What if We Enrich *day-ahead* Solar Irradiance Time Series Forecasting with Spatio-Temporal Context?

Oussama Boussif [*1] **Ghait Boukachab** [*1 2] **Dan Assouline** [*1] **Stefano Massaroli** [1] **Tianle Yuan** [3]
Loubna Benabbou [2] **Yoshua Bengio** [1]

Abstract

The global integration of solar power into the electrical grid could have a crucial impact on climate change mitigation, yet poses a challenge due to solar irradiance variability. We present a deep learning architecture which uses spatio-temporal context from satellite data for highly accurate day-ahead time-series forecasting, in particular Global Horizontal Irradiance (GHI). We provide a multi-quantile variant which outputs a prediction interval for each time-step, serving as a measure of forecasting uncertainty. In addition, we suggest a testing scheme that separates easy and difficult scenarios, which appears useful to evaluate model performance in varying cloud conditions. Our approach exhibits robust performance in solar irradiance forecasting, including zero-shot generalization tests at unobserved solar stations, and holds great promise in promoting the effective use of solar power and the resulting reduction of CO_2 emissions.

1. Introduction

Solar power is a vital renewable energy source with the potential to mitigate climate change effects by reducing greenhouse gas emissions (Doblas-Reyes et al., 2021; IEA, 2021). However, the variable nature of solar irradiance – the amount of solar radiation reaching the Earth's surface – poses a challenge for seamless integration of solar power into the electricity grid. Accurate solar irradiance forecasting can help grid operators manage this variability, leading to more efficient and reliable grid integration of solar power, and reducing the need for costly, environmentally damaging backup power sources.

Solar irradiance is influenced by multiple factors, including time of day, season, weather patterns, and the sun's position. Clouds cause the most variability as they block and scatter solar radiation. Therefore, accurate solar irradiance forecasting requires effective modeling of cloud cover.

Although prior research has leveraged time-series approaches to forecast solar irradiance (Yang et al., 2022), few have incorporated cloud cover (Nielsen et al., 2021; Bone et al., 2018; Si et al., 2021), particularly for the challenging task of **day-ahead forecasting**. When forecasting solar irradiance for a specific station, relying solely on local physical variables is insufficient due to the spatial variability of cloud cover. To accurately anticipate cloud impact on incoming solar radiation, it's important to consider the motion and trajectory of clouds within a larger spatial context.

We address these gaps by incorporating satellite imaging for solar irradiance forecasting and propose a multi-modal architecture capable of forecasting in principle any physical variable. We also highlight the limitations of conventional testing schemes which use metrics like MAE or RMSE on the entire dataset, as they fail to capture model performance in critical cloud-related scenarios. In order to alleviate this problem, we propose a new testing scheme based on multiple splits of the test data, separating particularly difficult examples from easy ones. Our primary contributions are:

- We develop CrossViViT, a deep learning architecture that uses spatio-temporal context (including satellite data) for highly accurate day-ahead time-series forecasting at any station, *even those unseen during training*, with a particular focus on GHI.
- We present a Multi-Quantile version of CrossViViT, which provides uncertainty estimation for each prediction, applicable to any forecasting task.
- We propose a testing scheme separating difficult and easy examples, allowing for a more nuanced evaluation of model performance.

[*]Equal contribution [1]Mila and Université de Montréal [2]Université du Québec à Rimouski [3]NASA Goddard Space Flight Center. Correspondence to: Oussama Boussif <oussama.boussif@mila.quebec>, Ghait Boukachab <ghait.boukachab@mila.quebec>, Dan Assouline <dan.assouline@mila.quebec>.

Accepted after peer-review at the 1st workshop on Synergy of Scientific and Machine Learning Modeling, SynS & ML ICML, Honolulu, Hawaii, USA. July, 2023.

2. Methodology

We propose a framework for predicting solar irradiance by integrating spatio-temporal context and historical data from various stations. This framework is influenced by recent video transformer models (Arnab et al., 2021; Feichtenhofer et al., 2022) and multi-modal models that use diverse data sources such as images and time series (Liu et al., 2023). Our architecture, CrossViViT, is detailed in the following part, including its main features and design principles.

2.1. Cross Video Vision Transformer for time-series forecasting (CrossViViT)

The overall methodology, depicted in Figure 1, can be summarized as follows:

1. **Tokenizing**: The video context $\mathbf{V} \in \mathbb{R}^{T \times C_{ctx} \times H \times W}$, with T frames for each of the C_{ctx} channels, and H and W respectively the height and width of the video images, is divided into N_p non-overlapping patches and linearly projected into a sequence of d-dimensional context tokens $\mathbf{z}^{ctx} \in \mathbb{R}^{T \times N_p \times d}$. We use the *Uniform frame sampling* ViViT scheme (Arnab et al., 2021) to embed the videos, the frames being concatenated along the batch dimension. The time series $\mathbf{t} \in \mathbb{R}^{T \times C_{ts}}$ are linearly projected into a sequence of d-dimensional time-series tokens $\mathbf{z}^{ts} \in \mathbb{R}^{T \times d}$. We augment the context tokens with ROPE (Su et al., 2021) , and a learnt positional encoding for the time-series tokens.

2. **Masking**: As a regularizing mechanism, we allow the model to mask a portion of the video context. During the training phase, a masking ratio m_{ctx} is randomly sampled from a uniform distribution $U(0, 0.99)$, and the corresponding patches are masked accordingly. We note that during inference, no masking is applied.

3. **Encoding:** We encode the time series and the past video context separately with two transformer architectures: a L-layer ViT for the video context, and a L-layer Transformer for the input time series.

4. **Mixing:** We combine the resulting context and time-series latents, respectively $\mathbf{z}_L^{ctx}$ and $\mathbf{z}_L^{ts}$, within L layers of a Transformer with Cross Attention (CA) (Vaswani et al., 2017). After adding ROPE, the two L-th layers are mixed with CA and passed through an MLP block. The output of each layer becomes a latent which is in turn mixed with the context latent $\mathbf{z}_L^{ctx}$ and again passed through a block of MLP. Formally, the following operations are performed respectively at the first layer ((1) and (2)) and on the remaining layers ((3) and (4)) of the CA:

$$\mathbf{y}_1^{mix} = \text{CA}(\text{LN}(\mathbf{z}_L^{ctx}, \mathbf{z}_L^{ts})) + \mathbf{z}_L^{ts} \quad (1)$$

$$\mathbf{z}_2^{mix} = \text{MLP}(\text{LN}(\mathbf{y}_1^{mix})) + \mathbf{y}_1^{mix} \quad (2)$$

$$\mathbf{y}_l^{mix} = \text{CA}(\text{LN}(\mathbf{z}_L^{ctx}, \mathbf{z}_l^{mix})) + \mathbf{z}_l^{mix} \quad (3)$$

$$\mathbf{z}_{l+1}^{mix} = \text{MLP}(\text{LN}(\mathbf{y}_l^{mix})) + \mathbf{y}_l^{mix} \quad (4)$$

5. **Decoding:** The sequence of mixed tokens $\mathbf{z}_L^{mix}$ returned by the layers of Cross Transformer is then passed through N layers of another Transformer as a decoder, before adding a learnt positional embedding to the token sequence. The output decoded sequence $\mathbf{z}_N$ is passed through a final MLP head to output the final predicted future time series $\mathbf{t}_{pred} \in \mathbb{R}^{T \times C_{ts}}$.

2.2. Multi-Quantiles: Extracting prediction intervals

We've adapted the CrossViViT architecture to predict intervals by replacing the original MLP head with multiple parallel MLPs, each predicting a specific quantile of the distribution per time step. We use distinct quantile loss functions for each MLP head, and their sum gives us the Multi-Quantile loss, which is the model's training objective. The quantile loss (Koenker & Hallock, 2001) $L_\alpha(y, \hat{y})$ for the α quantile is defined as:

$$L_\alpha(y, \hat{y}) = \max\{\alpha(\hat{y} - y), (1 - \alpha)(y - \hat{y})\} \quad (5)$$

The Multi-Quantile loss is then defined as: $MQL(y, \hat{y}) = \sum_{\alpha \in v_\alpha} L_\alpha(y, \hat{y})$. The selection of quantile heads v_α is a crucial hyperparameter that determines the density of the output distribution generated by the model. To achieve a 96% prediction interval while maintaining a sufficiently dense distribution, we set the list of quantiles as $v_\alpha = [0.02, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.9, 0.98]$.

3. Dataset

This section provides a description of the dataset designed for this study and shared publicly, including all the applied pre-processing steps.

Time series This study uses 15 years (2008-2022) of radiation data from six locations, collected at 30-minute intervals from the Baseline Surface Radiation Network (BSRN). The data captures diverse irradiance patterns and includes measurements of pressure, clear sky components, Direct Normal Irradiance (DNI), and Diffuse Horizontal Irradiance (DHI). Global Horizontal Irradiance (GHI) is calculated using DNI, DHI, and the sun's zenith angle, using the pvlib python library (Holmgren et al., 2020). The Ineichen model (Ineichen, 2016) from pvlib provides clear sky components.

Satellite images We use the EUMETSAT Rapid Scan Service dataset (Rothfuss, 2015), spanning 2008-2022, focusing on the 11 non-High Resolution Visible channels. These

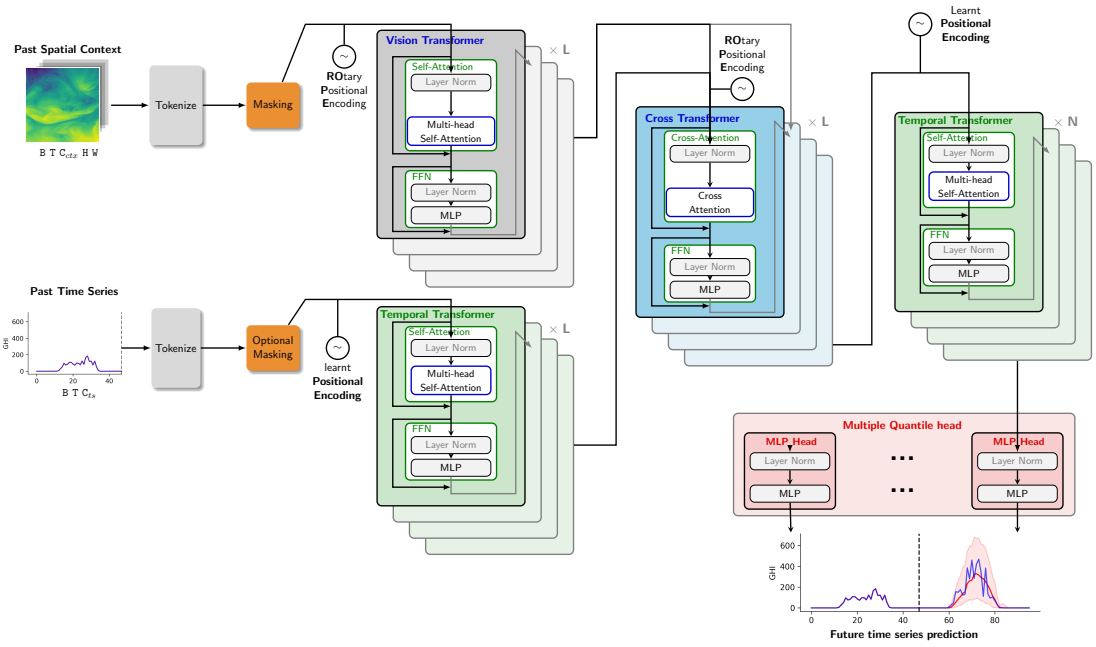


Figure 1. **CrossViViT architecture, in its Multi-Quantile version.**

channels, with a spatial resolution of 6-9km, cover the upper third of Earth, especially Europe, and include information from Infrared and Water vapor channels. The original data was reprojected onto the World Geodetic System 1984 (WGS 84) coordinate system (Rothfuss, 2015). We additionally compute the optical flow for each channel using the TVL1 algorithm (Sánchez Pérez et al., 2013), representing the cloud motion and add an elevation map as an additional feature. We downscale the pre-processed satellite data from a resolution of 512^2 to 64^2 for computational efficiency.

4. Experiments and Results

We compare our architecture with baselines and discuss the experiment setup, results, and comparisons under different test configurations. Our split methodology evaluates the model in difficult prediction situations, important for downstream tasks related to solar irradiance estimation. The models use a 9-year dataset (2008-2016) from IZA, CNR, and PAL stations for training, a separate 3-year dataset (2017-2019) from the PAY station for validation, and another 3-year dataset (2020-2022) from TAM and CAB stations for testing. The models employ a sliding window approach, using 24-hour historical input to predict the next 24-hour Global Horizontal Irradiance (GHI).

4.1. Baselines

We conduct a comprehensive comparison between our approach and several state-of-the-art deep learning architectures and propose tailored baselines for solar irradiance forecasting. These baselines include the **Persistence model**, which relies on the previous day's data for predictions, and the **Clear Sky baseline**, which utilizes computable clear sky components (Ineichen, 2016). Additionally, we employ **Fourier approximations** with different numbers of modes (3, 4, and 5) and apply a low-pass filter to generate Fourier-based baselines. Detailed information on these architectures and baselines can be found in Table 1.

4.2. "Hard" vs. "Easy" forecasting scenarios

We evaluate the model's ability in predicting cloud-induced GHI fluctuations by assessing it on different time splits at test stations. This helps identify the model's strengths and shortcomings compared to prior methods, giving insights into specific scenarios where CrossViViT excels or falls short. Given the Persistence baseline's effectiveness when GHI values are similar over consecutive days, we propose a time split approach based on GHI variation, categorizing examples as "Easy" or "Hard." "Easy" cases have minimal GHI changes over days, making Persistence effective, whereas "Hard" cases have significant GHI changes that challenge Persistence. We use a measure based on the area ratio under the GHI curve for two days to quantify similarity. This measure, denoted as $r = \left|\log \frac{y}{y_{\text{prev}}}\right|$, assigns equal importance to ratios such as 0.5 and 2. By using a threshold of $\left|\log\left(\frac{2}{3}\right)\right|$, we classify cases as "Easy" when r is below the threshold and as "Hard" otherwise.

Models	Parameters	CAB (2020-2022)						TAM (2017-2019)					
		All (9703)		Easy (5814)		Hard (3889)		All (2299)		Easy (2064)		Hard (235)	
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Persistence	N/A	63.57	131.44	52.56	109.05	80.04	159.14	**32.26**	94.71	**20.8**	59.47	132.92	238.12
Fourier$_3$	N/A	68.91	121.23	56.51	93.854	87.46	153.29	56.0	94.85	45.48	62.62	148.42	231.47
Fourier$_4$	N/A	65.74	123.15	53.82	96.38	83.56	154.76	44.02	92.22	33.15	57.56	139.56	232.61
Fourier$_5$	N/A	64.67	124.22	52.67	97.94	82.61	155.44	40.26	**91.36**	28.94	**55.57**	139.68	233.52
Clear Sky (Ineichen, 2016)	N/A	67.19	140.11	60.55	125.66	77.12	159.28	40.61	98.02	31.07	63.4	124.42	242.26
ReFormer (Kitaev et al., 2020)	8.6M	57.42	102.73	53.75	92.97	62.92	115.81	81.6	137.04	78.57	129.72	108.22	189.55
Informer (Zhou et al., 2021)	56.7M	72.26	122.89	70.85	118.85	74.35	128.69	83.43	140.38	82.6	138.46	**90.66**	**156.22**
FiLM (Zhou et al., 2022b)	9.4M	68.37	116.86	59.66	95.35	81.4	143.11	62.72	99.71	54.99	77.63	130.66	210.58
PatchTST (Nie et al., 2023)	9.6M	60.76	119.41	54.77	107.71	69.7	135.01	66.94	132.44	62.4	124.25	106.77	189.72
LighTS (Zhang et al., 2022)	32K	54.91	102.88	49.55	**89.28**	62.92	120.38	68.51	114.59	64.61	104.98	102.77	177.98
CrossFormer (Zhang & Yan, 2023)	227M	55.98	101.84	51.59	90.2	62.55	117.11	68.85	116.45	65.4	107.88	99.16	175.08
FEDFormer (Zhou et al., 2022a)	23.6M	56.38	**99.27**	53.08	90.13	61.31	**111.54**	92.12	146.52	91.13	142.83	100.82	175.64
DLinear (Zeng et al., 2022)	4.7K	75.01	121.01	65.21	99.72	89.65	147.21	75.54	115.40	69.04	98.74	132.67	211.28
AutoFormer (Wu et al., 2021)	50.4M	64.34	104.53	60.81	95.14	69.63	117.17	115.88	170.91	117.36	171.07	102.87	169.47
CrossViViT	145M	**50.35**	**99.18**	**47.04**	**89.6**	**55.30**	**112.00**	49.46	94.96	44.01	79.91	97.40	179.30
		MAE	p_t	MAE	p_t	MAE	p_t	MAE	p_t	MAE	p_t	MAE	p_t
Multi-Quantile CrossViViT	78.8M	61.80	0.91	57.03	0.93	68.94	0.90	81.20	0.71	78.93	0.70	101.18	0.75

Table 1. Comparison of model performances across **test stations** TAM and CAB, during **test years** (2020-2022) for CAB, and **val years** (2017-2019) for TAM. We report the MAE and RMSE for the easy and difficult splits presented in section 4.2 along with the number of data points for each split. We add the MAE resulting from the Multi-Quantile CrossViViT median prediction, along with p_t, the probability for the ground-truth to be included within the interval, averaged across time steps.

4.3. Performance on stations and years outside the training distribution

Table 1 provides a detailed comparison between CrossViViT, state-of-the-art timeseries models, and *dummy* baselines. The evaluation focuses on the test stations TAM and CAB, spanning the periods 2017-2019 and 2020-2022, respectively. We note however that since the 2020-2022 period is unavailable for TAM, we use 2017-2019 instead.

CrossViViT achieves the lowest MAE among the time-series models on the TAM station during the 2017-2019 period. However, the persistence baseline still outperforms our approach. This discrepancy can be attributed to the characteristics of the TAM station, located in a desert region known for clear and sunny days. The inclusion of cloud information in our model may occasionally lead to underestimation of GHI in such clear-sky conditions. Additionally, the training dataset primarily consists of data from a single "sunny" station (IZA), limiting exposure to clear-sky patterns. These findings suggest that for stations with low irradiance intermittency, a combination of persistence and clear-sky models may suffice. On the CAB station during the 2020-2022 period, CrossViViT surpasses all baselines across various time splits. This improvement can be attributed to the specific meteorological conditions of the CAB station, which experiences a higher frequency of cloudy days. It confirms CrossViViT's abilities under cloudier conditions.

Regarding Multi-Quantile CrossViViT, we include the MAE of the median prediction, along with the test confidence p_t obtained for the prediction interval: the probability for the ground-truth to be included within the interval, for each time step, averaged across the entire dataset. Note that the goal is not to provide the best prediction from the median but rather to provide confident prediction intervals, with a high p_t. The prediction intervals achieved a high level of confidence, surpassing 0.9, for the unseen CAB station. However, for the TAM station, the harsh environmental conditions of the desert posed a challenge for reliable estimation of prediction intervals. Although the median prediction results were comparatively inferior to those of a baseline method, it demonstrates consistent patterns in its variation across easy and difficult cases. Furthermore, the test confidence of the proposed method remains relatively constant across different splits, for both stations.

5. Conclusion, Limitations, and Future Work

We propose CrossViViT, an accurate day-ahead solar irradiance forecasting architecture that leverages spatio-temporal context through satellite data and enables prediction distribution extraction for each time step. Our testing scheme captures crucial scenarios, including varying cloud conditions, demonstrating the robustness of CrossViViT in solar irradiance forecasting which could contribute to the effective integration of solar power into the grid, even in zero-shot tests at unobserved stations and years. However, additional data would enhance the validation of our approach, and exploring different prediction and context horizons would

improve the model's robustness. Investigating a cropping methodology as a regularization technique for the context is also worth considering. Based on these promising results, we intend to investigate the applicability of CrossViViT to other context-dependent physical variables.

Broader impact

This research holds profound potential to positively influence society by promoting the effective integration of solar power into our electrical grid, a significant stride towards carbon neutrality and climate change mitigation. Accurate solar irradiance forecasts will enable more efficient grid management and reduced reliance on fossil-fuel reserves. Ethically, it is crucial that this research and its resulting technologies be accessible to all, not widening the gap between developed and developing nations in the transition to renewable energy giving the fact that our region of interest includes north Africa. Ensuring fair access to the benefits of this research will contribute to sustainable development goals and a more equitable energy future.

References

Net Zero by 2050: A Roadmap for the Global Energy Sector. OECD, May 2021. doi: 10.1787/c8328405-en. URL https://www.iea.org/reports/net-zero-by-2050.

Alzahrani, A., Shamsi, P., Dagli, C., and Ferdowsi, M. Solar Irradiance Forecasting Using Deep Neural Networks. *Procedia Computer Science*, 114:304–313, 2017. ISSN 1877-0509. doi: https://doi.org/10.1016/j.procs.2017.09.045. URL https://www.sciencedirect.com/science/article/pii/S1877050917318392.

Arnab, A., Dehghani, M., Heigold, G., Sun, C., Lučić, M., and Schmid, C. Vivit: A video vision transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 6836–6846, October 2021.

Baika, S. Meteorological synoptical observations from station Tamanrasset (2022-12), 2023. URL https://doi.org/10.1594/PANGAEA.954045. Backup Publisher: National Meteorological Office of Algeria Type: data set.

Bone, V., Pidgeon, J., Kearney, M., and Veeraragavan, A. Intra-hour direct normal irradiance forecasting through adaptive clear-sky modelling and cloud tracking. *Solar Energy*, 159:852–867, 2018. doi: 10.1016/j.solener.2017.10.037.

Boussif, O., Bengio, Y., Benabbou, L., and Assouline, D. Magnet: Mesh agnostic neural pde solver. In Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., and Oh, A. (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 31972–31985. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/cf4c7ee0734cdfe09a099cf6cd7b117a-Paper-Conference.pdf.

Brandstetter, J., Worrall, D. E., and Welling, M. Message passing neural pde solvers. *International Conference On Learning Representations*, 2022.

BSRN. Baseline surface radiation network. URL https://bsrn.awi.de/.

Cuevas-Agulló, E. Radiosonde measurements from station Izana (2014-08), 2014. URL https://doi.org/10.1594/PANGAEA.835518. Backup Publisher: Izaña Atmospheric Research Center, Meteorological State Agency of Spain Type: data set.

Doblas-Reyes, F., Sörensson, A., Almazroui, M., Dosio, A., Gutowski, W., Haarsma, R., Hamdi, R., Hewitson, B., Kwon, W.-T., Lamptey, B., Maraun, D., Stephenson, T., Takayabu, I., Terray, L., Turner, A., and Zuo, Z. *Linking Global to Regional Climate Change Supplementary Material*. 2021. URL Availablefromhttps://www.ipcc.ch/.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

Equer, L., Rusch, T. K., and Mishra, S. Multi-scale message passing neural pde solvers. *ARXIV.ORG*, 2023. doi: 10.48550/arXiv.2302.03580.

Feichtenhofer, C., Li, Y., He, K., et al. Masked autoencoders as spatiotemporal learners. *Advances in neural information processing systems*, 35:35946–35958, 2022.

Haeffelin, M. Basic measurements of radiation at station Palaiseau (2009-02), 2014. URL https://doi.org/10.1594/PANGAEA.830019. Backup Publisher: Laboratoire de Météorologie Dynamique du C.N.R.S., Ecole Polytechnique Type: data set.

Holmgren, W., Calama-Consulting, Hansen, C., Mikofski, M., Lorenzo, T., Krien, U., bmu, Anderson, K., Stark, C., DaCoEx, Driesse, A., konstant_t, mayudong, de León Peque, M. S., Heliolytics, Miller, E., Anoma, M. A., Boeman, L., jforbess, tylunel, Guo, V., Morgan, A., Stein, J., Leroy, C., R, A. M., JPalakapillyKWH, Dollinger, J., Anderson, K., MLEEFS, and

Dowson, O. pvlib/pvlib-python: v0.8.0, September 2020. URL https://doi.org/10.5281/zenodo.4019830.

Ineichen, P. Validation of models that estimate the clear sky global and beam solar irradiance. *Solar Energy*, 132:332–344, 2016. ISSN 0038-092X. doi: https://doi.org/10.1016/j.solener.2016.03.017. URL https://www.sciencedirect.com/science/article/pii/S0038092X16002048.

Jiang, C. M., Esmaeilzadeh, S., Azizzadenesheli, K., Kashinath, K., Mustafa, M., Tchelepi, H. A., Marcus, P., Prabhat, and Anandkumar, A. Meshfreeflownet: A physics-constrained deep continuous space-time super-resolution framework. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, SC '20. IEEE Press, 2020. ISBN 9781728199986.

Jønler, J. F., Brunø Lottrup, F., Berg, B., Zhang, D., and Chen, K. Probabilistic forecasts of global horizontal irradiance for solar systems. *IEEE Sensors Letters*, 7(1): 1–4, 2023. doi: 10.1109/LSENS.2022.3228783.

Kitaev, N., Kaiser, L., and Levskaya, A. Reformer: The efficient transformer. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=rkgNKkHtvB.

Knap, W. Horizon at station Cabauw, 2007. URL https://doi.org/10.1594/PANGAEA.669511. Backup Publisher: Koninklijk Nederlands Meteorologisch Instituut, De Bilt Type: data set.

Koenker, R. and Hallock, K. F. Quantile regression. *Journal of economic perspectives*, 15(4):143–156, 2001.

Kovachki, N., Li, Z., Liu, B., Azizzadenesheli, K., Bhattacharya, K., Stuart, A., and Anandkumar, A. Neural operator: Learning maps between function spaces, 2021.

Langley, P. Crafting papers on machine learning. In Langley, P. (ed.), *Proceedings of the 17th International Conference on Machine Learning (ICML 2000)*, pp. 1207–1216, Stanford, CA, 2000. Morgan Kaufmann.

Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Stuart, A., Bhattacharya, K., and Anandkumar, A. Multipole graph neural operator for parametric partial differential equations. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H. (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 6755–6766. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/file/4b21cf96d4cf612f239a6c322b10c8fe-Paper.pdf.

Li, Z.-Y., Kovachki, N. B., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., and Anandkumar, A. Fourier neural operator for parametric partial differential equations. *ArXiv*, abs/2010.08895, 2021.

Li, Z.-Y., Huang, D., Liu, B., and Anandkumar, A. Fourier neural operator with learned deformations for pdes on general geometries. *ARXIV.ORG*, 2022. doi: 10.48550/arXiv.2207.05209.

Liu, J., Zang, H., Cheng, L., Ding, T., Wei, Z., and Sun, G. A Transformer-based multimodal-learning framework using sky images for ultra-short-term solar irradiance forecasting. *Applied Energy*, 342:121160, 2023. ISSN 0306-2619. doi: https://doi.org/10.1016/j.apenergy.2023.121160. URL https://www.sciencedirect.com/science/article/pii/S030626192300524X.

Lu, L., Jin, P., and Karniadakis, G. Deeponet: Learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators. *ARXIV.ORG*, 2019.

Narvaez, G., Giraldo, L. F., Bressan, M., and Pantoja, A. Machine learning for site-adaptation and solar radiation forecasting. *Renewable Energy*, 167:333–342, 2021. ISSN 0960-1481. doi: https://doi.org/10.1016/j.renene.2020.11.089. URL https://www.sciencedirect.com/science/article/pii/S0960148120318395.

Nie, Y., Nguyen, N. H., Sinthong, P., and Kalagnanam, J. A time series is worth 64 words: Long-term forecasting with transformers, 2023.

Nielsen, A. H., Iosifidis, A., and Karstoft, H. Irradiancenet: Spatiotemporal deep learning model for satellite-derived solar irradiance short-term forecasting. *Solar Energy*, 228:659–669, 2021. ISSN 0038-092X. doi: https://doi.org/10.1016/j.solener.2021.09.073. URL https://www.sciencedirect.com/science/article/pii/S0038092X21008306.

Olano, X. Basic measurements of radiation at station Cener (2022-04), 2022. URL https://doi.org/10.1594/PANGAEA.943875. Backup Publisher: National Renewable Energy Centre Type: data set.

Pathak, J., Subramanian, S., Harrington, P., Raja, S., Chattopadhyay, A., Mardani, M., Kurth, T., Hall, D., Li, Z.-Y., Azizzadenesheli, K., Hassanzadeh, P., Kashinath, K., and Anandkumar, A. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. *ARXIV.ORG*, 2022.

Rothfuss, H. Data access at eumetsat, 2015 2015.

Rusch, T. K., Mishra, S., Erichson, N. B., and Mahoney, M. W. Long expressive memory for sequence modeling. *International Conference On Learning Representations*, 2022.

Sharda, S., Singh, M., and Sharma, K. Rsam: Robust self-attention based multi-horizon model for solar irradiance forecasting. *IEEE Transactions on Sustainable Energy*, 12(2):1394–1405, 2021. doi: 10.1109/TSTE.2020.3046098.

Si, Z., Yu, Y., Yang, M., and Li, P. Hybrid Solar Forecasting Method Using Satellite Visible Images and Modified Convolutional Neural Networks. *IEEE Transactions on Industry Applications*, 57(1):5–16, 2021. doi: 10.1109/TIA.2020.3028558.

Su, J., Lu, Y., Pan, S., Wen, B., and Liu, Y. Roformer: Enhanced transformer with rotary position embedding. *ARXIV.ORG*, 2021.

Sánchez Pérez, J., Meinhardt-Llopis, E., and Facciolo, G. TV-L1 Optical Flow Estimation. *Image Processing On Line*, 3:137–150, 2013. https://doi.org/10.5201/ipol.2013.26.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Vuilleumier, L. Meteorological synoptical observations from station Payerne (2015-01), 2018. URL https://doi.org/10.1594/PANGAEA.896350. Backup Publisher: Swiss Meteorological Agency, Payerne Type: data set.

Wang, H., Lei, Z., Zhang, X., Zhou, B., and Peng, J. A review of deep learning for renewable energy forecasting. *Energy Conversion and Management*, 198:111799, 2019. ISSN 0196-8904. doi: https://doi.org/10.1016/j.enconman.2019.111799. URL https://www.sciencedirect.com/science/article/pii/S0196890419307812.

Wen, Q., Zhou, T., Zhang, C., Chen, W., Ma, Z., Yan, J., and Sun, L. Transformers in time series: A survey. *arXiv preprint arXiv:2202.07125*, 2022.

Wu, H., Xu, J., Wang, J., and Long, M. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in Neural Information Processing Systems*, 34:22419–22430, 2021.

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., and Peters, I. M. A review of solar forecasting, its dependence on atmospheric sciences and implications for grid integration: Towards carbon neutrality. *Renewable and Sustainable Energy Reviews*, 161:112348, 2022. ISSN 1364-0321. doi: https://doi.org/10.1016/j.rser.2022.112348. URL https://www.sciencedirect.com/science/article/pii/S1364032122002593.

Zeng, A., Chen, M.-H., Zhang, L., and Xu, Q. Are transformers effective for time series forecasting? *ARXIV.ORG*, 2022. doi: 10.48550/arXiv.2205.13504.

Zhang, T., Zhang, Y., Cao, W., Bian, J., Yi, X., Zheng, S., and Li, J. Less is more: Fast multivariate time series forecasting with light sampling-oriented mlp structures, 2022.

Zhang, X., Zhen, Z., Sun, Y., Wang, F., Zhang, Y., Ren, H., Ma, H., and Zhang, W. Prediction interval estimation and deterministic forecasting model using ground-based sky image. *IEEE Transactions on Industry Applications*, 59 (2):2210–2224, 2023. doi: 10.1109/TIA.2022.3218758.

Zhang, Y. and Yan, J. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=vSVLM2j9eie.

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., and Zhang, W. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *The Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Virtual Conference*, volume 35, pp. 11106–11115. AAAI Press, 2021.

Zhou, T., Ma, Z., Wen, Q., Wang, X., Sun, L., and Jin, R. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting, 2022a.

Zhou, T., Ma, Z., xue wang, Wen, Q., Sun, L., Yao, T., Yin, W., and Jin, R. FiLM: Frequency improved legendre memory model for long-term time series forecasting. In Oh, A. H., Agarwal, A., Belgrave, D., and Cho, K. (eds.), *Advances in Neural Information Processing Systems*, 2022b. URL https://openreview.net/forum?id=zTQdHSQUQWc.

A. Appendix.

A.1. Related works

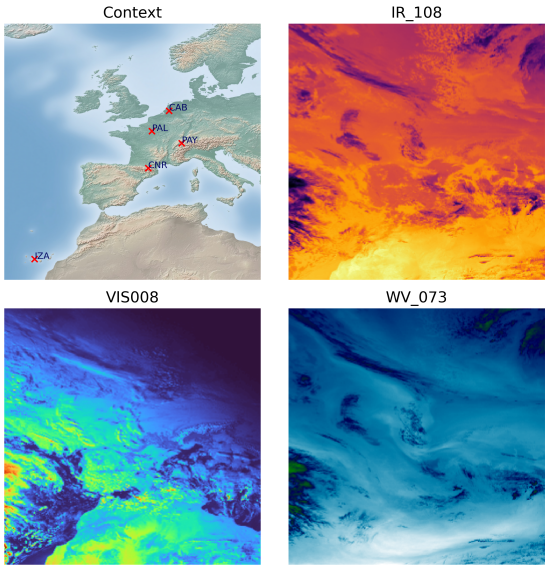
Machine Learning for time-series forecasting Deep learning approaches have gained popularity for time-series forecasting in recent years due to their ability to model complex nonlinear relationships and capture temporal dependencies. These approaches have demonstrated superior performance compared to traditional statistical methods, motivating further research in this area. In a recent survey (Wen et al., 2022), it was found that transformers, renowned for their success in natural language processing and computer vision, were also effective for time-series analysis. The authors discussed the strengths and limitations of transformers and compared the structure and performance of recent transformer-based architectures on a benchmark weather dataset (Zhou et al., 2021). The particular case of solar irradiance forecasting represents an interesting application for time-series models (Wang et al., 2019; Narvaez et al., 2021; Alzahrani et al., 2017). One recent study developed a multi-step attention-based model for solar irradiance forecasting that generates deterministic predictions and quantile predictions as well (Sharda et al., 2021). In a similar perspective, Jønler et al. (2023) developed a probabilistic solar irradiance transformer that incorporates gated recurrent units and temporal convolution networks, demonstrating strong performance for short-term horizons.

Context mixing / Multimodal learning for time-series forecasting Previous studies highlight the potential of time-series methods for solar irradiance forecasting, emphasizing the significance of short-term horizons in solar energy management. However, day-ahead forecasting remains challenging due to the influence of cloud cover on surface irradiance (Bone et al., 2018; Si et al., 2021), a problem which we aim to address in this paper. Thus, it is crucial to account for cloud effects in solar irradiance forecasting regardless of the chosen method. For instance, Zhang et al. (2023) investigated the impact of cloud movement on irradiance prediction and proposed an approach to automatically learn the relationship between sky image appearance and solar irradiance. A concurrent work (Liu et al., 2023) proposed a multimodal-learning framework for ultra-short-term (10min-ahead) solar irradiance forecasting. They used Informer (Zhou et al., 2021) to encode historical time-series data, then utilized Vision Transformer (Dosovitskiy et al., 2020) to handle sky images. Finally, they employed cross-attention to couple the two modalities. The studies discussed above highlight the potential of incorporating external data sources, such as sky images and satellite images, in combination with time-series approaches to improve the accuracy of solar forecasting.

Operator Learning Utilizing available satellite imagery to forecast GHI over a region presents limitations as it may not capture clouds that exist at a resolution beyond that of the satellite data. To ensure accurate forecasting of quantities of interest, the ability to query the model at any possible resolution and any point within the domain becomes crucial. Recent advancements have witnessed the rise of algorithms focusing on learning operators capable of mapping across functional spaces, with a focus on solving partial differential equations (PDE) (Lu et al., 2019; Li et al., 2021; Kovachki et al., 2021; Li et al., 2020). These operators can effectively map initial conditions to PDE solutions, making it possible to query the learned solution theoretically anywhere within its domain. Fourier Layers, developed by Li et al. (2021), enable zero-shot prediction on both uniform and non-uniform grids with learnable deformations (Li et al., 2022). Pathak et al. (2022) replace attention in ViT (Dosovitskiy et al., 2020) with Fourier layer mixing for competitive weather forecasting results with faster inference. MeshFreeflowNet (Jiang et al., 2020) learns high-resolution frames from corresponding lower resolution ones by querying the model at any point of the domain for irregular grids. Similarly, Boussif et al. (2022) employ message passing with a low-resolution graph for zero-shot super-resolution PDE learning. Additionally, message passing neural PDE solvers (Brandstetter et al., 2022) exhibit spatio-temporal multi-scale capabilities benefiting from long-expressive memory (Equer et al., 2023; Rusch et al., 2022). We note that while these approaches were developed for PDEs in mind, they can still be used for weather-related applications.

A.2. Additional visualisations

Figure 2 shows the location and characteristics of the 6 stations utilized within the study. Predictions visualisations can be seen in Figure 3, along with the comparison of the fourier spectra of our prediction, the ground truth and a strong baseline, CrossFormer.



Station	Latitude	Longitude	Elevation
Cabauw (CAB) (Knap, 2007)	$51°58'N$	$4°55'E$	$0m$
Cener (CNR) (Olano, 2022)	$42°48'N$	$1°36'W$	$471m$
Izana (IZA) (Cuevas-Agulló, 2014)	$28°18'N$	$16°29'W$	$2373m$
Palaiseau (PAL) (Haeffelin, 2014)	$48°42'N$	$2°12'E$	$156m$
Payerne (PAY) (Vuilleumier, 2018)	$46°48'N$	$6°56'E$	$491m$
Tamanrasset (TAM) (Baika, 2023)	$22°47'N$	$5°31'E$	$1385m$

Left: Location of the six meteorological stations considered in the study along with three of the eleven spectral channels considered. The IR_108, VIS_008 and WV_073 channels are infrared (10.8μ), visible ($0.8\mu m$) and water vapor ($7.3\mu m$) channels respectively. *Right*: Table summarizing the geographic coordinates and elevation of each station used in the paper.

Figure 2. **Stations and satellite data.**

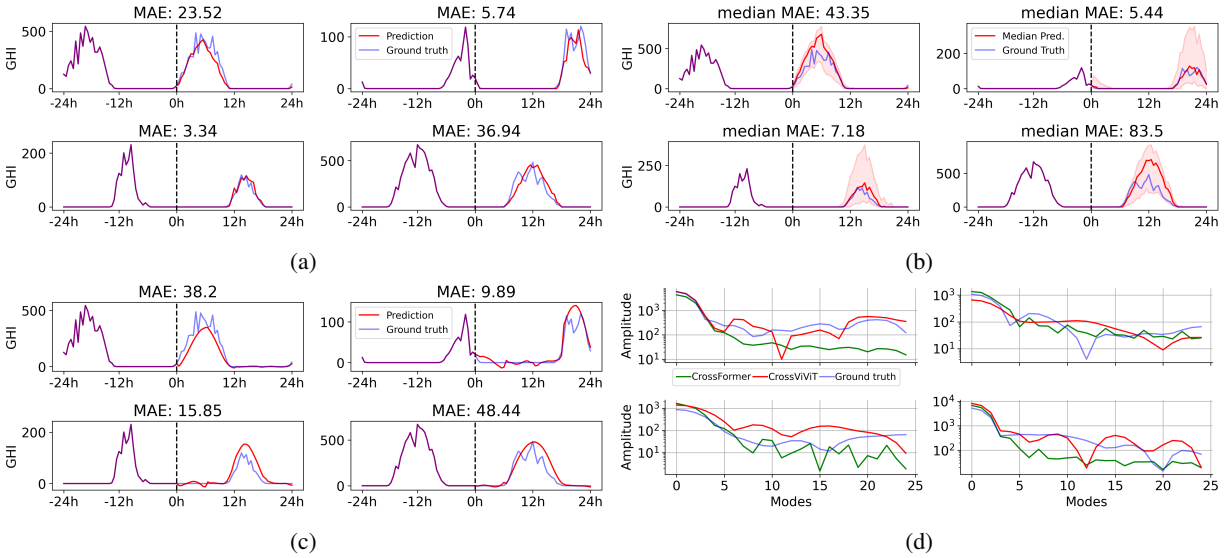


Figure 3. **Prediction visualisations from CrossViViT for four examples in CAB station, on the 2020-2022 test period.** (a) CrossViViT predictions. (b) Multi-Quantile CrossViViT median ($q_{0.50}$ quantile) predictions with $[q_{0.02}, q_{0.98}]$ prediction interval. (c) Predictions from a strong baseline, CrossFormer, (d) Fourier spectrum of the target, our prediction, and CrossFormer prediction. Figure (a) illustrates that CrossViViT closely aligns with the ground truth by effectively capturing cloud variations, whereas CrossFormer assumes a clear-sky pattern. This is confirmed by the Fourier spectra depicted in (d), where CrossFormer's spectrum exhibits a rapid decay in contrast to CrossViViT.