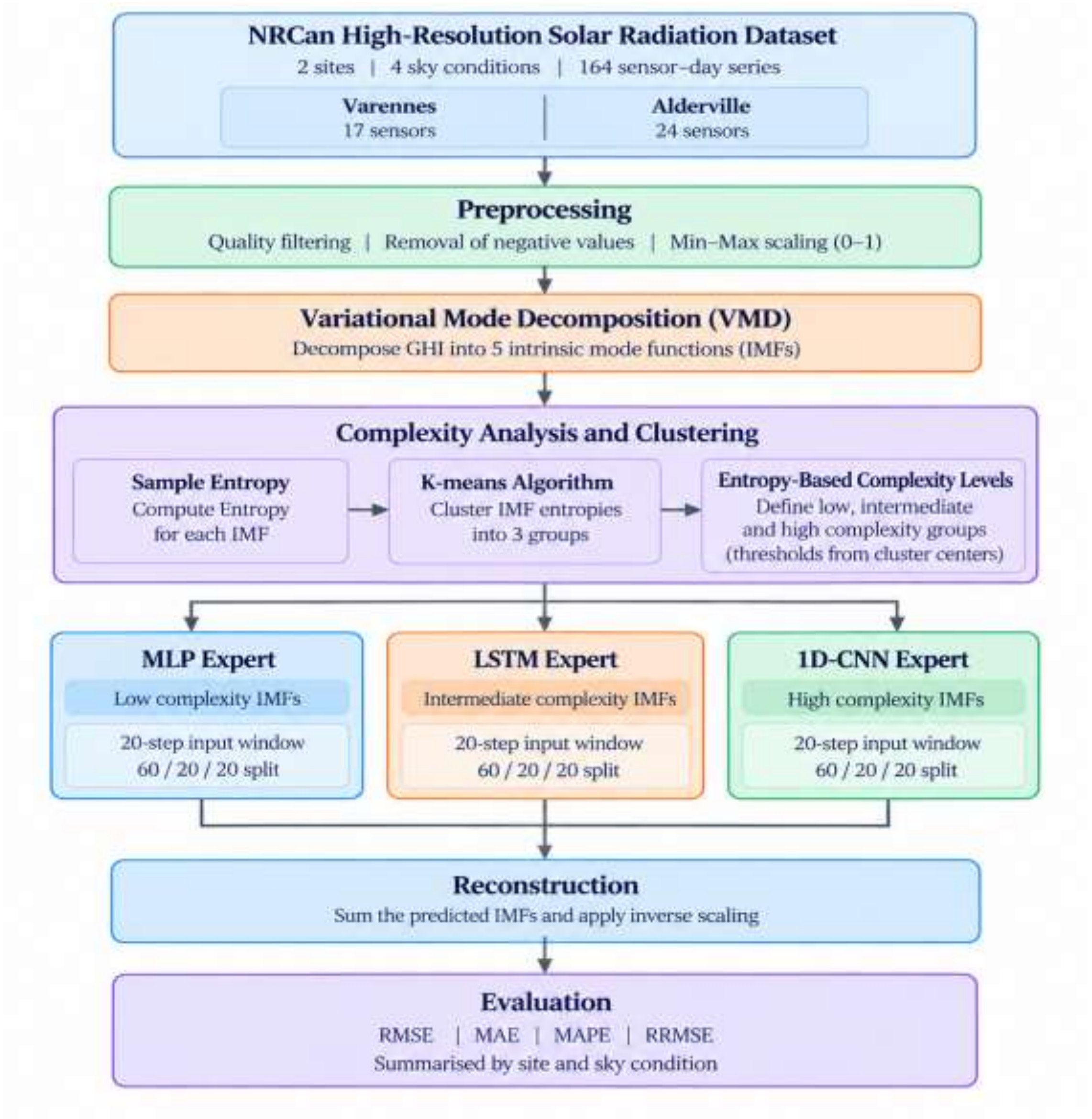

Figure 1. Workflow of the entropy-guided multi-expert framework, applied independently to each of the 164 sensor-day series.

Dataset and observation structure

The study uses the high-resolution solar radiation datasets published by NRCan [17]. The measurements were collected by CanmetENERGY-Varennes with 41 measurement units, each carrying two LI-COR LI-200S photodiode pyranometers. One pyranometer measures global horizontal irradiance and the other measures global tilted irradiance on a south-facing plane, tilted at 45° in Varennes and 30° in Alderville. Seventeen units are installed at Varennes, Quebec (about 45.62° N, 73.39° W), and 24 within a 1 MW sub-array of the Alderville First Nation solar farm in Ontario (about 44.19° N, 78.10° W). The units are synchronised by GPS. The same networks were used by Gagné et al. [4] to characterise short-term variability. NRCan provides one day per sky condition at each site, labelled clear sky, overcast, variable and very variable. Each day covers the period from sunrise to sunset, and each unit's record is a separate CSV file with the date, the time in Eastern Standard Time, the horizontal irradiance and the tilted irradiance in W/m². This study uses the irradiance column labelled G1 in the files, which the dataset documentation identifies as the global horizontal irradiance (GHI); the tilted channel was not used [17]. All eight available days were used, giving 164 sensor-day series (Table 1). Every series was processed and modelled independently of the others.

Table 1. Observation days used in the study and the number of stored records in each archive. Counts are rounded; the eight archives together contain 6,677,970 records.

Site	Date	Sky condition	Sensors	Records, all sensors	Mean records per sensor
Varennes	2014-07-17	Very variable	17	≈1,430,000	≈84,000
Varennes	2014-10-21	Overcast	17	≈13,000	≈760
Varennes	2014-12-30	Clear sky	17	≈87,000	≈5,100
Varennes	2015-02-26	Variable	17	≈250,000	≈14,700
Alderville	2015-02-08	Overcast	24	≈16,000	≈670
Alderville	2015-03-24	Clear sky	24	≈2,100,000	≈87,500
Alderville	2015-08-12	Very variable	24	≈2,150,000	≈89,600
Alderville	2015-10-08	Variable	24	≈621,000	≈25,900

The way the data were recorded determines how the prediction task can be defined. The logger measures irradiance every millisecond, averages it over 10 ms, and stores a new value only when irradiance has changed by more than 5 W/m² since the last stored value, or when one minute has passed [17]. The records are therefore event-driven. They are dense during cloud transitions and sparse, down to one record per minute, when the sky is stable. In this study each series is treated as an ordered sequence of observations $G(n)$, $n = 1, \ldots, N$, indexed by position rather than by clock time. The timestamps are read but not used in modelling. The task is one-step-ahead prediction: given the 20 most recent observations, the model predicts the next stored observation. The time between two consecutive observations is not fixed. It ranges from the 10 ms averaging interval during rapid fluctuations to about one minute under stable conditions, so the prediction step does not correspond to a fixed horizon in seconds.

The eight archives together contain 6,677,970 records, and their distribution across the days follows the recording rule directly. The two overcast days contribute roughly 13,000 and 16,000 records, because irradiance changed slowly enough that most values were written by the one-minute rule, whereas the two very variable days contribute about 1.43 and 2.15 million (Table 1). The measured intervals between consecutive records show the same pattern: the median interval is 0.08 s on the Alderville very variable day, with about 61% of intervals at or below 50 ms, 0.05 s on the Alderville clear-sky day and 0.07 s on the Varennes very variable day, but 60 s on both overcast days. Across all 164 files there are no missing, negative or duplicated values in either irradiance column and no invalid timestamps, so quality control for these data is a question of physical interpretation rather than of incomplete records. The record counts in Table 1 are counts of stored observations, not of elapsed time.Each record also carries the tilted irradiance measured by the second pyranometer of the unit. The two channels are strongly correlated, at approximately 0.98 and above on all days, but their ratio changes with season and solar geometry: the median tilted-to-horizontal ratio is about 1.41 on the Alderville variable day in October and about 2.78 on the Varennes clear-sky day at the end of December. The sensors of a network are likewise strongly but not perfectly related. Computed on the common one-minute records, the median pairwise correlation of horizontal irradiance is about 0.9999 on the Alderville overcast day and falls to about 0.934 on the Alderville very variable day, where moving cloud shadows reach different units at different times. This study uses only the horizontal channel and models each sensor-day series independently, so neither the tilted channel nor the between-sensor structure is used by the framework; both are noted here because they bear on the between-sensor differences reported in Section 5.

This dataset is particularly suitable for the present study because its event-driven acquisition preserves rapid irradiance changes that can be missed by regularly averaged measurements. At the same time, the availability of multiple co-located sensors and labelled observation days permits the framework to be examined across both sensor-level and sky-condition variability. The dataset therefore provides an appropriate test setting for studying how decomposition complexity and expert assignment behave under high-resolution irradiance observations.

Preprocessing

Preprocessing consisted of two cleaning steps and a normalisation. First, observations with negative irradiance were removed. Second, a conventional interquartile range (IQR) filter was applied to each series. With Q_1 and Q_3 the first and third quartiles of the series, observations outside $[Q_1 - 1.5 \cdot \mathrm{IQR}, Q_3 + 1.5 \cdot \mathrm{IQR}]$, where $\mathrm{IQR} = Q_3 - Q_1$, were removed. The retained observations were then re-indexed consecutively. The quartiles are computed from the whole daily series, so the retained range depends on the distribution of irradiance over that day. On days whose records are concentrated at high irradiance, the filter can also remove legitimate low-irradiance observations from the early morning and late afternoon. The cleaned series was then scaled to [0, 1] with min–max normalisation:

$$\tilde{G}(n) = \frac{G(n) - G_{\min}}{G_{\max} - G_{\min}} \qquad (1)$$

where $G_{\min}$ and $G_{\max}$ are the minimum and maximum of the cleaned series of that sensor and day.

Decomposition and complexity measurement

The normalised series was decomposed with VMD [7]. VMD represents a signal f as a sum of K modes uk, each compact around a centre frequency ωk, by solving the constrained problem

$$\min_{\{u_k\},\{\omega_k\}} \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \quad \text{subject to} \quad \sum_{k=1}^{K} u_k = f \qquad (2)$$

The problem is solved in the frequency domain with the alternating direction method of multipliers. The implementation used was the vmdpy package [23], with K = 5 modes, bandwidth penalty α = 2000, dual-ascent step τ = 0, no imposed DC mode, uniform initialisation of the centre frequencies and convergence tolerance 10^{-7}. The same settings were used for all 164 series. With τ = 0, the reconstruction constraint is not enforced exactly, so the sum of the five modes approximates the normalised series rather than reproducing it. The five modes are referred to as IMF1–IMF5, in the order returned by the implementation.

The complexity of each mode was measured with sample entropy [12]. For a series of length N, let B be the number of pairs of templates of length m whose Chebyshev distance is at most r, and A the number of such pairs of length m + 1, with self-matches excluded. Sample entropy is

$$\mathrm{SampEn}(m, r, N) = -\ln \frac{A}{B} \qquad (3)$$

Low values indicate regular, self-similar behaviour and high values indicate irregular behaviour. It was computed with the AntroPy package using $m = 2$ and $r = 0.2$ times the standard deviation of the mode, giving one value E_k for each of the five IMFs of a series.

Expert assignment and model architectures

The entropy values were used to assign each IMF to one of three expert models. The five values E1, …, E5 of a series were grouped into three clusters with K-means [24], using a fixed random state of 42. The cluster centres were sorted in ascending order, c1 < c2 < c3, and two thresholds were taken from them, T1 = c2 and T2 = c3. Each IMF was then assigned by the rule.

$$\text{expert}(k) = \begin{cases} \text{MLP}, & E_k < T_1 \\ \text{LSTM}, & T_1 \le E_k \le T_2 \\ \text{1D-CNN}, & E_k > T_2 \end{cases} \qquad (4)$$

K-means is used here only as a simple way to place five values into relative low, intermediate and high levels. It is not used to discover structure in a large sample. The rule has two practical consequences. First, all IMFs with entropy below the middle centre, including some from the intermediate cluster, go to the MLP. Second, intermediate IMFs pass to LSTM. Third an IMF goes to the CNN only if its entropy lies strictly above the highest centre. That can only happen when the highest cluster holds at least two IMFs with different entropy values. Because the thresholds are computed for each series, the assignment can differ between sensors and days. The pairing of experts with complexity levels follows a design rationale rather than an established result. The MLP provides a simple nonlinear mapping for regular, low-complexity modes. The LSTM models sequential dependence through gated memory [25]. The 1D-CNN learns local patterns within the input window. The study does not test whether this pairing is optimal.

Three complementary expert models, comprising feed-forward, recurrent, and convolutional neural networks, were employed to capture different temporal patterns within the decomposed Intrinsic Mode Functions (IMFs). For each IMF, a sliding window of 20 consecutive observations was used to predict the subsequent value. Each expert was trained independently for individual IMFs using the Adam optimization strategy with mean squared error (MSE) as the objective function. This IMF-specific modeling enables the experts to learn component-level dynamics before their predictions are integrated in the subsequent forecasting stage.

The expert assignment is based on a modelling hypothesis rather than an assumption that one architecture is universally optimal for a particular entropy level. Lower-complexity modes are expected to require a relatively simple nonlinear mapping, motivating the MLP. Intermediate-complexity modes are assigned to the LSTM because recurrent memory provides an explicit mechanism for modelling sequential dependencies. Higher-complexity modes are assigned to the 1D-CNN because local temporal patterns can be extracted through convolutional filters. Thus, Sample Entropy is used as a data-driven routing criterion, while the specific expert pairing remains an empirical design choice evaluated in this study.

Three complementary experts were selected to capture different temporal characteristics of the decomposed IMFs. The MLP provides a nonlinear mapping for relatively regular, low-complexity modes, the LSTM captures sequential dependencies through recurrent memory, and the 1D-CNN extracts local temporal patterns from the input window. Accordingly, sample entropy was used to characterize IMF complexity and guide the assignment of lower-, intermediate-, and higher-complexity modes to the MLP, LSTM, and 1D-CNN experts, respectively.

Prediction, evaluation and protocol

The reconstructed prediction was obtained by summing the five predicted VMD modes and applying the inverse normalization. For consistency with the implemented pipeline, the prediction was evaluated against the corresponding VMD reconstruction of the test period rather than directly against the original measured

GHI values. This choice evaluates the ability of the expert models to predict the decomposed signal representation, but it does not provide a direct measure of prediction error relative to the raw observations. Accordingly, the reported RMSE, MAE, RRMSE, and MAPE should be interpreted as reconstruction-domain performance under the present offline protocol rather than as conventional forecasting accuracy against measured irradiance. A direct evaluation against the original measured GHI would require an additional post-processing evaluation step and was not part of the implemented experimental pipeline.

The prediction for each test position is obtained by summing the predictions of the five experts and converting the result back to W/m²:

$$\hat{G}(n) = G_{\min} + (G_{\max} - G_{\min}) \sum_{k=1}^{5} \hat{u}_k(n) \qquad (5)$$

The reference series used for evaluation is formed in the same way, from the sum of the five true IMF values at each test position, rescaled to W/m². Predictions are therefore compared with the VMD reconstruction of the test period, not with the measured values $G(n)$ themselves. Four measures were computed over the M test positions of each series, where y_n is the reference value and $\hat{y}_n$ the prediction:

$$\mathrm{RMSE} = \sqrt{\frac{1}{M}\sum_{n=1}^{M}(y_n - \hat{y}_n)^2}, \qquad \mathrm{MAE} = \frac{1}{M}\sum_{n=1}^{M}|y_n - \hat{y}_n| \qquad (6)$$

$$\mathrm{MAPE} = \frac{100}{M}\sum_{n=1}^{M}\left|\frac{y_n - \hat{y}_n}{y_n + \varepsilon}\right|, \qquad \mathrm{RRMSE} = \frac{\mathrm{RMSE}}{\bar{y}} \qquad (7)$$

In these expressions, $\varepsilon = 10^{-8}$ prevents division by zero and $\bar{y}$ is the mean of the reference values over the test period. RMSE and the mean absolute error (MAE) are expressed in W/m², and MAPE in percent. The relative root mean square error (RRMSE) is a dimensionless ratio. RMSE, MAE and RRMSE are used as the main measures. MAPE is reported because it was part of the pipeline, but it is interpreted with caution: it is biased and unstable when the reference values approach zero [21], as happens at dusk and on overcast days. The measures were computed for each of the 164 series and then averaged over the sensors of each site and sky condition.

Figure 1 summarises the complete pipeline. Three properties of the protocol matter for interpreting the results. First, the IQR filter, the min–max scaling, the VMD and the sample entropy calculation all use the complete series of each sensor-day, before the chronological split. The inputs and targets in the test period are therefore derived from a decomposition that also saw the test observations. This is the form of look-ahead that X. Yang et al. [22] describe for decomposition-based models, so the evaluation is offline rather than a simulation of real-time forecasting. Second, the prediction step is one stored observation, not a fixed time interval. Third, the reference values are VMD reconstructions rather than raw measurements. These choices are reported as implemented and are revisited in the Discussion.

Experimental Setup

Experimental design

The framework described in Section 3 was applied under a single fixed configuration to each of the 164 sensor-day series. No hyperparameter search was performed. The decomposition settings, the entropy parameters, the clustering rule, the expert architectures, the input window and the partition proportions were fixed in advance and held constant across all runs, so that differences in the results reflect the data rather than changes to the model.

Each series was processed in isolation. The min–max scaling, the decomposition, the entropy values, the K-means thresholds and the expert models were all derived from that series alone, and nothing was carried over between sensors, days or sites. The 164 runs are therefore independent repetitions of the same procedure on different data, and each produced five trained experts.

Experimental variation comes from three strata of the dataset: the site, the sky condition of the day, and the individual sensor within a site-day. The eight site-day combinations listed in Table 1 form eight groups of 17 or 24 series. The days were not selected for this study. NRCan publishes one day per sky condition at each site, classified by the daily clear-sky index [4, 17], and all eight were used, together with every available sensor at both sites. No selection of days or sensors took place at any stage.

The purpose of this design is to characterise how the framework behaves across these strata. It is not a comparison between models. No baseline, no single-expert variant and no ablation of the decomposition or of the assignment rule was run, so the experiments do not support statements about relative performance. Section 6 returns to this limitation.

Model training and experimental settings

The three experts are small feed-forward, recurrent and convolutional networks built in Keras (Table 2). All take a window of 20 consecutive values of a single IMF as input and output the next value of that IMF. Each is trained separately for each IMF with the Adam optimizer and a mean squared error loss. A new model is therefore trained for every IMF of every series, which gives 820 trained models in total.

Table 2. Architecture of the expert models (input window of 20 values; parameter counts computed from the layer definitions).

Expert	Input shape	Layers	Trainable parameters
MLP	20	Dense 64 (ReLU) → Dense 32 (ReLU) → Dense 1 (linear)	3,457
LSTM	20 × 1	LSTM 64 (full sequence output) → LSTM 32 → Dense 1 (linear)	29,345
1D-CNN	20 × 1	Conv1D 32, kernel 3 (ReLU) → Conv1D 16, kernel 3 (ReLU) → Flatten → Dense 1 (linear)	1,937

Training samples were built separately for each IMF with a sliding window. For mode uk, the i-th sample has input [uk(i), …, uk(i + 19)] and target uk(i + 20), which gives N − 20 samples. The samples were split chronologically, without shuffling, into the first 60% for training, the next 20% for validation and the last 20% for testing. The same split points apply to all five IMFs of a series, so their test predictions are aligned in time. Because the windows are formed before the partition, the twenty windows spanning each boundary contain observations from both sides of it. The validation set was used only to monitor the loss during training. It was not used for model selection or early stopping.

All experts were trained with the Adam optimiser at its default learning rate of 10^{-3} and a mean squared error loss, in mini-batches of 32 samples, for a fixed budget of 10 epochs. The same settings were used for every expert, every mode and every series (Table 3). No learning-rate schedule, early stopping, checkpointing, dropout, batch normalisation or weight decay was applied. The validation set was supplied to the training loop so that the validation loss could be monitored, but no callback acted on it, and the weights retained are those reached at the end of the tenth epoch rather than those with the lowest validation loss. The epoch budget was set in advance and was not tuned, so the networks were not trained to a convergence criterion.

Table 3. Experimental configuration, held constant across the 164 sensor-day runs.

Item	Setting
Series processed	164 (8 site-days; 17 or 24 sensors each)
Models trained	820 (5 per series)
Input window	20 consecutive values of one mode
Prediction target	Next value of that mode
Partition	60 / 20 / 20 chronological, no shuffling
Modes per series	5
Optimiser	Adam, learning rate 10^{-3}
Loss	Mean squared error
Batch size	32
Epochs	10, fixed
Early stopping, scheduler, checkpointing	None
Regularisation	None
Weights retained	Final epoch
Use of validation set	Loss monitoring only
Random state	Fixed for K-means; not fixed for network initialisation
Runs per series	1

Computational environment and reproducibility

The pipeline was implemented in Python 3.9 and executed as a set of Jupyter notebooks, one for each site and sky condition. The decomposition used vmdpy [23], sample entropy was computed with AntroPy, K-means with scikit-learn, and the experts were built with the Keras interface to TensorFlow, NumPy and pandas were used for data handling and Matplotlib for the data visualization and figures. Jupyter notebook tool used for model development and training.

The K-means step uses a fixed random state, so the thresholds and the resulting expert assignments are determined by the series. The seeds governing weight initialisation and batch ordering in TensorFlow were not fixed, and each series was run once. The reported values are therefore those of a single execution and would differ slightly on a repeat run, although the settings given here are sufficient to reproduce the protocol. No repeated runs were performed, so no run-to-run variability is available.

Evaluation protocol and analysis

RMSE, MAE, MAPE and RRMSE were computed for each series over its test set as defined in Section 3, and then summarised within each of the eight groups. Both the mean and the standard deviation across the sensors of a group are reported. The standard deviation describes the spread between co-located sensors observing the same day; it is not an estimate of run-to-run variability, which the single-run design does not provide. Group comparisons are accordingly descriptive, and no significance test was performed. Because MAPE becomes unstable as the reference values approach zero [21]. Its median is reported alongside its mean wherever the two differ appreciably.

The four questions stated in Section 1 map onto the analyses as follows. Performance across sky conditions is assessed by comparing the group summaries for the four conditions, and performance across sites by comparing the Varennes and Alderville groups under matching sky labels. The assignment of modes to experts is examined by tabulating, for each group, the number of modes assigned to each expert, both in total and by mode index, which indicates whether the assignment responds to the individual series or is effectively fixed within a group. Variation between sensors is measured by the dispersion of the per-sensor measures within each group, summarised by the standard deviation, the coefficient of variation and the range between the best and the worst sensor.

RESULTS

Overall performance

Across the 164 sensor-day series, RMSE ranged from 0.21 to 9.69 W/m² with a mean of 3.62 and a median of 3.05 W/m², and MAE from 0.16 to 9.49 W/m² with a mean of 3.23 and a median of 2.62 W/m². RRMSE covered three orders of magnitude, from 0.0003 to 0.540 (equivalently 0.03% to 54.0%), with a median of 0.0071. The group means obtained here reproduce those produced by the aggregation step of the original pipeline for all eight groups and all four measures.

RMSE and MAE order the series almost identically (Spearman $\rho = 0.990$), so the two carry close to the same information at series level. Their ratio is more informative than either alone. It lies between 1.15 and 1.38 in six of the eight groups but falls to 1.02 in both overcast groups, which indicates an error distribution with very few large deviations in those two cases.

Because RRMSE is defined as RMSE divided by the mean of the reference values, the mean reference irradiance over each test set can be recovered exactly as the ratio of the two. This quantity is reported in Table 4, because it organises the results more effectively than the sky label does. It ranges from 18 W/m² for the Alderville overcast day to 711 W/m² for the Alderville clear-sky day.

Performance across sky conditions

Table 4 and Figure 2 give the results by group. The two error measures do not agree on the ordering of the sky conditions, and the disagreement is systematic.

Table 4. Prediction performance by site and sky condition. Values are mean ± standard deviation across the sensors of each group; the standard deviation describes between-sensor spread, not run-to-run variability. Mean reference irradiance is derived as RMSE / RRMSE.

Site	Sky condition	Series	Mean ref. irradiance (W/m²)	RMSE (W/m²)	MAE (W/m²)	RRMSE	MAPE mean / median (%)
Varennes	Clear sky	17	256	4.87 ± 2.13	4.06 ± 2.02	0.0230 ± 0.0186	65.12 / 1.57
Varennes	Overcast	17	34	6.16 ± 1.62	6.07 ± 1.63	0.1836 ± 0.0484	1039.70 / 563.48
Varennes	Variable	17	345	2.86 ± 1.06	2.50 ± 0.96	0.0083 ± 0.0032	0.80 / 0.79
Varennes	Very variable	17	491	3.02 ± 2.05	2.23 ± 1.62	0.0066 ± 0.0049	2.94 / 1.81
Alderville	Clear sky	24	711	0.57 ± 0.38	0.49 ± 0.35	0.0008 ± 0.0005	0.07 / 0.04
Alderville	Overcast	24	18	6.40 ± 1.64	6.28 ± 1.62	0.3600 ± 0.0917	472.95 / 485.52
Alderville	Variable	24	539	3.28 ± 0.87	2.59 ± 0.72	0.0063 ± 0.0021	0.57 / 0.48
Alderville	Very variable	24	626	2.54 ± 1.29	2.19 ± 1.22	0.0041 ± 0.0020	7.88 / 3.18

In absolute terms, the lowest errors occur on the Alderville clear-sky day (RMSE 0.57 W/m²) and the highest on the two overcast days (6.40 and 6.16 W/m²). Pooling the two sites, mean RMSE is 2.35 W/m² for clear sky, 2.74 for very variable, 3.10 for variable and 6.30 for overcast. The variable and very variable conditions are therefore inverted with respect to the ordering that increasing cloud variability would suggest. In the pooled absolute figures this inversion is driven by Alderville, where mean RMSE is 3.28 W/m² on the variable day against 2.54 on the very variable day; at Varennes the two days are close and in the expected order (2.86 against 3.02 W/m²).

In relative terms the ordering changes again. Pooled mean RRMSE is 0.0051 for very variable, 0.0071 for variable, 0.0100 for clear sky and 0.2869 for overcast. The clear-sky condition, which has the lowest absolute error, has the third-highest relative error, because the two clear-sky days differ greatly in irradiance level. On this measure the very variable day has the lower error than the variable day at both sites, so the inversion is consistent in relative terms even though it is not in absolute terms.

A single relationship accounts for most of this behaviour. Across all 164 series, the Spearman correlation between mean reference irradiance and RRMSE is −0.933, and that between mean reference irradiance and RMSE is −0.780. The measured error tracks the irradiance level of the series more closely than it tracks the sky label assigned to the day.

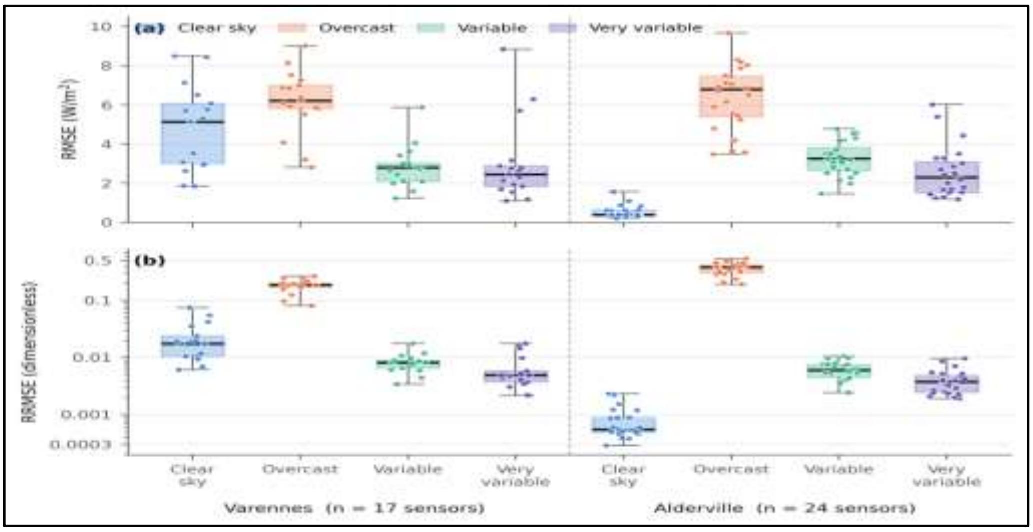

Figure 2. Distribution of per-sensor prediction error within each of the eight groups. (a) RMSE on a linear scale; (b) RRMSE on a logarithmic scale. Boxes show the median and interquartile range, whiskers span the full range, and each point is one sensor-day series.

MAPE behaves erratically in the low-irradiance groups, as anticipated in Section 3. Its mean reaches 472.95% on the Alderville overcast day and 1039.70% on the Varennes overcast day, with a maximum of 4,470.92% for a single sensor. On the Varennes clear-sky day the mean is 65.12% against a median of 1.57%. In total, 43 of the 164 series exceed 100% MAPE, and 41 of these belong to the two overcast groups; the remaining two are the two lowest-irradiance sensors of the Varennes clear-sky day. The measure is reported for completeness but is not used to rank the groups.

Performance across sites

Pooled over all four days, Alderville has the lower mean RMSE (3.20 against 4.23 W/m²) but the higher mean RRMSE (0.0928 against 0.0554). The two measures again point in opposite directions, and the reason is the Alderville overcast day, whose mean reference irradiance of 18 W/m² is the lowest in the study and inflates the relative error for that site.

Comparing matched sky labels is more informative. Alderville has the lower RMSE on the clear-sky and very variable days and the higher RMSE on the overcast and variable days, so the absolute comparison is evenly divided. On RRMSE, Alderville is lower for three of the four labels, the exception being the overcast day. The matched days also differ substantially in irradiance level: mean reference irradiance is 711 W/m² for the Alderville clear-sky day against 256 W/m² for the Varennes clear-sky day, and 626 against 491 W/m² for the two very variable days. The site comparison therefore rests on pairs of days that are not equivalent in irradiance, and no consistent site-level difference is established by these results.

Assignment of modes to experts

Across the 820 IMF instances, 371 (45.2%) were assigned to the MLP, 353 (43.0%) to the LSTM, and 96 (11.7%) to the 1D-CNN (Table 5a). The distribution varied across sites and sky conditions, with the MLP and LSTM accounting for most assignments, while the 1D-CNN was selected less frequently. The CNN assignment was highest under Alderville overcast conditions (23.3%) and Varennes variable conditions (20.0%), whereas no IMF was assigned to the CNN for the Alderville variable group.

Table 5a. Assignment of modes to experts, by group.

Site	Sky condition	Modes	MLP	LSTM	1D-CNN
Varennes	Clear sky	85	45 (52.9%)	35 (41.2%)	5 (5.9%)
Varennes	Overcast	85	36 (42.4%)	34 (40.0%)	15 (17.6%)
Varennes	Variable	85	34 (40.0%)	34 (40.0%)	17 (20.0%)
Varennes	Very variable	85	42 (49.4%)	33 (38.8%)	10 (11.8%)
Alderville	Clear sky	120	40 (33.3%)	66 (55.0%)	14 (11.7%)
Alderville	Overcast	120	34 (28.3%)	58 (48.3%)	28 (23.3%)
Alderville	Variable	120	72 (60.0%)	48 (40.0%)	0 (0.0%)
Alderville	Very variable	120	68 (56.7%)	45 (37.5%)	7 (5.8%)
All		820	371 (45.2%)	353 (43.0%)	96 (11.7%)

The assignment also varied systematically with IMF index (Table 5b). IMF1 was assigned to the MLP in all 164 sensor-day series. For IMF2, IMF3, and IMF4, the MLP assignments decreased to 73.8%, 39.6%, and 1.8%, respectively, while LSTM assignments accounted for an increasing proportion of the intermediate cases. The 1D-CNN assignments increased from 0.6% for IMF2 to 9.1% for IMF3 and 15.9% for IMF4, reaching 32.9% for IMF5. Overall, this distribution indicates that the entropy-guided assignment mechanism does not apply a fixed expert to a given IMF; instead, it adapts the expert selection according to the measured complexity of each mode.

Table 5b. Assignment of modes to experts, by mode index, pooled over the 164 series.

Mode	MLP	LSTM	1D-CNN
IMF1	164 (100.0%)	0 (0.0%)	0 (0.0%)
IMF2	121 (73.8%)	42 (25.6%)	1 (0.6%)
IMF3	65 (39.6%)	84 (51.2%)	15 (9.1%)
IMF4	3 (1.8%)	135 (82.3%)	26 (15.9%)
IMF5	18 (11.0%)	92 (56.1%)	54 (32.9%)

The CNN was used least often, as the threshold rule derived in Section 3 implies. It received no modes at all in the Alderville variable group, across all 120 assignments in that group. In two groups, Varennes variable and Alderville variable, the assignment was identical for every series: the same mode indices went to the same experts in all 17 and all 24 series respectively, so the rule produced no series-specific variation there. In the remaining six groups the assignment varied between sensors, most strongly at IMF3 and IMF5. One departure from the monotonic pattern is visible. IMF5 was assigned to the MLP in 18 series, 16 of which belong to the Alderville clear-sky group, indicating that the highest-frequency mode of a smooth clear-sky series can itself be regular enough to fall in the lowest entropy cluster.

Variation between sensors

Errors differ considerably between co-located sensors observing the same day (Table 6, Figure 2). The coefficient of variation of RMSE within a group ranges from 0.256 to 0.678, and the ratio between the worst and the best sensor of a group ranges from 2.8 to 8.0.

Table 6. Between-sensor variation within each group. CV is the coefficient of variation of RMSE across the sensors of the group.

Site	Sky condition	CV of RMSE	RMSE min (W/m^2)	RMSE max (W/m^2)	max / min
Varennes	Clear sky	0.437	1.85	8.51	4.6
Varennes	Overcast	0.262	2.83	9.03	3.2
Varennes	Variable	0.370	1.24	5.88	4.8
Varennes	Very variable	0.678	1.11	8.85	8.0
Alderville	Clear sky	0.660	0.21	1.58	7.5
Alderville	Overcast	0.256	3.48	9.69	2.8
Alderville	Variable	0.267	1.48	4.78	3.2
Alderville	Very variable	0.507	1.19	6.03	5.1

The two overcast groups are the most homogeneous (CV 0.256 and 0.262), and the very variable groups are the most heterogeneous at both sites (0.507 at Alderville, 0.678 at Varennes). The Alderville clear-sky group has a high coefficient of variation (0.660) but a very small absolute spread, with all 24 sensors between 0.21 and 1.58 W/m^2. Sensor performance is not stable across days. Comparing the RMSE ranking of the sensors between each pair of days at a site, none of the twelve pairwise rank correlations reached conventional significance levels, and the strongest was $\rho = -0.48$. This analysis was exploratory and no correction for multiple comparisons was applied, but the absence of any consistent ranking indicates that the between-sensor differences are specific to the day rather than to the sensor.

DISCUSSION

The small absolute errors in Section 5 call for cautious interpretation: 115 of the 164 series have an RMSE below 5 W/m^2, the threshold at which the logger stores a new value, and 21 fall below 1 W/m^2, the resolution at which irradiance is stored [17]. The evaluation reference is the VMD reconstruction rather than the original measured GHI, which may yield smaller errors because the reconstructed signal is smoother. In addition, scaling and decomposition used the complete sensor-day series before the chronological split, representing an offline look-ahead protocol similar to that described for decomposition-based models by X. Yang et al. [22]. The recording rule also limits the change between consecutive stored observations [17]. Thus, Table 4 characterizes behavior under this protocol rather than conventional real-time forecasting accuracy, and without a persistence reference no skill score can be given [20]. Direct evaluation against measured GHI using a strictly causal decomposition procedure remains an important direction for future work.

Error tracks the irradiance level of a series ($\rho = -0.933$ with RRMSE) more closely than the sky classification. The Varennes clear-sky day of 30 December averages 256 W/m^2, the Alderville clear-sky day of 24 March 711 W/m^2 two physical regimes sharing one label. NRCan publishes one day per sky condition at each site, classified by the daily clear-sky index [4], so sky regime is aliased with date and season, and the four conditions cannot be compared on cloud behaviour alone. The same account covers the inversion between the variable and very variable days, both mid-summer records with higher irradiance, and the large relative errors of the overcast groups, where mean reference irradiance is below 40 W/m².

The assignment rule gives a mixed picture. Sample entropy generally increased with mode index, and the assignment reflected this pattern, with IMF1 assigned to the MLP in all 164 series and progressively greater use of the LSTM and 1D-CNN for higher-index modes. Overall, the MLP, LSTM, and CNN accounted for 45.2%, 43.0%, and 11.7% of the assignments, respectively. The CNN was not selected in one group of 120 assignments because its allocation required entropy to exceed the highest K-means cluster center, a condition not met by the modes in that group. In two groups, the assignment pattern was identical across all series, limiting the adaptive contribution of the rule. Without an ablation study, the separate contribution of entropy-guided expert selection cannot be isolated.

Errors differed between co-located sensors by factors of 2.8 to 8.0 within a single site and day, so a study reporting one sensor could have landed anywhere in that band. The cause does not appear to be the sensors: some Varennes units stand near trees and power lines [17], which would give a stable ranking, yet the orderings on the four days of a site are essentially unrelated. A day-specific cause is more consistent with cloud shadow geometry.

Comparison with earlier results is not possible. Sanchez-Lopez et al. [18], the only prediction study we located on this dataset, predicts at fixed horizons of 15 s and 30 s from several correlated sensors, whereas this study predicts the next stored observation from one sensor's own history offline; the wider decomposition literature works at hourly or coarser resolution. Three constraints are specific to these data: the IQR filter discards genuine observations where irradiance is strongly skewed, most clearly in the overcast groups; VMD assumes uniform sampling, so mode centre frequencies are estimated over an observation index rather than over time; and MAPE is unreliable at low irradiance [21].

CONCLUSION

This study examined a decomposition-based multi-expert deep learning framework on high-resolution, event-recorded solar irradiance, where entropy-guided model assignment has rarely been documented. All eight publicly released days of the NRCan high-resolution datasets were used, one day for each of four sky conditions at each of two Canadian sensor networks, giving 164 sensor-day series of global horizontal irradiance. Each series was cleaned, scaled and decomposed into five modes by variational mode decomposition; sample entropy then grouped these modes by K-means into three complexity levels, assigning each to an MLP, LSTM or one-dimensional CNN expert, whose predictions were summed and rescaled. In total, 820 expert models were trained under a single fixed configuration.

Four findings follow. Error depended more on a series' irradiance level than on the day's sky label: the rank correlation between mean reference irradiance and relative error was −0.933, and the two clear-sky days, in late December and late March, differed almost threefold in irradiance and by over an order of magnitude in relative error. Consistently, the very variable days gave lower relative errors than the variable days at both sites. The assignment tracked mode ordering, with the lowest-frequency mode assigned to the MLP in all 164 series, while LSTM and CNN assignments varied across modes. Overall, the CNN accounted for 11.7% of modes, including none in one group of 120 assignments, while two of the eight groups showed identical assignments across all series. Errors between co-located sensors on the same day varied by factors of 2.8 to 8.0 within a group, that variation being day-specific rather than sensor-specific. Percentage error proved unusable at low irradiance, exceeding 100% in 43 of the 164 series, 41 of them overcast.

The contribution is descriptive rather than comparative: an integrated pipeline documented across both networks rather than a single station, with sensor-level errors and expert assignments reported rather than site averages alone, and evaluation conditions stated in full. Because the evaluation used VMD reconstructions rather than the original measured GHI, scaling and decomposition used the complete series before the chronological split, no reference model was computed, and each series was run once, the reported values describe the pipeline's behavior under this protocol rather than conventional causal prediction accuracy.

These limits define the next steps. A causal implementation, with scaling and decomposition using only past observations, would show how much measured performance survives a realistic protocol; a persistence reference and skill score would make results interpretable against the field standard; and an ablation removing the decomposition or replacing the assignment rule would isolate each component's contribution, the present results suggesting that a rule based on entropy quantiles or mode index would be simpler to justify and behave similarly. Replacing the distributional outlier filter with physically motivated quality control would avoid discarding genuine observations on low-irradiance days, and explicit treatment of the irregular sampling would give the extracted modes a defined relation to time. Finally, more than one day per sky condition is needed before differences between cloud regimes can be separated from differences between seasons.

REFERENCES

[1] International Energy Agency. (2024). Renewables 2024. IEA. https://www.iea.org/reports/renewables-2024

[2] Yang, D., Wang, W., Gueymard, C. A., Hong, T., Kleissl, J., Huang, J., Perez, M. J., Perez, R., Bright, J. M., Xia, X., van der Meer, D., & Peters, I. M. (2022). 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. https://doi.org/10.1016/j.rser.2022.112348

[3] Chu, Y., Li, M., Coimbra, C. F. M., Feng, D., & Wang, H. (2021). Intra-hour irradiance forecasting techniques for solar power integration: A review. iScience, 24(10), 103136. https://doi.org/10.1016/j.isci.2021.103136

[4] Gagné, A., Turcotte, D., Goswamy, N., & Poissant, Y. (2016). High resolution characterisation of solar variability for two sites in Eastern Canada. Solar Energy, 137, 46–54. https://doi.org/10.1016/j.solener.2016.07.042

[5] Ajith, M., & Martínez-Ramón, M. (2023). Deep learning algorithms for very short term solar irradiance forecasting: A survey. Renewable and Sustainable Energy Reviews, 182, 113362. https://doi.org/10.1016/j.rser.2023.113362

[6] Barancsuk, L., Groma, V., & Kocziha, B. (2025). Hybrid ultra-short term solar irradiation forecasting using resource-efficient multi-step long-short term memory. Renewable Energy, 247, 122962. https://doi.org/10.1016/j.renene.2025.122962

[7] Dragomiretskiy, K., & Zosso, D. (2014). Variational mode decomposition. IEEE Transactions on Signal Processing, 62(3), 531–544. https://doi.org/10.1109/TSP.2013.2288675

[8] Wang, Y., Zhang, C., Fu, Y., Suo, L., Song, S., Peng, T., & Nazir, M. S. (2023). Hybrid solar radiation forecasting model with temporal convolutional network using data decomposition and improved artificial ecosystem-based optimization algorithm. Energy, 280, 128171. https://doi.org/10.1016/j.energy.2023.128171

[9] Liu, Z., & Zhao, X. (2026). LSTM-Transformer with decomposition–reconstruction for enhanced solar irradiance forecasting incorporating meteorological variables. Renewable Energy, 258, 124971. https://doi.org/10.1016/j.renene.2025.124971

[10] Sarango, K., Maldonado-Correa, J., Torres-Cabrera, J., & Samaniego-Ojeda, C. (2026). VMD-LSTM-transformer: A hybrid approach for robust solar irradiance forecasting in complex terrains with

application to photovoltaic systems. Next Energy, 13, 100855. https://doi.org/10.1016/j.nxener.2026.100855

[11] Sibtain, M., Li, X., Saleem, S., Qurat-ul-Ain, Shi, Q., Li, F., Saeed, M., Majeed, F., Shah, S. S. A., & Saeed, M. H. (2022). Multifaceted irradiance prediction by exploiting hybrid decomposition-entropy-spatiotemporal attention based Sequence2Sequence models. Renewable Energy, 196, 648–682. https://doi.org/10.1016/j.renene.2022.07.041

[12] Richman, J. S., & Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. American Journal of Physiology–Heart and Circulatory Physiology, 278(6), H2039–H2049. https://doi.org/10.1152/ajpheart.2000.278.6.H2039

[13] Zemouri, N., Mezaache, H., Zemali, Z., La Foresta, F., Versaci, M., & Angiulli, G. (2025). Hybrid AI-based framework for renewable energy forecasting: One-stage decomposition and sample entropy reconstruction with least-squares regression. Energies, 18(11), 2942. https://doi.org/10.3390/en18112942

[14] Liu, J., Huang, X., Li, Q., Chen, Z., Liu, G., & Tai, Y. (2023). Hourly stepwise forecasting for solar irradiance using integrated hybrid models CNN-LSTM-MLP combined with error correction and VMD. Energy Conversion and Management, 280, 116804. https://doi.org/10.1016/j.enconman.2023.116804

[15] Xue, S., & Li, L. (2026). Photovoltaic power forecasting based on secondary decomposition strategy and hybrid model. Scientific Reports, 16, 12915. https://doi.org/10.1038/s41598-026-42896-z

[16] Mehmood, F., Asim, M. N., Mehmood, A., Mahmood, W., Vollmer, S., Ul Haq, M., Ali, F., & Dengel, A. (2026). Solar irradiance forecasting meta-review: An in-depth systematic meta-review on solar irradiance forecasting datasets, predictive pipelines and performance. Results in Engineering, 32, 112449. https://doi.org/10.1016/j.rineng.2026.112449

[17] Natural Resources Canada. (2025, January 16). High-resolution solar radiation datasets. Government of Canada. https://natural-resources.canada.ca/energy-sources/renewable-energy/high-resolution-solar-radiation-datasets

[18] Sanchez-Lopez, J. E., Solís-García, J., & Riquelme, J. C. (2024). Semi-real-time decision tree ensemble algorithms for very short-term solar irradiance forecasting. International Journal of Electrical Power & Energy Systems, 158, 109947. https://doi.org/10.1016/j.ijepes.2024.109947

[19] Hendrikx, N. Y., Barhmi, K., Visser, L. R., de Bruin, T. A., Pó, M., Salah, A. A., & van Sark, W. G. J. H. M. (2024). All sky imaging-based short-term solar irradiance forecasting with long short-term memory networks. Solar Energy, 272, 112463. https://doi.org/10.1016/j.solener.2024.112463

[20] Yang, D., Alessandrini, S., Antonanzas, J., Antonanzas-Torres, F., Badescu, V., Beyer, H. G., Blaga, R., Boland, J., Bright, J. M., Coimbra, C. F. M., David, M., Frimane, Â., Gueymard, C. A., Hong, T., Kay, M. J., Killinger, S., Kleissl, J., Lauret, P., Lorenz, E., . . . Zhang, J. (2020). Verification of deterministic solar forecasts. Solar Energy, 210, 20–37. https://doi.org/10.1016/j.solener.2020.04.019

[21] Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Computer Science, 7, e623. https://doi.org/10.7717/peerj-cs.623

[22] Yang, X., Li, J., & Jiang, X. (2024). Research on information leakage in time series prediction based on empirical mode decomposition. Scientific Reports, 14, 28362. https://doi.org/10.1038/s41598-024-80018-9

[23] Carvalho, V. R., Moraes, M. F. D., Braga, A. P., & Mendes, E. M. A. M. (2020). Evaluating five different adaptive decomposition methods for EEG signal seizure detection and classification. Biomedical Signal Processing and Control, 62, 102073. https://doi.org/10.1016/j.bspc.2020.102073

[24] Lloyd, S. (1982). Least squares quantization in PCM. IEEE Transactions on Information Theory, 28(2), 129–137. https://doi.org/10.1109/TIT.1982.1056489

[25] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

[26] Yan, Z., Lu, X., Wu, L., Zhang, H., Liu, F., Wang, X., Xu, W., & Liu, W. (2025). Enhancing short-term solar radiation forecasting with hybrid VMD and GraphCast-based machine learning models. Expert Systems with Applications, 285, 128042. https://doi.org/10.1016/j.eswa.2025.128042