

Solar irradiance prediction with machine learning

Candidate Number: 1072827

Project Number: AO15

Supervisor: Dr Rui Song & Prof Don Grainger

Abstract

Accurate, local solar irradiance forecasts are essential for solar energy production. This project investigates solar irradiance forecasting with machine learning by using a spatio-temporal Fourier transformer architecture. Coarse atmospheric information has been combined with local cloud data to generate predictions which address weaknesses in the ECMWF's Artificial Intelligence Forecasting System (AIFS). A site-specific and general model were developed to predict solar irradiance at 15 minute intervals up to an hour ahead for specific coordinates. Results show that both models achieve a lower root-mean-square error than AIFS, with the site-specific model showing particularly strong performance during high cloud cover.

1 Introduction

Accurate prediction of solar irradiance is essential when forecasting energy production from solar farms. Solar irradiance exhibits large temporal fluctuations owing to cloud movement and the diurnal cycle, which have significant effects on the energy produced by photovoltaic plants. Access to accurate forecasts allows grid operators to anticipate these variations in energy supply and thus balance the supply and demand of the grid, creating greater reliability. Furthermore, knowledge of these fluctuations allows solar farms to charge and discharge their energy stores more efficiently. As a result, reliable solar irradiance forecasts are essential in increasing solar energy uptake, as they allow suppliers to improve efficiency and reduce operational uncertainty during the energy transition. The International Energy Agency predicts that between 2024 and 2030, solar energy will account for 80% of the growth in global renewable capacity [1], which could increase further if fluctuations become more predictable, as it facilitates easier integration into national grids. Thus, creating a more accurate forecast for solar irradiance is an important factor in increasing the uptake of renewable energy and combating climate change.

However, despite the increasing importance of solar energy, forecasting remains challenging. Most robust weather forecasts include solar irradiance predictions, and have traditionally been based on numerical models. Numerical models attempt to capture the underlying dynamics of weather into complex systems of equations that are solved into the future to give predictions. Although these are fairly reliable, they require massive computational resources to run and cannot fully capture the complex nature of climatic systems. In recent years, a new machine learning approach has been trialled in weather forecasting; this gives faster compute times and greater ability to capture more complex patterns. Although artificial intelligence (AI) forecasts do not rely on solving a theoretical climate model, they integrate historical data with physical models to produce physics-based predictions with a high level of accuracy. Most recently, the European Centre for Medium-Range Weather Forecasts (ECMWF) has released a new AI Forecasting System (AIFS) which has been operational since February 2025 [2]. This forecast outperforms the ECMWF's numerical weather prediction models on many variables, demonstrating the effectiveness of using machine learning for weather forecasting.

Although AIFS performs strongly in many aspects, for use in solar plants, it has many limits. Firstly, because AIFS predicts over 100 variables, it has no particular focus. This is a useful attribute for general weather forecasting, but means it lacks specialised solar irradiance capabilities. Secondly, the resolution of AIFS is a 0.25° grid (roughly 25 km). Solar farms are generally smaller than this and need more localised forecasts to accurately predict fluctuations in energy production. Finally, AIFS is trained using the ECMWF Reanalysis v5 (ERA5) database, which is a consistent and complete dataset, combining

model data with observations globally. Using ERA5 for training reinforces the resolution issue, as this exists on a 0.25° grid, and limits the accuracy of AIFS by not considering important physical features that affect solar irradiance, such as cloud cover.

This project aims to create a forecast which predicts solar irradiance to a higher temporal resolution than AIFS, and with closer alignment to the ground-truth as measured by weather stations. To achieve this, a Spatio-temporal Fourier Transformer machine learning architecture will be used to build on AIFS’s model, by combining it with satellite data on cloud cover to achieve a more robust prediction. For training, historic data from the ERA5 database and data from METEOSAT second-generation (MSG) satellites will be used, which combine to give a more expansive foundation for prediction.

2 Methods

2.1 Machine Learning Weather Forecasts

Since the first computer-assisted weather forecasts in 1950, the field has developed dramatically [3]. Currently, the ECMWF’s primary numerical weather predictor is the Integrated Forecast System (IFS), which produces forecasts up to 10 days ahead, at a spatial resolution of 25 km and at 91 vertical levels, twice a day. The IFS combines millions of observations from several sources to produce the best estimate of the current state of the climate. Using these as initial conditions, the IFS solves the fundamental equations governing atmospheric dynamics on a global grid, including parametrisations for more complex processes such as clouds and radiation. The IFS is a highly trusted forecast; however, it still suffers from limitations inherent in numerical weather predictions. Since the forecast is built on physical atmospheric models, the equations being solved include approximations, for example, the finite resolution means small-scale processes are excluded. This creates uncertainty, especially when considering extreme weather scenarios and coupled systems. In addition, the huge datasets that are fed into the complex equations behind the forecasts require powerful computers to perform the predictions. This limits the temporal resolution of the forecasts by having to factor in the computation time between observation and release [4].

Rather than using physical models to compute future atmospheric states from initial conditions, AI based models use vast stores of historic data to learn patterns in the data to produce predictions. AI forecasts are unpopular among some because of their black-box nature, in that they have limited physical reasoning behind predictions which complicates interpretability and transparency [5]. However, machine learning models are able to address many of the limitations of numerical weather predictors. Although training AI models requires enormous computation, once that has been completed, the computations to produce predictions are much lighter. So forecasts can be released on an operational basis at higher frequencies than numerical models. Additionally, by learning from large historic datasets, rather than relying on approximated models, AI forecasts are able to better capture the complex dynamics of weather systems.

AIFS has been operational since February 2025 and already has shown improvements compared to IFS [2]. AIFS is a strong forecasting tool, however, it still has some limitations when applied to solar energy generation. Generally, solar irradiance forecasts are used in solar plants to improve stability of operations. Forecasts are needed to pre-empt fluctuations in generation which inform efficient battery charging and scheduling of alternative energy generation to ensure sufficient energy supply in the grid. For these purposes, localised forecasts, on sub-hourly scales, are needed. AIFS operates on a 0.25° grid and at 6 hour intervals meaning it is not ideal for the use of forecasting at solar plants. In addition, AIFS predicts up to 2 weeks in advance, however, for application to solar energy generation, forecasts are only needed in the immediate future. In this project two models were created, which predict an hour in advance in quarter-hourly increments; a general model, trained on all available data, and a site-specific model, trained only on a single location. The improved temporal resolution and location targeting means it is more suitable for this use. The models use AIFS as a baseline to give a rough estimate for solar irradiance at a location and then predicts a correction upon AIFS that is more refined to a specific coordinate.

2.2 Training Data

AI forecasting models use training data to learn patterns from historic data for future prediction. For the models to improve upon AIFS, the training data must be carefully selected. The resolution and data type of the training data informs where the model learns patterns from, so must be chosen to specifically address the shortcomings of AIFS.

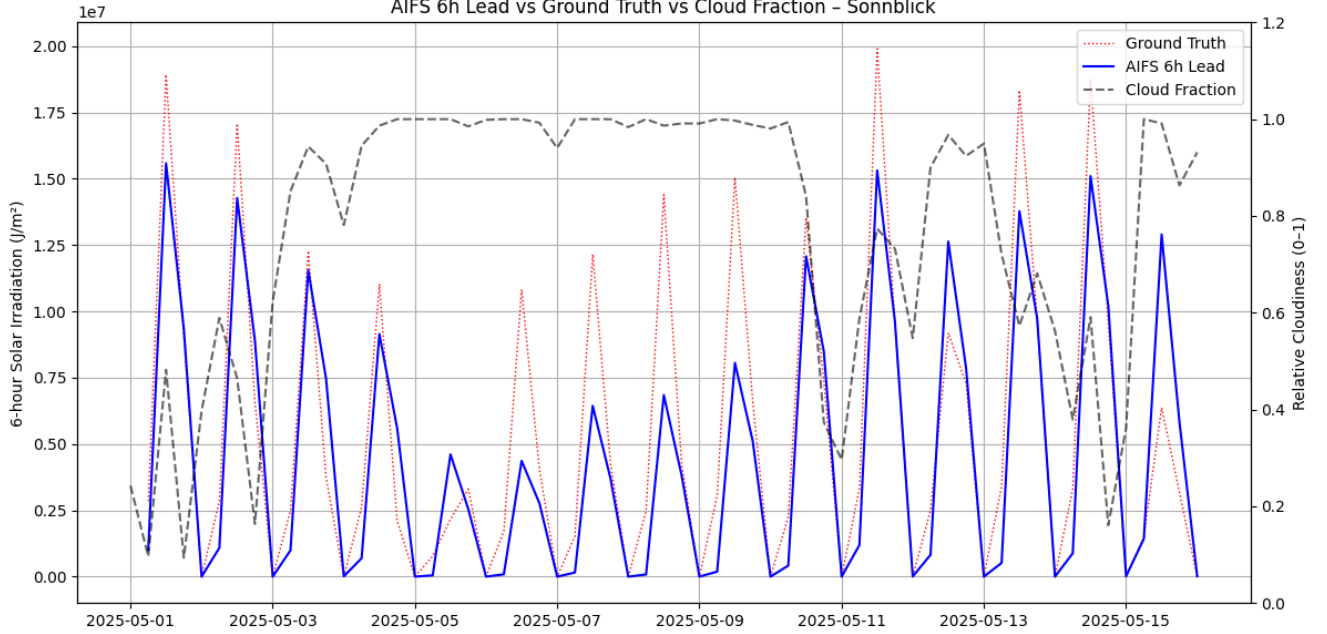


Figure 1: Cumulative solar irradiation over 6 hour period as predicted by AIFS with 6 hour lead compared to measured values at Sonnblick Observatory and cloud fraction.

In Figure 1, we can see solar irradiation over a six hour period as predicted six hours before by AIFS, compared to measured values at the Sonnblick weather station, Austria, at 47.0540 °N, 12.9577 °E. AIFS performs best when predicting results close to when the forecast was calculated, thus this is the best prediction that AIFS can produce. We also have the six hour average cloud fraction; this is defined as the fraction of points on a 13 x 13 3 km grid centred on Sonnblick, which have cloud cover, averaged over 6 hours. From this graph, it can be seen when the model is performing badly, based on the difference between AIFS and ground truth plots. Greater misalignment can generally be seen when there is a high cloud fraction (defined as cloud fraction ≥ 0.75), particularly between 2025-05-06 and 2025-05-09. This suggests that AIFS struggles to predict solar irradiation accurately under cloudy conditions. This is because AIFS is trained and outputs on a 0.25 °(roughly 25 km) grid, meaning features below this scale, which may influence solar irradiance, are underutilised. The presence of clouds massively suppresses the solar radiation reaching the surface, however, cloud structures are on a smaller scale than the AIFS grid, so can't effectively be accounted for. Thus, the model will use cloud data from MSG satellites to address this.

For training, this project's models rely on two sources of input that compliment each other in terms of spatial and temporal scale. Firstly, the ERA5 database is used to gain coarse atmospheric information on the local region. ERA5 provides a reconstruction of the Earth's atmosphere from 1940 to near real time by combining historical observations with modern numerical weather prediction models. The database offers hourly data on a 0.25° grid for over 300 distinct variables. Within the models, ERA5 data provides general atmospheric context. For each forecast initialisation time, the models are trained on a large 5 x 5 grid of points, as seen in Figure 2, hourly, for the 5 hours before initialisation time, as seen in Figure 3. For each of these points, the solar short-wave radiation downwards, 2 m temperature and 2 m dewpoint temperature are input into the model. The solar short-wave radiation downwards is used because this is the forecast output, and so gives the models information on how this is evolving prior to our forecasting

window. The 2 m temperature is the surface air temperature and correlates closely with cloud cover, while also capturing the diurnal cycle. 2 m dew-point temperature reflects the near-surface moisture content, which influences likelihood of cloud formation and presence of fog clouds. These two variables supplement solar short-wave radiation downwards by giving greater information on data relating to cloud formation, allowing the models to capture a more rich understanding of the evolution of the atmosphere prior to forecasting. Since ERA5 data is of hourly resolution, this is able to give us insights beyond the 6 hour resolution of AIFS, and thus improve its temporal capabilities.

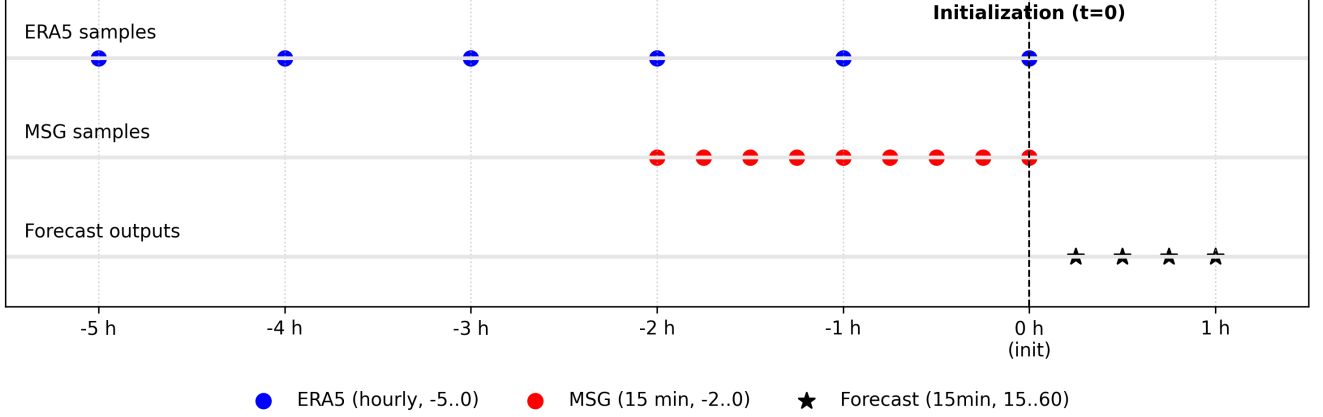


Figure 3: Temporal distribution of training data and predictions

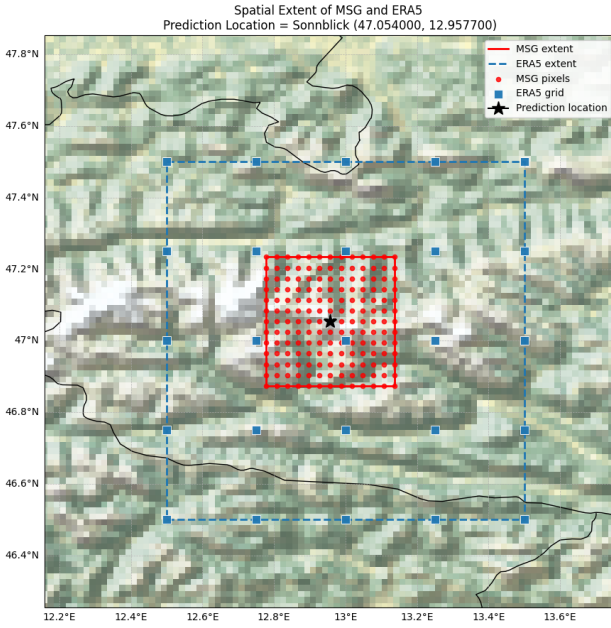


Figure 2: Spatial distribution of training data for predictions at Sonnblick Observatory, Austria

ance as it provides information on whether the sunlight is being blocked before reaching the surface. The higher spatial and temporal resolution enhance the models ability to reason beyond the course resolution of AIFS and achieve accurate forecasts for specific locations at hourly intervals.

The final input being used is the initialisation time. This contributes to the models understanding of the diurnal cycle, by giving the model an awareness of when in the cycle the forecast is being computed. Although this is indirectly encoded through the ERA5 and MSG time series, the explicit inclusion helps reinforce the day-night cycle. Thus, the model gains a broad understanding of the atmospheric state

through ERA5 and a local understanding of cloud patterns from MSG, which is complemented by the initialisation time, to capture the diurnal cycle.

During training, the models learn how patterns in data correlate with measured results (ground truth). For the ground truth, solar irradiance data from the Baseline Surface Radiation Network (BSRN) is used, which is a database of solar irradiance measurements from weather stations across the world. More information about the BSRN data used in this project can be found in Appendix A. Based on correlations between training data and ground truth, the models predict a correction to AIFS forecasts. For a machine learning model to effectively capture patterns, the quantity of training data must be maximised. With more training data, a model encounters different atmospheric situations with greater frequency during training and is thus able to adapt to capture complex dynamics with greater accuracy. Ideally, the model would be trained using many years of data, and across a dense spatial grid for the general model, however, in reality the training data is limited by times and locations in which AIFS predictions, ERA5 and MSG data and ground truth measurements align. AIFS has been operational since 26 February 2025, and ground truth measurements have only been released by BSRN up to the end of 2025 in most locations, thus giving the training data’s temporal extent. There are spatial limits due to MSG, which covers an area of -67.5° to $+67.5^{\circ}$ in longitude and latitude, and the BSRN weather station locations. The locations of the training data can be seen in Figure 4.

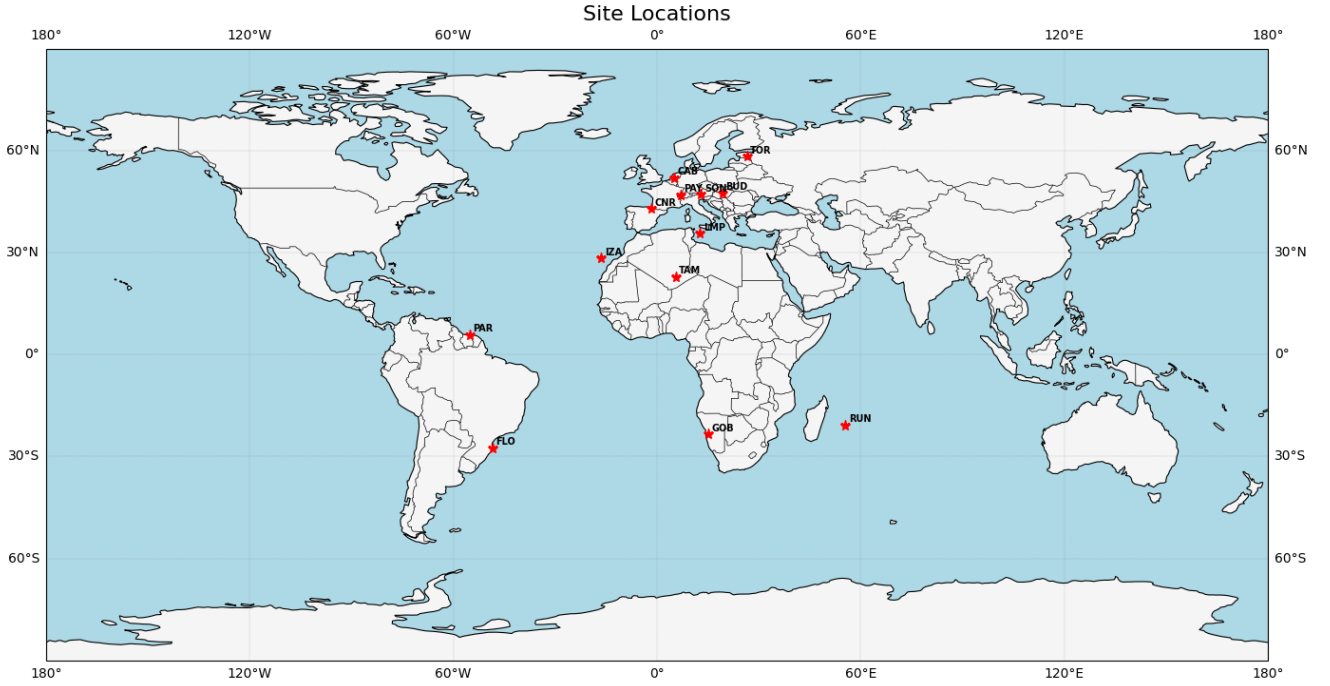


Figure 4: Locations of stations where ground truth is measured

2.3 Spatio-temporal Fourier Transformer Architecture

A transformer is a neural network architecture that models interactions between inputs by allowing each input element to learn how strongly it is influenced by all other inputs; this is called attention. In general, transformers learn patterns by applying linear transformations and attention to matrix representations of input data. After several transformations, which result in an output representation, a loss function is calculated based on the difference between the desired result and the produced result. Using this, the parameters of the linear transformations and attention mechanisms, called weights, are modified to minimise the loss. After several iterations of this process upon a large set of training data, the model should learn the correct weightings to generalise and convert unseen input data into accurate predictions. In practice, different data sources within the training data are first processed in separate channels, where they can learn internal interactions, before being combined to learn how the different data types influence

each other and produce a prediction. This allows the model to capture interactions both within and between data sources.

More specifically, the models in this project are built on a Spatio-temporal Fourier Transformer (StFT) architecture. It is a state of the art architecture proposed in March 2025 specifically for simulating dynamics of multi-scale and multi-physics systems, for which the models can act as a proof of concept [6]. Where regular transformers only take the raw spatial and temporal data as inputs, and learn patterns from these, StFT models apply a Fourier transform to the spatial data to obtain spectral channels. Only the low frequency modes of the spectral representations are retained, so the model can learn long-range interactions more freely. The spectral and spatial channels are both kept, and run in parallel before recombination, giving the model the ability to learn interactions across different scales.

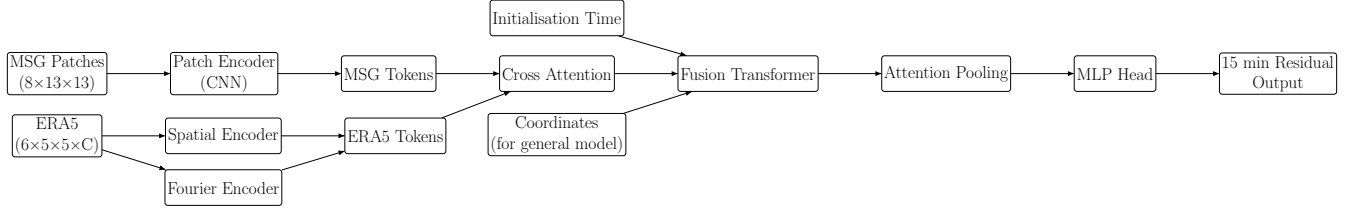


Figure 5: Spatio-temporal Fourier Transformer (StFT) architecture used for quarter hourly solar irradiance residual forecasting.

Figure 5 shows a schematic of how data passes through the models to produce predictions. Starting with the MSG inputs, the 13x13 grid of points is converted into sub-patches made of overlapping 5x5 points, before being fed into the model. This allows small areas with cloud features to be interpreted by the model rather than single points. These patches then pass through the encoder, which converts the raw data into units the model can operate on, called tokens. During this encoding, the MSG data learns internal interactions before being output as tokens. Simultaneously, the ERA5 data is split into a spatial channel and Fourier channel, and encoded before these two channels are recombined into a pool of ERA5 tokens, containing both spatial and Fourier information. The MSG tokens and ERA5 tokens interact by cross attention. This is the process of the separate data sources learning relations from each other. Specifically, in my model, the MSG tokens query the ERA5 tokens to see how the large scale atmospheric conditions influence cloud processes locally to our prediction location. Next, the processed MSG and ERA5 data are combined with initialisation time and forecasting location information in the fusion transformer, before passing through attention pooling, where the interactions are once again learned. Finally, MLP (Multi-Layer Perceptron) Head converts these tokens into a final output, in the form of 4 quarter-hourly predictions for the coming hour.

3 Results

3.1 Site-Specific Model

The site-specific model uses StFT architecture to create predictions up to one hour ahead, in quarter-hourly intervals, for specific coordinates. AIFS, ERA5 and MSG data are available everywhere within the $\pm 67.5^\circ$ longitude and latitude region of MSG, thus the site-specific model can be trained for any location within this region with solar irradiance measurements. The site-specific model in this project was trained on data from the Sonnblick weather station, Austria, however, for use in solar energy production, the model would be trained on data at the location of the solar farm of interest. The International Electrotechnical Commission has defined the international standard for monitoring the performance of photovoltaic systems, which requires the installation of irradiance sensors on-site [7]. Therefore, this model can be used at any photovoltaic plant which adheres to international standards.

As previously stated, input data for the model is only available between 26 February and 31 December

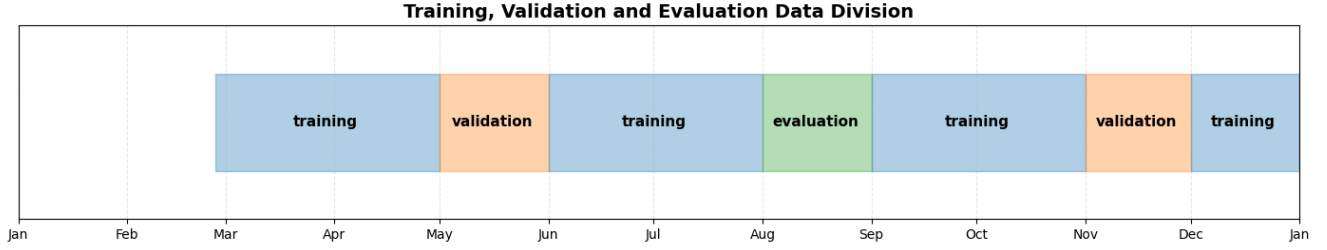


Figure 6: Temporal division of data for site-specific model into training, validation and evaluation data

2025. This must be divided into a training, validation, and evaluation dataset, which are non-overlapping datasets with unique purposes in training the model. The training dataset is where the model learns general patterns in data for prediction, and is the largest. During training, the training dataset is passed through the model several times; each run is called an epoch. During each epoch, the model generates predictions, and the mean-square error between these predictions and the ground truth is calculated. After each epoch, the parameters of the model are modified to minimise this loss, resulting in closer correlation between model predictions and the ground truth. If all available data was in the training dataset, the model would learn the training data too closely, including random noise and fluctuations, which is called over-fitting. In this case the model cannot generalise to unseen scenarios, and so the validation dataset is needed. After each epoch the validation data is passed through the model and the mean-square error is calculated. Once the model has completed training, the parameters of the epoch with the lowest validation lost is saved, as these correspond to when the model was best able to recreate unseen data during training. The evaluation dataset is used once the model has been trained to test the performance of the model on totally unseen data. Given that during training the model minimises the RMSE of the training data, and only keeps the best parameters of the epoch with the lowest validation RMSE, both of these are invalid for evaluating model performance.

For the site-specific model, the validation data is from May and November, the evaluation data is from August and the remainder is the training data resulting in a ratio of roughly 70:20:10 for training to validation to evaluation, as can be seen in Figure 6. Ideally, the training set would contain at least one full year worth of data, so that the model encounters the complete seasonal cycle during training, however, this is impossible due to the limited data available. For that reason, the validation and evaluation data is interspersed across the year; this means that although the data is unseen, similar seasonal scenarios have been encountered during training, so the model can generalise to the validation and evaluation data.

Metric	All	Day	Night	Cloud < 0.25	0.25–0.5	0.5–0.75	≥ 0.75
Baseline RMSE	128005.88	172874.14	55219.33	110568.00	129908.56	112105.66	143632.00
Site-Specific Model RMSE	105941.01	145187.39	38562.71	89279.48	122243.23	105682.92	110783.81
Absolute Difference	22064.87	27686.75	16656.62	21288.52	7665.33	6422.74	32848.19
Improvement (%)	17.23	16.02	30.16	19.25	5.90	5.73	22.87

Table 1: Performance comparison between AIFS baseline and site-specific model trained on Sonnblick, Austria, August 2025. (All RMSE in units of $\text{J/m}^2/15 \text{ mins}$)

Table 1 shows the RMSE of AIFS predictions and the site-specific model predictions, along with the absolute change in RMSE and percentage difference. Overall, the site-specific model has a 17.23 % lower RMSE than AIFS predictions. Therefore the aim of creating a more spatially localised forecast, with a higher temporal resolution than AIFS has been achieved. In particular, the greatest improvement in RMSE can be seen when cloud fraction ≥ 0.75 . Therefore, the inclusion on cloud mask has addressed the weakness of AIFS during cloudy conditions. This can in particular been seen on 2025-08-01 in Figure 7, where AIFS has overestimated the solar irradiance during high cloud cover, and the site-specific model has corrected towards the ground truth. In addition, there is a large improvement in RMSE when cloud fraction ≤ 0.25 , which also results from including cloud mask. The site-specific model has local cloud information, which AIFS does not, allowing it to explicitly distinguish between cloudy and clear scenarios. This drives improvements during all cloud levels, but most at low and high cloud cover when

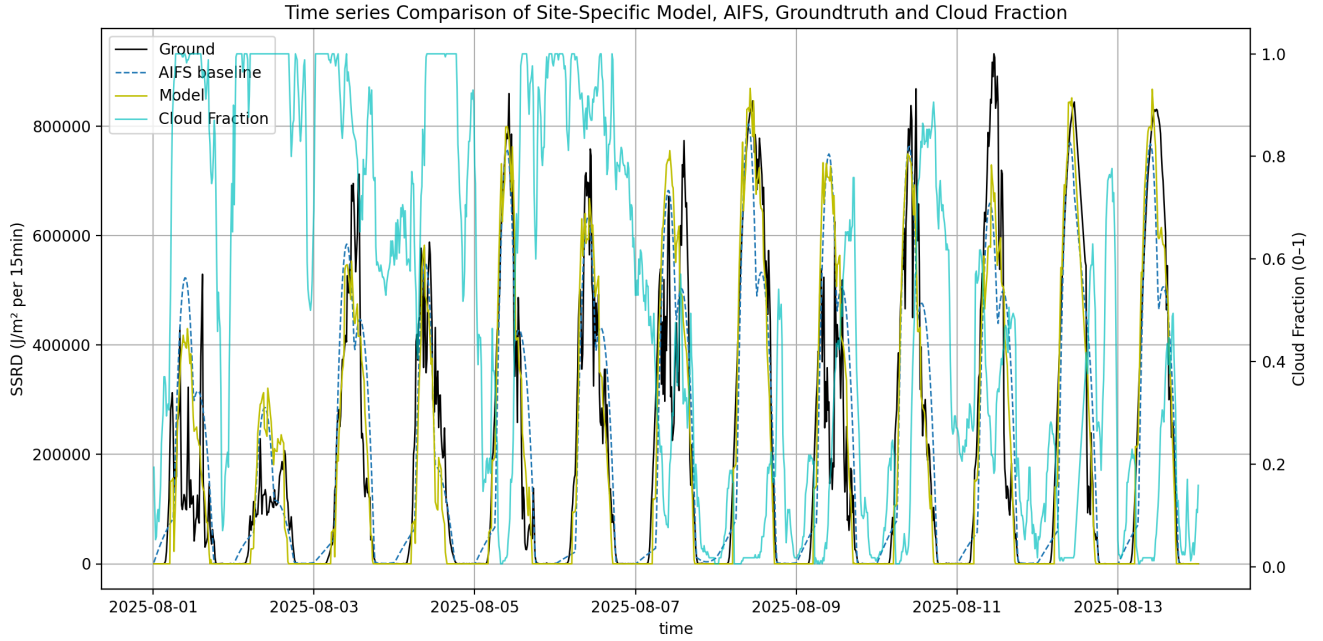


Figure 7: Time series comparison of site-specific model predictions, interpolated AIFS predictions, ground truth and cloud fraction at Sonnblick, Austria, August 2025.

the suppression of solar irradiance is uniform. Furthermore, the site-specific model improves on systematic weaknesses of the AIFS interpolation. For example, in Figure 7 AIFS begins to increase before the ground truth does everyday; suggesting a weakness in early mornings. The site-specific model captures the sharp rise in solar irradiance every morning, which can be seen both in Figure 7 and in the large percentage improvement in predictions during the night (defined as 7 pm - 7 am). It is important to note that the training period is below a year, meaning that the model is not exposed to the full range of seasonal variations; the performance could be improved with a greater backlog of training data.

3.2 General Model

The aim of the general model is to achieve improved predictions even in locations without ground truth for site-specific training. The general model has the same architecture and outputs, as the site-specific model, however, also includes coordinates as an input parameter. The main difference is in training, where all available data, shown in Figure 4, is used rather than just from a single location. Data from Sonnblick, Austria has been excluded from the training data so the performance of the general model on an unseen location can be compared to the site-specific model.

Metric	All	Day	Night	Cloud < 0.25	0.25–0.5	0.5–0.75	≥ 0.75
AIFS Baseline RMSE	110185.69	145473.01	57169.39	99084.39	113239.66	120173.12	115582.49
General Model RMSE	88749.80	119189.10	40638.92	74498.72	85731.92	98602.42	99850.32
Absolute Difference	21435.90	26283.91	16530.47	24585.67	27507.74	21570.70	15732.17
Improvement (%)	19.45	18.07	28.92	24.82	24.29	17.95	13.61

Table 2: Overall performance comparison between AIFS baseline and general model across all locations within training data, August 2025. (All RMSE in units of $\text{J/m}^2/15 \text{ mins}$)

Table 2 outlines the performance of the general model on locations within its training data set and Table 3 outlines the performance when making predictions at an unseen location, in particular Sonnblick, Austria. Table 2 shows overall the general model has a 19.45 % lower RMSE than AIFS for locations within the training set. However, the smallest improvement occurs during cloudy conditions, so improvements don't

Metric	All	Day	Night	Cloud < 0.25	0.25–0.5	0.5–0.75	≥ 0.75
Baseline RMSE	128005.88	172874.14	55219.33	110568.00	129908.56	112105.66	143632.00
Model RMSE	114385.34	157663.72	38116.66	94358.61	127675.79	103760.15	125839.95
Absolute Difference	13620.53	15210.43	17102.67	16209.39	2232.77	8345.51	17792.05
Improvement (%)	10.64	8.80	30.97	14.66	1.72	7.44	12.39

Table 3: Performance comparison between AIFS baseline and general model for a location not included in training data (Sonnblick, Austria), August 2025. (All RMSE in units of $\text{J/m}^2/15 \text{ mins}$)

entirely come from the inclusion of MSG cloud mask data, but rather from the correction of systematic biases within AIFS. Some of these systematic biases are results of the polynomial interpolation of AIFS required to change the temporal resolution from the original 6 hours to 15 minutes; therefore a core success of the general model is reducing the temporal granularity of AIFS to fit with the aims of this project. The model focuses on these corrections due to the diverse training locations, meaning that the model can most easily correct AIFS deficiencies that occur at all locations, such as early morning over-prediction, seen in Figure 8.

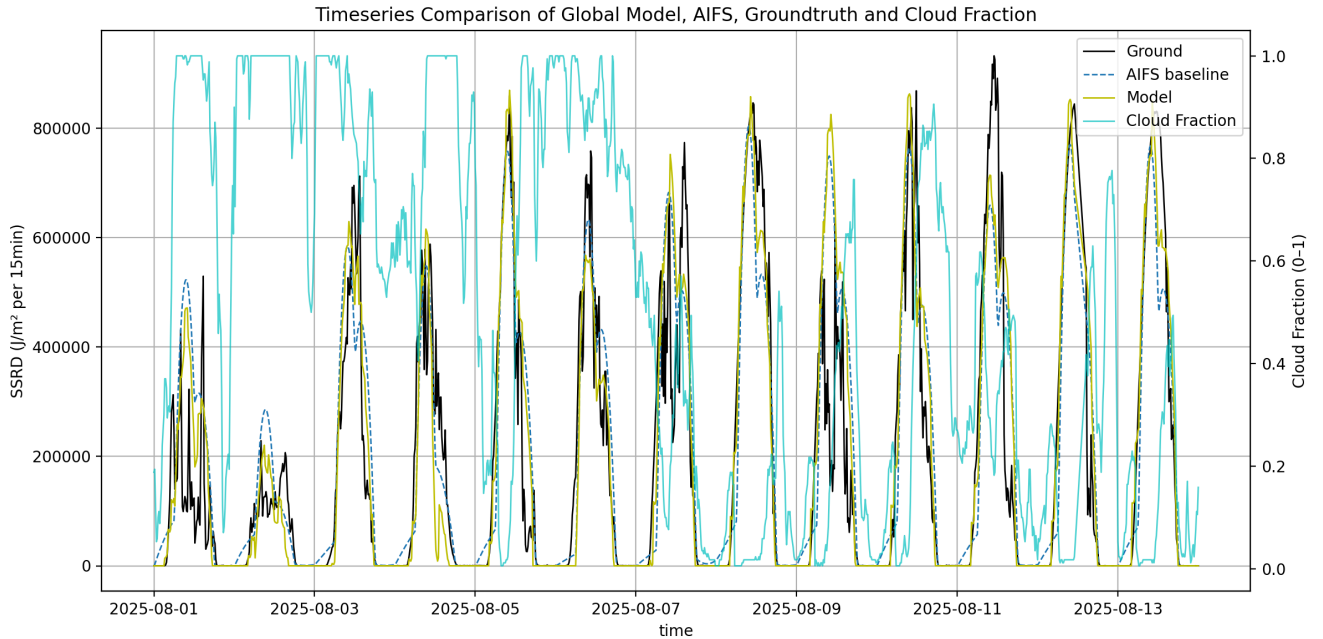


Figure 8: Time series comparison of general model predictions, interpolated AIFS predictions, ground truth and cloud fraction at Sonnblick, Austria, August 2025.

Table 3 shows the performance of the general model for Sonnblick, Austria in August 2025. Sonnblick was not included in the training data, and so Table 3 shows the general models ability to generalise to unseen locations. The general model has a 10.64 % lower RMSE than AIFS, demonstrating the model has not only learnt location specific behaviours, but general corrections to AIFS. There are improvements during high cloud conditions, as can be seen on 2025-08-01 and 2025-08-02 in Figure 8, where the general model corrects an overestimate by AIFS. However, there is a greater absolute difference in RMSE during the night (7 pm - 7 am) than during the day (7 am - 7 pm), which once again suggests the general model mainly addresses systematic biases in AIFS. If the model was mostly correcting issues in the physical reasoning of AIFS, rather than recurring biases, it would be expected that the largest absolute difference in RMSE would occur in conditions where AIFS struggles most. In the site-specific model this has been achieved, where the largest improvements are during high cloud cover. Whereas, the general model has the largest absolute improvement during the night, when AIFS already performs well. Therefore, the general model mostly corrects for recurring non-physical biases in AIFS.

4 Conclusion

In conclusion, by using a Spatio-Temporal Fourier Transformer machine learning architecture, the models have successfully combined coarse atmospheric data from ERA5 and local cloud information from MSG, to generate solar irradiance forecasts which are better suited to photovoltaic plant applications than AIFS. AIFS is unsuitable for predictions in solar energy production due to its coarse spatial and temporal resolution, and weakness during cloudy conditions. The models within this project address these limits by utilising training data with a finer resolution in space and time, and incorporating cloud mask data; this is used to generate quarter-hourly predictions up to an hour ahead, for specific coordinates. Overall, the site-specific model achieved a 17.23 % lower RMSE than AIFS, with the greatest improvements seen during situations with cloud cover over 0.75. In contrast, the general model saw a 10.64 % lower RMSE than AIFS, with improvements mostly appearing due to corrections of interpolation and systematic biases, rather than addressing the physics-base shortcomings of AIFS.

Currently, improvements arise due to physics-based corrections, seen in better predictions during high cloud cover, and systematic bias and interpolation corrections. Both are valuable, however, to disentangle the contributions would require further work such as comparison against a purely statistical bias-correction baseline. In the future, the models could be improved in several ways while still keeping the core architecture unchanged. Primarily, as more data is released by BSRN, more training data can be obtained. This allows the models to see a wider range of atmospheric scenarios during training, and thus have greater predictions capabilities. In addition, Meteosat Third Generation (MTG) satellites have recently been made operational and will supply better imaging and atmospheric sounding data than MSG. Using this improved data, cloud opacity and height could be added as input parameters, which would give the models better understanding of cloud structure. Currently, the only cloud information inputted is cloud mask, which is only a binary value. Although this is very valuable, cloud opacity would give information on how much solar radiation is transmitted through the clouds and cloud height would give information on indirect solar radiation reaching any particular location, thus improving the models reasoning capacity with relation to clouds.

References

- [1] International Energy Agency. Solar pv. <https://www.iea.org/energy-system/renewables/solar-pv>, 2023.
- [2] Gabriel Moldovan, Ewan Pinnington, Ana Prieto Nemesio, Simon Lang, Zied Ben Bouallègue, Jesper Dramsch, Mihai Alexe, Mario Santa Cruz, Sara Hahner, Harrison Cook, Helen Theissen, Mariana Clare, Cathal O'Brien, Jan Polster, Linus Magnusson, Gert Mertes, Florian Pinault, Baudouin Raoult, Patricia de Rosnay, Richard Forbes, and Matthew Chantry. An update to ecmwf's machine-learned weather forecast model aifs, 2025.
- [3] George W. Platzman. The eniac computations of 1950—gateway to numerical weather prediction. *Bulletin of the American Meteorological Society*, 60(4):302 – 312, 1979.
- [4] Nedjeljka Žagar. A global perspective of the limits of prediction skill of nwp models. *Tellus A: Dynamic Meteorology and Oceanography*, 69(1):1317573, 2017.
- [5] Muhammad Waqas, Usa Wannasingha Humphries, Bunthid Chueasa, and Angkool Wangwongchai. Artificial intelligence and numerical weather prediction models: A technical survey. *Natural Hazards Research*, 5(2):306–320, 2025.
- [6] Da Long, Shandian Zhe, Samuel Williams, Leonid Oliker, and Zhe Bai. Stft: Spatio-temporal fourier transformer for long-term dynamics prediction, 2025.
- [7] International Electrotechnical Commission. IEC 61724-1:2021 Photovoltaic system performance – Part 1: Monitoring. <https://www.iec.ch>, 2021.

A BSRN Weather Station Information

Budapest:

Station Code: BUD

Longitude: 19.1822

Latitude: 47.4291

Reference: Fekete, Dénes (2022): Basic and other measurements of radiation at station Budapest (2025-02 et seq) [dataset publication series]. Hungarian Meteorological Service, PANGAEA, <https://doi.org/10.1594/PANGAEA.946379>

Cabauw:

Station Code: CAB

Longitude: 51.9680

Latitude: 4.9280

Reference: Knap, Wouter (2022): Basic and other measurements of radiation at station Cabauw (2025-02 et seq) [dataset publication series]. Koninklijk Nederlands Meteorologisch Instituut, De Bilt, PANGAEA, <https://doi.org/10.1594/PANGAEA.940531>

Cener:

Station Code: CNR

Longitude: 42.8160

Latitude: -1.6010

Reference: Olano, Xabier (2021): Basic measurements of radiation at station Cener (2025-02 et seq) [dataset publication series]. National Renewable Energy Centre, PANGAEA, <https://doi.org/10.1594/PANGAEA.931893>

Florianopolis:

Station Code: FLO

Longitude: -27.6047

Latitude: -48.5227

Reference: Starke, Allan R; da Silva, Alexandre K; Colle, Sergio (2025): Basic measurements of radiation at station Florianopolis (2025-02 et seq) [dataset publication series]. Universidade Federal de Santa Catarina, PANGAEA, <https://doi.org/10.1594/PANGAEA.974493>

Gobabeb:

Station Code: GOB

Longitude: -23.5614

Latitude: 15.0420

Reference: Vogt, Roland (2024): Basic and other measurements of radiation at station Gobabeb (2025-02 et seq) [dataset publication series]. Atmospheric Sciences, Department of Environmental Sciences, University of Basel, PANGAEA, <https://doi.org/10.1594/PANGAEA.968639>

Izaña:

Station Code: IZA

Longitude: 28.3093

Latitude: -16.4993

Reference: Torres, Carlos J; Cuevas-Agulló, Emilio; García Cabrera, Rosa D (2024): Basic and other measurements of radiation at station Izaña (2025-02 et seq) [dataset publication series]. Izaña Atmospheric Research Center, Meteorological State Agency of Spain, PANGAEA, <https://doi.org/10.1594/PANGAEA.968676>

Lampedusa:

Station Code: LMP

Longitude: 35.5180

Latitude: 12.6300

Reference: Meloni, Daniela; Sferlazzo, Damiano Massimiliano (2024): Basic measurements of radiation at station Lampedusa (2025-02 et seq) [dataset publication series]. National Agency for New Technologies, Energy and Sustainable Economic Development, PANGAEA, <https://doi.org/10.1594/PANGAEA.973604>

Paramaribo:

Station Code: PAR

Longitude: 5.8060

Latitude: -55.2146

Reference: Piters, Ankie (2025): Basic measurements of radiation at station Paramaribo (2025-02 et seq) [dataset publication series]. Koninklijk Nederlands Meteorologisch Instituut, De Bilt, PANGAEA, <https://doi.org/10.1594/PANGAEA.979681>

Payerne:

Station Code: PAY

Longitude: 46.8123

Latitude: 6.9422

Reference: Vuilleumier, Laurent; Heimo, Alain (2022): Basic and other measurements of radiation at station Payerne (2025-02 et seq) [dataset publication series]. Federal Office of Meteorology and Climatology MeteoSwiss, Payerne, PANGAEA, <https://doi.org/10.1594/PANGAEA.946384>

Reunion Island:

Station Code: RUN

Longitude: -20.9014

Latitude: 55.4836

Reference: Morel, Beatrice; Delsaut, Mathieu (2023): Basic measurements of radiation at station Reunion Island, University (2025-02 et seq) [dataset publication series]. Université de La Réunion, PANGAEA, <https://doi.org/10.1594/PANGAEA.962049>

Sonnblick:

Station Code: SON

Longitude: 47.0540

Latitude: 12.9577

Reference: Olefs, Marc (2022): Basic measurements of radiation at station Sonnblick (2025-02 et seq) [dataset publication series]. Zentralanstalt für Meteorologie und Geodynamik, Wien, PANGAEA, <https://doi.org/10.1594/PANGAEA.946389>

Tamanrasset:

Station Code: TAM

Longitude: 22.7903

Latitude: 5.5292

Reference: Baika, Sidi; Saanoune, Abdessadak; Mimouni, Mohamed (2024): Basic measurements of radiation at station Tamanrasset (2025-02 et seq) [dataset publication series]. National Meteorological Office of Algeria, PANGAEA, <https://doi.org/10.1594/PANGAEA.965124>

Toravere:

Station Code: TOR

Longitude: 58.2641

Latitude: 26.4613

Reference: Kallis, Ain (2022): Basic and other measurements of radiation at station Toravere (2025-02 et seq) [dataset publication series]. Tartu Observatoorium, Toravere, PANGAEA, <https://doi.org/10.1594/PANGAEA.946387>