Deep learning methods for intra-day cloudiness prediction using geostationary satellite images in a solar forecasting framework

Franco Marchesoni-Acland[a,d,*], Andrés Herrera[b], Franco Mozo[b], Ignacio Camiruaga[c], Alberto Castro[b,c], Rodrigo Alonso-Suárez[a,*]

[a] *Laboratorio de Energía Solar, Universidad de la República, Av. L. Batlle Berres km 508, Salto, Uruguay*
[b] *Instituto de Ingeniería Eléctrica, Facultad de Ingeniería, UDELAR, J. H. y Reissig 565, Montevideo, Uruguay*
[c] *Instituto de Computación, Facultad de Ingeniería, UDELAR, J. H. y Reissig 565, Montevideo, Uruguay*
[d] *Université Paris-Saclay, ENS Paris-Saclay, CNRS, Centre Borelli, 91190, Gif-sur-Yvette, France*

Abstract

Accurate solar resource forecasting remains a challenge. Electricity grid applications require both days-ahead and intra-day prediction. Satellite-based methods are known to be the best option for hourly intra-day solar forecasts up to some hours ahead. An adapted Deep Learning (DL) method has been recently reported to outperform the traditional Cloud Motion Vectors (CMV) strategy. This article analyzes the utilization of a well-documented computer vision DL architecture, the U-Net in various forms, for the satellite Earth albedo forecast problem (cloudiness), a straightforward proxy for solar irradiance forecast. It is shown that the U-Net performs better than advanced and optimized CMV techniques and previous art IrradianceNet, setting it at the state-of-the-art. The tests are done over the Pampa Húmeda region of southeast South America, an area in which challenging cloud conditions are frequent. The data for this study are GOES-16 visible channel images. These images present a finer spatial ($\simeq$ 1 km/pixel) and temporal (10 minutes) resolution than previously explored data sources for solar forecasting. Moreover, the image size used here is $\times 4$ bigger (1024×1024 pixels) and the predictions reach further into the future (5 hours) than in previous works. The analysis includes several ablation studies, involving different architectures, optimization objectives, inputs, and network sizes. The U-Net is optimized for direct and differential image prediction, being the latter a better-performing option. More notably, the U-Net models are shown to be able to predict cloud extinction, something that has been a barrier for CMV methods.

Keywords: Solar forecast, U-Net, deep learning, satellite images, GOES-16 satellite.

1. Introduction

Solar photovoltaic (PV) is the main technology to convert the renewable power provided by the Sun into electricity (REN21, 2021). Due to the systematic cost reduction of this technology, its share in the world's

*Corresp. authors: R. Alonso-Suarez and F. Marchesoni-Acland, {r.alonso.suarez, marchesoniacland}@gmail.com

energy mix is continuously growing (IRENA, 2021). Regardless of its availability, high technology maturity level, and low prices, solar energy is still a small portion of the world's energy mix, partly due to the low predictability of the solar resource (Voyant et al., 2017). The optimal operation of an electricity grid that includes weather-dependent energy sources, such as solar PV and wind power, requires their forecast in several time horizons. Solar forecasting in particular is mainly dependent on cloudiness forecasting, as the other atmosphere variable components, for instance, water vapor and aerosol content, have a significantly lower variability. The other source of solar variability is related to its well-known intra-day and seasonal geometrical trend, governed by the Sun's apparent movement, which can be effectively removed or modeled (Lauret et al., 2022) by using a clear sky model (Lefèvre et al., 2013; Rigollier et al., 2004), leaving cloudiness as the main driver of uncertainty in solar energy prediction.

Clouds undergo various transformations, such as formation, deformation, displacement, and dissipation, which make predicting cloud cover a challenging task. Numerical Weather Prediction (NWP) models have been historically used to provide solar irradiance forecasts by using their cloudiness prediction at different levels in the atmosphere. These models numerically solve the equations that govern the atmosphere dynamic and still are the only option to provide quality day(s)-ahead solar forecasts, typically augmented by post-processing techniques (Yang et al., 2022). However, for solar irradiance intra-day forecasts, it has been shown that pure satellite-based methods can achieve a better performance than regular NWP runs (i.e. not hourly updated) for horizons up to 4-5 hours ahead (Kühnert et al., 2013; Perez et al., 2010). These satellite-based methods have been mainly based on the two-dimensional cloud motion field estimation (Lorenz et al., 2004; Peng et al., 2013; Cros et al., 2014; Urbich et al., 2019) known as Cloud Motion Vectors (CMV), from which the clouds' motion is extrapolated into the future. A detailed comparison between the main variants of the CMV approach is presented in Aicardi et al. (2022), including the baseline block-matching technique of Lorenz et al. and different types of optical flow methods (Horn and Schunck, 1981; Lucas and Kanade, 1981; Farnebäck, 2003; Zach et al., 2007) that are common in the computer vision field. Further, Aicardi et al. bridged the gap between cloudiness and solar irradiance forecast performance evaluation, at least for the inspected region (which is the same as the present work), showing that the best satellite-based forecasting models at the cloudiness level are also typically the best at the solar irradiance level. The cloudiness level in this context refers to the Earth's albedo visible channel images (planetary reflectance), where clouds are easily distinguishable. The albedo information is the common input for satellite-based solar irradiance estimation methods (Perez et al., 2002; Rigollier et al., 2004; Qu et al., 2017; Laguarda et al., 2020). These methods use updated formulations of the well-known satellite cloud index (Cano et al., 1986), calculated, mainly, from visible channel images.

The CMV strategy presents some limitations at the current state of the art, such as the inherent two-dimensional modeling of a three-dimensional problem, the motion extrapolation based on a static motion field, and the inability to accurately forecast clouds' formation and extinction, among others. For instance, pure CMV strategies can only reallocate pixels of a satellite image but they are unable to create new values. While this area advances, for instance through better ways to physically model the cloudiness evolution from satellite images, data-intensive approaches have arisen in the field (Berthomier et al., 2020; Su et al., 2020; Nielsen et al., 2021b). This is not surprising: deep learning (DL) methods have been recently introduced in many meteorology-related areas (Ren et al., 2021; Ravuri et al., 2021; Espeholt et al., 2021; Schneider et al., 2022) and DL methods excel when abundant pairs of inputs and targets are available, especially in the case of automatically annotated data. The forecasting problem is inherently self-supervised, i.e. the targets are naturally contained in the historical records and no manual annotations are needed. Visible channel satellite images also provide large amounts of data, typically of dozens of TBs, from which the daylight cloudiness behavior can be observed, modeled, and learned at a high rate (current geostationary meteorological satellite scans the same Earth disk at a regular rate of 10-15 minutes). Due to all these, intra-day satellite cloudiness forecasting is especially suited for deep learning methods.

1.1. Related Work

With the advent of deep learning, some research groups have started its utilization for black-box atmospheric modeling in problems related to cloudiness and solar irradiance forecasting. This subsection describes the state-of-the-art work.

The U-Net architecture (Ronneberger et al., 2015) was used by Berthomier et al. (2020) for intra-day cloud cover binary forecast over France up to 90 minutes ahead with a 15-minute time step. Berthomier et al. showed that this type of U-Net cloud forecast is better than the forecast provided by the AROME NWP model (Seity et al., 2011; Brousseau et al., 2016). Furthermore, it also compared favorably to different types of neural networks, such as Recurrent Neural Networks (RNNs) and other Convolutional Neural Networks (CNNs). The algorithms' training and validation were conducted for the binary cloudiness forecasting problem (cloud vs. no cloud) using the Meteosat Second Generation (MSG) cloud classification product as the input/output data, which is not directly translatable into a solar irradiance forecast. The satellite images containing France's territory were cropped into 256×256 px patches that were processed separately due to computational restrictions. Separated training and testing periods were considered, using one year and a half for the former and six months for the latter. Even though the work did not aim to be used for solar irradiance forecasting, as the use of binary cloud cover is limited for this task, it represented the first deep learning attempt to cloud forecasting. In recent work, Nielsen et al. (2021b) proposed an adapted deep learning architecture for solar irradiance forecast based on the MSG satellite information. In this case, the European SARAH-2.1 data set (Uwe et al., 2019) of satellite-derived global solar horizontal irradiance was used. The architecture is similar to the one proposed by Tan et al. (2018) and was trained and tested for a 4-hours ahead forecast with hourly time step. The deep learning algorithm outperformed the optical flow TVL1 method (Zach et al., 2007; Sánchez et al., 2013) with the same satellite information, which has been reported to be one of the best optical flow strategies for cloud and solar irradiance forecasting (Urbich et al., 2019; Aicardi et al., 2022). The satellite images had European coverage with a 512×512 px resolution and a 30-minute refresh rate. As in the work of Berthomier et al., the images were cropped, this time into 128×128 px patches, and processed separately to avoid a high computational cost. The forecasts were validated against the solar irradiance derived from satellite images and against four solar ground measurement sites of the Baseline Solar Radiation Network (BSRN, https://bsrn.awi.de/) in Europe. An important period was considered, using 5 years of images for training, 1 year for validation, and 2 years for testing. The final results correspond to the evaluation across the latter split.

These works suggest that deep learning is currently the state-of-the-art best option for satellite intra-day cloud cover and solar forecast. Furthermore, the two most performing architectures are the U-Net and the IrradianceNet, which compare favorably to other architectures (Tan et al., 2018; Su et al., 2020). It is not clear from the literature what is the best model for satellite Earth albedo and solar forecast, but it is clear that they have outperformed NWP and CMV methods in each analyzed task and for the European region. At the same time, well-performing optical flow CMV methods were inspected in Aicardi et al. (2022), showing that a simple horizon-dependent spatial blurring of the CMV prediction can enhance performance. As this performance booster was not considered in the previous deep learning vs CMV comparisons, it is contemplated in the present work as an additional contribution. The present work benchmarks these three techniques (U-Net, IrradianceNet, blurred CMV) for a region in South America with several ablation studies and a common forecasting objective, showing that the U-Net is an upgrade of the current intra-day solar forecasting performance. The U-Net is a performing and still widely used computer vision architecture (Baranchuk et al., 2022; Rombach et al., 2022; Croitoru et al., 2023). It uses GOES-16 albedo images as a regression target, as opposed to a classification target obtained by quantization (e.g. binarization). Finally, it provides a novel scheme for the U-Net utilization in this framework (U-Net Diff), which resulted in the best-performing strategy from the ones considered in this work, although closely followed by the traditional U-Net.

In summary, IrradianceNet is the current performance lead on satellite-based solar forecasting using deep learning, in fact being the only work providing in-depth evaluations. The present work addresses current limitations, namely (i) moderate training and inference speed, (ii) limited image size, (iii) improvable benchmarking, and (iv) laborious end-user implementation. These limitations are overcome by the U-Net which is faster, extremely well-documented, can handle larger images (Figure 6), and is compared in this work against blurred CMV, a tougher benchmark than plain CMV strategies (Figure 3). On top of overcoming these limitations, the U-Net outperforms IrradianceNet in this region, setting it at the state-of-the-art and providing a framework for future implementations of deep learning methods for this task.

1.2. *Article's outline*

This work builds upon previous works and compares past and concurrent research efforts in the field, with a common ground in solar forecasting. It considers deep learning methods previously proposed in different contexts and one of the best-performing CVM methods, including the important spatial blurring step. The analysis is done with one complete year of satellite images over the Pampa Húmeda region of South America, an area with challenging clouds' behavior and intermediate solar irradiance variability. The results confirm that, for this region, deep learning methods are currently preferable over CMV methods. Additionally, it is shown that the simple and lightweight U-Net architecture yields a strong performance that surpasses all other methods. Regardless of its good performance, the main advantage remains practical: the U-Net is a battle-tested architecture featured in many tutorials and articles in diverse research areas, hence it is simpler to implement, train and deploy than custom architectures. Notwithstanding, there are many details about the data processing, the training procedure, and the particular solar forecasting topic that still need to be considered. Knowing which well-documented architecture performs well and how to adapt it to the specific forecasting problem, makes deep learning solutions much more accessible to other research teams and regions. The contributions of this work are as follows:

1. Illustrating the successful use of deep learning methods for 5-hours ahead albedo forecasting using GOES-16 satellite images, being this the first comprehensive work aimed at solar satellite-based forecasting using this new satellite and over a region other than Europe. The only exception, to the best of our knowledge, is the preliminary work of Alonso-Suárez et al. (2021), which presents the evaluation of the operational CMV solar forecasting system developed and operated by our R&D group using GOES-16 satellite images.
2. Showing that the U-Net deep learning model achieves the best performance for the albedo forecasting problem when compared to Persistence, CMV, blurred CMV, and IrradianceNet. This benchmarks all these models against the same data set over the same region, periods, and forecast horizons. The findings demonstrate that the simpler U-Net architecture achieves the highest reported performance in the field. This is also the first application of this algorithm in the specific framework of cloud albedo prediction for solar irradiance forecasting.
3. Providing numerous ablation experiments for the specific problem and the U-Net strategy, regarding the inference modality, the objective function, data augmentation, inner architecture, network size, and alternative inputs. One of the tested U-Net configurations resulted in the best option.
4. Training and applying the previously proposed IrradianceNet algorithm over South America, thus corroborating that it improves over the CMV and showing that it can handle different geographical regions and image types with solid performance.
5. Displaying qualitative examples illustrating deep learning models being capable of predicting complex

phenomena such as cloud extinction, which regular CMV strategies are unable to perform.

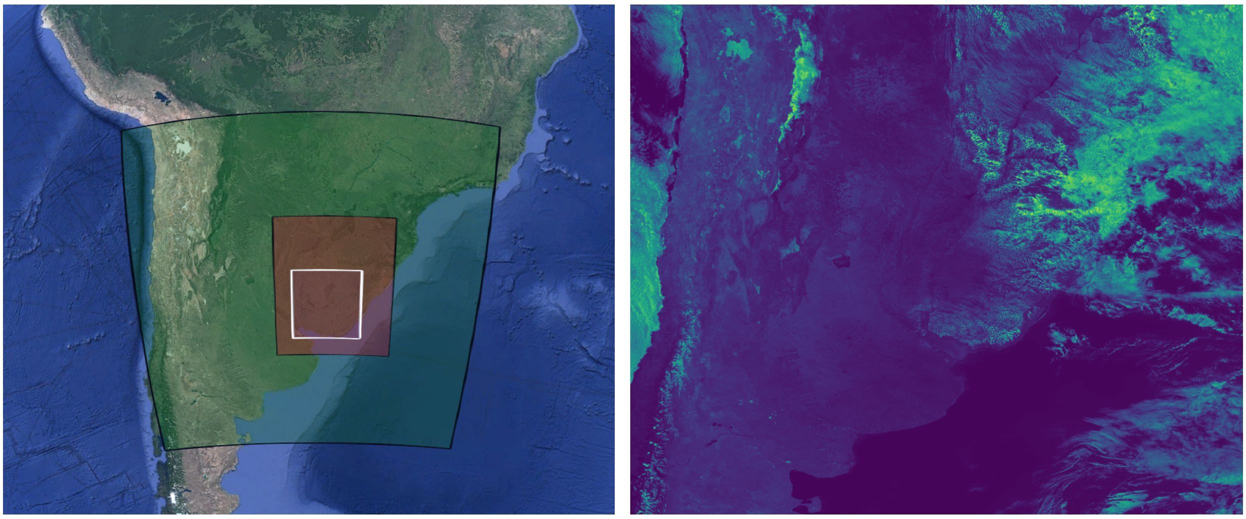
The rest of this article is organized as follows. Section 2 presents the satellite data set, including its filtering and pre-processing. The metrics required for the work are introduced in Section 3, and the considered models along with their experiments are described in Sections 4 and 5, respectively. This includes training details and ablation studies. Section 6 presents the final results over the test set, which is left aside from the previous analysis. It also discusses some case studies in which the U-Net deep learning method was able to forecast cloud extinction. Finally, Section 7 summarizes the main conclusions of the work and provides some ideas on how to further advance in this field, based on our experience and the observed limitations of the techniques.

2. Data

2.1. Satellite images

The data set used in this work is managed and stored by the Solar Energy Laboratory (LES) of the Universidad de la República (Udelar), Uruguay. It is composed of visible channel images taken from the GOES-16 satellite's C02 channel, which has its central wavelength at 640 nm. This is a meteorological satellite administrated by the US's National Oceanic and Atmospheric Administration (NOAA). It is located in the geostationary position known as GOES-East (-75°W over the Earth's equator) and its images are freely available via different download mechanisms. This satellite provides a higher frame rate and resolution for South America than the previous generation, with the important added value of schedule regularity. The images are available every 10 minutes with a nominal resolution of 500 m at the satellite's nadir. The visible channel is useful for cloud detection as they typically reflect more sunlight than the ground, appearing brighter in the images. This assumption does not hold for high albedo terrains, i.e. areas containing snow or salt flats, which are not present in the portion of the images used in this work.

A large orthorectified crop of the satellite full disk images is processed by the LES lab for solar resource assessment and forecasting, covering the southeast of South America as shown in the green rectangle of Figure 1a. An example of a satellite image from this database is shown in Figure 1b, being of 2501×3001 px and ranging 25 degrees in latitude and 30 degrees in longitude. However, due to computational restrictions of the deep learning training, a reduced crop of 1024×1024 px is used in this work, shown in Figure 1a as a red rectangle with a black frame. Note that this region contains at least $4\times$ more pixels than any previous work involving DL in this context. All methods will be evaluated inside this region, thus having the same space input information. This implies a built-in difference concerning the borders' effect. In some borders, and especially for the larger time horizons, the CMV is unable to provide predictions, as this would require pixels from outside the image. Such regions are not considered for the CMV evaluation. The option of providing a bigger frame for CMV forecast would result in this method having more spatial information than the others. On the other hand, deep learning methods, by construction, provide a forecast for the whole image including these borders. This requires deep learning methods to figure out the most likely forecast based on the previous history of those pixels in similar situations. In this way, each method is evaluated in the pixels that it can predict and the space information given to all the methods is the same. Figure 1a also shows a smaller white-framed crop inside the red rectangle that will be used only for some comprehensive experiments on the U-Net architecture and related implementation details (Section 5). The size of this crop was 512×512 pixels, representing a region of 5×5 degrees in latitude and longitude. These crops are referred to as *smaller-crop data set*, which should not be assumed unless specified in the text. The images have a regular uniformly-spaced grid in the latitude-longitude domain with an average pixel side of $\simeq 1$ km in all these regions. The work was conducted with 10-minute images taken from the whole year 2020.



(a) Location of the satellite crops. (b) Example image of the LES satellite database.

Figure 1: Satellite area considered in the work. It represents the southeast of South America, known as Pampa Húmeda.

Complementary to the satellite image data, an elevation map of the region is used for some experiments, as it is recommended by Nielsen et al. for the IrradianceNet method. Note that the southeast side of the South American continent is mostly flat. This region is referred to as Pampa Húmeda and it is considered warm, temperate, and humid, with hot summers, being Cfa in the updated Köppen-Geiger climate classification (Peel et al., 2007). The area is characterized by challenging mesoscale convective systems that tend to peak during daytime (Salio et al., 2007; Rasmussen et al., 2014) and has no rainy or dry season, being the cloudiness distributed throughout the year. In terms of ground-level solar irradiance, the short-term intermittency is intermediate (Alonso-Suárez et al., 2020), meaning that clear-sky, partly cloudy, and overcast conditions are all present and frequent.

2.2. *Filtering and preprocessing*

The original data set comprises 52314 images (daytime and nighttime). These cover almost all the 10-minute intervals over the year (only 390 images are missing). The original data set was filtered in two steps: (i) images with any pixel with solar elevation lower than 10° are removed, leaving only images with all their pixels in the daytime, and (ii) manual inspection over consecutive images with large mean differences, as a way to detect images with corrupted pixels (bad acquisition or missing sectors). The first filter removes images with any pixel at night but also ensures that the remaining images do not have pixels at sunrise and sunset, which are moments of very low solar elevation over the horizon that appear distorted in the images after the geometrical normalization with the cosine of the solar zenith angle (θ_z). The second filter separates images for possible reconstruction. The separated images with less than 5% corrupted pixels were corrected by inpainting. Specifically, the Navier-Stokes method of Bertalmio et al. (2001) was run. This method was found to be the fastest of those in the OpenCV library. The percentage of images affected by the inpainting was 1.2%, but only 0.12% had more than 1% of the pixels modified. Since less than 0.02% of the pixels were inpainted a negligible impact is expected from the choice of the inpainting algorithm. After filtering and reconstruction, the final data set is composed of 20842 valid images.

The images require a last preprocessing before their utilization. The original images contain reflectance factor information (instead of Earth albedo), a dimensionless quantity that scales the measured radiance recorder by the satellite's radiometer by the maximum value that the sensor can detect. This includes the correction to account for the variable Sun-Earth distance across the Earth's orbit. This magnitude has a dependence on the incidence angle of the sunlight to Earth, thus having geometrical spatial information. Normalization is done by dividing each pixel by its corresponding $\cos(\theta_z)$ at each moment. The normalization removes the deterministic geometrical variability coming from the apparent movement of the Sun relative to the Earth and allows the comparison of images taken at different moments. All pixels exceeding the initial range are clamped and the range of values is linearly mapped to $[0, 1]$. By doing this normalization, albedo images are calculated from the original images and the data set is ready for utilization.

2.3. *Data set split*

The data set was divided into two subsets following (Nielsen et al., 2021b; Sønderby et al., 2020; Su et al., 2020), one for training (75%) and the other for testing (25%). The random split was done day-wise and distributed across the whole year. This resulted in 274 randomly selected days going into the training split. The testing data set is then composed of 92 days that are unseen during the architecture definition and parameters' training. The validation data set, used for architecture decisions, included 40 random days (15%) from the training set. The K-fold cross-validation methodology was not implemented due to its high computational cost, being unfeasible for this type of satellite data set given the training times.

3. Metrics

The metrics used in this article are the usual ones in solar forecasting (Yang et al., 2020) and machine learning literature. These metrics can be used either for performance evaluation or as optimization targets. A short discussion of this topic will be presented in Subsection 5.1.3 for the U-Net architecture. For performance evaluation, the usual metrics in the solar forecasting field are favored, as they are directly related to the satellite albedo forecasting problem and were used in Aicardi et al. to bridge the gap between the forecast evaluation of both quantities. The basic common metrics are:

$$\text{Mean Bias Error} \qquad \text{MBE} = \frac{1}{N}\sum_{i=1}^{H}\sum_{j=1}^{W}(\hat{y}_{ij} - y_{ij}),$$

$$\text{Mean Absolute Error} \qquad \text{MAE} = \frac{1}{N}\sum_{i=1}^{H}\sum_{j=1}^{W}|\hat{y}_{ij} - y_{ij}|,$$

$$\text{Mean Squared Error} \qquad \text{MSE} = \frac{1}{N}\sum_{i=1}^{H}\sum_{j=1}^{W}(\hat{y}_{ij} - y_{ij})^2,$$

$$\text{Root Mean Square Error} \qquad \text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{H}\sum_{j=1}^{W}(\hat{y}_{ij} - y_{ij})^2},$$

where $\hat{y}$ and y are the predicted and ground truth images, respectively. Each image is of dimensions (H, W) and N denotes the number of valid pixels in the summation (which may be lower than $H \times W$). These metrics are calculated image-wise for each forecast horizon and can be expressed as a percentage of the image's average value. These normalized metrics will be denoted with a % symbol. The performance evaluation is done by computing the average for each one of these metrics across all images in the test set, both for the normalized and not normalized cases. Note that RMSE is more sensitive to outliers than MAE (as it weighs more large variances) and that minimizing MAE or RMSE does not necessarily imply taking the MBE to zero.

An extended performance metric for solar irradiance (and meteorological) deterministic forecast is the forecasting skill, defined as:

$$\text{Forecasting Skill (\%)} \qquad \text{FS} = 100 \times \left(1 - \frac{\text{RMSE}_{fcs}}{\text{RMSE}_{per}}\right),$$

where fcs refers to the forecast being evaluated and per refers to the persistence. This metric measures the gain, in terms of RMSE, of the forecast being evaluated relative to a baseline performance reference given by the persistence naive procedure. A positive (negative) metric means that the forecast is better (worse) than the baseline. The persistence procedure used in this work for the albedo satellite images is described in Subsection 4.1. All of the metrics are evaluated for each forecast horizon independently.

Finally, another metric that is used in this work is the Structural Similarity Index Measure (SSIM), introduced by Wang et al. (2004). It measures the structural degradation between a distorted image and a reference, thus it can quantify textures and perceptual similarity of an image relative to the reference. This is a widely-adopted metric in the image processing community (Fan et al., 2019). However, it combines texture and alignment in a single score, which was ultimately uninformative in most of the forecasting tests. Its further utilization for this forecasting problem requires specific studies. This metric is introduced here because it is used for an interesting study regarding the training metrics, which is presented in Subsection 5.1.3.

4. Models

This section briefly describes the baseline models and the deep learning architectures considered in the work. For more detailed information on the methods, the reader is referred to the original articles. The experiments on the U-Net and IrradianceNet architectures are discussed in Section 5. The CMV description, implementation, and optimization are fully presented in this section.

4.1. Persistence

Regular persistence procedures use the last observation as the prediction. In this work, persistence is implemented by simply maintaining the time t albedo image constant across all forecast horizons, providing a simple baseline performance reference. Note that the aim of the work is not to assess to which extent the models outperform a given reference, but to compare different models that are known to have a better performance than naive forecasting procedures. In this sense, persistence provides a general reference and a common ground to calculate the forecasting skill metric.

4.2. U-Net architecture

The U-Net (Long et al., 2015; Ronneberger et al., 2015) is a well-known deep learning architecture for semantic segmentation and pixel-wise prediction. It involves two stages, an encoder and a decoder in a U-shaped scheme, both with several convolutional layers and skip connections between them. Through the encoder, the network learns features and patterns in the image sequence. In the decoder, the prediction is built by extracting and upsampling the features learned at different levels of the encoder's convolutional layers. The skip connections help the decoder reconstruct the output by providing additional information, and make the learning more stable by reducing the risk of vanishing gradients. The number of trainable parameters in this architecture is normally several million, whose adjustment can be done in modern computing facilities equipped with Graphics Processing Units (GPUs). It has to be noted that the computationally expensive part of the process is the training stage, both in terms of processing time and memory allocation, but once the architecture is trained, its utilization can be performed in a few seconds without the need for high computational resources. As this is an image-to-image deep learning technique, readers are referred to the specific bibliography of the field for further details, which includes convolutional neural networks and autoencoders. The architecture has been used on a plethora of problems (Falk et al., 2019; Wei et al., 2019; Du et al., 2020; Smith et al., 2020; Kang et al., 2022) and has yielded a solid performance across a wide variety of domains, from medical imaging to remote sensing.

The basic architecture has led to different variants of the U-Net. Three of them are here revisited, namely, the Attention U-Net (Oktay et al., 2018), the Nested U-Net (Zhou et al., 2020), and the Recurrent Residual U-Net (R2U-Net) (Alom et al., 2018). In the Attention U-Net, the main modification is to add the so-called soft-attention layers, which allow the detection of the most relevant regions in the input images to assign more importance to those while processing. The Nested U-Net modifies the skip connections by adding more convolutional layers between the encoder and decoder, referred to as dense skip connections, which are expected to reduce the semantic gap between the feature maps and the predictions. Finally, the R2U-Net uses recurrent convolutional layers to feedback residuals in the training process instead of the original convolutional layers. Further details on these variants can be found in their corresponding articles. The use of these architectures is tested in Subsection 5.1.1.

4.3. IrradianceNet

IrradianceNet is a custom convolutional long short-term memory (ConvLSTM) neural network-based prediction model with a two-stage separated encoder-decoder scheme designed by Nielsen et al. (2021b) for solar irradiance forecasting using geostationary satellite information. This method is reportedly the best-performing of this family of forecasting techniques. It uses three ConvLSTM layers in both the encoder and the decoder networks and employs a patch-based approach similar to Sønderby et al. (2020) due to computational restrictions. The proposal uses four previous images as input to generate a prediction. Apart from the satellite images, the authors introduced other sources of information as input: temporal information as the hour, day, and month, and spatial information as the longitude, latitude, and elevation. A different model is trained for each forecast horizon from 1 to 4 hours ahead. For the implementation of IrradianceNet in this work, two versions of the model are considered, one with only satellite images and the other with satellite images and geographic information, i.e. the longitude, latitude, and elevation. Full details on this architecture are given in the article by Nielsen et al..

4.4. CMV

Cloud Motion Vectors (CMV) methods estimate the cloudiness velocity field from the last two consecutive satellite images and then use it to generate future images, i.e. pixels are projected to their future position by using the velocity field and the time t image. Several techniques have been applied to estimate the cloud motion field, being the optical flow methods the most recent and best performing, as discussed in Section 1.

In this work, due to ease of implementation and performance, the Farnebäck optical flow method is used with its OpenCV 3.x implementation (calcOpticalFlowFarneback function). This method requires some input parameters that were locally optimized over the training set. Some parameters refer to the mathematical formulation of the method, like the window size in which the polynomial expansion of this method is done. Other parameters refer to its computational implementation, like the down-scaling levels that are used to obtain the dense motion estimation from lower to higher resolution images, with a multi-level pyramid strategy. This method's parameters and their optimized values are presented in Table 1 with their library names. More information about this method can be found in Farnebäck (2003); Aicardi et al. (2022) and in its OpenCV documentation. It shall be noticed that the values of the parameters winsize and levels are similar to those found in Aicardi et al. for the same region, but with 2016-2017 GOES-13 albedo images, that have a different spatial resolution and time rate. This previous work only optimizes these two parameters, leaving the others as default.

Table 1: Optimized parameters for the Farnebäck optical flow method.

parameter	pyr_scale	levels	winsize	iterations	poly_n	poly_sigma
value	0.3987	4	22	3	5	0.8480

The projection algorithm being used is the common backward search, in which the predicted image is constructed by using the opposite vector flow at each given pixel and scaled by the time interval. Note that the scaling is required as the CMV is estimated with a 10-minutes difference between images and the forecast horizons are hourly. The scaled and inverted CMV is used to obtain the value to assign to each pixel in the predicted image from the previous image via a bi-linear interpolation. This procedure is iterated for all forecast horizons by taking the basis in the previous predicted image at each stage, starting with the time t real image. In Aicardi et al. it is shown that this iterative procedure is the best option to obtain the predicted images for these kinds of CMV algorithms.

Aicardi et al. also showed that running a spatial blurring on the predicted images improves the forecasting performance. Further, it showed that the blur window size should be horizon-dependent, as this provides better performance than a fixed spatial blur across all forecast horizons. The blurring implemented here is based on an isotropic Gaussian kernel. The size of this kernel was optimized over the training data to minimize the RMSE% between the predicted images and the corresponding ground truth for each forecast horizon. The analysis is shown in Figure 2. As can be seen, there is a flat optimum value of kernel size for each forecast horizon that increases and flattens with increasing lead time. The behavior of this plot is the same as that of Aicardi et al., which uses a simple average value in a square spatial window of variable side length. This blurred model is the best-performing CMV strategy and is called here Blurred CMV.

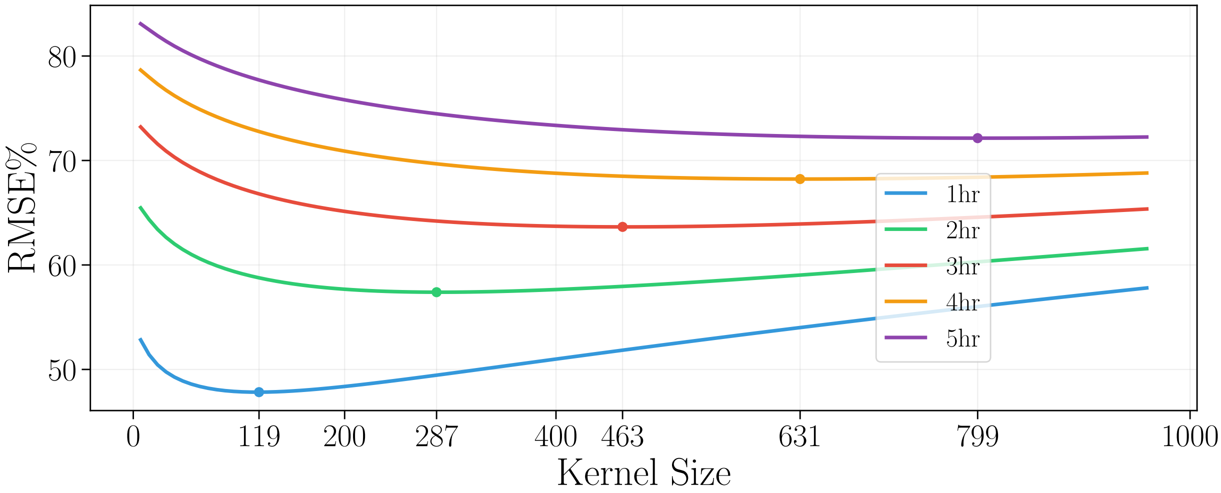


Figure 2: Kernel size optimization for the Blurred CMV strategy.

5. Ablation

This section describes several implementations of DL models, some of which yield promising results over the validation set. These preliminary best-performing architectures were selected for the final assessment with the test data set. The experiments presented in this section were obtained with the validation data set and a forecast horizon of 1 hour, favoring the metrics that are closer to the machine learning field. The red region of Figure 1a was used for most experiments, with the only exception of the variations of the U-Net architecture experiment (Subsection 5.1.1) that was performed on the smaller white bounding box of Figure 1a due to computational restrictions. All the experiments and runs required for this work were done in the ClusterUY center (Nesmachnow and Iturriaga, 2019), a national supercomputing infrastructure in Uruguay.

5.1. U-Net

Our proposed U-Net architecture uses 16 filters in the first layer (that will be simply referred to in the following as "filters") and MAE as the training loss. Several variations were tested before arriving at this configuration. Experiments were done regarding the network capacity, the training loss function, and the architecture itself. The batch size was set as the maximum possible in each case. These experiments are described in different subsections in the following. This section also presents results regarding data augmentation and the inclusion of extra input information. Although not all of the experiments yielded positive improvements, they are briefly described here as they may be of interest to other researchers in the field. The baseline input information for the networks are the last three available images, this is, the images at times $(t-k)$ with $k=\{0,1,2\}$.

5.1.1. U-Net variations

The original U-Net and its variants, Attention U-Net, Nested U-Net, and R2U-Net, were compared to each other. These architectures were tested in the smaller-crop data set with 64 filters, the larger amount that was computationally possible. The variants (other than the original U-Net) have around 36 million trainable parameters, approximately twice the number of the largest U-Net considered in this work. This limits the maximum image size, directly related to the training time and required memory allocation.

The results of these tests are presented in Table 2. It shows the training and validation MAE and the validation MSE (a proxy for the RMSE), and it is sorted by validation performance. It is observed that the performance of the variants at their maximum capacity does not improve the performance of the original U-Net, neither in training nor validation. It also shows that the U-Net variants tend to overfit more, as they have a poorer generalization performance relative to their training performance. This may suggest that the data set information is small relative to the networks' capacity. In light of the previous analysis and the added complexity of the U-Net variants, only the original U-Net architecture was considered in the following tests of this section.

Table 2: Validation metrics for the different architecture variations of the original U-Net model with 64 filters. The test was done for 1h ahead forecast over the smaller-crop data set and MAE as training loss.

architecture	**training MAE**	**validation MAE**	**validation MSE**
original U-Net	0.0555	0.0780	0.0182
attention U-Net	0.0597	0.0870	0.0198
nested U-Net	0.0746	0.0931	0.0221
R2U-Net	0.0614	0.1053	0.0283

5.1.2. Number of parameters

In the original U-Net architecture the number of parameters is directly proportional to the number of filters. Three networks with different numbers of filters (16, 32, and 64) were trained for a prediction horizon of one hour. The first two rows of Table 3 show the configurations that were tested, in which the relation between the number of filters and parameters can be observed. From these tests, it can be concluded that the U-Net with 16 filters achieves better training and validation performance than the U-Nets with 32 and 64 filters. Note that the metrics of Table 2 and Table 3 do not match exactly due to different images' sizes and coverage, although their behavior and order of magnitude are the same. The U-Net configuration with 16 filters is the baseline U-Net used in the next experiments.

Table 3: Validation metrics for the U-Net model with different number of filters and 1h ahead forecast.

filters	parameters	training MAE	validation MAE	validation MSE
16	1080929	0.0532	0.0631	0.0119
32	4318401	0.0591	0.0766	0.0171
64	17262977	0.0587	0.0920	0.0238

5.1.3. Impact of the training metric

When training a machine learning model, an optimization objective must be chosen, also known as loss function. Naturally, a model trained to minimize a specific metric can perform sub-optimally under another metric (Zhao et al., 2017). Which one of the metrics is the most adequate for the solar forecasting problem is still an open question, as the value of the forecast is directly related to its ability to influence the decision-making processes (Yang et al., 2022). However, there is agreement that a set of independent metrics can assess different aspects of the quality of a forecast (Yang et al., 2020) if chosen carefully. A related question is which should be the training metric for models aimed at cloudiness and solar deterministic forecast, with the added complexity of spatial representation, which is different from the single-location time-series analysis. An analysis in this sense is provided in this subsection, inspecting mainly the MAE and MSE as training metrics, to understand the effect of the loss function choice.

Three U-Net architectures with 16 filters were trained with the only difference of its training loss function, being respectively, the MAE, MSE, and SSIM. The test was done in the same conditions as the previous subsection (for one hour ahead forecast with the 1024×1024 px images). MAE and MSE were selected as they represent classic metrics in the solar forecast field that are typically used for the adjustment of machine learning methods. These metrics weigh the forecast errors differently and quantify different aspects of the forecast quality. SSIM metric was also included as an image quality metric, mainly as an exploratory option. This metric has a quite different conception and objective than the MAE and MSE. The performance of the three models was assessed over the validation set with the same three metrics, obtaining the double-entrance 3×3 matrices of Table 4. The left-hand side of this table presents the three validation metrics (in absolute terms) when each of them is used as a loss function for training. The right-hand side shows the same information but centered by subtracting the optimum value of each metric across the three tests (the minimum for MAE and MSE, and the maximum for SSIM) and expressed as a percentage of it. Please note that MAE and MSE are negative-oriented metrics while SSIM is positive-oriented. As expected, each model performs better when evaluated with the same metric used for training. It is observed that both MAE and MSE can be used as loss functions without much loss in the other metrics. MAE as the optimization target has slightly less impact on the SSIM than the MSE. On the other hand, using MAE as the loss function degrades the MSE more (8.2%) than the reverse situation, in which using MSE degrades MAE by 4.1%.

The use of the SSIM as the loss function increases importantly the MSE validation metric and has less impact on the MAE validation metric. This is consistent with the sharper visual results of the MAE in comparison to the MSE, as this latter tends to generate predicted images with higher blur. All in all, as the choice of MAE or MSE as the loss function has a different relative impact (higher or lower) in the other two metrics, the analysis is not conclusive. The MAE metric was then favored as it is common ground in the machine learning field. If the RMSE is chosen as the evaluation metric, one may expect an improvement of $\simeq 4\%$[1] from the results presented in this article (Section 6) by using instead the MSE as loss function for the U-Nets' training.

Table 4: Validation metrics for the U-Net model and 1h ahead forecast when trained using three different loss functions.

validation metric	**cost function** MAE	MSE	SSIM	**percentage difference**	**cost function** MAE	MSE	SSIM
MAE (abs.)	**0.0631**	0.0657	0.0669	MAE (Δ%)	**0.0%**	4.1%	6.0%
MSE (abs.)	0.0132	**0.0122**	0.0153	MSE (Δ%)	8.2%	**0.0%**	25.4%
SSIM (abs.)	0.5905	0.5838	**0.5979**	SSIM (Δ%)	-1.2%	-2.4%	**0.0%**

The previous analysis provides a first study on the relationship between the training loss function and the target performance metrics, an issue that has not been extensively addressed in the field so far. Of course, if one particular metric is of interest for whatever problem-specific reason, then it should be used as the optimization target for the training stage. An interesting discussion of one part of this problem can be found in Section 2.1 of Yang et al. (2020).

5.1.4. U-Net Diff

The U-Net Diff model consists of changing the target to be the difference between the last image (time t) of the input sequence and the desired objective for the given time horizon. In this way, the U-Net Diff is trained to predict the changes between the actual and the future image, and not the future image itself. This can be seen as a naive way to remove the image's background. To allow for negative values in the output, the last activation function was changed to a hyperbolic tangent instead of a sigmoid. To test this modification, the baseline U-Net and the U-Net Diff were trained for prediction horizons from 1 to 5 hours ahead with the 1024×1024 px images using MAE as the loss function. The validation results show similar performance for the two networks, with the U-Net Diff achieving marginally better results in four of the five forecast horizons. As the results over the validation set are very similar to the results over the final test set, only the latter are presented in Section 6.

[1]This was assessed from Table 4 by calculating RMSE $= \sqrt{\text{MSE}}$.

5.1.5. Extra ablations

This subsection summarizes some other tests that caused no improvement but might be of interest. The preliminary studies about the optimal input data pipeline showed that no performance improvement came from (i) using rotations as a data augmentation strategy (a strategy that is used in some contexts), (ii) adding the date and time as input information for the network, or (iii) using the CMV method as extra input information. The CMV information was included in the form of forecasts or vector fields, either as additional channels or at a later network stage, and none of them produced improvements. Another experiment regarding implementation showed that recursively using a 10-min single-horizon U-Net resulted in worse performance than using the horizon-specific models. This was the expected behavior, as the U-Net with a 10-min horizon was not trained to deal recursively with its blurry outputs as inputs or to minimize the recursive error on larger horizons, but was tested for completeness.

5.2. IrradianceNet

Two types of IrradianceNet models were trained. One only uses images as input and the other the images with added spatial (latitude and longitude) and topographic data (elevation map). The elevation map was normalized by its absolute maximum inside the corresponding crop. The other spatial inputs (coordinates) were mapped to the $[0, 1]$ interval. The network using these extra inputs is called IrradianceNet GEO and was recommended by Nielsen et al. (2021b). As it is trained over patches of the images due to its computational cost, the spatial information may provide the network with knowledge from where the patch is taken. A 10-minute prediction step was too computationally expensive, so a 30-minute time step was used instead, following the original article by Nielsen et al.. Also, to be consistent with the original article, both models use four previous images as input to generate the prediction (times $(t - k)$ with $k = \{0, 1, 2, 3\}$), although in this case, they are separated by 10-minute steps. Note that, in this way, this algorithm uses one more previous image than the U-Nets. A related issue is that, operationally, the algorithm needs to wait for one daylight image more than the versions with three inputs, delaying 10 minutes (in this case) its first forecast of the day. This is not an important cost when using 10-minute images, but it certainly would be an issue with 30-minute images.

The training configuration for IrradianceNet and IrradianceNet GEO is similar to the ones used for the U-Net: 100 epochs, Xavier initialization (Glorot and Bengio, 2010) for the weights, and Adam optimizer (Kingma and Ba, 2015) with a variable learning rate scheduled to be reduced in half if validation performance does not improve for 15 epochs. However, there are two changes to be consistent with the training configuration used in the original article. These are the initial learning rate equal to 2×10^{-3} and the selection of MSE as the training loss. The location of the patches taken during training is random for each batch. During validation, the patches are fixed to cover the whole image without overlapping.

The results from running the trained models on the validation data set showed that the IrradianceNet

GEO was not able to exploit the additional input data, as it performed very similarly to the basic IrradianceNet. These results could be explained by the low value that the geographic information adds for the South American Pampa Húmeda (which is mostly flat grassland) in comparison to the region of the original article (Europe). IrradianceNet GEO is included in the final results for completeness.

6. Results

This section contains the final results of the performance evaluation and the method's behavior over the test data set, considering the proposed and selected models. This test data set has been unseen in all previous analyses and optimization. Subsection 6.1 presents the final performance assessment and metrics, while Subsection 6.2 shows some selected predicted sequences with interesting insights. For clarity, some of the quantitative information is presented in Appendix B.

6.1. Quantitative results

Figure 3 shows the MAE, RMSE, and FS metrics across the hourly forecast horizons up to 5h ahead. The evaluation includes three reference algorithms: the Persistence and the CMV in its two versions, pixel-wise and with optimized spatial blurring. The two pairs of DL models being evaluated are included, the U-Net and U-Net Diff, and the IrradianceNet with and without spatial information. As explained in Section 1, the CMV algorithm is evaluated only over its valid pixels. For the rest of the models, the evaluation is conducted over the entire 1024×1024 px images. Please note that the metrics' values do not need to match those of Section 5 as the data sets are not the same (validation vs test data set). However, it is a sanity check to observe that the orders of magnitudes are the same. A direct qualitative comparison with the results of Nielsen et al. (2021a) or Nielsen et al. (2021b) over Europe is not feasible due to different reasons. In Nielsen et al. (2021a) the algorithm is run with a satellite cloud classification product, so the reported metrics at the image level are in accordance with the objectives of the study, being different from the ones here. In Nielsen et al. (2021b) the evaluation is performed for the hourly predicted irradiation, not at image level (which would be in that case for the effective cloud albedo, a satellite cloud index used for satellite-to-irradiance conversion), and, apart from MAE, the metrics are different. In any case, IrradianceNet is included in this article and evaluated with the same data set along with the other methods, thus providing a fair comparison in the target region. The performance differences that may arise due to different regions and climates should be addressed in future benchmark studies.

The largest difference between performances in Figure 3 is seen when evaluating with MAE, where the U-Net and U-Net Diff significantly outperform all the other models. This is expected as these are the only two models trained to optimize this metric. It can be seen that the two IrradianceNet versions have almost the same performance in this metric and outperform the persistence and pixel-wise CMV. In comparison

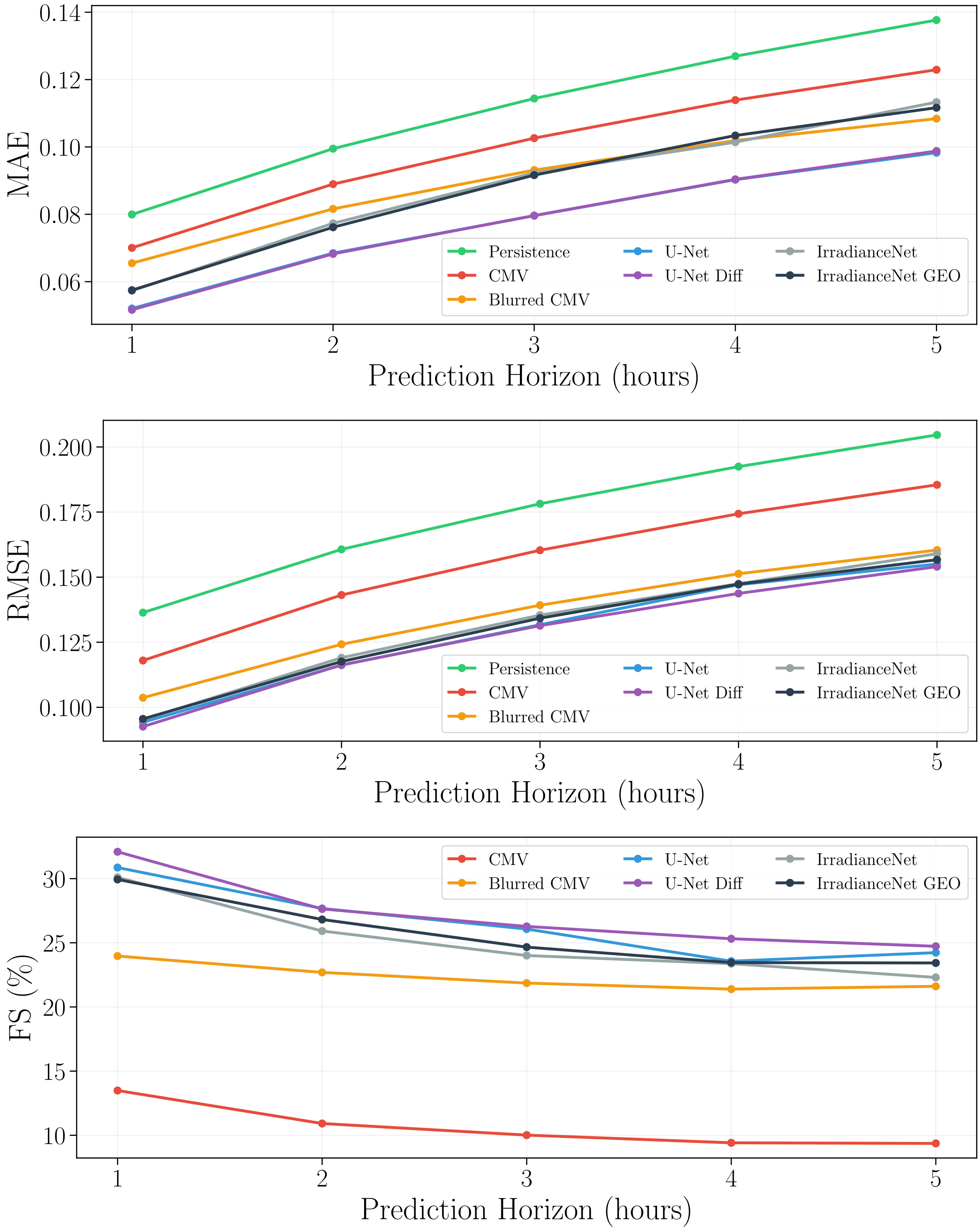


Figure 3: Main performance metrics vs forecast horizon for the reference methods and DL models.

with the blurred CMV, the IrradianceNet performance is similar for 3 and 4 hours ahead, being better in the 1h and 2h forecast horizons and worse in the last one. It shall be recalled at this point that both the Blurred CMV and IrradianceNet networks were optimized to minimize the training MSE.

Observing the RMSE performance plot of Figure 3 the differences are smaller than in the MAE analysis, and they are better observed in the Forecasting Skill plot. The IrradianceNet variations perform better than the blurred CMV for all forecast horizons under this metric. The best-performing architectures are the U-Nets, being the U-Net Diff the one with the highest FS% for all forecast horizons. The fact that the U-Net is still superior in RMSE when using a different training loss is remarkable, as per Table 4 this performance could be further improved in about 4% if MSE were to be used as a loss function. The Forecasting Skill plot also allows seeing the important improvement that is gained with a simple spatial blurring in the CMV output, coinciding with the results of Aicardi et al. (2022) with the former GOES-East satellite images. The work of Aicardi et al. also obtained a decreasing trend in the forecasting skill with the forecast horizon with a higher drop in the transition between the first and the second hour, and with similar values. The MBE% and RMSE% plots are shown and discussed in Appendix B, and also have similar values to that of Aicardi et al..

6.2. Qualitative results

It is interesting to explore some differences between the U-Nets, IrradianceNet, and CMV predictions. Two comparisons are made: (i) the U-Net vs CMV, showing some selected cases in which the U-Net was able to predict cloud extinction, a feature that current baseline CMV strategies are unable to perform, and (ii) the spatial distribution of the errors of the U-Net and IrradianceNet networks, specifically addressing the artifacts observed by using the patch processing strategy. Point (i) also includes a visualization of the border's effect of the CMV methods. The U-Net Diff is selected for the following discussion.

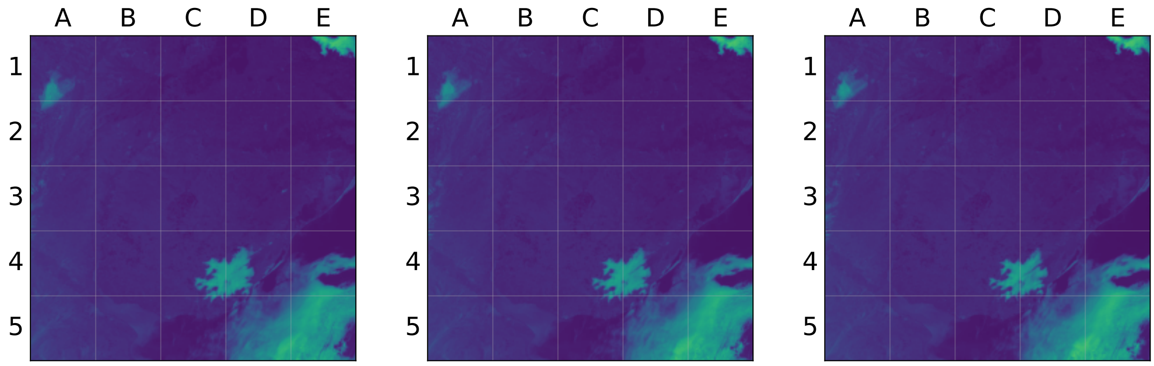
6.2.1. Cloud extinction case studies

Figures 4 and 5 show two examples in which the U-Net was able to correctly forecast cloud extinction, while the CMV, as expected, was not. The first row shows the three lagged images used for the U-Net forecast, and the one on the right is the time t image. Note that the CMV only uses the last two of these three images. The second and third rows show the predictions of the U-Net and the CMV along with the corresponding ground truth image on the left side. In the first sequence of Figure 4 (August 4th, 2020, time t being 14:40 UTC-0) the clouds in 4C-4D, 1A-2A, and 1E cells disappear after 2 hours, as can be seen in the ground truth. The U-Net forecasts these extinctions accurately in the three cases, including the intermediate stage at one hour ahead, being a remarkable feature of this method. The CMV, in change, maintains these clouds in its prediction. The second sequence of Figure 5 (October 14th, 2020, time t being 12:40 UTC-0) shows a similar behavior but in a more complex situation. In this case, the extinction of the clouds located in cells 2D-2E and 4D occurs after three hours. The U-Net, again, manages to predict its reduction and later disappearance, but in this case in a moving context. In particular, the U-Net detects the drift of the 2D-2E clouds and it is capable of predicting that these clouds will detach from the main system in the top-left corner, and gradually vanish for larger horizons. Similar behavior and accurate prediction by the U-Net are observed for the cloud in the 4D cell. Of course, the CMV is unable to predict any of these phenomena, and these identified clouds wrongly remain in all its predictions. This second sequence also allows seeing the border's effect of the CMV prediction, as this sequence includes important movement in the scene. This effect increases with the time horizon, being a drawback of the CMV strategy that can be mitigated by using larger images, with the corresponding computational cost.

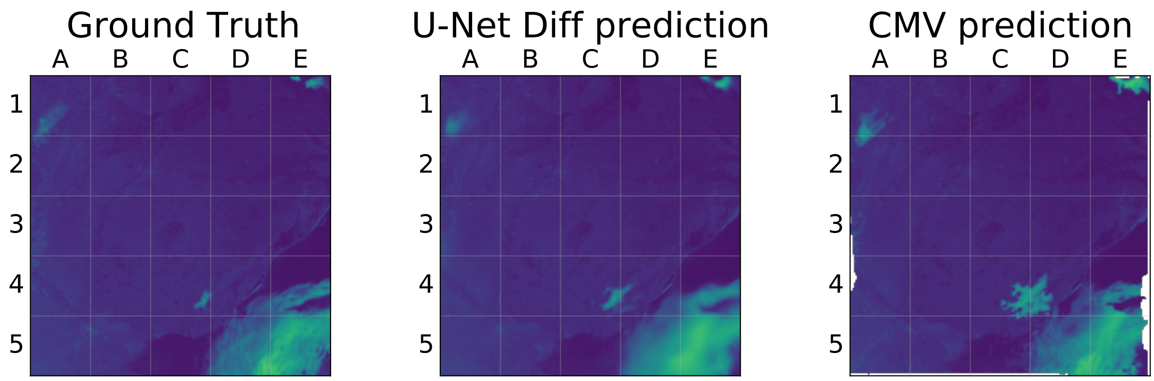
The previous features of the U-Net do not come along without any cost. As can be seen in both sequences the U-Net predictions tend to be blurred, a trick that is learned by the network to reduce the cost of high errors. As in any deterministic forecasting problem, there is a trade-off between having an overall good performance and taking risks in the prediction. There is then important room for further studies to understand the separation between image blurring and predicting clouds' movement, deformation, and formation/extinction, some of which are part of our current work.

6.2.2. Spatial dependence of the error

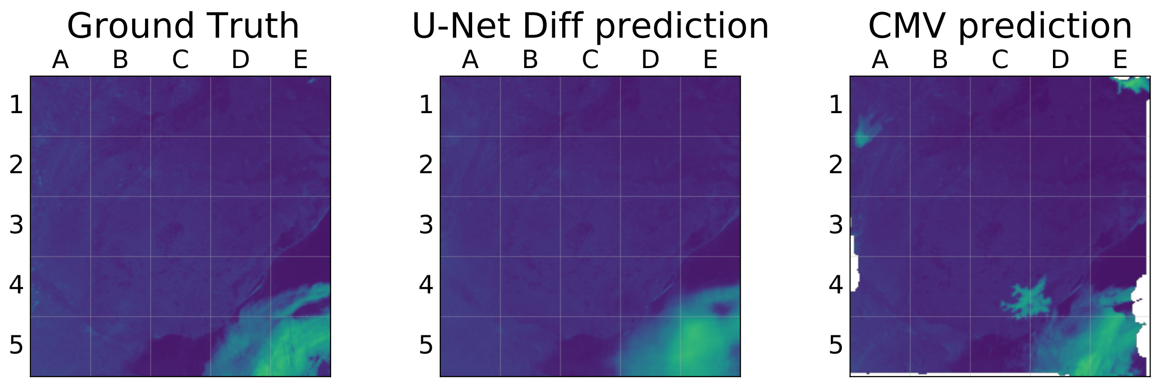
It is expected that the models perform better close to the center of the image, as there is less uncertainty about the clouds that could move into or out of that area. In change, the prediction of incoming clouds through the image borders is a much harder task. However, this may not be true, if clouds are more difficult to predict in a given sub-region. Also, predicting the borders of the image is not a symmetric problem in average terms, as the atmosphere circulation makes clouds' phenomena have preferred directions in different parts of a given territory. This analysis was done with three models: the Persistence, the U-Net Diff, and IrradianceNet methods. Figure 6 presents the spatial per-pixel distribution of the RMSE over the test data set for each of the 5 hourly time horizons. These error maps show that all models make, on average, the largest errors in the top right corner (northeast) of the region under study, being the Persistence procedure the most disadvantaged. This analysis reveals that the climate and geographical characteristics of a region can be an important driver of errors' spatial distribution. For this region, in particular, the clouds in the northeast region are harder to predict, being associated with the typical circulation and behavior of cloudiness in that area. These areas with the highest errors are consistent through the 5-time horizons, with the difference of a natural error increase in the whole map as the prediction horizon grows. When visualizing the error maps of IrradianceNet, the patch-based prediction is observed as a grid-like layout throughout the region. Apart from the grid-like effect, the underlying error maps of the IrradianceNet predictions are similar to the ones of the U-Net Diff.



(a) Last three images (times $(t-k)$ with $k = \{0, 1, 2\}$), from left (14:20 UTC0) to right (14:40 UTC0).

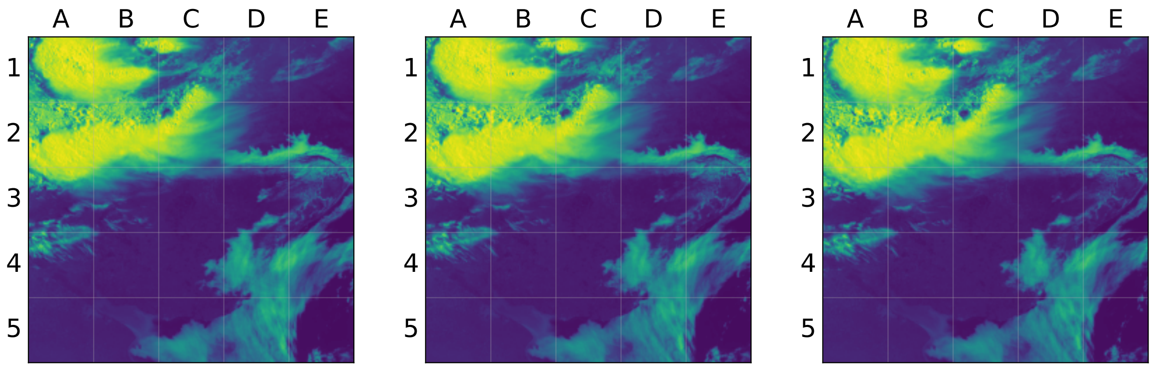


(b) 1 hour ahead ground truth and predictions.

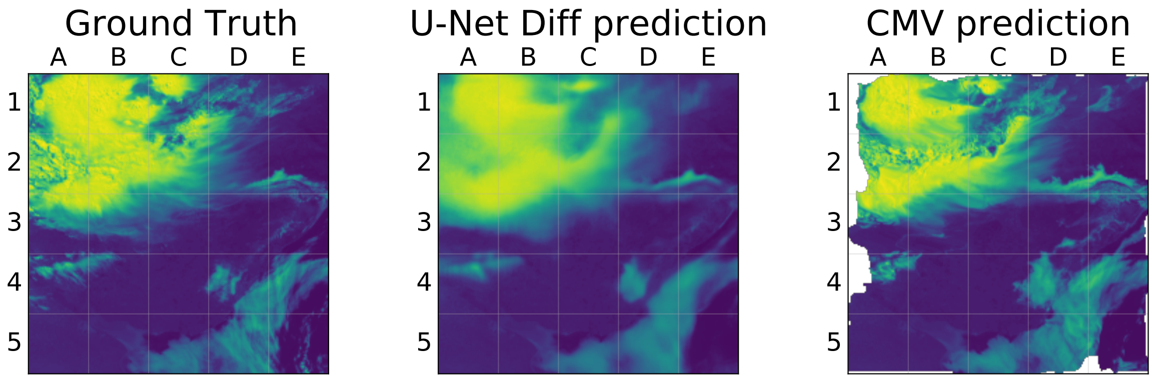


(c) 2 hours ahead ground truth and predictions.

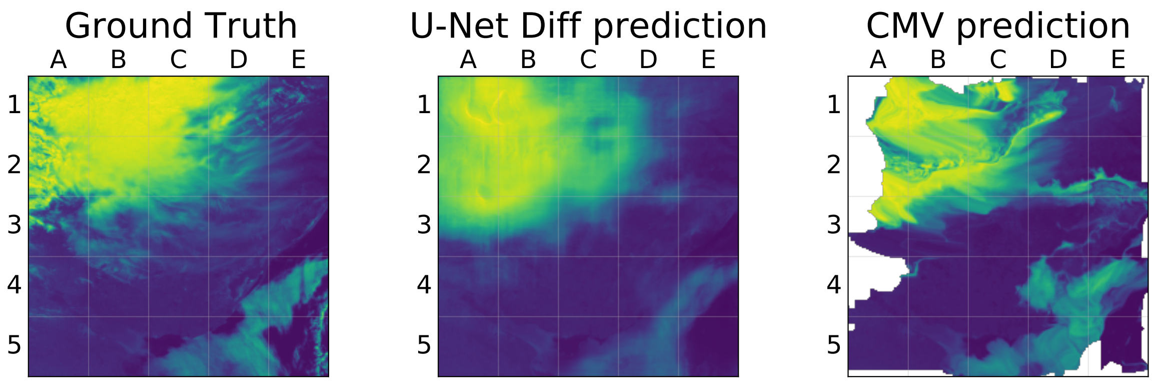
Figure 4: Image sequence captured on August 4th, 2020. Current time: 14:40 UTC0 (top-right image).



(a) Last three images (times $(t-k)$ with $k = \{0, 1, 2\}$), from left (12:20 UTC0) to right (12:40 UTC0).



(b) 1 hour ahead ground truth and predictions.



(c) 3 hours ahead ground truth and predictions.

Figure 5: Image sequence captured on October 14th, 2020. Current time: 12:40 UTC0 (top-right image).

7. Conclusions

This article analyzed Deep Learning (DL) techniques applied to satellite-based cloudiness prediction (Earth's albedo) up to 5 hours ahead. This is the first stage of satellite intra-day solar irradiance forecasting.

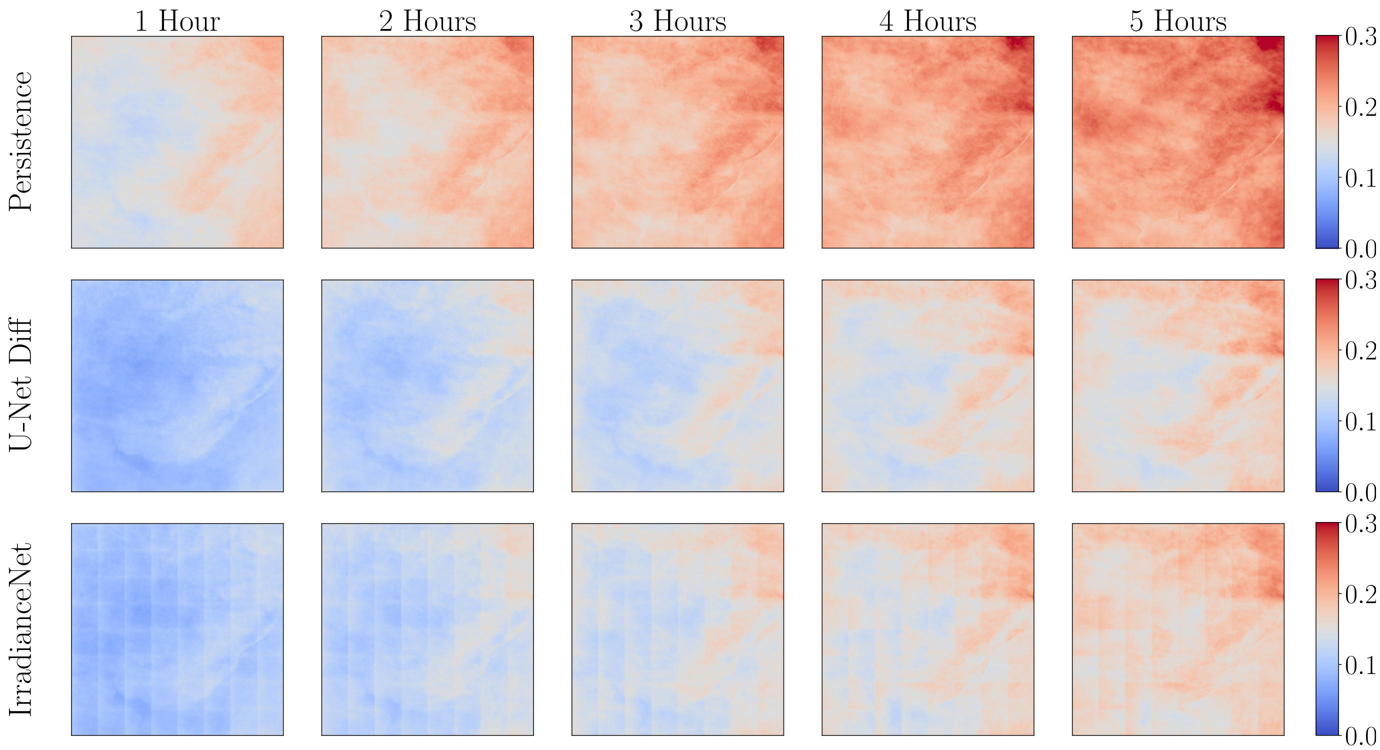


Figure 6: Root mean squared error per pixel for all five-time horizons.

The work used the 10-minute GOES-16 visible channel satellite images for the southeast part of South America, an area known as Pampa Húmeda. This is a region in which the convective systems' evolution is challenging and the solar irradiance variability is at an intermediate level, as clear sky, overcast, and partly cloudy conditions alternate. The utilization of the original U-Net DL network was proposed and tested for this purpose along with the up-to-date IrradianceNet DL algorithm. Two baseline methods were also included, namely, the satellite cloudiness persistence and an advanced Cloud Motion Vectors (CMV) strategy with optimized spatial blurring. The U-Net optimization was analyzed in detail, providing for it different ablation studies. Two U-Nets with 16 filters were implemented; the regular U-Net that aims to predict the next image and the U-Net Diff that aims to predict the difference with the last available image (time t). A different network was trained for each time horizon. This is a common practice in the field, however, it was confirmed in this work as best practice, in opposition to the recurrent utilization of a single one-lead-time DL network. None of the U-Net variants (Attention U-Net, Nested U-Net, and R2U-Net) were found to outperform the original U-Net. Both final U-Nets showed the remarkable feature of predicting cloud extinction, which is one of the harder issues in satellite-based solar forecasting.

The DL methodologies presented better performance than the baseline methods, including the blurred CMV, which sets a very exigent performance bound. The preexisting architecture, IrradianceNet, as proposed by Nielsen et al., was adapted successfully when retrained and evaluated on a different geographical region and larger images. In this study, at least for this region, both U-Net architectures outperformed IrradianceNet, and the U-Net Diff was the best performing. One of the strengths of the U-Net architecture is its wide utilization and ease of implementation, as it is considered a light, robust, and extensively tested DL method. This work illustrated the U-Net utilization over images with a larger size and a higher resolution than in previous works dealing with DL methods for this purpose. Regarding training time, the IrradianceNet had higher requirements than the U-Nets. However, once the DL networks are trained, prediction times are low, making any of them suitable for real-time operation. In particular, none of the final U-Net models are computationally expensive to use, being possible to generate a prediction in less than a second using a single GPU.

A. Limitations

There is plenty of room for improvement and this work is far from comprehensive. For instance, data augmentation via random cropping was not considered for the U-Net. This method was used to train IrradianceNet and can help escalate the models to larger regions (Espeholt et al., 2021). In addition, the optimization objective remains arbitrarily defined, for there is not a single metric of interest. This introduces optimization compromises, e.g. when the optimization is done with the MSE loss function, the predictions appear with higher blur than when using MAE. There is still an unsolved (and not yet fully understood) trade-off between overall accuracy and risky variability prediction, that the networks learn to mitigate by blurring. Moreover, the preprocessing could also be enhanced by including image background removal in sophisticated ways. This could be especially impactful for salty/snowy regions of land, not present in our images. The impact of such a transformation on performance is yet unknown and it was not specifically addressed in this work, being part of our current work. Lastly, the results suggested no improvement when adding the GEO information to the IrradianceNet. Although this can be explained by the low variability in altitude in the studied area compared to Europe, the question of what information can be effectively fed to the neural network is raised. Another way to include meaningful information would be to directly provide the neural network with relevant regional data, such as typical cloudiness variability or mean wind direction.

B. Complementary assessment information

For completeness and clarity of presentation in the main text, the MBE% and RMSE% plots are provided here. Figure B.7 shows these two metrics for the assessment of Section 6. The MBE% are between $\simeq \pm 10\%$, similar to Aicardi et al. (2022), in which all tested CMV strategies and the persistence tend to decrease (with its sign) with the forecast horizons. In particular, persistence shows very similar behavior. The differences here are observed with the IrradianceNet variations, whose MBE% increases with the forecast horizons, being positive for 3 to 5 hours ahead. The RMSE% plot contains the same information as the RMSE plot of Figure 3 but is normalized by each image's mean value before averaging the metric across all images in the test set. It is interesting to note that the order of magnitude in this plot is similar to that of Aicardi et al..

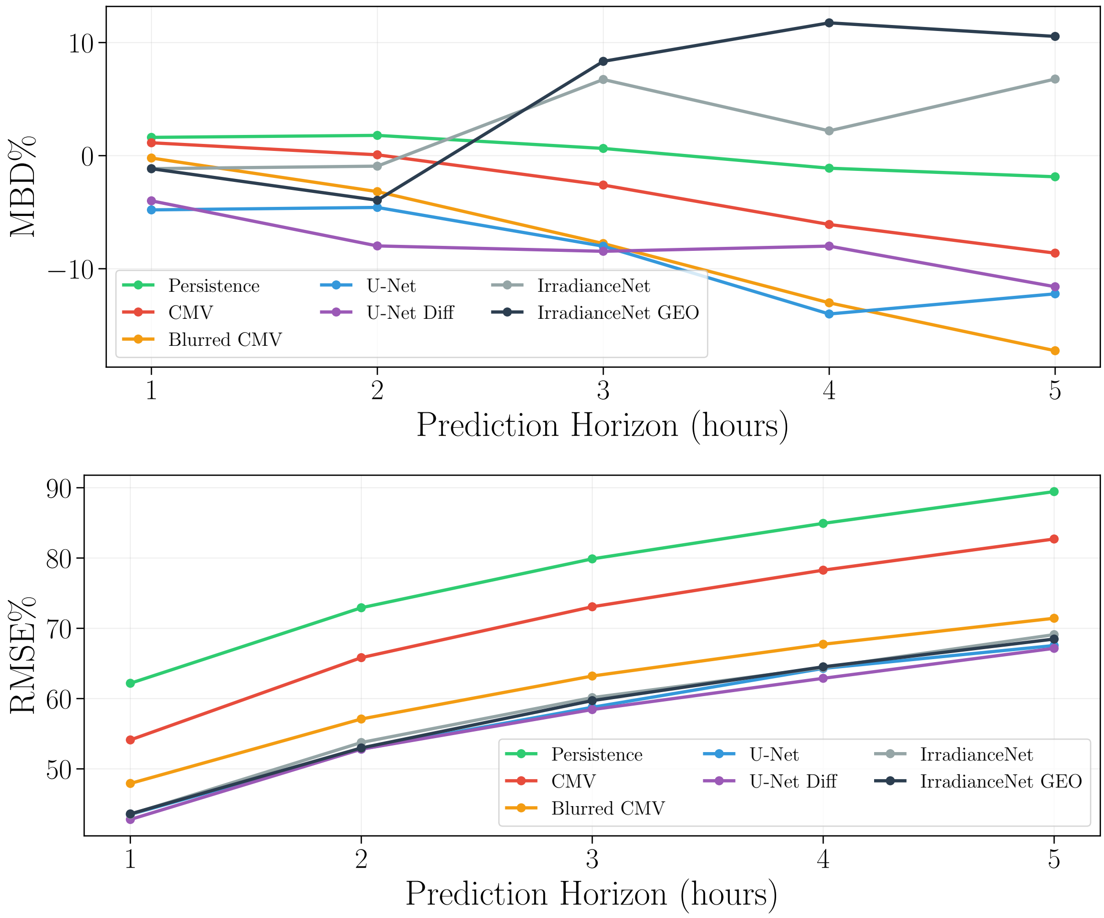


Figure B.7: Complementary performance metrics (%) vs forecast horizon for the reference methods and DL models.

Author contributions

Franco Marchesoni-Acland contributed with conceptualization, methodology development, writing - original draft, writing - review and editing, supervision, and project administration. **Andrés Herrera, Franco Mozo, and Ignacio Camiruaga** contributed with conceptualization, methodology development, software, validation, formal analysis, investigation, resources, data curation, writing - original draft, writing - review and editing, and visualization. **Alberto Castro** contributed with resources, writing - review and editing, supervision, and project administration. **Rodrigo Alonso-Suárez** contributed with conceptualization, methodology, resources, writing - original draft, writing - review and editing, supervision, and project administration.

Acknowledgments

Data was made available by the Solar Energy Laboratory (LES) of the Universidad de la República. The experiments presented in this article were performed using the ClusterUY infrastructure (https://cluster.uy). R. Alonso-Suárez acknowledges partial financial support from the CSIC Group's Program, Udelar, Uruguay.

References

Aicardi, D., Musé, P., and Alonso-Suárez, R. (2022). A comparison of satellite cloud motion vectors techniques to forecast intra-day hourly solar global horizontal irradiation. *Solar Energy*, 233:46–60.

Alom, M. Z., Hasan, M., Yakopcic, C., Taha, T. M., and Asari, V. K. (2018). Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation. *CoRR*, abs/1802.06955.

Alonso-Suárez, R., David, M., Branco, V., and Lauret, P. (2020). Intra-day solar probabilistic forecasts including local short-term variability and satellite information. *Renewable Energy*, 158:554–573.

Alonso-Suárez, R., Marchesoni, F., Dovat, L., and Laguarda, A. (2021). Satellite-based operational solar irradiance forecast for Uruguay's solar power plants. In *2021 IEEE URUCON*, pages 182–187.

Baranchuk, D., Rubachev, I., Voynov, A., Khrulkov, V., and Babenko, A. (2022). Label-efficient semantic segmentation with diffusion models. Published as a conference paper at ICLR 2022, arXiv:2112.03126.

Bertalmio, M., Bertozzi, A., and Sapiro, G. (2001). Navier-Stokes, fluid dynamics, and image and video inpainting. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, volume 1, pages I–I.

Berthomier, L., Pradel, B., and Perez, L. (2020). Cloud cover nowcasting with deep learning. In *Proceedings of the 2020 Tenth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 1–6.

Brousseau, P., Seity, Y., Ricard, D., and Léger, J. (2016). Improvement of the forecast of convective activity from the AROME-France system. *Quarterly Journal of the Royal Meteorological Society*, 142(699):2231–2243.

Cano, D., Monget, J., Albuisson, M., Guillard, H., Regas, N., and Wald, L. (1986). A method for the determination of the global solar radiation from meteorological satellite data. *Solar Energy*, 37(1):31–39.

Croitoru, F.-A., Hondru, V., Ionescu, R. T., and Shah, M. (2023). Diffusion models in vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–20. arXiv:2209.04747.

Cros, S., Sébastien, N., Liandrat, O., and Schmutz, N. (2014). Cloud pattern prediction from geostationary meteorological satellite images for solar energy forecasting. In *Proceedings of SPIE - The International Society for Optical Engineering*, volume 9242.

Du, G., Cao, X., Liang, J., Chen, X., and Zhan, Y. (2020). Medical image segmentation based on u-net: A review. *Journal of Imaging Science and Technology*, 64(2):020508–1–020508–12.

Espeholt, L., Agrawal, S., Sønderby, C. K., Kumar, M., Heek, J., Bromberg, C., Gazen, C., Hickey, J., Bell, A., and Kalchbrenner, N. (2021). Skillful twelve hour precipitation forecasts using large context neural networks. *CoRR*, abs/2111.07470.

Falk, T., Mai, D., Bensch, R., Çiçek, O., Abdulkadir, A., Marrakchi, Y., Böhm, A., Deubner, J., Jäckel, Z., Seiwald, K., Dovzhenko, A., Tietz, O., Bosco, C. D., Walsh, S., Saltukoglu, D., Tay, T. L., Prinz, M., Palme, K., Simons, M., Diester, I., Brox, T., and Ronneberger, O. (2019). U-Net: deep learning for cell counting, detection, and morphometry. *Nature Methods*, 16:67–70.

Fan, L., Zhang, F., Fan, H., and Zhang, C. (2019). Brief review of image denoising techniques. *Visual Computing for Industry, Biomedicine, and Art*, 2(1):1–12.

Farnebäck, G. (2003). Two-frame motion estimation based on polynomial expansion. In Bigun, J. and Gustavsson, T., editors, *Image Analysis*, pages 363–370, Berlin, Heidelberg. Springer Berlin Heidelberg.

Glorot, X. and Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. *Journal of Machine Learning Research - Proceedings Track*, 9:249–256.

Horn, B. K. P. and Schunck, B. G. (1981). Determining optical flow. *Artif. Intell.*, 17(1-3):185–203.

IRENA (2021). Renewable energy statistics.

Kang, Z., Huang, Z., and Lu, C. (2022). Speech enhancement using u-net with compressed sensing. *Applied Sciences*, 12(9).

Kingma, D. P. and Ba, J. (2015). Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*.

Kühnert, J., Lorenz, E., and Heinemann, D. (2013). Chapter 11 - Satellite-based irradiance and power forecasting for the german energy market. In Kleissl, J., editor, *Solar Energy Forecasting and Resource Assessment*, pages 267–297. Academic Press, Boston.

Laguarda, A., Giacosa, G., Alonso-Suárez, R., and Abal, G. (2020). Performance of the site-adapted cams database and locally adjusted cloud index models for estimating global solar horizontal irradiation over the pampa húmeda. *Solar Energy*, 199:295–307.

Lauret, P., Alonso-Suárez, R., Le Gal La Salle, J., and David, M. (2022). Solar forecasts based on the clear sky index or the clearness index: Which is better? *Solar*, 2(4):432–444.

Lefèvre, M., Oumbe, A., Blanc, P., Espinar, B., Qu, Z., Wald, L., Homscheidt, M. S., and Arola, A. (2013). McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions. *Atmospheric Measurement Techniques, European Geosciences Union*, 6:2403–2418.

Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3431–3440, Los Alamitos, CA, USA. IEEE Computer Society.

Lorenz, E., Hammer, A., and Heinemann, D. (2004). Short term forecasting of solar radiation based on satellite data. In *EUROSUN2004 (ISES Europe Solar Congress)*, pages 841–848, Freiburg, Germany.

Lucas, B. D. and Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2*, IJCAI'81, pages 674–679, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Nesmachnow, S. and Iturriaga, S. (2019). Cluster-UY: Collaborative Scientific High Performance Computing in Uruguay. In Torres, M. and Klapp, J., editors, *Supercomputing*, pages 188–202, Cham. Springer International Publishing.

Nielsen, A. H., Iosifidis, A., and Karstoft, H. (2021a). CloudCast: A Satellite-Based Dataset and Baseline for Forecasting Clouds. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:3485–3494.

Nielsen, A. H., Iosifidis, A., and Karstoft, H. (2021b). IrradianceNet: Spatiotemporal deep learning model for satellite-derived solar irradiance short-term forecasting. *Solar Energy*, 228:659–669.

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B., and Rueckert, D. (2018). Attention U-Net: Learning where to look for the pancreas. In *Medical Imaging with Deep Learning*.

Peel, M. C., Finlayson, B. L., and Mcmahon, T. A. (2007). Updated world map of the köppen-geiger climate classification. *Hydrology and Earth System Sciences Discussions*, 11:1633–1644.

Peng, Z., Yoo, S., Yu, D., and Huang, D. (2013). Solar irradiance forecast system based on geostationary satellite. In *2013 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 708–713.

Perez, R., Ineichen, P., Moore, K., Kmiecik, M., Chain, C., George, R., and Vignola, F. (2002). A new operational model for satellite-derived irradiances: description and validation. *Solar Energy*, 73(5):307–317.

Perez, R., Kivalov, S., Schlemmer, J., Hemker, K., Renné, D., and Hoff, T. E. (2010). Validation of short and medium term operational solar radiation forecasts in the US. *Solar Energy*, 84(12):2161–2172.

Qu, Z., Oumbe, A., Blanc, P., Espinar, B., Gesell, G., Gschwind, B., Klüser, L., Lefèvre, M., Saboret, L., Schroedter-Homscheidt, M., and Wald, L. (2017). Fast radiative transfer parameterisation for assessing the surface solar irradiance: The Heliosat-4 method. *Meteorologische Zeitschrift*, 26(1):33–57.

Rasmussen, K. L., Zuluaga, M. D., and Houze Jr., R. A. (2014). Severe convection and lightning in subtropical South America. *Geophysical Research Letters*, 41(20):7359–7366.

Ravuri, S., Lenc, K., Willson, M., Kangin, D., Lam, R., Mirowski, P., Fitzsimons, M., Athanassiadou, M., Kashem, S., Madge, S., et al. (2021). Skilful precipitation nowcasting using deep generative models of radar. *Nature*, 597(7878):672–677.

Ren, X., Li, X., Ren, K., Song, J., Xu, Z., Deng, K., and Wang, X. (2021). Deep learning-based weather prediction: A survey. *Big Data Research*, 23:100178.

REN21 (2021). Renewables 2021: Global status report. Technical report, Renewable Energy Policy Network for the 21st Century, 1 rue Miollis, Building VII, 75015 Paris, France.

Rigollier, C., Lefevre, M., and Wald, L. (2004). The method Heliosat-2 for deriving shortwave solar radiation from satellite images. *Solar Energy*, 77(2):159–169.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10674–10685.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In Navab, N., Hornegger, J., Wells, W. M., and Frangi, A. F., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham. Springer International Publishing.

Salio, P., Nicolini, M., and Zipser, E. J. (01 Apr. 2007). Mesoscale convective systems over southeastern south america and their relationship with the south american low-level jet. *Monthly Weather Review*, 135(4):1290–1309.

Schneider, R., Bonavita, M., Geer, A., Arcucci, R., Dueben, P., Vitolo, C., Le Saux, B., Demir, B., and Mathieu, P.-P. (2022). ESA-ECMWF report on recent progress and research directions in machine learning for Earth system observation and prediction. *npj Climate and Atmospheric Science*, 5(1):1–5.

Seity, Y., Brousseau, P., Malardel, S., Hello, G., Bénard, P., Bouttier, F., Lac, C., and Masson, V. (2011). The AROME-France convective-scale operational model. *Monthly Weather Review*, 139(3):976–991.

Smith, A. G., Petersen, J., Selvan, R., and Rasmussen, C. R. (2020). Segmentation of roots in soil with U-Net. *Plant Methods*, 16.

Su, X., Li, T., An, C., and Wang, G. (2020). Prediction of Short-Time Cloud Motion Using a Deep-Learning Model. *Atmosphere*, 11(11).

Sánchez, J., Meinhardt-Llopis, E., and Facciolo, G. (2013). TV-L1 optical flow estimation. *Image Processing Online (IPOL).*

Sønderby, C. K., Espeholt, L., Heek, J., Dehghani, M., Oliver, A., Salimans, T., Agrawal, S., Hickey, J., and Kalchbrenner, N. (2020). MetNet: A neural weather model for precipitation forecasting. *arXiv preprint arXiv:2003.12140*.

Tan, C., Feng, X., Long, J., and Geng, L. (2018). FORECAST-CLSTM: A new convolutional LSTM network for cloudage nowcasting. In *2018 IEEE Visual Communications and Image Processing (VCIP)*, pages 1–4.

Urbich, I., Bendix, J., and Müller, R. (2019). The Seamless Solar Radiation (SESORA) Forecast for Solar Surface Irradiance—Method and Validation. *Remote Sensing*, 11(21).

Uwe, P., Steffen, K., Trentmann, Jörg, Rainer, H., Petra, F., Johannes, K., and Martin, W. (2019). Surface radiation data set - heliosat (sarah) - edition 2.1.

Voyant, C., Notton, G., Kalogirou, S., Nivet, M.-L., Paoli, C., Motte, F., and Fouilloy, A. (2017). Machine learning methods for solar radiation forecasting: A review. *Renewable Energy*, 105:569–582.

Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, E. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612.

Wei, S., Zhang, H., Wang, C., Wang, Y., and Xu, L. (2019). Multi-temporal sar data large-scale crop mapping based on u-net model. *Remote Sensing*, 11(1).

Yang, D., Alessandrini, S., Antonanzas, J., Antonanzas-Torres, F., Badescu, V., Beyer, H. G., Blaga, R., Boland, J., Bright, J. M., Coimbra, C. F., David, M., Âzeddine Frimane, Gueymard, C. A., Hong, T., Kay, M. J., Killinger, S., Kleissl, J., Lauret, P., Lorenz, E., van der Meer, D., Paulescu, M., Perez, R., Perpiñán-Lamigueiro, O., Peters, I. M., Reikard, G., Renné, D., Saint-Drenan, Y.-M., Shuai, Y., Urraca, R., Verbois, H., Vignola, F., Voyant, C., and Zhang, J. (2020). Verification of deterministic solar forecasts. *Solar Energy*, 210:20–37. Special Issue on Grid Integration.

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. (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.

Zach, C., Pock, T., and Bischof, H. (2007). A duality based approach for realtime TV-L1 optical flow. In *Proceedings of the 29th DAGM Conference on Pattern Recognition*, pages 214–223, Berlin, Heidelberg. Springer-Verlag.

Zhao, H., Gallo, O., Frosio, I., and Kautz, J. (2017). Loss functions for image restoration with neural networks. *IEEE Transactions on Computational Imaging*, 3(1):47–57.

Zhou, Z., Siddiquee, M., Tajbakhsh, N., and Liang, J. (2020). UNet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Transactions on Medical Imaging*, 39(6):1856—-1867.