Revista Brasileira de Geografia Física v.19, n.01 (2026) 018-032.

Revista Brasileira de Geografia Física

ISSN:1984-2295

Homepage:https://periodicos.ufpe.br/revistas/rbgfe

Optical Flow-Based Forecasting of Surface Irradiance Using Cloud Motion Vectors in Brazil

Rodrigo Souza Xavier[1], Lucas Ramos de Pretto[2], André Luiz Vizine Pereira[3], Fernando Ramos Martins[4]

[1] MSc. Interdisciplinary Sciences, Instituto do Mar, Universidade Federal de São Paulo. PhD. Student at Universidade Federal do Ceará, Rua 5, 100, CEP 60355-636, Fortaleza, Ceará. (85) 3366-9500. rodrigoxavierlt@gmail.com (corresponding author). ORCID: 0009-0004-5957-8250
[2] Postdoctoral fellow at Nuclear Center for Lasers and Applications, Nuclear and Energy Research Institute IPEN-CNEN/SP, São Paulo, São Paulo, Brazil. Lucas.de.pretto@gmail.com. ORCID: 0000-0001-7783-2137
[3] Professor Adjunto III, Instituto do Mar, Universidade Federal de São Paulo, Rua Dr. Carvalho de Mendonça, 144 CEP 11070-100, Santos, São Paulo. (11) 3385-4255. andre.vizine@unifesp.br. ORCID: 0000-0002-5658-7221
[4] Professor Dr. Associado I, Instituto do Mar, Universidade Federal de São Paulo, Rua Carvalho de Mendonça, 144, CEP 11070-100, Santos, São Paulo'. (11) 3385-4255. fernando.martins@unifesp.br. ORCID: 0000-0002-7618-4462
Artigo recebido em 16/02/2025 e aceito em 23/09/2025

R E S U M O
A transição energética está impulsionando a inserção de fontes renováveis de energia que apresentam uma relação intrínseca com as condições meteorológicas e com o clima. Nesse contexto, é importante desenvolver ferramentas que forneçam suporte ao crescimento da participação de fontes renováveis intermitentes, como a energia solar, na rede elétrica. Esta pesquisa desenvolveu um modelo de previsão de curto prazo para prever a irradiância solar incidente na superfície em quatro locais diferentes no Brasil. O modelo de previsão usa o método baseado em fluxo óptico de Farnebäck para rastrear o movimento das nuvens em imagens de satélite do GOES-16. Comparamos os resultados previstos com dados reais coletados em estações da rede SONDA durante diferentes estações do ano. A metodologia apresentou os menores valores de RMSE durante o inverno na localidade de Cachoeira Paulista (SP), com 0,13 para horizonte de previsão de 30 minutos, e 0,11 em 120 minutos a frente. O maior desvio RMSE foi observado durante o outono, com 0,39 em 30 minutos e 0,41 em 120 minutos na mesma localidade. Essas descobertas são consistentes com pesquisas em regiões tropicais da América do Sul, indicando a utilidade potencial dessa abordagem na área.
Palavras-chave: sensoriamento remoto da atmosfera, fluxo ótico, vetor de movimento de nuvens, irradiância solar global horizontal incidente na superfície.

Optical Flow-Based Forecasting of Surface Irradiance Using Cloud Motion Vectors in Brazil

A B S T R A C T
The energy transition is driving the integration of renewable energy sources with an intrinsic relationship with meteorological and climatic conditions. In this context, it is essential to develop tools that support the growth of the share of intermittent renewable sources, such as solar energy, in the electrical grid. This research developed a short-term forecasting model to predict surface solar irradiance at four different locations in Brazil. The forecasting model uses the Farnebäck optical flow method to track cloud motion in GOES-16 satellite images. We compared the predicted results with actual data collected at SONDA network stations during different seasons of the year. The methodology showed the lowest RMSE values during winter in Cachoeira Paulista (SP), with 0.13 for a 30-minute forecast horizon and 0.11 for a 120-minute horizon. The highest RMSE deviation was observed during autumn, with 0.39 in 30 minutes and 0.41 in 120 minutes at the same site. These findings are consistent with research in tropical regions of South America, indicating the potential utility of this approach in the field.
Keywords: Remote sensing of the atmosphere, optical flow, cloud motion vectors, surface horizontal global irradiance.

Introduction

Renewable energy technologies such as solar photovoltaic power systems are deemed essential to mitigate the effects of climate change, reducing dependence on fossil fuels (Olabi & Abdelkareem, 2022; Change, 2022) and offering an alternative to meet the commitments of the International Conference of the Parties (IPCC). Implementing proven sustainable energy sources is environmentally relevant for energy security and sustainable economic development (de Campos *et al.*, 2021; de Lima *et al.,* 2024). This is especially relevant for countries like Brazil, which has an installed hydroelectric generation capacity close to the limit of its energy potential. (Raupp and Costa, 2021).

As ethical energy transition advances, with greater participation of intermittent sources such as solar generation, the demand for accurate forecasts of solar irradiation also increases, reducing the impact their intermittency may cause on electricity transmission and distribution systems (Luiz *et al.*, 2018). Several methodologies are continuously developing to forecast Global Horizontal Irradiance (GHI) (Guermoui *et al.*, 2020; Singla *et al.*, 2021; Rahimi *et al.,* 2023). These methodologies intrinsically relate to the required forecast horizon and the available data infrastructure. Data collected by cloud cover imaging cameras installed on the surface (all-sky cameras) are relevant for very short-term forecasts (Rodríguez-Benítez *et al.*, 2021; Arraiz *et al.*, 2022). In contrast, satellite imagery is appropriate for time horizons of a few hours (Cheng *et al.*, 2022; Luiz *et al.*, 2018), and numerical models encompass from one to several days in the future (Benavides *et al.*, 2022). Numerical models require more time for their complex computational calculations and are subject to uncertainties related to the numerical parameterizations adopted to simulate the dynamics of the atmosphere (Polo *et al.*, 2019).

Satellite imagery is critical given the context of availability /quality of information, spatial coverage, and acquisition frequency. Several studies use satellite images to forecast surface solar irradiance. Widely studied approaches, such as those proposed by Hammer *et al.* (1999) and Escrig *et al.* (2013), use consecutive satellite images to extract information on the vector field of motion in the cloud cover pattern. Despite the methodology's limitations (such as its disregard for physical processes such as cloud formation and dissipation), it still provides the basic structure of the models used in recent generations. With the increased computational capacity to handle and analyze large amounts of data in the last decade, machine-learning methodologies have become popular in forecasting applications for the energy sector (Marchesoni-Acland *et al.*, 2023). However, these purely statistical models (with no link to the physical processes at work) make identifying and understanding the sources of uncertainties in their forecasts is challenging.

The tropical atmosphere responds directly to the difference in surface heating, making cloud formation and dissipation processes more dynamic. Therefore, it remains to be seen whether these methods, consolidated in medium/high-latitude countries, show reliability and equivalent performance in regions close to the tropics and the equator. Verbois *et al.* (2018) combined the WRF physical model with statistical tools, such as principal components and linear regression, to produce surface solar irradiance forecasts a day in advance for Singapore. Lima *et al.* (2016) combined the WRF model with artificial neural networks over Northeastern Brazil, making forecasts up to 24 hours. Guarnieri *et al.* (2006) used the ETA model and artificial neural networks to reduce the bias of their forecasts for Southern Brazil over a daily time horizon. However, the literature on applying forecasting methods in tropical countries still needs to be enriched.

This study focuses on developing a short-term forecast model of surface solar irradiance in the Brazilian territory based on cloud motion vector (CMV) identified in satellite images using the optical flow methodology, which is based on the following premises: 1) the apparent speed of reflectance patterns can be directly identified from the movement of the surfaces in the image, remaining constant during such movement; 2) the image has a flat surface and uniform illumination and; 3) reflectance varies smoothly without spatial discontinuities in the image, making its brightness mathematically differentiable.

Urbich *et al.* (2018) offer a more in-depth reading of this approach. They used the optical flow algorithm over Germany to produce valid forecasts over a time horizon of up to two hours. A similar methodology has also been applied in countries such as Finland (Kallio-Mayers *et al.*, 2020) and the humid pampas (Aicardi *et al.*, 2022), an area that covers Argentina, southern Brazil, Uruguay, and Paraguay.

The present study describes a short-term irradiance forecast model (a time horizon of up to two hours) for four locations in Brazil using visible channel GOES-16 satellite images and an optical flow algorithm to identify the movement pattern of their cloud cover. Section 2 details the research methodology, including a description of climate characteristics at the ground measurement locations, the satellite and ground-truth databases, the pre-processing steps performed before using the optical flow algorithm, and the validation procedure to evaluate the forecast model performance. Section 3 shows the GHI forecasts, pointing out which locations show better results and comparing results with similar studies. Finally, section 4 outlines some conclusions and suggestions for improvement.

Methodology

Overview

Figure 1 illustrates the methodology steps taken in this research, with a strong emphasis on validation. We began by processing the raw data from the GOES-16 satellite Cloud Moisture Imagery (CMI) product, focusing on four locations within the Brazilian territory where ground measurement sites are part of the SONDA network (www.sonda.ccst.inpe.br/). To evaluate cloud coverage, we applied a hierarchical clustering algorithm to establish clear and overcast sky thresholds for each location. We then evaluated the effective cloud cover index for each image in all four locations. Next, we estimated the Cloud Motion Vectors (CMV) using an optical flow algorithm, providing the necessary information for forecasting cloud coverage over time horizons ranging from 10-15 minutes to 2 hours. The reliability of these forecasts was ensured through a comparison between the forecast outputs and observed satellite images (Validation 1) using the Costanza approach. The final step involved forecasting and validating Global Horizontal Irradiance (GHI) (Validation 2), comparing GHI forecasts with groundtruth data observed in each location.

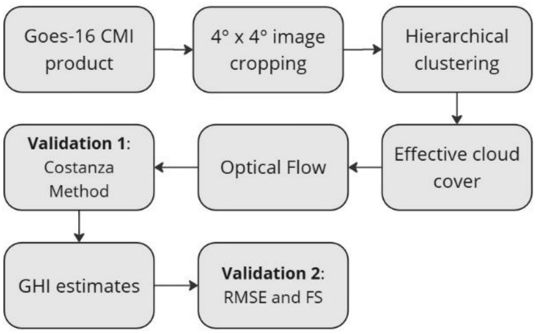

Figure 1. Methodology diagram applied in this study.

Satellite Data

The satellite images used in this study originate from the Cloud and Moisture Imagery, made available by GOES-16 in different spectral bands (including the one chosen for applying the methodology in this study: band 2, a visible band with a 0.5-km spatial resolution). The temporal resolution was 15 minutes before April 2019 and 10 minutes since then. In each measurement location, four months - January, April, July, and October - were used to forecast cloud motion in different atmospheric and seasonal climate conditions. Then, we applied the forecasted cloud coverage image into a radiative transfer model to understand the performance of the proposed CMV approach in forecasting the downward surface solar irradiance.

Groundtruth measurements

In total, four strategically chosen locations with distinct climate classifications (Alvares *et al*., 2013) were selected in different regions of Brazil. These locations, Brasília (Central region of Brazil), Cachoeira Paulista (Brazilian Southeastern region), Petrolina (Brazilian Northeastern region), and São Martinho da Serra (South of Brazil), were chosen for their unique climate conditions and the availability of surface solar irradiance data observed with class 1 pyranometer in the ground site measurements sites operating in the SONDA network. The GOES-16 satellite images used as the input parameter for the model were shaped into four by 4-degree boxes with a solarimetric surface station at their center (Figure 2).

Brasília (brb), marked as “1” in Figure 2, has a tropical climate with a dry winter. Maximum precipitation occurs in November and December, and the maximum incidence of solar radiation occurs in September and October. Similarly, Cachoeira Paulista (cpa) has a dry winter and rainy summer, with milder temperatures than Brasília in the winter. The rainfall regimes of both municipalities are significantly influenced by the South American monsoon, a regional climate system (Reboita *et al*., 2010; Pereira *et al*., 2017).

Petrolina (ptr) lies in the Brazilian semi-arid, resulting in considerably lower monthly rainfall rates. The climate characteristic in the measurement location and its latitude contribute to its high levels of global horizontal irradiance reaching the surface throughout the year, especially during the second half. São Martinho da Serra (sms) has a humid subtropical climate without a defined dry season; its monthly rainfall averages range from 100 to 200 mm. Its geographical position and the influence of the passage of frontal systems, which carry a more significant amount of cloud cover, render it the location with the lowest annual average of surface downward solar irradiance out of the other three locations (Pereira *et al*., 2017; de Albuquerque Cavalcanti *et al*., 2021).

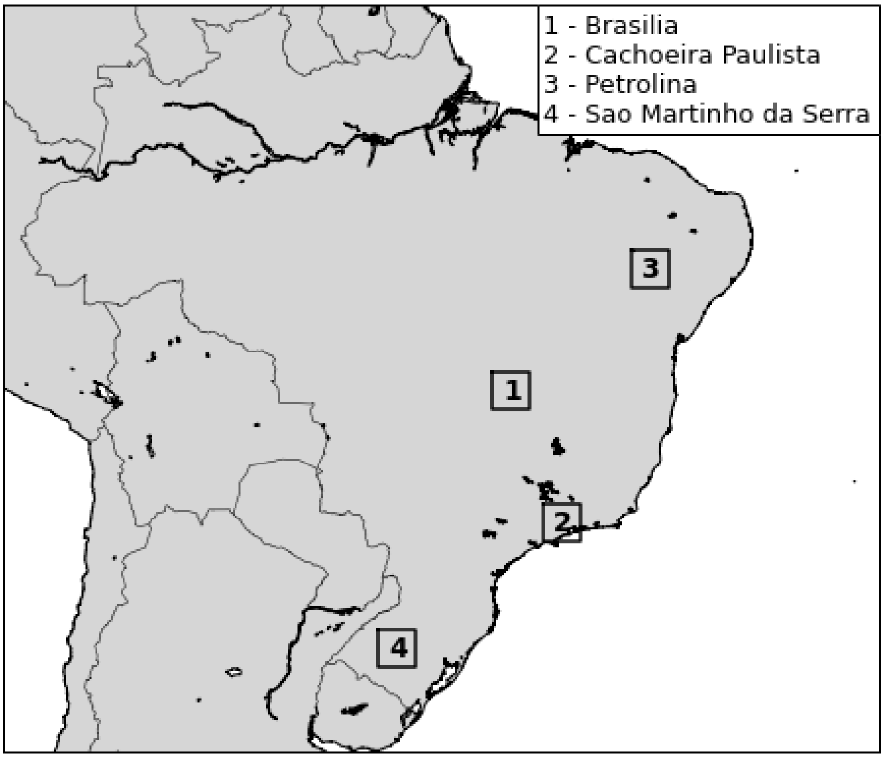

Figure 2. Location of the four ground measurement sites in Brazilian territory. The square boxes represent the 4° x 4° dimensions in GOES-16 satellite images around the ground measurement geographical coordinates.

Effective cloud cover evaluation

Effective cloud cover (C_{eff}) is a dimensionless index used by several numerical modeling approaches to forecast surface downward solar irradiance (Cano *et al.*, 1986; Lorenz *et al.*, 2003; Martins *et al.*, 2008a; Urbich *et al.*, 2018). It is based on the premise that cloud coverage over a given area constitutes the primary modulating factor of surface solar irradiance (Martins *et al.*, 2003). The C_{eff} index is calculated from the reflectivity of the clouds in satellite images.

$$C_{eff} = \frac{\rho - \rho_{clear}}{\rho_{cloud} - \rho_{clear}} \quad (1)$$

where ρ is the instantaneous reflectivity of the surface obtained from a satellite image, ρ_{clear} refers to the ρ value associated with the reflectivity of the surface in clear sky conditions, and ρ_{cloud} represents the ρ value observed by satellite in a completely overcast sky condition. Therefore, to determine the value of C_{eff}, it is necessary to establish the reference thresholds for both clear and overcast sky conditions. This was achieved by utilizing Ward's (1963) hierarchical clustering algorithm, which uses images captured daily at the same time within a month. The 30-day time window was selected because the geometry of the Sun-Earth system undergoes a slight change during this period, causing a minimal impact on the pixel reflectance.

The clustering methodology categorizes images into four groups: pixels in clear sky (no cloud cover), completely cloudy pixels (fully cloudy sky with high reflectivity clouds), and two groups of partially cloudy pixels: partial cloud cover and highly variable cloud cover, based on criteria proposed by da Rocha *et al.* (2022). Thus, the proposed categorization can identify the numerical thresholds that define the values of ρ_{clear} and ρ_{cloud} to calculate C_{eff}. Figure 3 shows a schematic diagram representing the determination of ρ_{clear} and ρ_{cloud} values at the four ground measurement locations. This process begins with a dataset consisting of a sample of 30 grayscale satellite images of each location, captured at the same time on different days.

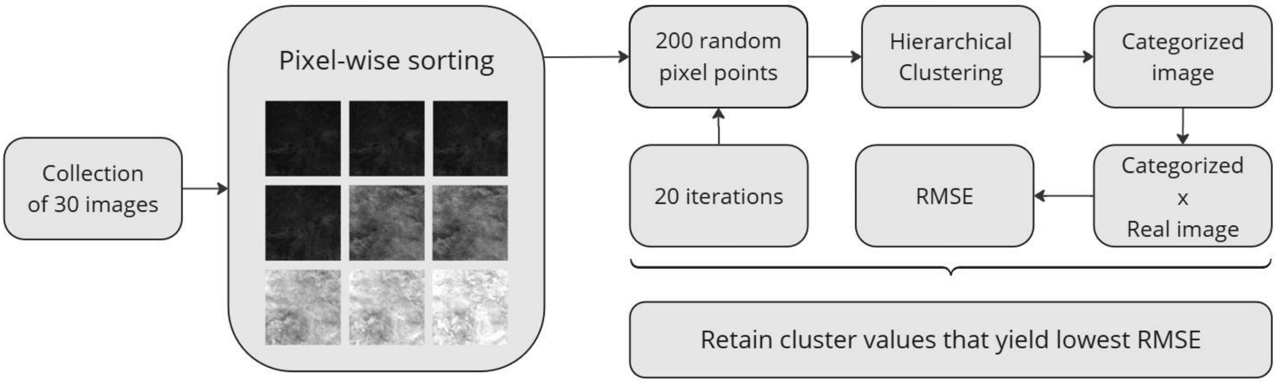

Figure 3. Clear and overcast sky thresholds with hierarchical clustering.

Each image is represented as an array of pixel intensity values, where zero ("0") corresponds to black (low reflectivity) and one ("1") to white (high reflectivity). To structure the dataset, the pixel values in each image are sorted in ascending order while preserving their spatial arrangement. This process results in a sequence of 30 images ordered from the least to the most reflective, effectively representing a gradient of cloud cover conditions.

To classify cloud cover, a random selection of 200 pixels is extracted from each image. Hierarchical clustering is applied to these sampled pixels, grouping them into the four aforementioned categories. The centroids of these clusters serve as threshold values, which are then used to categorize all pixels in the original images. Using these thresholds, the original satellite images are transformed by replacing each pixel with its corresponding cluster centroid value. This process effectively removes surface features, and the result is a categorized representation of cloud cover, isolating atmospheric features from the underlying surface. To assess the accuracy of this process, the categorized images are compared to the original grayscale images using pixel-wise root mean square error (RMSE). This process is repeated 20 times to optimize the classification thresholds, selecting the set of values that minimizes RMSE and provides the most accurate representation of cloud patterns. By systematically refining the threshold values, we ensure that the categorized images retain essential cloud features while minimizing discrepancies with the original dataset.

Optical flow algorithm

After obtaining the values for C_{eff} at each location at different times of the year, the images were used as an input parameter for the optical flow algorithm using the open-source Computer Vision library (Bradski, 2000; Bradski & Kaehler, 2008) (in Python) and Farnebäck's (2003) methodology. This strictly statistical approach ignores physical parameters such as wind shear, cloud types, and thermodynamic conditions in the atmospheric column as it generates forecast images. Before running the optical flow algorithm, it was necessary to convert the image information to C_{eff} values, as discussed in the previous topic.

Table 1 shows the input parameters for setting up the Farnebäck method. Different winsize and iteration values in two-hour forecasts and a time series with seven days of satellite imagery for each location were evaluated. The combination of parameters that generated the surface solar irradiance estimates with the lowest uncertainties was established as the final parameters' configuration of the algorithm. The choice of the seven days was adopted considering two aspects: the high computational cost of running the algorithm and achieving a statistical sample set representative of each location.

Table 1. Input parameters of the Farnebäck's optical flow algorithm.

Parameter	Location			
	brb	cpa	ptr	sms
Pyr_scale	0.5	0.5	0.5	0.5
levels	5	5	5	5
winsize	25	25	35	25
iterations	7	3	5	3
Poly_n	5	5	5	5
Poly_sigma	1.2	1.2	1.2	1.2

Validation 1: Costanza method

To verify the confidence of the forecasted cloud coverage generated by the optical flow algorithm, we utilized the method proposed by Costanza (1989) for evaluating the image pattern similarities between a forecasted cloud coverage image and a satellite-acquired image.

The validation procedure determines the fit index (F_t) described in equation (3). Firstly, we establish a sample window size, which starts at 1x1 and progressively expands to match the size of the original array representing the entire image. The following equation determines the fit of a particular sampling window F_W:

$$F_W = \frac{\sum_{s=1}^{t_w}\left[1-\frac{\sum_{i=1}^{p}|a_{1i}-a_{2i}|}{2w^2}\right]_s}{t_w} \quad (2)$$

in which F_W is the adjustment value of a *w*-size window; a_{ki} is the number of cells of a category *i* in a scene *k* in the sampling window; *p* is the number of categories in the sampling window; *s* is the *w* by *w* dimension sampling window, which moves across the scene one cell at a time; and t_w is the total number of sample windows in the scene in a *w*-sized window.

After adjusting the windows from size "1 by 1" to the original image dimensions, it is necessary to aggregate the fit values for all windows and obtain the value of the overall goodness of (F_t), as in the equation below:

$$F_t = \frac{\sum_{w=1}^{n} F_W e^{-k(w-1)}}{\sum_{w=1}^{n} e^{-k(w-1)}} \quad (3)$$

Equation 4, a balanced approach, is a weighted average of all the fits of several window sizes. It provides greater weight to smaller windows, while still considering the larger ones. The k-value, a constant, determines the weight of smaller windows relative to larger ones. If *k = 0*, all window sizes are equally weighted; if *k = 1*, only the smaller windows are significant. In our study, due to the relatively large image sizes, we chose a value of *k = 0.6*, which strikes a fair balance between smaller and larger windows.

Finally, the coefficient z determines the statistical significance of the result of validating the future image forecast:

$$z = \frac{F_t/\mu}{\sigma} \quad (4)$$

where μ and σ correspond to the mean and standard deviation of the F_t for 20 random scenarios, respectively. The random images required to estimate z are generated by shuffling the categories in the original scene, preserving the number of occurrences of each category; at least 20 random scenarios are required. The forecast images are considered statistically significant if z is greater than 10.

GHI estimates

The GHI forecast is based on Equation 5, which assumes a linear relationship between effective cloud cover and downward surface solar irradiance at the surface (Siqueira *et al.*, 2025; Martins *et al.*, 2003). The average C_{eff} in a 19 by 19-pixel area around the ground measurement sites was used to reduce geometric effects when comparing estimates at grid points with point values measured at the surface. This configuration produced the best results and closely resembled the geometry used in mapping solar irradiance on the surface of the Brazilian territory using images from the visible channel of the GOES-12 satellite (Martins *et al.*, 2008b).

$$GHI = G_0\{(\tau_{clear} - \tau_{cloud})(1 - C_{eff}) + \tau_{cloud}\} \quad (5)$$

where G_o is the downward solar irradiance at the top of the atmosphere, and τ_{clear} and τ_{cloud} are the atmospheric transmittances under clear and completely cloudy sky conditions obtained by solving the radiative transfer equation in 30 atmospheric layers and 135 spectral intervals, respectively.

The global surface solar irradiance was calculated using Equation 5, assuming a 0.05 for the cloud optical depth (τ_{cloud}). This calculation was based on solar irradiance mapping experience with the Brazil-SR model (Gonçalves *et al.*, 2020). The clear sky optical depth (τ_{clear}) was determined using the 'NREL' method available in the pvlib library in Python (Anderson *et al.*, 2023), based on the geographic coordinates of the solarimetric stations. Following the previous steps, the forecasted effective cloud cover (C_{eff}) was obtained from forecasted images. Surface downward global horizontal irradiance (GHI) estimates were made at three different times - 9 am, 12 pm, and 3 pm (Universal Time Clock) - for January, April, July, and October, resulting in 3,960 estimates for each location.

Validation 2: CMV-based Model validation

In this second validation step, we compared the GHI outputs from the CMV-based model with the forecasts provided by the persistence model using the Forecast Skill Score (FS) metric described in equation (6). The persistence forecast model assumes that the atmospheric transmittance remains constant for future horizons, and the surface downward solar irradiance (GHI_p) will be adjusted considering only the daily cycle of solar irradiance at the top of the atmosphere, as described in equation (7).

$$FS = 1 - \frac{nRMSE_{forecast}}{nRMSE_{persistence}} \quad (6)$$

where $nRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(GHI_{fs} - GHI_{obs})^2}}{\overline{GHI_{obs}}}$, GHI_{fs} and GHI_{obs} are the forecasted and observed GHI, respectively, and $\overline{GHI_{obs}}$ represents the average of observed GHI.

$$GHI_p(t+h) = \left[\frac{GHI_{obs}(t)}{G_0(t)}\right] G_0(t+h) \quad (7)$$

where h is the forecast horizon from 15 to 120 min ahead of the current time.

Availability of the computer codes and datasets

All numerical codes developed during the research were developed in Python and are available for public access in the *GITHUB* repository (https://github.com/rod-xavier/optical_flow). The database of downward surface solar irradiance collected at the four ground sites of the SONDA network can be accessed at http://sonda.ccst.inpe.br/. The GOES-16 satellite imagery can be downloaded from https://home.chpc.utah.edu/~u0553130/Brian_Blaylock/cgi-bin/goes16_download.cgi.

Results and discussion

Optical flow-based forecast images

In this section, we present the absolute deviation of the modeled C_{eff} in the images generated after applying the optical flow algorithm in two distinct scenarios. The first scenario exhibits a relatively low absolute deviation, while the second highlights the limitations of the chosen approach more prominently. Figure 4 displays the cloud pattern present in Cachoeira Paulista during the summer in the southern hemisphere, indicating that the optical flow algorithm is performing well, especially in the earliest time steps, with smaller and more isolated high-deviation areas. From t+105 onwards, these regions have become more widespread, indicating that errors accumulate over time, likely due to increasing uncertainties in cloud motion prediction.

The second scenario (Figure 5) displays the absolute deviation during the spring season in São Martinho da Serra. In this example, cloud formation processes are rapidly taking place due to atmospheric instability, which is typical during this time of the year. This, in turn, decreases the performance of the optical flow algorithm, especially from forecast horizons t+50 onwards.

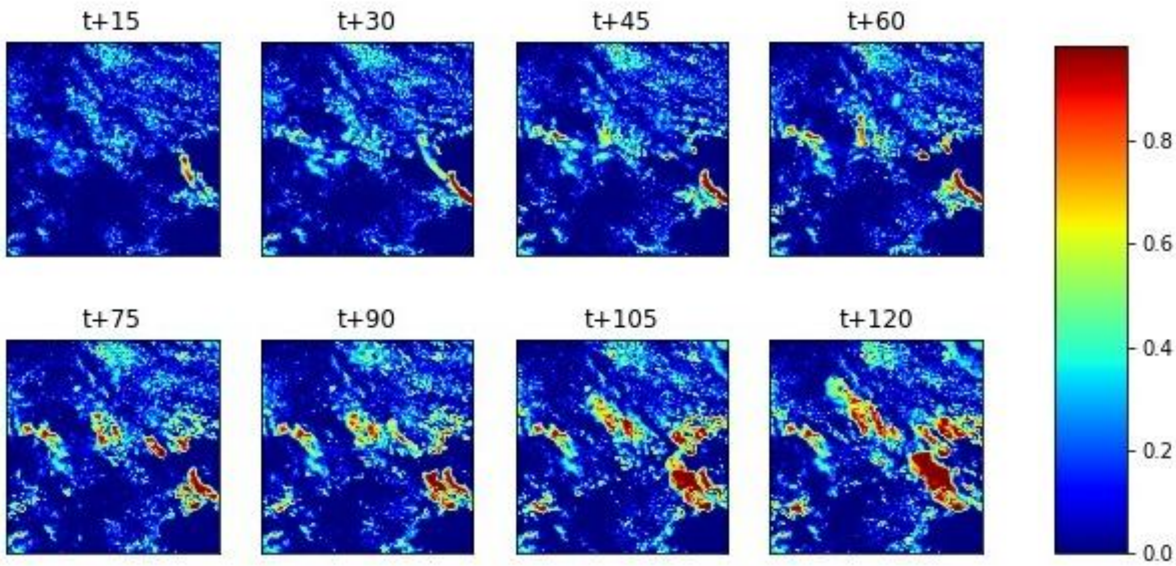

Figure 4: The absolute deviation (adimensional) of the effective cloud cover between forecast and observed images during the austral summer in Cachoeira Paulista.

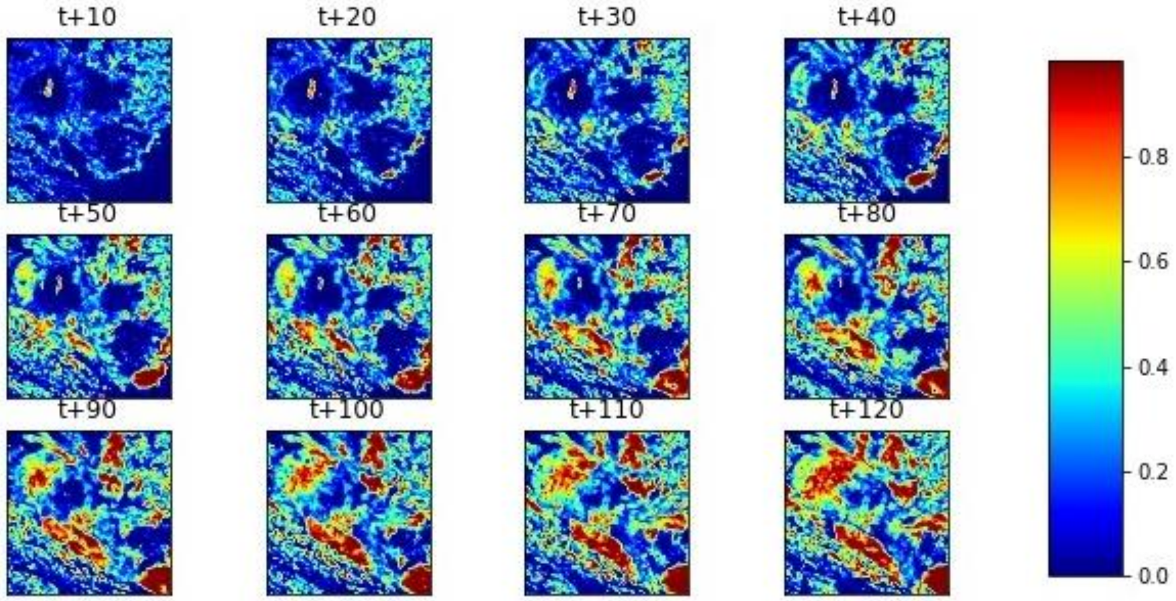

Figure 5: The absolute deviation (adimensional) of the effective cloud cover between forecast and actual images during the austral spring in São Martinho da Serra.

Costanza Method

Figure 6 shows the F_t values (y-axis) and statistical significance z (grayscale) for each location, grouped in dates according to the atmospheric conditions of the austral summer. The numbering from one to eight on the x-axis refers to each forecast horizon, starting at 15 min and ending at 120 min ahead. Timesteps with no values of F_t and z indicate a lack of observed satellite images.

During the summer season (Figure 6), much of the Brazilian territory is strongly influenced by convective rainfall, primarily driven by atmospheric instability. As a result, the behavior of F_t follows a similar pattern across all locations, meaning that the quality of the forecasted images tends to degrade at comparable rates. However, despite this decline in accuracy, the forecasts remain statistically significant throughout the entire forecast window ($z > 10$), indicating that the optical flow algorithm maintains useful predictive capability even as errors accumulate over time.

In contrast, the fall season (Figure 7) exhibits a different pattern. As the short rainy season ends in Petrolina, cloud cover becomes less frequent, leading to a more stable F_t value. This stability arises from the drier atmospheric conditions, which reduce the occurrence of cloud formation. However, convective activity remains more prevalent in the remaining locations, contributing to persistent cloud cover and increased atmospheric variability. This, in turn, negatively affects the performance of the optical flow algorithm, as it struggles to accurately predict the evolving cloud structures (cloud formation and dissipation).

During winter (Figure 8), the primary weather systems responsible for rainfall in all locations, except Petrolina, are cold fronts. Unlike convective systems, cold fronts are more organized and stable, making their movement easier to predict using optical flow. Since this approach is purely statistical, its effectiveness improves when dealing with well-defined, large-scale weather patterns. This effect is particularly evident in Cachoeira Paulista and São Martinho da Serra, where F_t values remain more stable over time compared to the other locations, suggesting a higher degree of predictability.

In the austral spring (Figure 9), atmospheric conditions exhibit characteristics of both summer and winter. For instance, in Brasília and Cachoeira Paulista, both frontal systems and convective activity can still be observed due to lingering atmospheric instability. This seasonal transition results in a mixed pattern, where both organized frontal movements and harder-to-predict convective cloud formations impact optical flow performance. The decrease in forecast accuracy is especially noticeable in Cachoeira Paulista and São

Martinho da Serra, where the presence of both weather regimes creates additional challenges for the algorithm, leading to greater variability in F_t values.

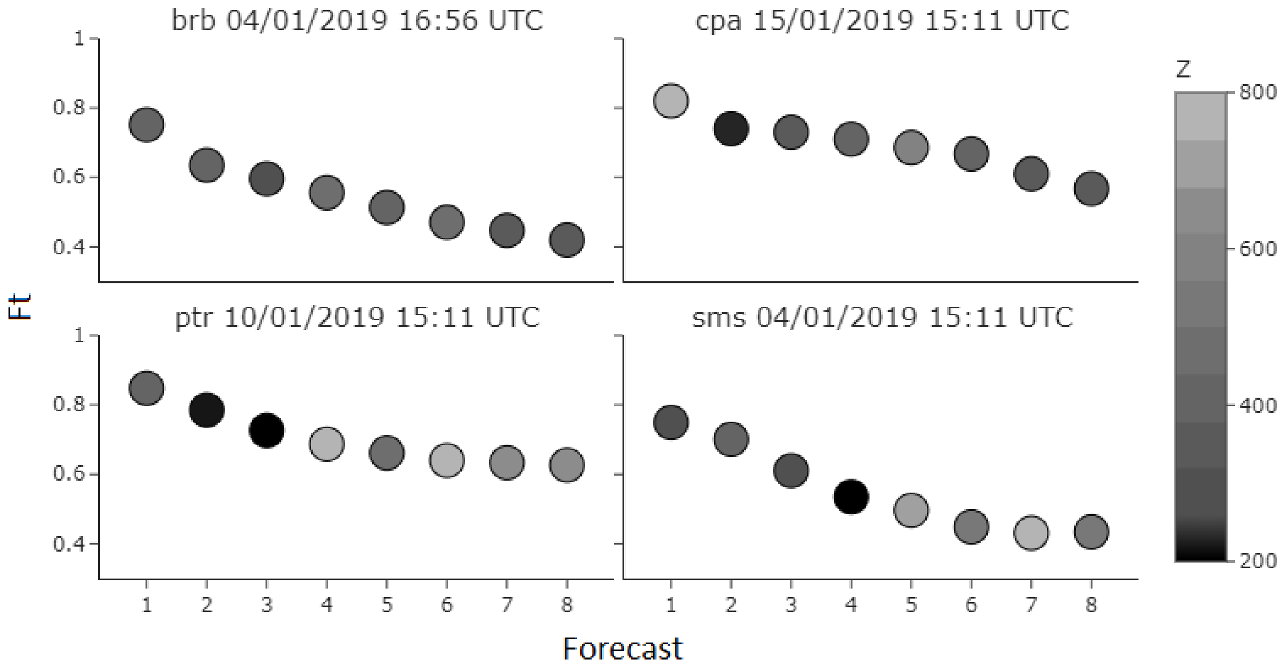

Figure 6.F_t values obtained from the Costanza method (y-axis) and its statistical significance, z (grayscale color bar), in the four ground measurement locations during summer. The values on the x-axis refer to the forecast horizons at 15-minute intervals, starting at min and ending at 120 min.

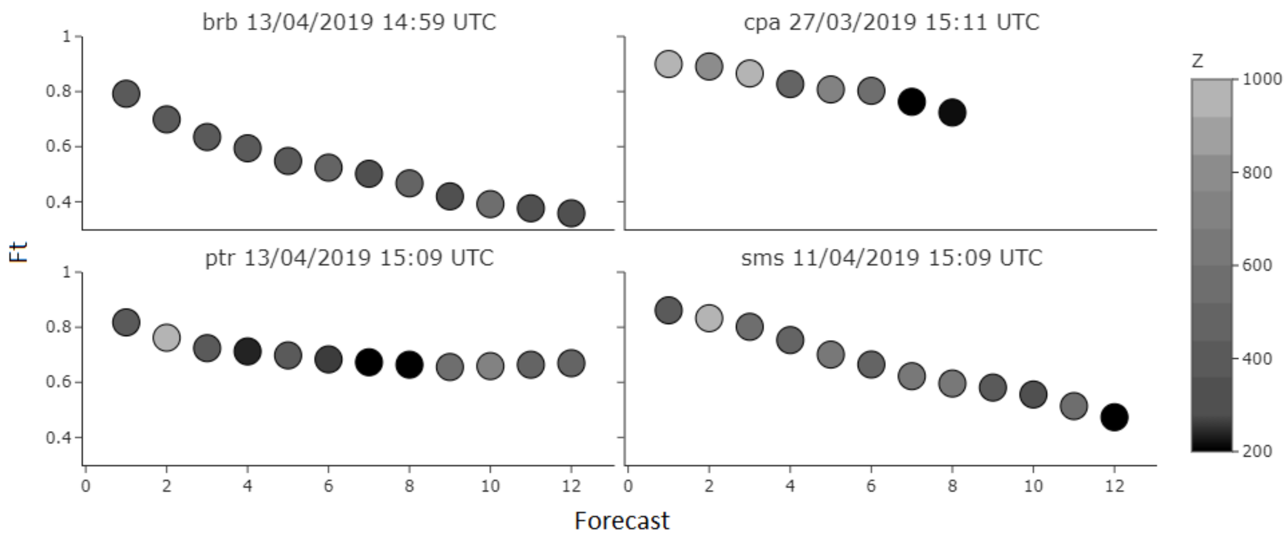

Figure 7. F_t values obtained from the Costanza method (y-axis) and its statistical significance, z (grayscale color bar), in the four ground measurement locations during fall. The values on the x-axis refer to the forecast horizons at 10-minute intervals, starting at min and ending at 120 min.

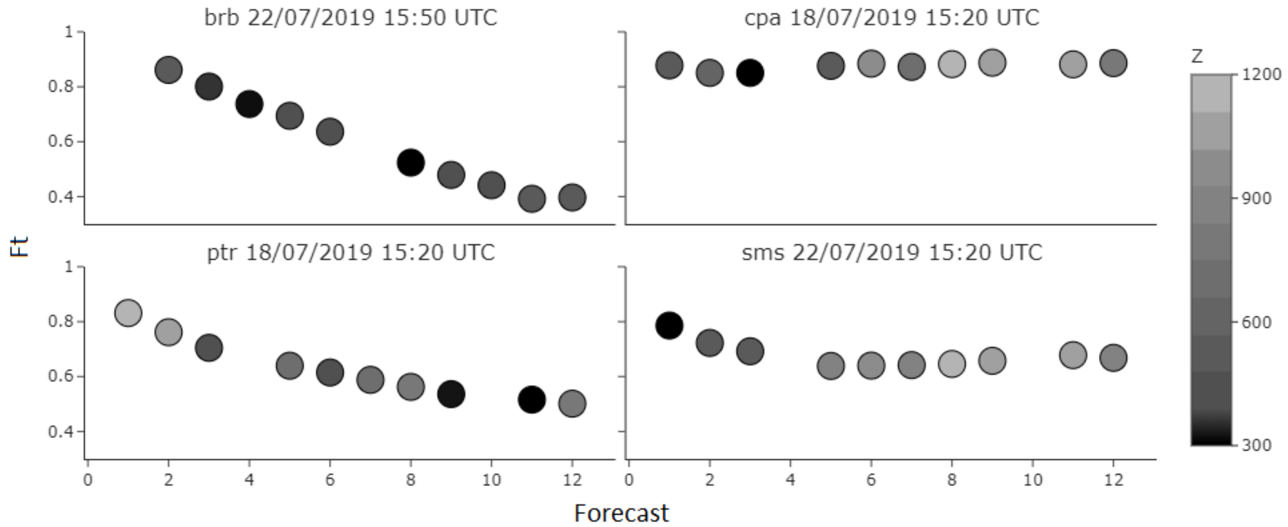

Figure 8. F_t values obtained by the Costanza method (y-axis) and its statistical significance, z (grayscale color bar), in the four ground measurement locations during winter. The values on the x-axis refer to the forecast horizons at 10-minute intervals, starting at min and ending at 120 min.

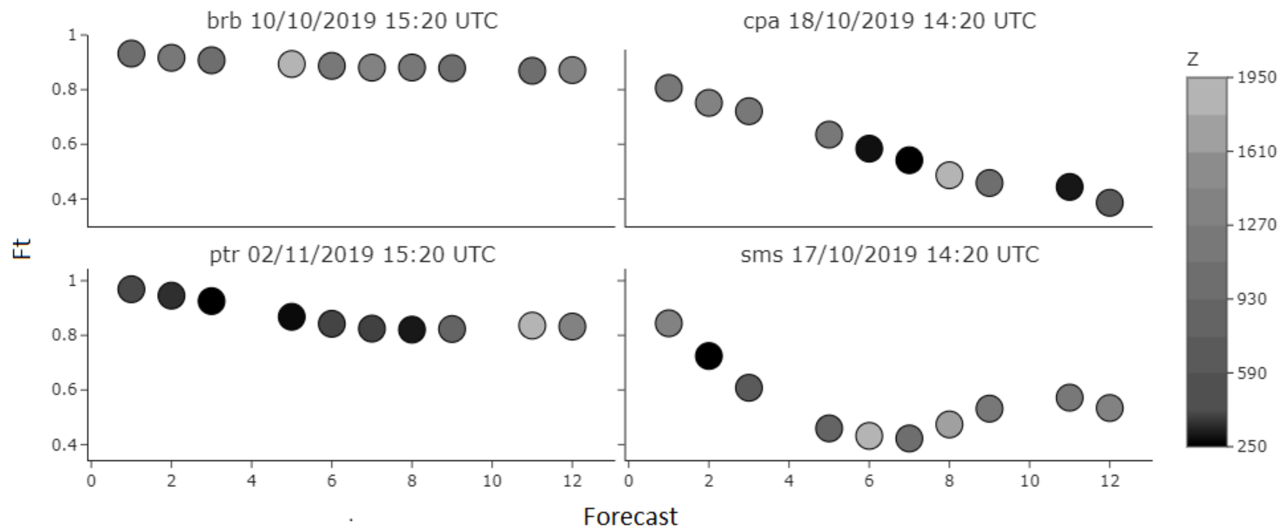

Figure 9. F_t values obtained by the Costanza method (y-axis) and its statistical significance, z (grayscale color bar), in the four ground measurement locations during winter. The values on the x-axis refer to the forecast horizons at 10-minute intervals, starting at min and ending at 120 min.

Forecast Skill Score

Figure 10 shows the evolution of the forecast skill score index, FS, for the predictions of GHI in the four locations. As no ground truth measurements are available for autumn and winter in Petrolina, this study could not validate its GHI forecast results for both seasonal periods.

For the Summer season, the performance of the forecast methodology based on CMV is lower than the persistence forecast for all sites except Petrolina, in which the model estimates achieve positive FS after about 45 minutes (the upper left plot in Figure 10).

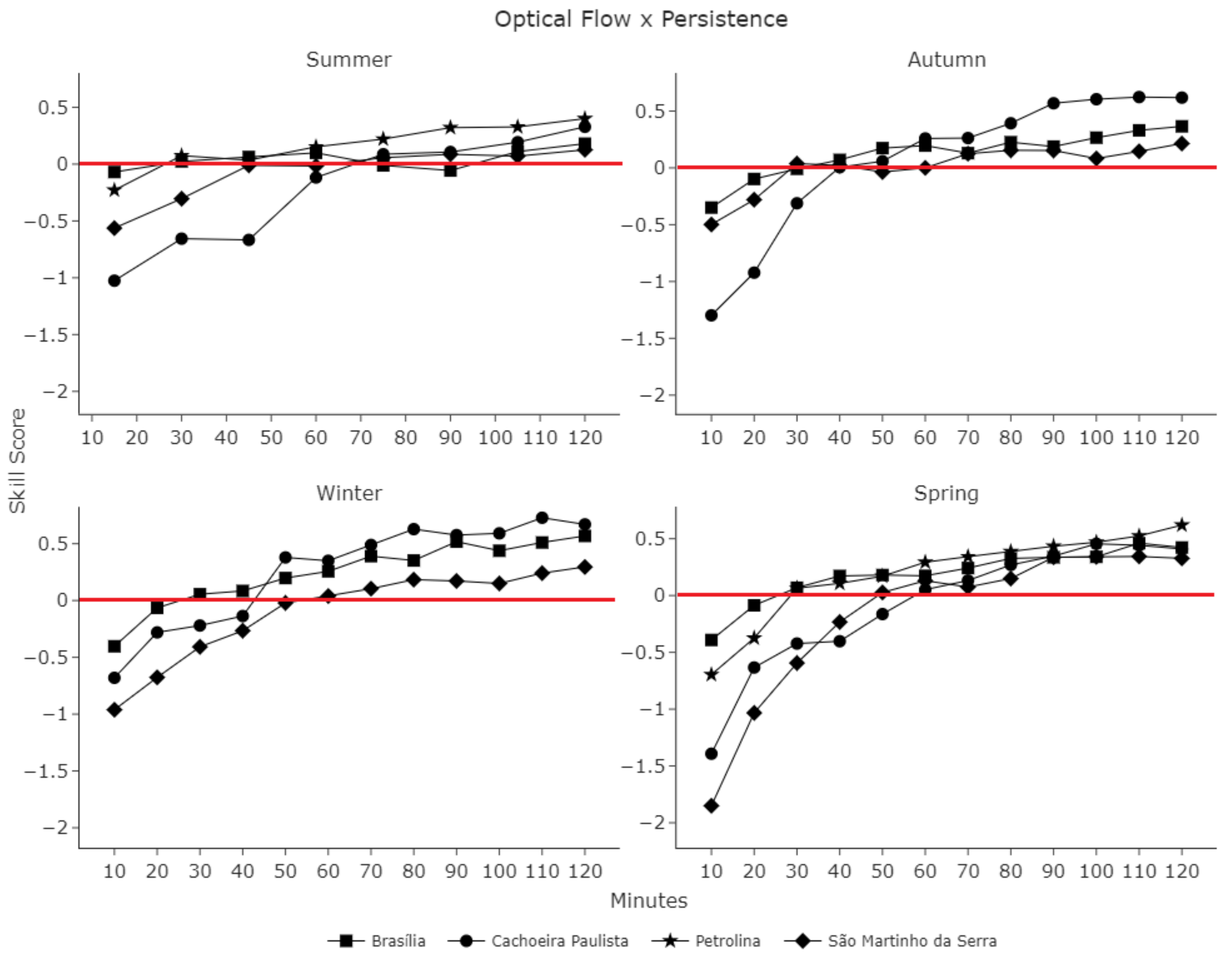

Figure 10: Skill Score seasonal plots (y-axis) on each forecast time step (x-axis) for ground measurement location to compare the CMV optical flow-based model and persistence-based model.

The persistence method also offers better results for the 10-min and 20-min forecast horizons in Autumn at Brasília and Cachoeira Paulista (see the upper right plot in Figure 10). However, subsequent forecast horizons show a better forecast of the CMV-based model. In São Martinho da Serra, the CMV-based model performed better for horizons longer than 1 hour, producing predictions with RMSE deviations 60% lower than the persistence model for 120 minutes ahead. Overall, the CMV-based forecast model performs poorer than the persistence method at all locations on the initial forecast horizons, outperforming the latter in later time steps.

The CMV-based model demonstrates its prowess in the winter season, as depicted in the lower left plot of Figure 10. It consistently outperforms persistence forecasts after approximately 50 minutes ahead in all locations. The proposed model also delivers satisfactory results for spring in all locations. However, as in the previous season, the persistence method initially outperforms the CMV-based model during the first forecast horizon.

Similar time behavior was found in recent studies. Yang et al. (2023) proposed a different approach using raw satellite data (area average of satellite images) as input in a Random Forest (RF) machine learning model to generate 10-minute timesteps forecasts over a 180-minute window for three locations in Northwest China. Their RF model yielded a positive FS for every timestep compared to a climatology and persistence model, ranging from 0.13 (10-min) to 0.29 (120-min). This RF-based approach improves the initial forecast time steps compared to our forecast model. However, the CMV-based model yields better results on more advanced timesteps, consistently reaching FS scores close to 0.5. It is important to consider that the climatology of the four study locations in Brazil and China is distinct, presenting different challenges for the performance of the proposed forecast methodologies.

Table 2 shows the average relative root mean square errors (rRMSE) for the GHI forecasts in 30, 60, 90, and 120 minutes ahead for different locations and seasons. The results support Figure 10, showing that the lowest forecast errors occur for more extended horizons (up to two hours ahead) during the winter for Brasília (rRMSE = 0.15%) and Cachoeira Paulista (rRMSE = 12%) and during the spring for Petrolina and São Martinho da Serra (rRMSE = 25% and 23% respectively). These findings align with previous research results. For instance, Urbich *et al.* (2018) found rRMSE values of 19% using Farnebäck's method in Germany for up to two hours of forecasting. On the other hand, the highest deviations tend to occur in the summer in all locations, with an rRMSE of 36% in São Martinho da Serra, consistent with similar studies. Kallio-Myers *et al.* (2020) and colleagues applied Farnebäck's algorithm in Finland, obtaining rRMSE values ranging from 29% to 53% for different locations over time horizons of up to two hours. Aicardi *et al.* (2022) used various methods for GHI forecasts over the South American humid pampa, with Farnebäck's algorithm showing deviations of 26% for forecasts of up to two hours.

Table 2. Average seasonal nRMSE of the CMV-based forecast model for all locations. The values are presented for forecast horizons in a 30-minute step.

	Summer				**Autumn**			
	brb	cpa	ptr	sms	brb	ptr	cpa	sms
t+30	33	22	36	36	33	-	26	39
t+60	28	26	34	38	31	-	21	37
t+90	32	22	31	37	41	-	18	40
t+120	31	25	28	36	37	-	20	41
	Winter				**Spring**			
t+30	11	13	-	28	12	26	14	28
t+60	12	12	-	26	15	22	15	20
t+90	11	11	-	31	20	23	18	20
t+120	15	12	-	32	22	25	16	23

Conclusions

This study produced 15,840 global horizontal irradiance estimations over four climatically distinct locations in the Brazilian territory at different times of the year. The estimations were generated using a CMV model, a key component of which is the optical flow methodology. This method, based on the equations of the Brazil-SR model, plays a crucial role in the accurate estimation of global horizontal irradiance. Results point to a clear difference in the forecast model's response across locations, with the model performing better in Brasília, Cachoeira Paulista, and Petrolina and obtaining similar results to the persistence method in São Martinho da Serra.

Most research on optical flow focuses on medium/high-latitude territories. The horizontal temperature gradient in these locations gives rise to jet streams (intense westerly winds in the upper troposphere), the evolution of which constitutes the main factor determining weather conditions; atmosphere-surface interactions play a secondary role in this scenario. Due to the weak horizontal temperature gradient in the tropical region (where most of the Brazilian territory is situated), the atmosphere is much more sensitive to external forces. Thus, it directly responds to differences in surface heating.

It is crucial to understand the limitations of the CMV-based model, particularly in convective cloud patterns. The optical flow method may have errors due to the formation and dissipation of clouds. When there are more convective clouds, like during the summer in Brasília and Cachoeira Paulista, the cloud cover pattern changes quickly. This makes the CMV-based model less effective than the persistence method, especially for short-term forecasts. However, when the weather is influenced by frontal systems, such as winter and spring, the model can provide accurate GHI estimates close to what is observed at surface measurement sites. The proposed model also showed reliable performance for Petrolina for predictions for horizons from one to 2 hours ahead, likely due to its drier climate and the cloud coverage pattern caused by the Intertropical Convergence Zone during the rainy season. Acknowledging these limitations is the first step towards further research and improvement in the field.

This study's unique focus on applying the optical flow method in Brazil significantly contributes to the field. It not only provides valuable insights but also opens up new avenues for research and development. Future research could explore alternative solutions to achieve superior results, particularly in forecasting the cloud movement vector in satellite images over Brazilian territory. This potential for improvement encourages further exploration and development in the field, especially in the initial forecast moments where the developed model shows the lowest performance. It's an exciting time for the field, with plenty of opportunities for innovation and advancement.

Acknowledgments

The authors would like to thank the LabREN/INPE researchers for operating and coordinating the Sistema de Organização Nacional de Dados Ambientais (SONDA – National Environmental Data Organization System) and making available the data from solarimetric stations, without which this study would be impossible. The authors also thank CNPq for funding them with scholarships (process 133091/2020-1). The authors would also like to thank PPG-ICTMAR and Unifesp for their support and infrastructure for developing this research.

References

Aicardi, D., Musé, P., & 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. DOI: 10.1016/j.solener.2021.12.066.

Alvares, C. A., Stape, J. L., Sentelhas, P. C., Gonçalves, J. D. M., & Sparovek, G. (2013). Köppen's climate classification map for Brazil. *Meteorologische zeitschrift*, *22*(6), 711-728. DOI: 10.1127/0941-2948/2013/0507.

Anderson, K. S., Hansen, C. W., Holmgren, W. F., Jensen, A. R., Mikofski, M. A., & Driesse, A. (2023). pvlib python: 2023 project update. *Journal of Open Source Software*, *8*(92), 5994. DOI: 10.21105/joss.05994.

Arrais, J. M., Martins, B. J., Chaves, T. Z. L., Cerentini, A., Netto, S. L. M., & von Wangenheim, A. (2022). Systematic Literature Review on Ground-Based Cloud Tracking Methods for Nowcasting and Short-Term Forecasting. DOI: INCoD/LAPIX.11.2022.E.01

Benavides Cesar, L., Amaro e Silva, R., Manso Callejo, M. Á., & Cira, C. I. (2022). Review on spatio-temporal solar forecasting methods driven by in situ measurements or their combination with satellite and numerical weather prediction (NWP) estimates. *Energies*, *15*(12), 4341. DOI: 10.3390/en15124341.

Bradski, G.(2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2236121.

Bradski, G., & Kaehler, A. (2008). Learning OpenCV: Computer vision with the OpenCV library. " O'Reilly Media, Inc."

Change, C. (2022). Mitigating Climate Change. *Working Group III contribution to the sixth assessment report of the Intergovernmental Panel on Climate Change*.

Cheng, L., Zang, H., Wei, Z., Ding, T., & Sun, G. (2021). Solar power prediction based on satellite measurements–a graphical learning method for tracking cloud motion. *IEEE Transactions on Power Systems*, *37*(3), 2335-2345. DOI: 10.1109/TPWRS.2021.3119684.

Costanza, R. (1989). Model goodness of fit: a multiple resolution procedure. *Ecological modelling*, *47*(3-4), 199–215. DOI: 10.1016/0304-3800(89)90001-X.

da Rocha, V. R., Costa, R. S., Martins, F. R., Gonçalves, A. R., & Pereira, E. B. (2022). Variability index of solar resource based on data from surface and satellite. *Renewable Energy*, *201*, 354-378. DOI: 10.1016/j.renene.2022.11.101.

de Albuquerque Cavalcanti, I. F., & Ferreira, N. J. (2021). *Clima das regiões brasileiras e variabilidade climática*. Oficina de Textos.

de Lima, F. J. L., Gonçalves, A. R., Costa, R. S., Pes, M. P., Dos Santos, A. P. P., Orsini, J. A. M., ... & Martins, F. R. (2024). Reducing uncertainties of climate projections on solar energy resources in Brazil. *Scientific Reports*, *14*(1), 23586. DOI: 10.1038/s41598-024-23586-7.

de Lima, Francisco José Lopes et al. (2019). The seasonal variability and trends for the surface solar irradiation in northeastern region of Brazil. *Sustainable Energy Technologies and Assessments, 35*, 335-346. DOI: 10.1016/j.seta.2019.08.006.

Escrig, H., Batlles, F. J., Alonso, J., Baena, F. M., Bosch, J. L., Salbidegoitia, I. B., & Burgaleta, J. I. (2013). Cloud detection, classification and motion estimation using geostationary satellite imagery for cloud cover forecast. *Energy*, *55*, 853-859. DOI: 10.1016/j.energy.2013.04.037.

Farnebäck, G. (2003). Two-frame motion estimation based on polynomial expansion. In *Image Analysis: 13th Scandinavian Conference*, SCIA 2003 Halmstad, Sweden, June 29–July 2, 2003 Proceedings 13 (pp. 363-370). Springer Berlin Heidelberg. DOI: 10.1007/3-540-45103-X_50.

Gonçalves, A. R., Assireu, A. T., Martins, F. R., Casagrande, M. S. G., Mattos, E. V., Costa, R. S., Passos, R. B., Pereira, S. V., Pes, M. P., Lima, F.

J. L., & Pereira, E. B. (2020). Enhancement of Cloudless Skies Frequency over a Large Tropical Reservoir in Brazil. *Remote Sensing, 12(17),* 2793. DOI: 10.3390/rs12172793.

Guarnieri, R. A., Pereira, E. B., & Chou, S. C. (2006). Solar radiation forecast using artificial neural networks in South Brazil. *Proceedings of the 8th ICSHMO*, 24-28.

Guermoui, M., Melgani, F., Gairaa, K., & Mekhalfi, M. L. (2020). A comprehensive review of hybrid models for solar radiation forecasting. *Journal of Cleaner Production*, *258*, 120357. DOI: 10.1016/j.jclepro.2020.120357.

Hammer, Annette *et al*. (1999). Short-term forecasting of solar radiation: a statistical approach using satellite data. *Solar Energy, 67*, 139-150. DOI: 10.1016/S0038-092X(00)00038-4.

IPCC, (2022). *Climate Change 2022: Impacts, Adaptation, and Vulnerability.* Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (eds.)]. Cambridge University Press. Cambridge University Press, Cambridge, UK and New York, NY, USA, 3056 pp., DOI:10.1017/9781009325844.

Kallio-Myers, V., Riihelä, A., Lahtinen, P., & Lindfors, A. (2020). Global horizontal irradiance forecast for Finland based on geostationary weather satellite data. *Solar Energy*, *198*, 68-80. DOI: 10.1016/j.solener.2020.01.048.

Lima, F. J., Martins, F. R., Pereira, E. B., Lorenz, E., & Heinemann, D. (2016). Forecast for surface solar irradiance at the Brazilian Northeastern region using NWP model and artificial neural networks. *Renewable Energy*, *87*, 807-818. DOI: 10.1016/j.renene.2015.09.062.

Luiz, E. W., Martins, F. R., Gonçalves, A. R., & Pereira, E. B. (2018). Analysis of intra-day solar irradiance variability in different Brazilian climate zones. *Solar Energy*, *167*, 210-219. DOI: 10.1016/j.solener.2018.03.048,

Marchesoni-Acland, F., Herrera, A., Mozo, F., Camiruaga, I., Castro, A., & Alonso-Suárez, R. (2023). Deep learning methods for intra-day cloudiness prediction using geostationary satellite images in a solar forecasting framework. *Solar Energy*, *262*, 111820. DOI: 10.1016/j.solener.2023.111820.

Martins, F. R., M. P. Souza, and E. B. Pereira. (2003) Comparative study of satellite and ground techniques for cloud cover determination." *Advances in Space Research 32 (11*), 2275-2280. DOI: 10.1016/S0273-1177(03)90553-4.

Martins, F. R., Pereira, E. B., Silva, S. A. B., Abreu, S. L., & Colle, S. (2008). Solar energy scenarios in Brazil, Part one: Resource assessment. *Energy Policy, 36*(8), 2853-2864. DOI: 10.1016/j.enpol.2008.03.040.

Martins, F. R., Silva, S. A. B., Pereira, E. B., & Abreu, S. L. (2008). The influence of cloud cover index on the accuracy of solar irradiance model estimates. *Meteorology and Atmospheric Physics, 99(3),* 169-180. DOI: 10.1007/s00703-007-0276-3.

Olabi, A. G., & Abdelkareem, M. A. (2022). Renewable energy and climate change. *Renewable and Sustainable Energy Reviews*, *158*, 112111. DOI: 10.1016/j.rser.2022.112111.

Pereira, E. B., Martins, F. R., Gonçalves, A. R., Costa, R. S., De Lima, F. J. L., Rüther, R., ... & De Souza, J. G. (2017). Atlas Brasileiro de Energia Solar. 2ª Edição. São José dos Campos: INPE. DOI: 10.34024/9788517000893.

Polo, J., Martín-Pomares, L., & Sanfilippo, A. (Eds.) (2019). *Solar Resources Mapping: Fundamentals and Applications*. Springer. DOI: 10.1007/978-3-030-25504-5.

Rahimi, N., Park, S., Choi, W., Oh, B., Kim, S., Cho, Y. H., ... & Lee, D. (2023). A comprehensive review on ensemble solar power forecasting algorithms. *Journal of Electrical Engineering & Technology*, *18*(2), 719–733. DOI: 10.1007/s42835-022-01160-0.

Raupp, I., & Costa, F. (2021). Hydropower expansion planning in Brazil-environmental improvements. *Renewable and Sustainable Energy Reviews*, *152*, 111623. DOI: 10.1016/j.rser.2021.111623.

Reboita, M. S., Gan, M. A., Rocha, R. P. D., & Ambrizzi, T. (2010). Regimes de precipitação na América do Sul: uma revisão bibliográfica. *Revista Brasileira de Meteorologia*, *25*, 185-204. DOI: 10.1590/S0102-77862010000200004.

Rodríguez-Benítez, F. J., López-Cuesta, M., Arbizu-Barrena, C., Fernández-León, M. M., Pamos-Ureña, M. Á., Tovar-Pescador, J., ... & Pozo-Vázquez, D. (2021). Assessment of new solar radiation nowcasting methods based on sky-camera and satellite imagery.*Applied Energy*, *292*, 116838. DOI: 10.1016/j.apenergy.2021.116838.

Singla, P., Duhan, M., & Saroha, S. (2021). A comprehensive review and analysis of solar forecasting techniques. *Frontiers in Energy*, 1-37. DOI: 10.1007/s11708-021-0722-7.

de Siqueira, R. A., Gonçalves, A. R., Costa, R. S., & Martins, F. R. (2025). Sensitivity analysis of solar irradiance estimates over a tropical region by cloud index models. *Solar Energy*, *287*, 113176.DOI: 10.1016/j.solener.2023.113176.

Urbich, I., Bendix, J., & Müller, R. (2018). A novel approach for the short-term forecast of the effective cloud albedo. *Remote Sensing*, *10*(6), 955. DOI: 10.3390/rs10060955.

Verbois, H., Huva, R., Rusydi, A., & Walsh, W. (2018). Solar irradiance forecasting in the tropics using numerical weather prediction and statistical learning. *Solar Energy*, *162*, 265-277. DOI: 10.1016/j.solener.2018.01.059.

Yang, L., Gao, X., Li, Z., & Jia, D. (2023). Intra-day solar irradiation forecast using machine learning with satellite data. *Sustainable Energy, Grids and Networks*, *36*, 101212. DOI: 10.1016/j.segan.2023.101212.

Ward, J. H. (1963). “Hierarchical Grouping to Optimize an Objective Function”, *Journal of the American Statistical Association, 58,* 236–244. DOI: 10.1080/01621459.1963.10500845.