

Article

Development of a Low-Cost Data Acquisition System for Very Short-Term Photovoltaic Power Forecasting

Guilherme Fonseca Bassous *, Rodrigo Flora Calili  and Carlos Hall Barbosa 

Graduate Programme in Metrology, Pontifical Catholic University of Rio de Janeiro—PUC-Rio, Rio de Janeiro 22451-900, Brazil; calili@puc-rio.br (R.F.C.); hall@puc-rio.br (C.H.B.)

* Correspondence: gfbassous@gmail.com

Abstract: The rising adoption of renewable energy sources means we must turn our eyes to limitations in traditional energy systems. Intermittency, if left unaddressed, may lead to several power-quality and energy-efficiency issues. The objective of this work is to develop a working tool to support photovoltaic energy forecast models for real-time operation applications. The current paradigm of intra-hour solar-power forecasting is to use image-based approaches to predict the state of cloud composition for short time horizons. Since the objective of intra-minute forecasting is to address high-frequency intermittency, data must provide information on and surrounding these events. For that purpose, acquisition by exception was chosen as the guiding principle. The system performs power measurements at 1 Hz frequency, and whenever it detects variations over a certain threshold, it saves the data 10 s before and 4 s after the detection point. A multilayer perceptron neural network was used to determine its relevance to the forecasting problem. With a thorough selection of attributes and network structures, the results show very low error with R^2 greater than 0.93 for both input variables tested with a time horizon of 60 s. In conclusion, the data provided by the acquisition system yielded relevant information for forecasts up to 60 s ahead.

Keywords: solar energy; neural networks; sky-camera; forecasting; renewable energy; energy quality; multilayer perceptron; computer vision; short-term forecasting; metrology



Citation: Bassous, G.F.; Calili, R.F.; Barbosa, C.H. Development of a Low-Cost Data Acquisition System for Very Short-Term Photovoltaic Power Forecasting. *Energies* **2021**, *14*, 6075. <https://doi.org/10.3390/en14196075>

Academic Editors: Marcin Kamiński and Angel A. Juan

Received: 19 August 2021
Accepted: 18 September 2021
Published: 24 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the past few decades, the world has experienced considerable growth in environmental awareness, especially regarding climate changes. This rise, allied with an ever-increasing population and limitations to fossil fuels, stimulates the development of Renewable Energy Systems (RES). To reduce greenhouse gas emissions, energy matrices must be composed of more low-carbon sources as opposed to the current fossil-reliant paradigm. Solar photovoltaic (PV) and wind are the future of energy systems if the world is to meet the goals set by the Paris Agreement [1,2].

In addition to the climate-specific Paris Agreement, the 2030 Agenda for Sustainable Development [3] proposes 17 general sustainable development goals with 169 associated global targets. Of these goals, the access to affordable and reliable renewable energy sources is directly aligned with three goals: 7—to ensure access to affordable, reliable, sustainable and modern energy for all; 9—to build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation; and 12—to ensure sustainable consumption and production patterns.

However, each of the mentioned energy sources has its own limitations, such as geographical location and unreliability, mostly regarding weather. In the case of solar energy, particularly PV energy conversion to produce electricity, it possesses high variability from various sources (e.g., weather, Earth's rotation and translation movements).

Solar energy's inherent intermittency creates several economical, technical and political barriers against larger penetration [4–6]. Most of the variability components are

deterministic in nature, meaning that they can be easily forecasted and addressed, provided that it is technically possible to do so.

One of the most detrimental variability components is the presence of clouds, which filter the solar radiation and decrease the amount of energy available for photovoltaic conversion. Particularly on days with partial cloudiness and fast-moving clouds, the insolation variation in one solar plant output can reach well over 50% in one minute [7,8]. These fast variations in such a short time may cause technical problems in plant and grid operation, such as voltage variations and current harmonics [5,6,9–11]. To address these variations, it is necessary to forecast them. In the work reported in Reference [8], the need for power system operators to be able to address generation and load profiles over short time-scales was stressed due to the stochastic variations caused by fast cloud transients. Numerous methods for short-term insolation or power forecasting exist; however, for plant and grid operation, conventional statistic forecasting methods based on time series are not well suited [12]. The most widely used physical methods for short-term predictions are sky-image based [13].

Tropical countries, in general, boast higher solar photovoltaic potential in comparison with temperate regions, and, in contrast, these countries also possess lower development indexes. That leads to increased difficulty in acquiring specific equipment for conducting research in solar modeling and forecasting, even more so if such equipment is necessary for implementing mitigation strategies. Figure 1 shows the discrepancy in solar resource availability and, in contrast, its utilization. The red scale represents daily average global tilted irradiance (GTI) [14], and the yellow sun symbols represent the installed capacity normalized by a country's area. In some areas, such as Europe and Central America, only the most important solar generators were kept on the map for clearer data representation.

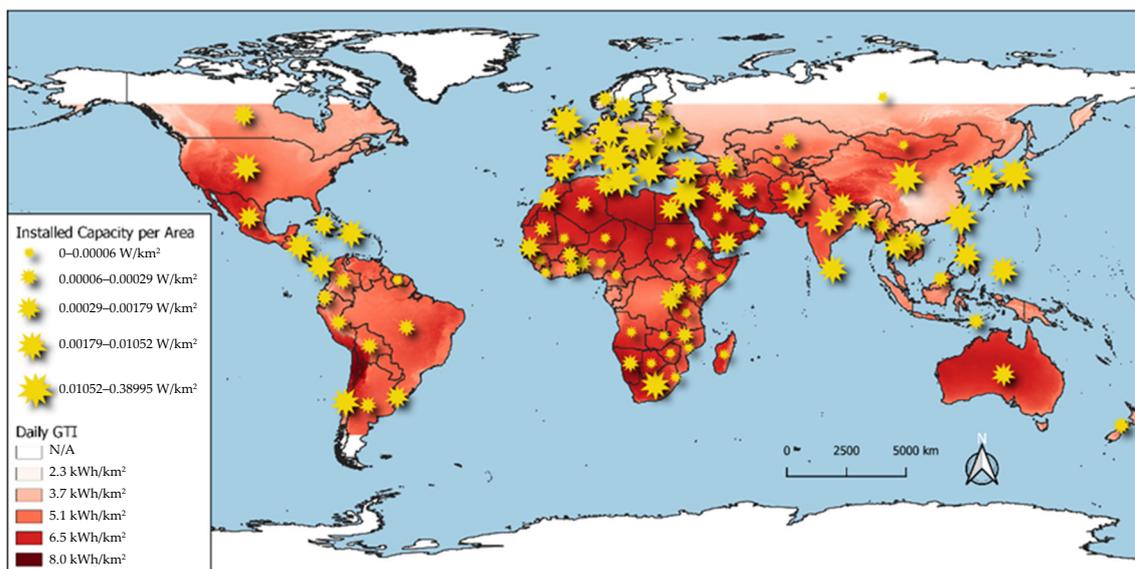


Figure 1. Daily GTI and installed capacity per country [15,16].

The first discrepancy between solar potential and utilization is clearly between African and European countries. Despite having two to three times more average daily global tilted irradiance, most African countries have a couple orders of magnitude less PV installed capacity. Another interesting comparison can be made between Mexico and the United States, because, despite being in the same continent, both have very different development levels, and that is more correlated to the installed capacity than solar resource availability. A similar comparison can be made between Brazil and Uruguay, Morocco and South Africa, and Spain and the United Kingdom. This makes clear the necessity for lower cost equipment, because, by reducing cost barriers, these countries can look to solar energy infrastructure to support their industrial development. Addressing these discrepancies has

been recognized as an important step in achieving the sustainable development goals set for 2030 [3,17].

Aside from the solar resource availability, forecasting is essential to energy generation and distribution. As mentioned in Reference [8], system operators need better information about the stochastic behavior of cloud-induced variability, to increase reliability. Several time horizons and resolutions are necessary to meet the demands of each specific aspect in PV energy management. The focus of this work is on very short-term forecasting to bolster PV plant operation capabilities, reliability, grid integration and grid operation in a scenario of high penetration.

In Reference [12], different irradiance forecasting methods are explored with the objective of proposing a small-scale insular grid forecasting system. Small isolated grids have less system inertia, therefore are more susceptible to the negative effects of RES, especially those caused by PV systems. Each different model available has its advantages and disadvantages and, for a holistic forecasting system, different models should be used in parallel.

Persistence and image-based models fit well, for short-term forecasts, in terms of horizon, frequency and spatial resolution. Other statistical models, as named in Reference [12], also encompass various regression models and learning algorithms, such as artificial neural networks (ANN).

In recent years there has been a rise in research work on sky-image based PV or insolation forecasting [18,19]. Sky-image models keep improving the reliability of very short-term forecasting, as shown in Reference [13]. This tendency points towards the superiority of using sky-images over what Diagne et al. [12] refer to as statistical models. In the study conducted by Kow et al. [20] it becomes apparent just how powerful sky-image based forecasting can be, achieving a detection rate of over 90% of power fluctuation events and mitigation of almost 80% of power fluctuation events with minimal energy loss.

While being a powerful tool, forecasting alone cannot solve the issues caused by high-frequency variability. However, coupled with other systems, such as energy storage systems and power electronics, especially in progressively smarter grids, forecasting can be a valuable aid in increasing PV penetration [9–11,21–24]. The results presented by Kow et al. [20] depict the beneficial effect that short-term forecasting can have on the operation of PV plants.

As some authors have shown, even lower-cost equipment can yield trustworthy results when comprehensively developed and tested [25]. This serves as encouragement for research institutions in developing and less-developed countries to work on their own equipment to provide their scientific and industrial needs.

Looking at the case for Brazil, which meets the criteria for solar resource abundance and developing economy, increasing accessibility to research equipment aligns with the country's goals set for the UN sustainable development goals prioritized for its 2030 agenda. Oliveira et al. [26] point out that Brazilian relay targets, highly influential as well as dependent within the agenda, can be directly impacted by increased affordability in solar power research. Goals such as resource efficiency, upgraded infrastructure, education and institutional capacity on climate change, and renewable energy depend on other goals, but also impact several others.

Reduction of costs associated with determinant goals such as research and development, innovation and economic growth have a high potential of impacting the relay goals previously mentioned [26]. More specifically, the affordability of newer renewable energy technology and their development align with the following targets: 7.1—"ensure universal access to affordable, reliable and modern energy services"; 7.2—"increase substantially the share of renewable energy in the global energy mix"; 7.3—"double the global rate of improvement in energy efficiency"; 7.b—"expand infrastructure and upgrade technology for supplying modern and sustainable energy services for all in developing countries, in particular least developed countries, small islands developing States and landlocked developing countries [. . .]"; 9.1—"develop quality, reliable, sustainable and resilient

infrastructure, including regional and transborder infrastructure, to support economic development and human well-being, with a focus on affordable and equitable access for all”; 9.2—“promote inclusive and sustainable industrialization [. . .]”; 9.5—enhance scientific research, upgrade the technological capabilities of industrial sectors in all countries in particular developing countries [. . .]”; 12.2—“achieve the sustainable management and efficient use of natural resources”; and 12.a—“support developing countries to strengthen their scientific and technological capacity to move towards more sustainable patterns of consumption and production”.

With these possible impacts in mind, the objective of this work is to present and validate a low-cost system for monitoring and modeling short-term variability developed during a Master’s course [27].

2. Short-Term Forecasting

As stated in the previous section, accurate very short-term forecasting is the first step in adding reliability to PV plant operation. The first step in forecasting is to build a model that describes the behavior of the studied phenomenon. To that end, many different models can describe or learn the behavior of PV conversion, some more accurately than others. Table 1 presents the terminology regarding forecasting horizons and their applications, based on the concepts used in References [12,28].

Table 1. Forecast horizon categories, granularity and applications.

Category	Time Horizon	Resolution	Applicability
Very short-term	Up to 15 min ahead	Up to 1 min	Plant operation Ramping events Power quality control
Short-term	15 min to 1 h ahead	1 to 5 min	Load following Grid operation planning
Medium-term	1 h to 6 h	Hourly	Load following Grid operation planning
Long-term	One day ahead	Hourly	Unit commitment Transmission scheduling Day ahead markets

Within the statistical category mentioned in Reference [12], persistence models are the best fit for the spatial and temporal requirements of very short-term forecasting for a single PV plant. However, it is a naïve predictor, serving as a baseline for more complex models. It assumes the predicted value \hat{X}_{t+1} to be best described by its value at a previous time X_t . In this case, the modeling and prediction are one and the same; it does not take into consideration the several variables that affect the behavior of real-world PV panels, and that is why it is considered a trivial predictor.

Still, within linear models, the regression models addressed by Diagne et al. [12] use historical data either from irradiance or clear-sky index to make predictions. While better than the previous, naïve predictors in terms of fidelity to the real world, it is still unable to provide forecasts in the required time horizon and resolution. These models, however, fare well from 15 min to hourly forecasts [29]. In the 5 min resolution, results were mixed among the models tested by Reikard [29], but the autoregressive integrated moving average (ARIMA) model started to be outperformed, especially by neural networks. The author also pointed that the ARIMA model exhibited large errors at intermittent intervals, corresponding to the fast cloud transients that deeply impact PV reliability. These intermittent large errors are the events successfully predicted in the work by Kow et al. [20].

Switching over to the non-linear models addressed by Diagne et al. [12], neural-network models attempt to simulate the computational and learning process of the human brain [30]. The complexity, nonlinearity and parallel computational power excel in pattern recognition and perception. The networks are composed of simple processing units commonly referred to as neurons. The network can acquire “knowledge” through a learning process that acts in the interconnection of the neurons, just as synapses would in a biological brain [30].

Neural networks, in their many architectures and sizes, are able to learn from data, in both supervised and non-supervised processes, and apply this knowledge to new data [30]. They are well suited to model complex problems, especially when involving complex relationships between the variables [30], such as forecasting energy conversion dependent on cloud passage, location, time and meteorological variables [31,32].

As mentioned previously, neural networks start faring better against other forecast methods at higher temporal resolution [29]; however, by looking at other studies into the subject, there appears to be a time-resolution limitation in these machine-learning methods for short-term forecasts. Even in the most recent state of the art works with intra-hour forecasting, using time series prediction of irradiance or other atmospheric parameters, the minimum resolution is still 5 min [33,34], which still falls short of the necessary frequency to properly characterize the local solar variability [35]. Still, within the 5 min time horizon, sky images can be used to boost forecasting accuracy when coupled with machine learning models and historic irradiance or power data [36].

The conclusion that can be drawn from the consistent number of time-series models limited to the 5-min time horizon is that the fault is in the type of data used to characterize the relationships involved in the high variability of solar irradiance. As explained before, these models aim to predict the future state of a certain aspect of solar variability. The approaches using cloud tracking in sky images, as proposed by Chow et al. [37] and Kow et al. [20], add components of physical and geometrical modeling of cloud systems. Since the main actor in short-term variability is related to passing clouds, relevant information on their dynamic provides a more comprehensive characterization [38].

The trend in researching sky-based approaches to very short-term solar forecasting began with the work by Chow et al. [37], despite not being the first to approach the subject [38]. The goal behind it is to use physical information from cloud systems, extracted from sky images captured by hemispheric cameras.

Initially, researchers used already existing sky imagers developed for meteorological purposes other than estimating solar quantities [38]. In more recent years, other lower-cost alternatives have been developed for the specific purpose of estimating solar quantities [25,38]. These newer, specific systems are fully programmable and expandable, leaving room for development and expansion, as well as being suitable for use with a plethora of different forecasting models [25,39].

Amongst the already mentioned advantages, specifically designed systems have proven to yield superior results to other non-specific sky imaging systems [25,40,41]. Most likely this superior performance is due to the higher data-acquisition frequency which provides better insight into local short-term solar variability [35]. Another significant difference is that these specific devices do not have a shadow band to occlude the solar disk and part of the circumsolar region. This fact positively impacts the amount of information available for intra-minute forecasts.

Throughout the research process that laid the theoretical foundations of this work, several key works stood out and greatly influenced the work developed here. Table 2 contains these important works in chronological order with their objectives, whether it is forecasting or modeling, and the materials and methods used in the pursuit of these objectives.

Table 2. Important works that shaped this research.

Work	Objective	Materials and Methods
Chow et al. (2011) [37]	Forecast of GHI from 30 s to 5 min ahead	Sky images obtained from a Total Sky Imager (TSI) every 30 s; Clear Sky Library (CSL) + Sunshine Parameter + Red-Blue Ratio (RBR) cloud classification; Cloud tracking through cross-correlation GHI deterministically calculated.
Gohari et al. (2013) [41]	Forecast of Clear Sky Index up to 15 min ahead in 30 s intervals	Comparison between TSI and UCSD-developed USI; Sky images every 30 s + irradiance measurements every second; Geometric cloud tracking; Solar ray tracing.
Chu et al. (2013) [42]	Forecast of 1-min-average DNI 5 min and 10 min ahead	TSI images every 20 s + DNI every 30 s; CLS + RBR adaptive threshold cloud classification; Cloudiness indices from gridded image + time lagged DNI as inputs for NN.
Marquez and Coimbra (2013) [43]	Forecast of 1-min-average DNI 3 min to 15 min ahead	TSI images every minute + 30 s averaged DNI; Cloud tracking, using Particle Image Velocimetry software; Hybrid threshold algorithm for cloud pixel classification; Grid of cloudiness indices used to deterministically calculate DNI.
Quesada-Ruiz et al. (2014) [44]	Forecast of 1-min-average DNI from 3 to 20 min ahead	TSI images every 20 s + 1 min averaged DNI; Hybrid threshold algorithm for cloud pixel classification; Cloud tracking, using grid cloud fraction change; DNI estimation, using grid cloud fraction.
West et al. (2014) [45]	Forecast of DNI from 0 to 20 min ahead in 10 s resolution and updated every 10 s	Sky images from internet protocol (IP) camera + DNI every 10 s; Cloud pixel detection, using NN; Cloud tracking through pixel-wise optical flow; Image regions averaged and total cloudiness as feature to be forecasted and derived into DNI.
Chu et al. (2015a) [46]	Forecast of 10 min ahead GHI and DNI	Images from 2 IP sky cameras every 60 s + irradiance every 30 s; Adaptive threshold cloud detection; Gridded cloudiness + time lagged irradiance as inputs for NN.
Alonso-Montesinos and Battles (2015) [47]	Modeling of GHI, DNI and DIF	TSI images every 60 s + GHI + DNI every 60 s; Correlations of digital image channels to model irradiance.
Alonso-Montesinos et al. (2015) [48]	Forecast of GHI, DNI and DIF from 1 to 180 min, at 15 min resolution	TSI images every 60 s; Cloud tracking, using cloud motion vectors (CMV); Pixel-wise cloud detection; Pixel-wise irradiance, using correlation of digital channel information.
Cazorla et al. (2015) [49]	Methodology for cloud detection	SONA sky imager + GHI + DIF; Multi-exposure (High Dynamic Range—HDR) images every 5 min; Adaptive RBR threshold method for cloud detection.

Table 2. Cont.

Work	Objective	Materials and Methods
Chu et al. (2015) [50]	Forecasting of prediction interval for 1-min-average DNI 5, 10, 15 and 20 min ahead	USI images provide parameters for hybrid model; Hybrid estimation/forecast model based on bootstrapped-ANN selected by SVM classifier, using mean RBR, RBR standard deviation and entropy + time-lagged DNI and DIF measurements as inputs; SVM for sky classification and model selection (high vs. low cloud-derived variability).
Chu et al. (2015b) [51]	Forecast of PV power 5, 10 and 15 min ahead	2 TSI providing images every 30 s; 3 methods as inputs for ANN reforecasting (deterministic based on cloud tracking, ARMA and kNN); Preliminary forecast by one of the 3 methods followed by reforecast, using ANN to enhance performance; Genetic algorithm to select ANN inputs; among several time-lagged power measurements and preliminary power forecasts for each of the horizons.
Lipperheide et al. (2015) [52]	Forecast of power ramp events 20 s to 180 s ahead with 20 s resolution	1 Hz power data from PV panels used in 4 different methods; Persistence and ramp persistence forecast based on detection from PV panels within plant; Cloud speed persistence forecast based on cloud motion vectors detected by PV panel power fluctuation; Second-order autoregressive forecast model based on the modified covariance method.
Pedro and Coimbra (2015) [53]	Forecast of GHI and DNI from 5 to 30 min ahead	5-min-averaged irradiance data; IP camera images every 60 s; Digital image channel individual information and relationships' properties, such as mean, standard deviation and entropy; kNN forecast model with images vs. without images vs. persistence.
Xu et al. (2015) [54]	Forecast of GHI from 1 to 15 min ahead	TSI images every 20 s; Complex cloud detection and tracking; Pixel-wise classification using RGB values, RBR and Laplacian of Gaussian (LoG); Cloud-type classification through texture metrics and kNN classifier; Comparison of persistence, linear regression and Support Vector Regression (SVR) with image inputs and NWP variables.
Cervantes et al. (2016) [39]	Forecast of 5 min ahead DNI negative ramp events	Low-cost sky-imager; Cloud detection through RBR; Cloud tracking with optical flow; Shadow mapping, using Cloud Base Height (CBH) data.
Mejia et al. (2016) [55]	Cloud optical depth modeling	2 USI providing images every 30 s; Estimation of irradiance from calibrated pixel values; Usage of deterministic models to obtain optical depth from digital image channels, solar position, pixel position and clear-sky library.

Table 2. Cont.

Work	Objective	Materials and Methods
Rana et al. (2016) [56]	Forecast of PV power from 5 to 60 min ahead, with 5 min resolution	5 min power average + meteorological data; Univariate (solely power measurements) vs. multivariate models NN ensemble vs. SVR vs. persistence.
Sanfilippo et al. (2016) [57]	Forecast of 1-min-average clearness index from 1 to 15 min ahead	GHI, DNI and DHI measurements every 60 s; Modeling of solar zenith-independent clearness index; SVR, persistence and autoregressive models of different orders used for forecasting.
Schmidt et al. (2016) [19]	Forecasts of GHI from 15 s to 25 min GHI forecasts in grid form for the surrounding area, updated every 15 s with 15 s resolution	Sky images every 15 s from custom imager + GHI every 1 s from 99 pyranometers + CBH measurements averaged over 10 min; Area of study of 10 km × 12 km; RBR with clear-sky images for cloud pixel classification; SVC cloud type classification from several features; CMV cloud tracking.
Soubdhan et al. (2016) [58]	Forecast of PV power and GHI 1, 5, 10, 30 and 60 min ahead	PV power data every 1 s + percentage cloud cover + ambient temperature + GHI every 1 s; Persistence and smart persistence baselines; Forecasting by Kalman filter with initialized parameters, using expectation-maximization (EM) algorithm vs. autoregressive (AR) estimation; Comparison between with and without exogenous inputs.
Ai et al. (2017) [59]	Forecast of 30-s-average GHI 1, 2, 3 min ahead	Sky images every 30 s from IP camera; SVM-determined clear-sky model; Adaptive threshold cloud detection; Optical flow cloud tracking; GHI deterministically determined, using cloud fraction and clear-sky model.
Blanc et al. (2017) [60]	Forecast of 1-min-average DNI map 15 min ahead with up to 10 m × 10 m spatial resolution	Stereoscopic IP sky cameras providing images every 30 s; CBH estimation from stereography; Cloud-layer CMV for each class of altitude; Estimation of projection-pixel-wise DNI, using beam clear-sky indexes computed per class of cloud combined with physical and geometrical information.
Cheng (2017) [61]	Detection of irradiance ramp down events 5, 10, 15 and 20 min ahead	Sky images every 60 s from Santa Barbara; Instrument Group + 1 min averaged GHI; Cloud detection and tracking through feature point clusters.
Elsinga and Van Sark (2017) [62]	Forecasts of 1 min average GHI from 1 to 30 min ahead for multiple sites	202 rooftop PV systems acting as a sensor grid; PV power data averaged every 1 min from inverter data every 2 s and then converted into GHI; Hourly interpolated ambient temperature deterministically calculated; GHI converted into clearness index Peer-to-Peer (P2P) forecasting method, using correlations between the rooftop PV systems to determine time lag between correlated sites.

Table 2. Cont.

Work	Objective	Materials and Methods
Ni et al. (2017) [63]	Forecast of power interval 5 min ahead	Ensemble of single-layer feed-forward NN (weights assigned, using a least-squares method in 1 step); Data from 3 kW micro-grid with 3 PV systems + photosynthetically active radiation + ambient temperature + relative humidity + wind speed + wind direction + GHI and precipitation (all averaged over 5 min).
Richardson et al. (2017) [25]	Forecast of GHI 10 and 15 min ahead	Images from a PiCamera; Cloud detection, using RBR; Optical flow cloud tracking; Ray tracing for GHI forecast, using a fixed ramp rate and clear sky GHI.
Kow et al. (2018) [20]	Forecast of PV power 30 s ahead coupled with mitigation system	GHI every 1 s + ambient temperature every 1 s and PV system modeled power; Self-organizing incremental neural network (M-SOINN) with active learning for forecasting power; Non-supervised method capable of forecasting power output of PV system 30 s ahead.
Kuhn et al. (2018) [64]	Forecast of 1-min-average GHI from 0 to 15 min ahead	Cloud segmentation, detection and georeferencing, using 4 sky cameras (WobaS-4cam) and 4-dimensional CSL; Irradiance maps validated with ground irradiance sensors and shadow camera; GHI and DNI obtained from geo-located shadow map and radiometer measurements at previous time steps.
Bouzgou and Gueymard (2019) [65]	Forecast of GHI from 5 min to 3 h ahead	Mutual information feature selection from time series of recent GHI; Extreme learning machine (ELM) for investigating the relationship between the historical variables and the future value, and also for determining the best combination of variables.
Kumler et al. (2019) [33]	Forecast of GHI 5, 15, 30 and 60 min ahead	Cloud albedo and fraction modeling based on GHI; Cloud optical thickness deterministically calculated; Forecast based on 5 min exponential weighed moving average of cloud fraction, used to determine albedo and GHI.

As seen in Reference [35], data resolutions of 30 s or less are essential for representing local solar variability. Moreover, most of the works presented in Table 2 do not meet this time resolution constraint, and those that do, possess higher cost systems often with multiple cameras and other sensors. As stated in Section 1, this limits the conduction of higher-resolution studies in less-developed countries, due to this fact, the goals of this work were to try and achieve high-resolution data acquisition and modeling, using low-cost equipment.

3. Materials and Methods

This section presents the hardware and methodology used for data collection, as well as the methods used for data analysis.

3.1. Data-Acquisition System

Image-based cloud tracking has become very popular for high-frequency photovoltaic modeling and short-term forecasting [13,37,45,54,66,67]. Cameras have massive variations of price depending on purpose, sensitivity, sturdiness, resolution and other various characteristics. The three main constraints for the camera used in this work were 180-degrees field of view to acquire total-sky images, enough resolution to provide all the information required for modeling the sky and the ability to be controlled by a low-cost embedded system. The system uses an ELP-USBFHD01M-L180 camera, which has a CMOS OV2710 sensor able to provide images with 1920 pixels by 1080 pixels resolution and is controlled directly via USB cable.

A 20 cm by 15 cm, 6 W solar panel was added to the data-acquisition system to provide a solar quantity to develop the image-based model. Power calculations were derived from voltage and current measurements provided by an Adafruit INA219 DC sensor. Another source of data was the panel temperature, provided by a Maxim Integrated DS18B20 temperature sensor attached to the back of the panel. Tying the system together, a Raspberry Pi 3B+ controls the camera and both sensors, synchronizing the data acquisition. Single board computers are ideal for this application due to their price, high tolerance to temperature variations and enough computing power for data acquisition in this scale.

In total, this equipment costs under US\$ 100.00, if compared with more traditional sky imagers, such as the Yankee TSI 880, which costs thousands of US dollars, or even some of the lower-cost equipment developed by some research institutions, using IP cameras that can reach hundreds of US dollars [25]. A sky imager expected to be used as part of control strategies for PV plant operation must be low in cost to be commercially attractive for investors.

A 3-dimensional rendering of the equipment is shown in Figure 2.

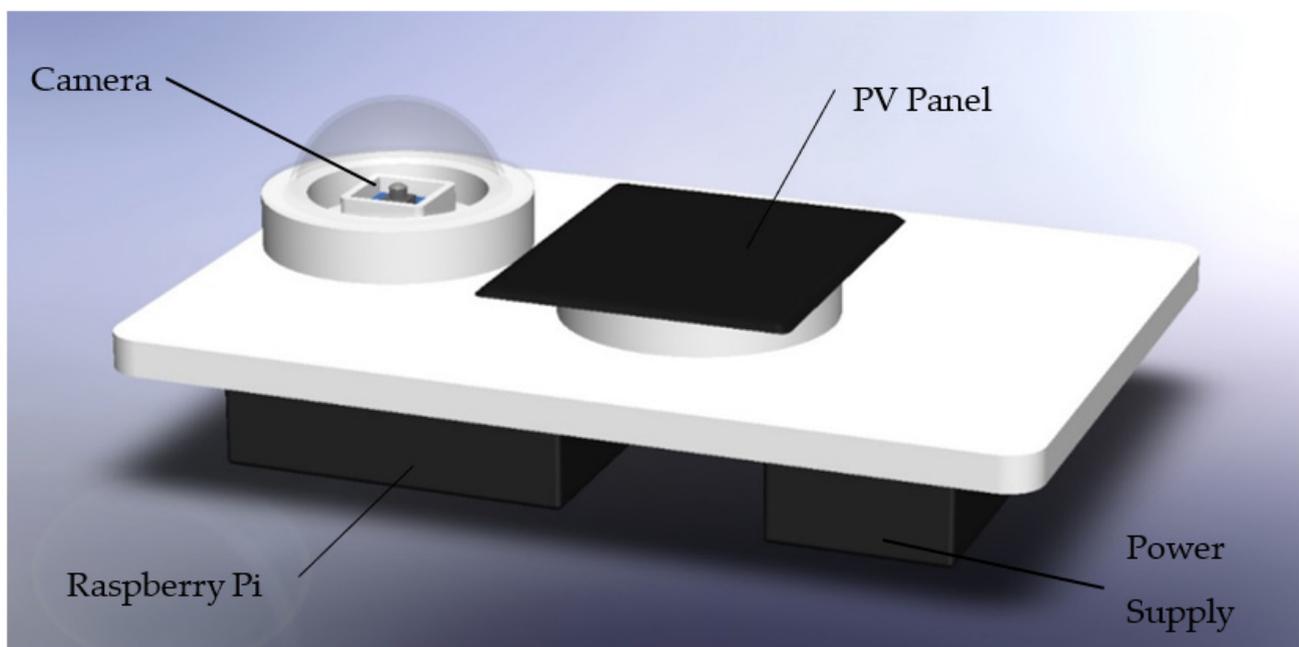


Figure 2. Three-dimensional model of the data-acquisition system.

The image acquisition software was developed using Python 3 with OpenCV 4 in conjunction the “pi_ina219” [68] and “w1thermsensor” [69] libraries for accessing the power and temperature sensors with Python.

To be able to generate power, the PV panel must be in a closed circuit with a load component. The initial goal was to use a ceramic resistor; however, during the testing

process, when higher currents were applied to the resistor, it started to overheat, so a dichroic light bulb was used instead.

The thermometer was placed under the PV panel enclosed by the fins from an aluminum heat exchanger pad with the flat part attached to the bottom of the panel. It was then covered by thick dense foam to act as a heat insulator between the thermometer and the environment. Both thermometer and heat exchange pad were assumed to possess higher heat-transfer coefficients than the panel and both have significantly less mass, meaning that they have lower thermal inertia. This causes the thermometer to quickly follow changes in panel temperature, which is a key variable in PV conversion efficiency [70].

As for the INA 219 sensor, it measures both circuit voltage and determines current by measuring voltage across a $0.1\ \Omega$ shunt resistor. It is capable of measuring voltages up to 26 V and currents up to 3.2 A at a maximum ADC resolution of 12 bit. Both sensors have well developed Python libraries for use with the Raspberry Pi, which will be presented in the next section, along with all the software components used by the DAS.

Both sensors are supplied by 3.3 V DC provided by the Raspberry Pi's 3V3 pin. The INA 219 communicates, via I2C protocol, with the Pi through the SDA and SCL pins, located on the GPIO2 and GPIO3 pins respectively. Voltage and current are measured between the V_+ connector and ground. The current enters the INA 219 through the V_+ connector, passes through the internal measurement circuit and exits through the V_- connector, then through the dichroic light bulb.

The DS18B20 uses the 1-Wire communication protocol through GPIO4 pin. It requires a pull-up resistor of $10\ \text{k}\Omega$ to stabilize the signal when not communicating with the Pi. Figure 3 presents the measurement circuit schematics for temperature, voltage and current measurements. The green lines indicate connected terminals, and the camera was not included in this schematic because it uses a simple USB connection.

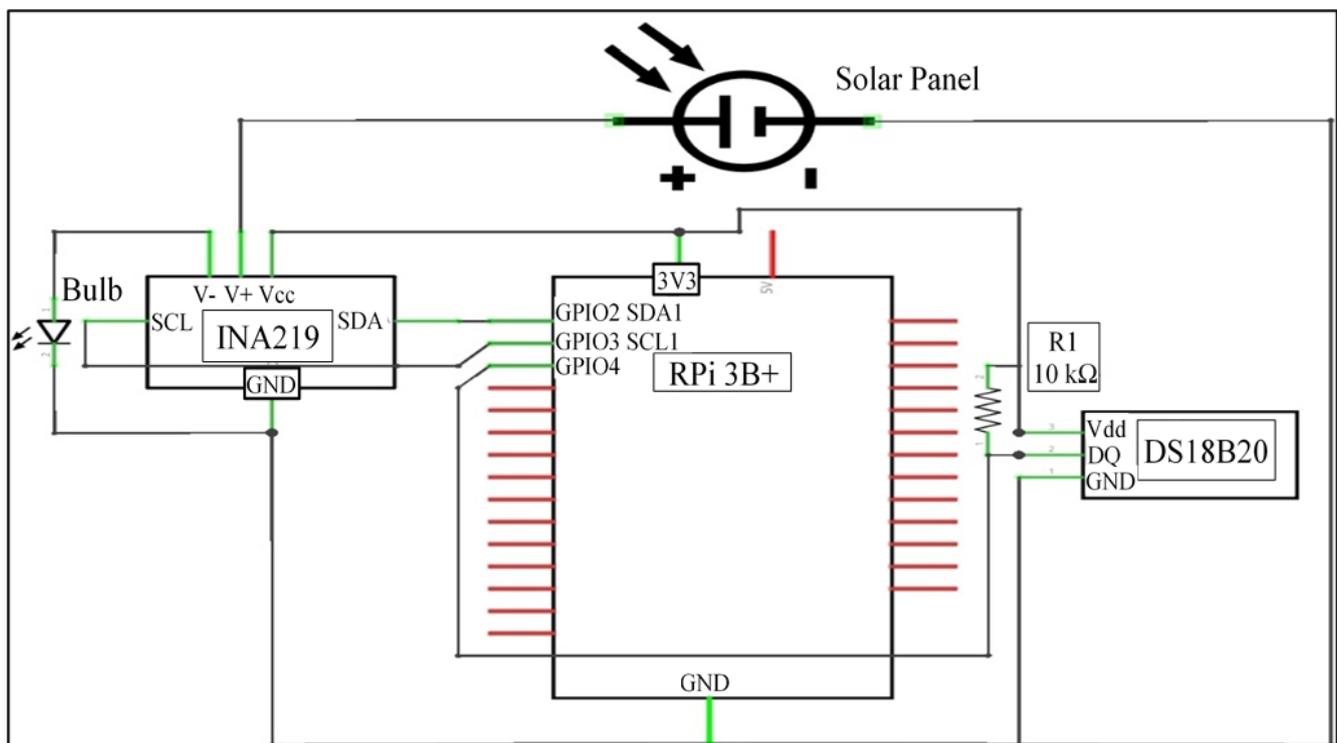


Figure 3. Measurement circuit schematic.

3.2. Acquisition Strategy

Due to the extremely high frequency of variations caused by cloud transients on PV power systems, the acquisition frequency must be high enough to provide information

on such variations [7,8,35]. An acquisition frequency of 1 Hz was chosen since it has been shown to provide important information on very short-term solar variability [35]. This approach is important to provide information surrounding fast ramp events; however, it generates so much data not pertaining to such events that it may hinder their study, especially when such a high volume of images must be analyzed.

To focus on the ramp events, the approach used is called acquisition by exception [71]. It consists of monitoring one or more variables of interest and only saving data when an event of interest is detected. In this case, whenever the power measurement would vary beyond a certain threshold, the system would save data pertaining to this event. In practice, the acquisition software continuously acquired data during the daytime at 1 Hz and temporarily stored this information using a queue structure (first in, first out). This queue had a maximum of 10 elements at a given time, and for every iteration where no variation event was detected, the oldest entry was deleted, making room for a new set of measurements. Each element was measured 1 s apart and was comprised of one sky image, one voltage and one current measurement as well as the calculated power from the PV panel.

In order to detect a variation event a moving average of the previous 3 power values—at t_{-3s} , t_{-2s} and t_{-1s} —are calculated and compared with the most recent value, t_0 . If there is a variation greater than a certain threshold, either up or down, the program enters the data-saving routine. It keeps acquiring data for 4 more seconds— t_{+1s} . . . t_{+4s} —and then it saves these 15 s worth of data, as well as one temperature measurement representative of this period. This structure of 15 s of measurements is henceforth referred to as an “event”. After recording an event, the system goes back into listening mode to detect other variation events.

The reason behind using only one temperature measurement is that, if the system were to include temperature measurements every time step, each iteration would take longer than 1 s, making it impossible to reach the desired 1 Hz acquisition frequency. Upon testing, this did not impact the quality of the data generated, due to the thermal inertia from the panel. Significant changes in panel temperature came at much lower frequencies than 1 Hz. The variation threshold was determined through experimentation and manual analysis of the quantities of interest. Using this strategy, it was possible to acquire data surrounding such ramp events, with images and power measurements taken 10 s before and 4 s after, totaling 15 s of data points per event. In the case of temperature, only one measurement was taken per detected event due to the sampling time from the sensor. Figure 4 presents a flowchart of the decision process and data flow from the data-acquisition system (DAS) software.

3.3. Measured Data

In the end of the data-collection phase, 500 event structures were recorded. Each event instance has 15 s worth of data saved into files: 15 JPG files containing the captured images and one text file containing the rest of the measured data. The text file is structured so every second of the event is one line (separated by a newline character “\n”), and the measured quantities are separated by commas. Table 3 shows an example of the numeric data measured during one event.

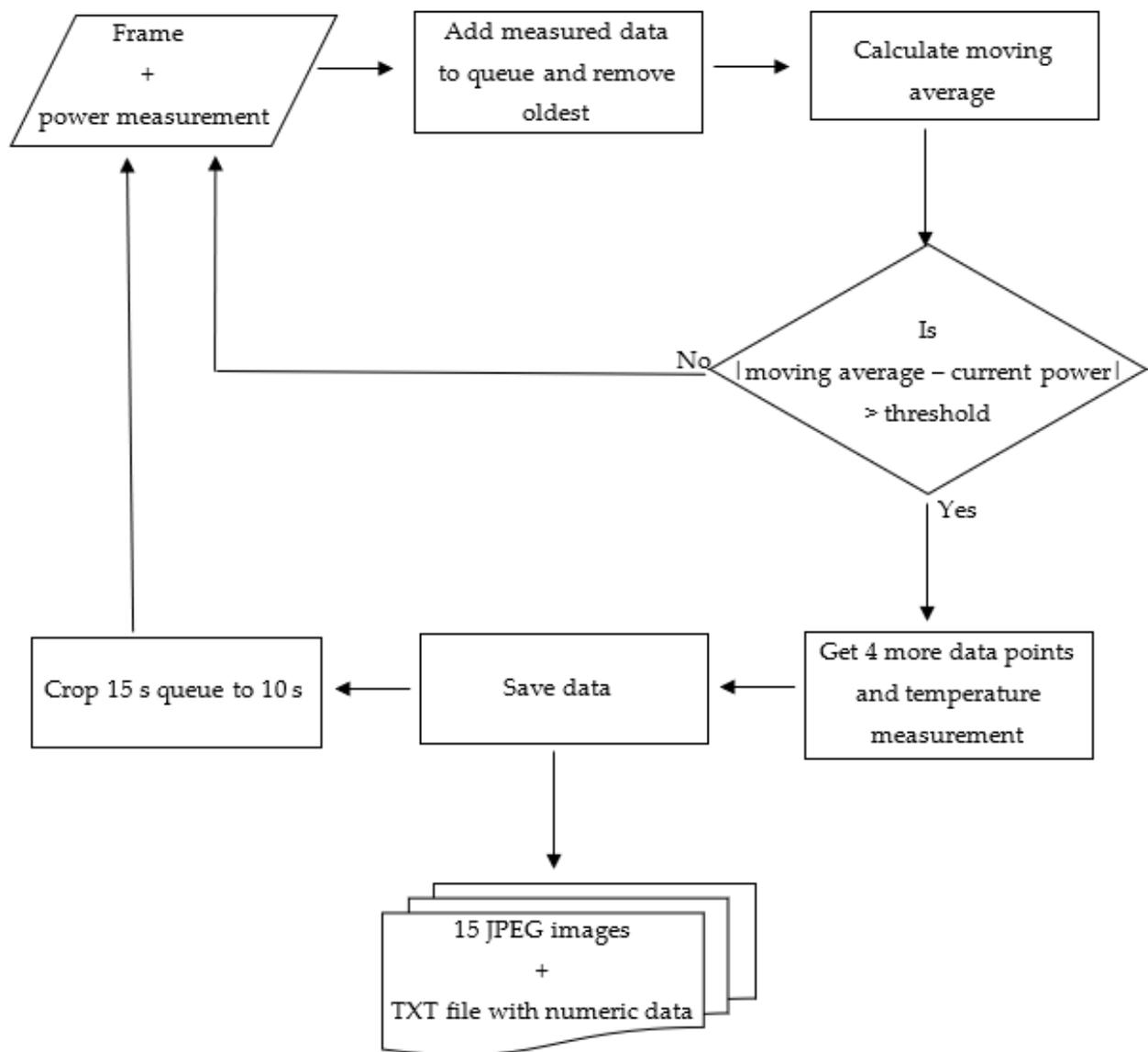


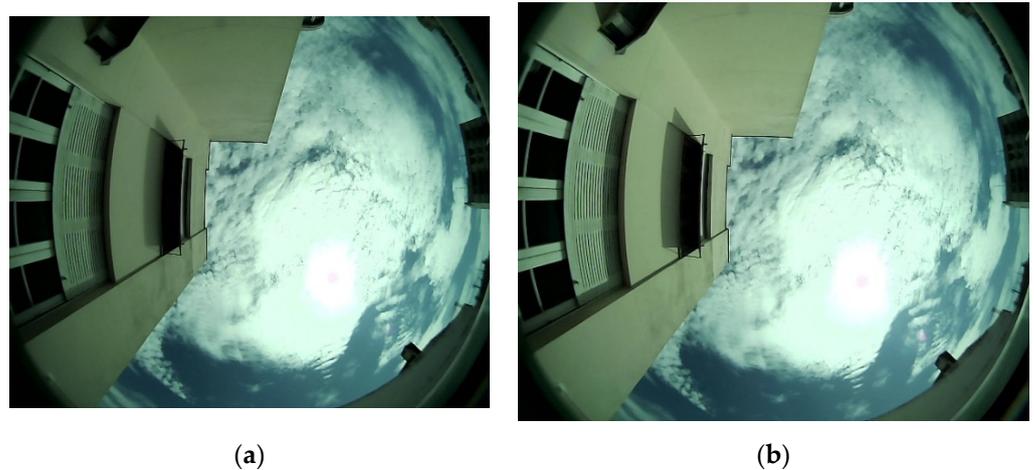
Figure 4. Flowchart of decision process and data flow within DAS software.

The first column contains the time stamps from each measurement point in the event. Each component is separated by an underscore, following the “hh_mm_ss_YYYY_MM_DD” format. The last component is a Boolean value, indicating whether daylight savings time is in effect. This is in local time and is used to determine the solar position angles.

Next are temperature measurements taken once per event, if, as with this example, there is more than one value in one event, it is because events were detected close to one another and some of the time stamps intersect. The negative values were used to represent “no data”, and those were replaced by linear interpolations for the data-modeling phase. The last three values are the measured voltage, current and power, respectively. One such file was generated per detected event. As for the captured images, Figure 5 shows two examples of raw images, one from the beginning of the event and the second from the end.

Table 3. Example of numeric data pertaining to one event structure.

Time Stamp	Temperature (°C)	Voltage (V)	Current (A)	Power (W)
12_27_34_2019_03_17_0	48.0	0.241	0.614	0.148
12_27_33_2019_03_17_0	−1000	0.246	0.626	0.154
12_27_32_2019_03_17_0	−1000	0.251	0.632	0.158
12_27_31_2019_03_17_0	−1000	0.247	0.633	0.156
12_27_30_2019_03_17_0	−1000	0.246	0.629	0.155
12_27_29_2019_03_17_0	−1000	0.242	0.621	0.150
12_27_28_2019_03_17_0	−1000	0.236	0.609	0.144
12_27_27_2019_03_17_0	−1000	0.232	0.601	0.139
12_27_26_2019_03_17_0	−1000	0.227	0.595	0.135
12_27_25_2019_03_17_0	−1000	0.226	0.589	0.133
12_27_24_2019_03_17_0	48.0	0.222	0.584	0.130
12_27_23_2019_03_17_0	−1000	0.222	0.581	0.129
12_27_22_2019_03_17_0	−1000	0.217	0.576	0.125
12_27_21_2019_03_17_0	−1000	0.216	0.569	0.123
12_27_20_2019_03_17_0	−1000	0.212	0.561	0.119

**Figure 5.** Raw image examples from the beginning of an event (a) and from the end (b).

3.4. Visual Analysis

The first step in analyzing the obtained data was by performing image subtraction to visually assess how much change occurred between different time points. This approach showed no information close to the solar disk because of image saturation, and a neutral density filter was used to try and reduce the saturation problem. This did reduce the saturated area, but there was still no visible information regarding the ramp events. Figure 6 shows the result of the image subtraction with the saturated region highlighted.

In the case of image subtraction, since both images had a similar saturated region, the result after subtraction are black regions in the image. On Figure 6 the rightmost image has a smaller saturated region due to the use of the neutral density filter. Since a visual analysis was not sufficient to determine whether the data obtained were useable for modeling purposes, the next step was to perform a linear correlation analysis from image features and power data.

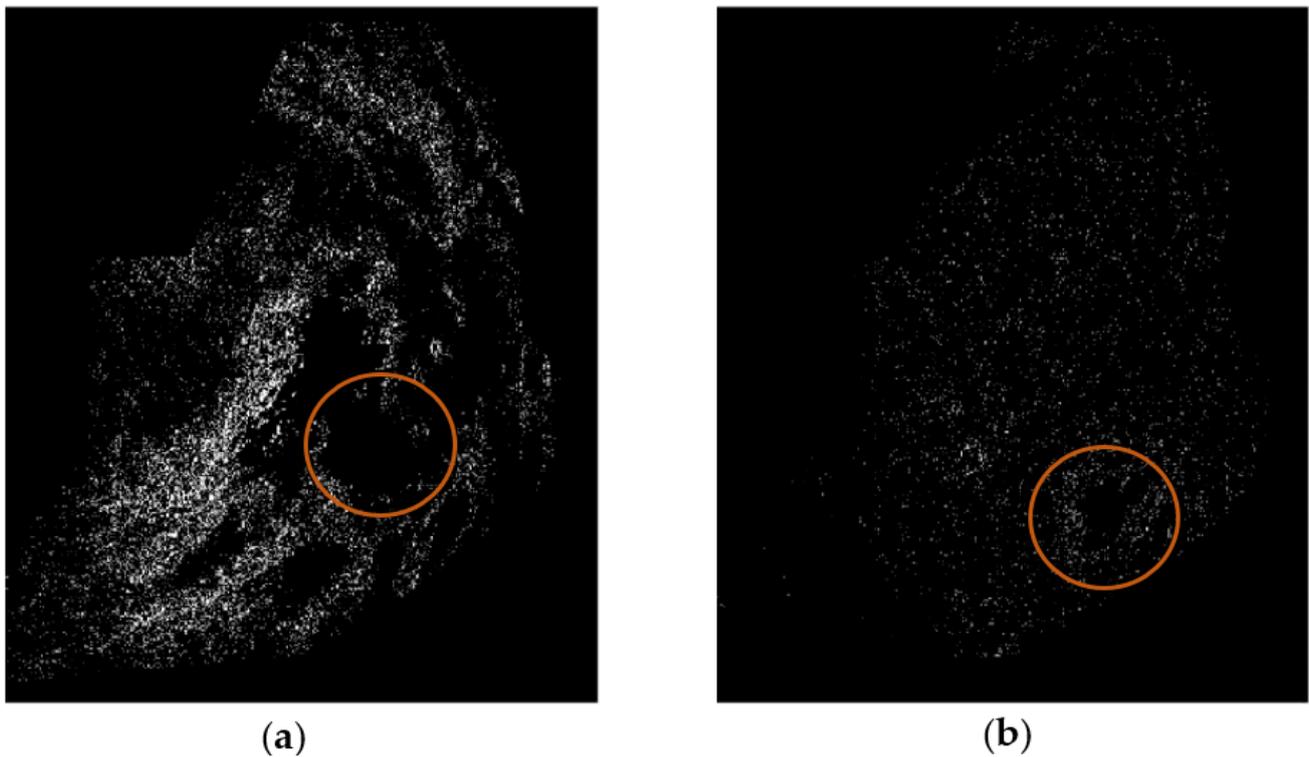


Figure 6. Image subtraction examples with the solar region highlighted. (a) Before installing the neutral density filter and (b) after installation.

3.5. Correlation Analysis

This goal in this step was not only to determine if there was a linear correlation between image features and the power ramp events, but to determine at which time intervals they were higher. Because of the detection-per-ramp event, the data are not contiguous in their entirety; however, during several periods, the variations occurred close enough for the data to overlap and create an almost continuous set of data points.

To unbiasedly determine which time interval would be more adequate for modeling the ramp events, several different values were analyzed. The power variation between points was calculated for each possible pair of data points that fits in the different intervals. Due to the disconnection between the data points, as the intervals grew larger, less points fit in a certain interval, so the maximum interval used was 90 s. The corresponding variable for the correlation analysis is obtained directly from the image. It was obtained by subtracting the corresponding images from each two data points used for calculating the power difference. After subtraction of each digital channel (RGB), the energy (image energy is calculated by summing the individual pixel values in an image or ROI) was calculated for a circular Region of Interest (ROI) around the sun. Different ROI radii (distance in pixels) were used to take into consideration cloud movement (speed) in a given interval (time).

Aside from the power difference and subtracted image energy, the instant power measurements and temperature measurement were also analyzed. In total, 84 combinations of time intervals $\Delta t = \{1; 2; 5; 8; 10; 15; 20; 30; 45; 60; 75; 90\}$ and ROI radii $r = \{25; 50; 75; 100; 150; 200; 250\}$ were analyzed in this step. To be able to present the results in a concise form, each combination of interval and radius was assigned an index that will be used to identify them throughout this work. Table 4 contains the keys to identify the combinations from their respective indexes.

Table 4. Indexes used to identify the combinations of Δt and ROI radius.

ROI Radius (pixels)	Δt (s)											
	1	2	5	8	10	15	20	30	45	60	75	90
25	1	8	15	22	29	36	43	50	57	64	71	78
50	2	9	16	23	30	37	44	51	58	65	72	79
75	3	10	17	24	31	38	45	52	59	66	73	80
100	4	11	18	25	32	39	46	53	60	67	74	81
150	5	12	19	26	33	40	47	54	61	68	75	82
200	6	13	20	27	34	41	48	55	62	69	76	83
250	7	14	21	28	35	42	49	56	63	70	77	84

Correlation coefficients were calculated for each combination of the target variables (power at t_0 , P_0 ; and power difference between t_0 and $t_{0-\Delta t}$: ΔP) and the aforementioned variables (power at $t_{0-\Delta t}$, P_{-1} ; temperature at t_0 , T_0 ; and ROI energy differences between t_0 and $t_{0-\Delta t}$). Correlation coefficients are used to measure linear proportionality between data pairs, which will show if a linear regression model would suffice for this problem.

3.6. Neural Network Modeling

To validate the obtained data, first a baseline regression performance was defined by performing multivariate linear regression to model P_0 and ΔP as a function of P_{-1} , T_0 and the image attribute of the blue channel, previously introduced. Only one color channel was used to prevent a collinearity issue from adversely affecting the model regression. To evaluate the regression performance, the coefficient of determination (R^2) was employed, as it measures how well the model represents the data used for regression.

All attempted linear regressions presented low R^2 , despite showing low error, most likely due to the extremely low variation rates in the data presented. This aligns with the information obtained from the correlation analysis, where for shorter time intervals, P_0 and P_{-1} showed high correlation coefficients. This fact does not suffice to produce a good regression model. The other variables were statistically insignificant to the model, despite being relevant in theory. This pointed to the possible suitability of a nonlinear model, and for that step a regression neural network was chosen.

Artificial neural networks aim to mimic a brain's neuronal structure by assigning weights to the individual interconnections between neurons, and thus are capable of solving complex, non-linear problems [30]. Despite the correlation analysis only looking into linear correlation between pairs of variables, most likely there are more complex relationships between these variables, and by increasing size and complexity of a neural network, it should be able to model these relationships.

A multilayer perceptron (MLP) network was used for the purpose of validating the acquired data and selected image features. The network used in this work had fully connected neurons to map underlying relationships between the selected variables. If a certain connection does not prove to be relevant to the problem, the learning process will assign low synaptic weights to them. The selected training algorithm was through feed-forward backpropagation [30].

In it, the function signals resulting of the response of the activation function move forward through the interconnected neurons biased by the synaptic weights until they reach the output layer. The result is compared to a previously known value and the error values are propagated backwards through the network and the synaptic weights are adjusted to minimize the error values. This process may take several iterations depending on the complexity of the model and the network [30].

This process has the potential to overfit the model to the presented data, rendering it unsuitable for interpreting new data. In order to avoid this, the data provided need to be of sufficient size and pertinent to the problem, a suitable architecture and size of network

must be used, the problem must not be complex beyond what the model can handle, and the training process must be stopped before the model is overfitted to the training data. This process may take several iterations depending on the complexity of the model and the network [30].

For the first issue, in the context of this work, the data-acquisition procedure and feature selection were tailored to the problem at hand, so the representativeness of the dataset should be sufficient. As for sample size, the system acquired data for as long as it could, until the camera failed, most likely due to humidity damage to the circuitry or ultraviolet (UV) damage to the camera sensor.

Regarding the second issue, the MLP network was tested with several sizes and architectures to produce the highest accuracy and generalization possible. As for the complexity of the problem, that cannot be changed, but the representativeness of the variables used should provide the network with enough valuable information. Again, that is also a result of the tailoring of the data-acquisition procedures to the very short-term forecast problem.

Finally, regarding overfitting by overtraining, a cross-validation approach [30] was used to the back-propagation learning. This means that the training sample was split into two subsets, one to perform the actual learning with error backpropagation and synaptic weights adjustment, and the other was used to validate the error on a fresh set of data that the model could not have been overfitted to. By comparing the network performance on both subsets, it is clear when the model starts to get overfitted. Whilst the training set would keep reducing errors, the validation set would start to see increasing errors. This would mean that the model was overfitted to the training set and was losing generalization capability.

4. Results

This section presents the results for the data-processing steps introduced previously. To visually aid in the comprehension of the plots, Figure 7 shows how to interpret the x -axis of figures using the aforementioned index.

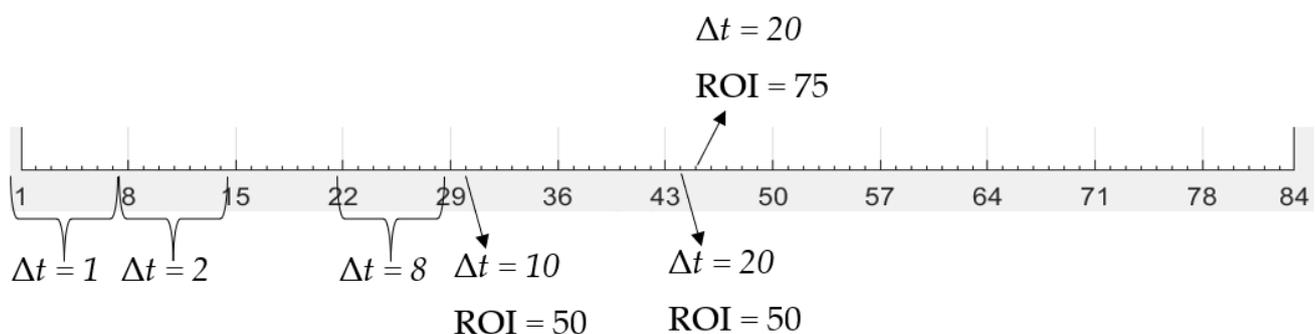


Figure 7. Visual guide to aid in the interpretation of the x -axis of the analysis plots with multiple networks.

Starting with the larger ticks, with the showing indexes, these mark the start of a new Δt value. As for the smaller ticks those correspond to the different ROI radii used within each group of Δt groupings.

4.1. Correlation Analysis

The first set of correlation analysis results, between P_0 and the evaluated variables, is presented in Figure 8.

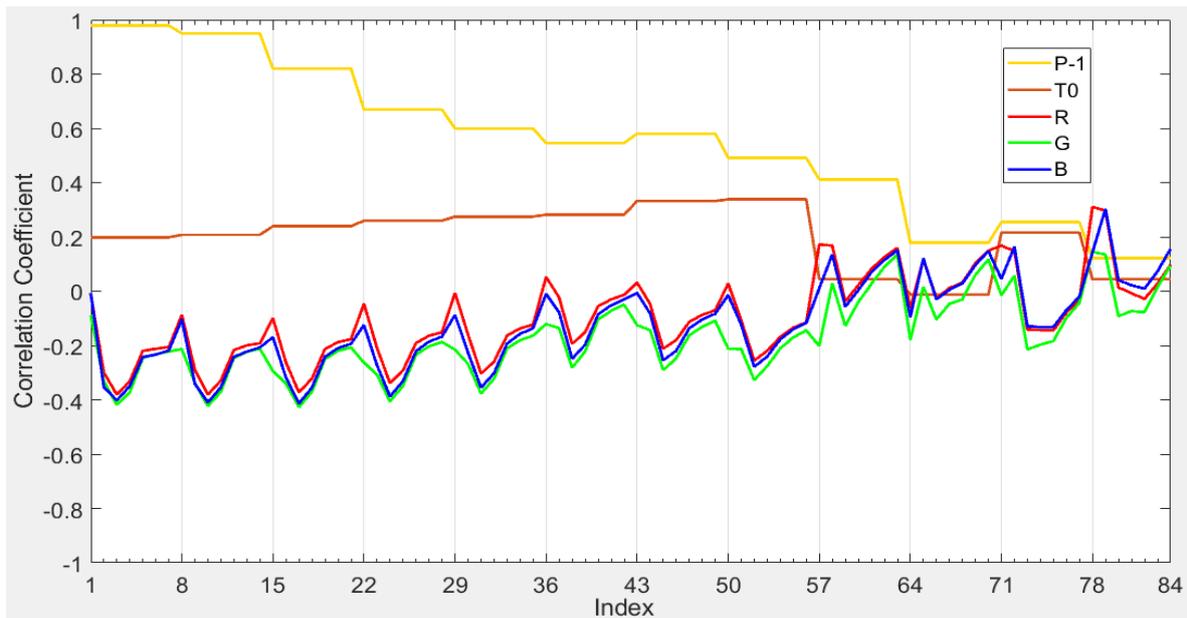


Figure 8. Correlation between P_0 and the evaluated variables.

At shorter time intervals, there is a high correlation between P_0 and P_{-1} ; however, as Δt increases, the correlation coefficient approaches zero. Temperature and image attributes present correlation coefficient magnitudes under 0.4, which means that a linear regression model is unsuitable to represent these data. Next, the correlation coefficients between ΔP and the same 5 variables in the previous analysis is presented in Figure 9.

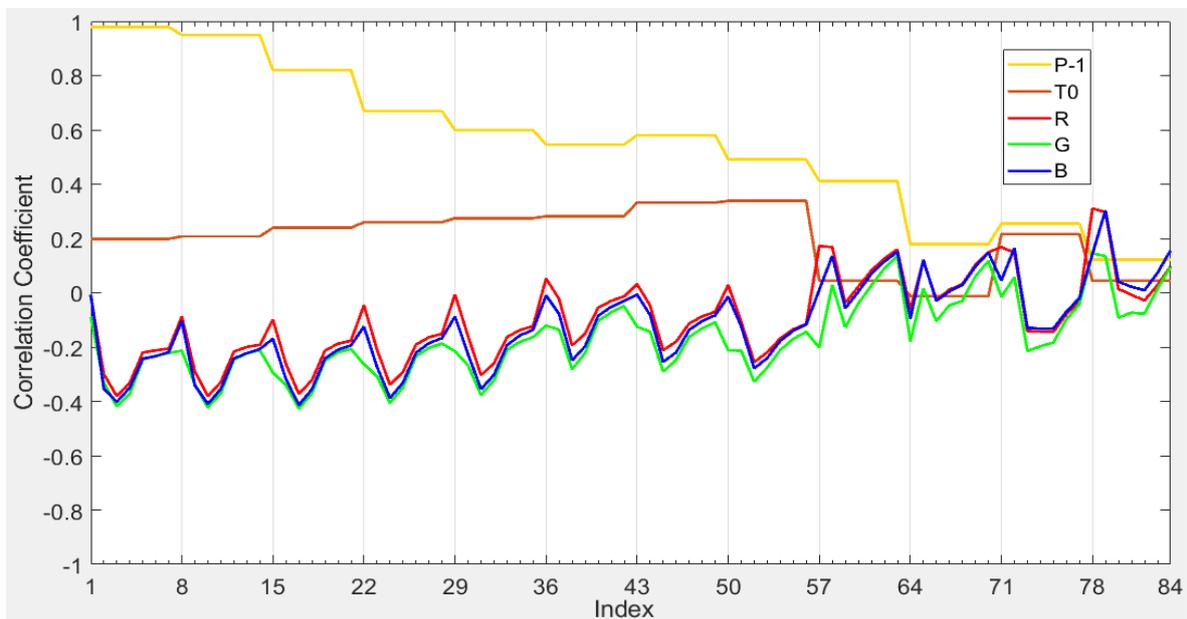


Figure 9. Correlation between ΔP and the evaluated variables.

The results for ΔP present rather different relationships between the chosen variables. As Δt increases the correlation coefficient magnitudes mostly increase as opposed to the results with P_0 . In the case of the image attributes, it reaches a value of about 0.7 at 75 s before drastically falling and becoming negative. P_{-1} does correlate better to ΔP than to P_0 ; however, it is still too low for a proper linear regression model. What these results point out is that a linear model is unsuitable for representing this phenomenon through

these data. The next step is to use a nonlinear model to attempt this representation, and the chosen model for that was a MLP (multilayer perceptron) artificial neural network.

4.2. Neural Network Architecture

Different networks with different architectures were trained for each combination of Δt and ROI radius to determine the best architecture for this model. Unlike the default random split employed by the Matlab neural network training tool, for this application an interleaved division algorithm was used to ensure that data from every day were available for training and validation, thus ensuring maximum representativeness. The proportion of training data was 70% of the set and consequently 30% was used for validation. Other splits were preliminarily tested; however, this proportion showed less variation among results when run multiple times. Normalization is an important process for neural network training, framing all values between 0 and 1, so that the gradients applied to the synaptic weights' updates are always decreasing [30].

The Matlab neural network training tool is highly customizable, but some of the default values for data-fitting problems, such as these, were left unchanged: the specific type of backpropagation algorithm, Levenberg–Marquardt; the mean-squared-error performance metric; and the hyperbolic tangent sigmoid (tansig) transfer function for the neurons. This was performed because these default values yielded solid results and were beyond the machine learning scope of this work.

Training was performed for both target variables, P_0 and ΔP , since both had very different behaviors and none were successfully represented by linear models. A diagram of the relationships between the inputs, neurons and modeled variables is presented in Figure 10.

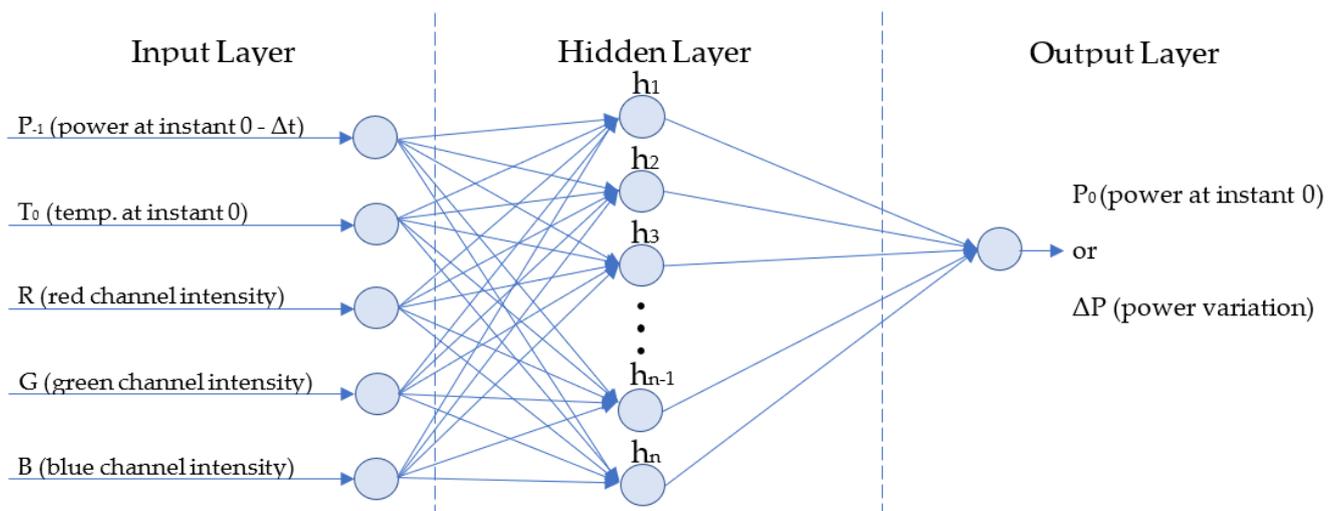


Figure 10. Diagram of inputs, outputs and layers in the tested networks.

First, the P_0 coefficient of determination is presented in Figure 11 for the different architectures and combination of Δt and ROI radius.

The model was trained with all five input variables previously used for the correlation analysis and linear regression (P_{-1} , T_0 and the image attributes from all three channels). Each line on the plot represents a different network architecture, with either one or two hidden layers and several layer sizes listed in the legend. Thicker lines represent networks with two hidden layers.

For the first two Δt values, all plot lines are indistinguishably close and boast good coefficients of determination, this being consistent with the results from the correlation analysis and linear regression. After this point, there is a dip in regression performance consistent with the linear evaluations; it then starts improving again, reaching even higher R^2 than the initial Δt range.

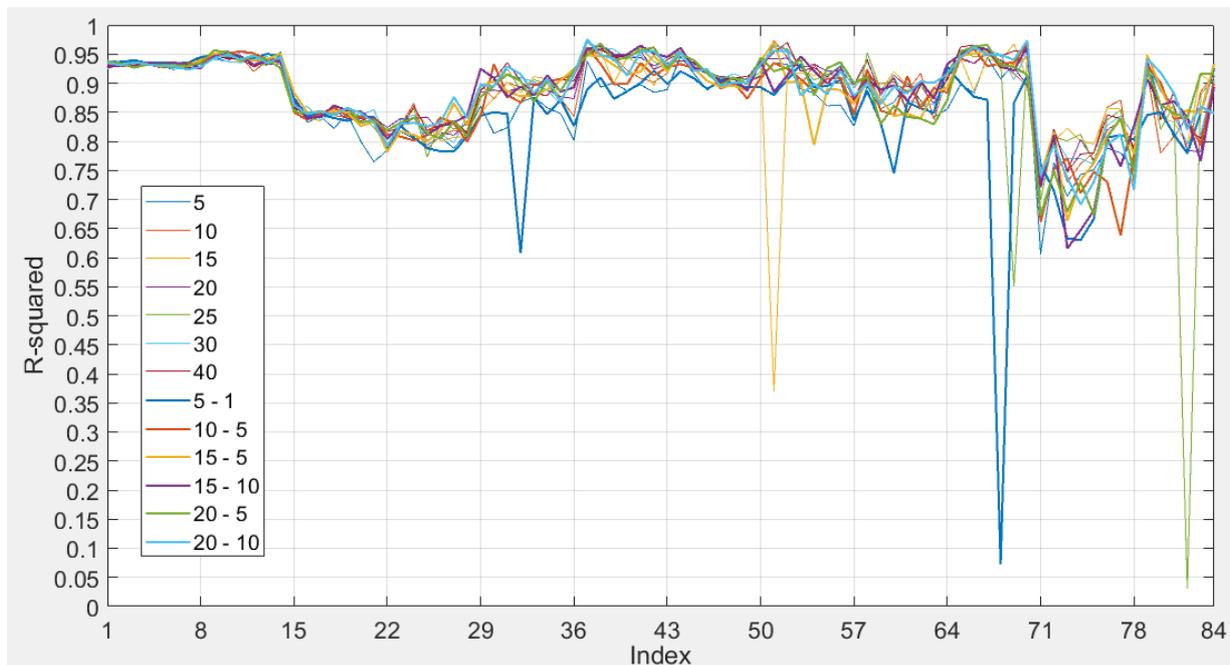


Figure 11. R^2 values for neural network regression models for P_0 .

Networks with five neurons on the first hidden layer seem to yield the worst results. Other architectures vary and not one architecture seems significantly better than another. That said, networks with two hidden layers seem to be very similar to one another in most cases, as well as seem to vary less in amplitude than networks with a single hidden layer. The performance starts decreasing again for the last two Δt values, which may be due to a less relevant relationship between input and output or due to less training samples availability. This occurs because the data are not contiguous, and therefore, with larger time intervals, the amount of data points that can be related decreases. These results show that P_0 is more accurately modeled with a nonlinear method, such as neural networks.

The same method, variables and architectures were used for modeling ΔP , and the R^2 values for this step are depicted in Figure 12.

This result showed that, for the first four Δt values, neither a linear nor a non-linear method was capable of properly fitting these data. As of the fifth Δt value, the neural network model starts presenting good R^2 , around 0.9. Similar to the previous plot, it is clear that networks with five neurons in the first hidden layer are inferior to the other tested architectures for most data points. After the fifth Δt value, the R^2 behaves similarly in both plots, reaching the highest coefficient of determination for $\Delta t = 60$ s, closely followed by $\Delta t = 15$ s. For both intervals, there are small peaks around ROI radius = 75 and 200 pixels.

One significant difference in both is the lack of the four drastically lower coefficient of determination points in the ΔP models, but that may be due to the lack of outlier analysis prior to model training. Since these regressions were made with all input variables, it was necessary to see if all variables were relevant to represent the target data. For this reason, the same training processes were performed for both target variables but varying the inputs. The chosen architecture was with two hidden layers with 15 and 10 neurons, respectively, which seemed to have some of the highest R^2 values and varied less than others. The inputs used for each model are displayed in Table 5. Each row represents one model, and each column represents one of the five variables. The Xs mark when a variable is used.

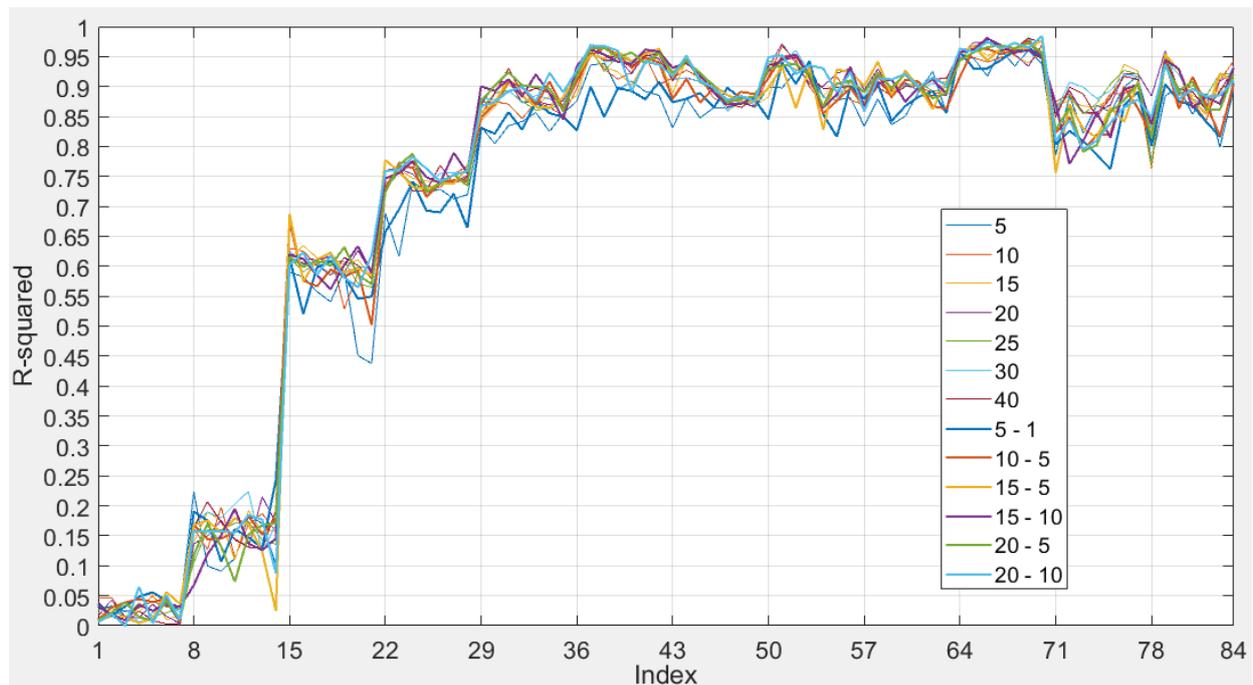


Figure 12. R^2 values for neural network regression models for ΔP .

Table 5. Variables used in the second step of NN modeling.

Model	Variables				
	P_{-1}	T_0	R	G	B
1	X	X	X	X	X
2	X	X	X	X	O
3	X	X	X	O	O
4	X	X	O	O	O
5	X	O	O	O	O
6	O	X	X	X	X
7	X	X	O	X	O
8	X	X	O	O	X
9	X	X	O	X	X
10	X	X	X	O	X
11	X	O	X	X	X

The results from this training with different input variables for target variable ΔP are shown in Figure 13. Each line represents one line on the plot, and to save some room in the legend, the variables P_{-1} (power at instant $0-\Delta t$) and T_0 (temperature at instant zero), were shortened to P and T , respectively. The red, blue, and green channel attributes were represented by R , G and B , respectively. In order to reduce some of the randomness attributed to the initialization of the variables and data division, each network was trained five times, and the best result was selected.

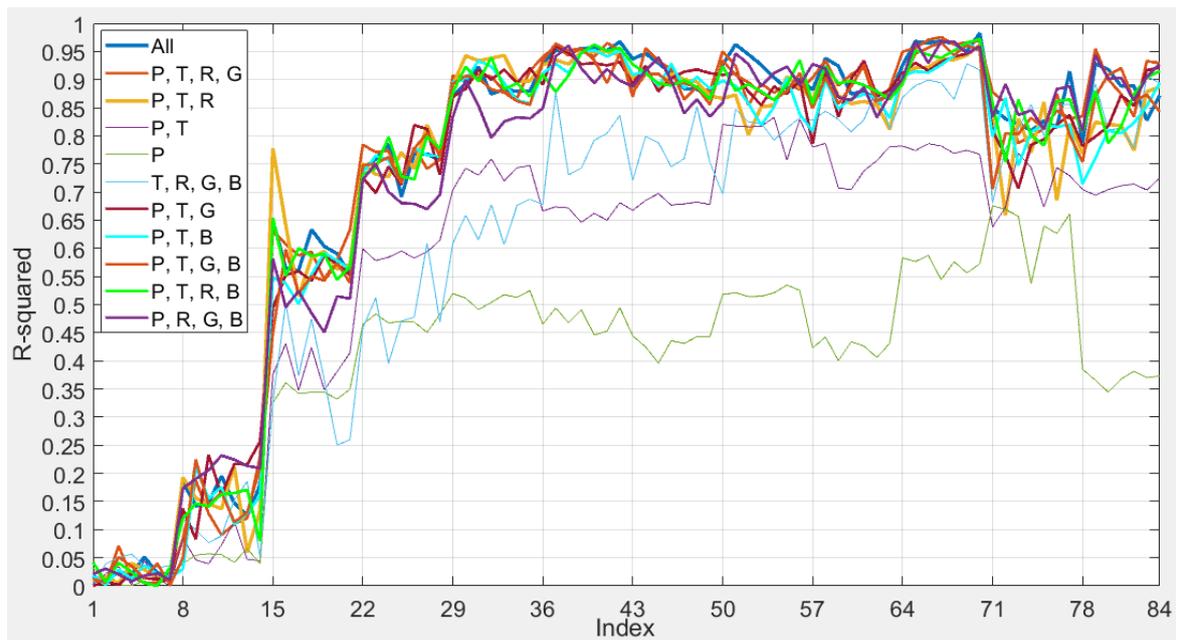


Figure 13. R^2 values for neural network regression models for ΔP with varying input variables.

The first information to stand out in this plot is the three lower performing models with either missing power, temperature or image attributes (P , T or $[R, G, B]$). The best result was considered to be with all variables, reaching the highest R^2 value (>0.98) and being the best result for several points. This result was achieved for $\Delta t = 60$ s and ROI radius = 250 pixels.

The same methodology was applied to training with P_0 as the target, with the same combinations of variables and the results shown in Figure 14.

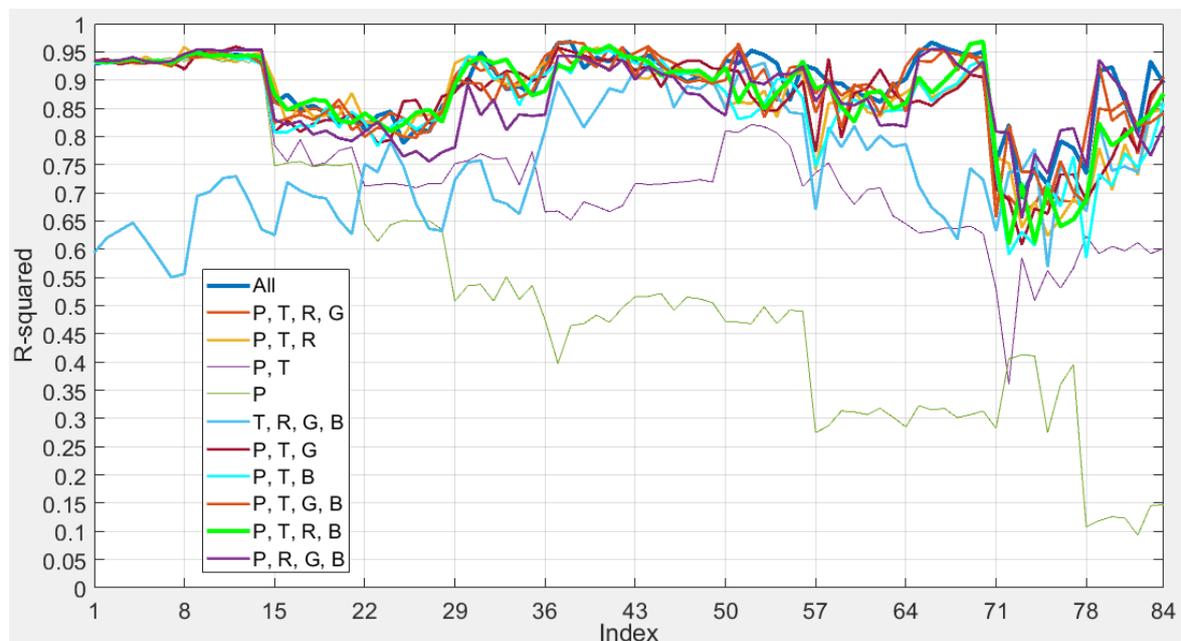


Figure 14. R^2 values for neural network regression models for P_0 with varying input variables.

Similar to Figure 13, the three worst variable selections have either power, temperature or image attribute missing, but in this case, power made a bigger impact. However, as

Δt increases, the importance of P_{-1} decreases, not just linearly as previously thought. Moreover, for P_0 the difference between using two or three image attributes is lower than for ΔP , but with all input variables, the models seem to fare overall slightly better, with ones using just the red and blue channels closely behind. The highest R^2 (≈ 0.97) is with just red and blue (P, T, R, B) at $\Delta t = 60$ s and ROI radius = 250 pixels.

It is safe to say both variables were successfully modeled by using neural networks, especially compared with linear models. For both cases, previous power, temperature and image attributes from image subtraction proved to be important to model the targets.

4.3. Best Neural Network Results

Finally, the selected architecture of two layers with 15 and 10 neurons, respectively, was trained several times, using all five input variables with data from the $\Delta t = 60$ s step and ROI radius = 250 pixels to provide further insight on their performance and finish the validation step.

First, the data were tested modeling P_0 , using all five input variables. The coefficient of determination obtained was $R^2 = 0.94$ for the validation process. This means that, when presented with data which were not used to train the network, it still was capable of estimating the output close to the real measured value.

Figure 15 presents the regression plot from this model, the blue line represents the model and the points are the pairs of estimated value versus real value for each input sample.

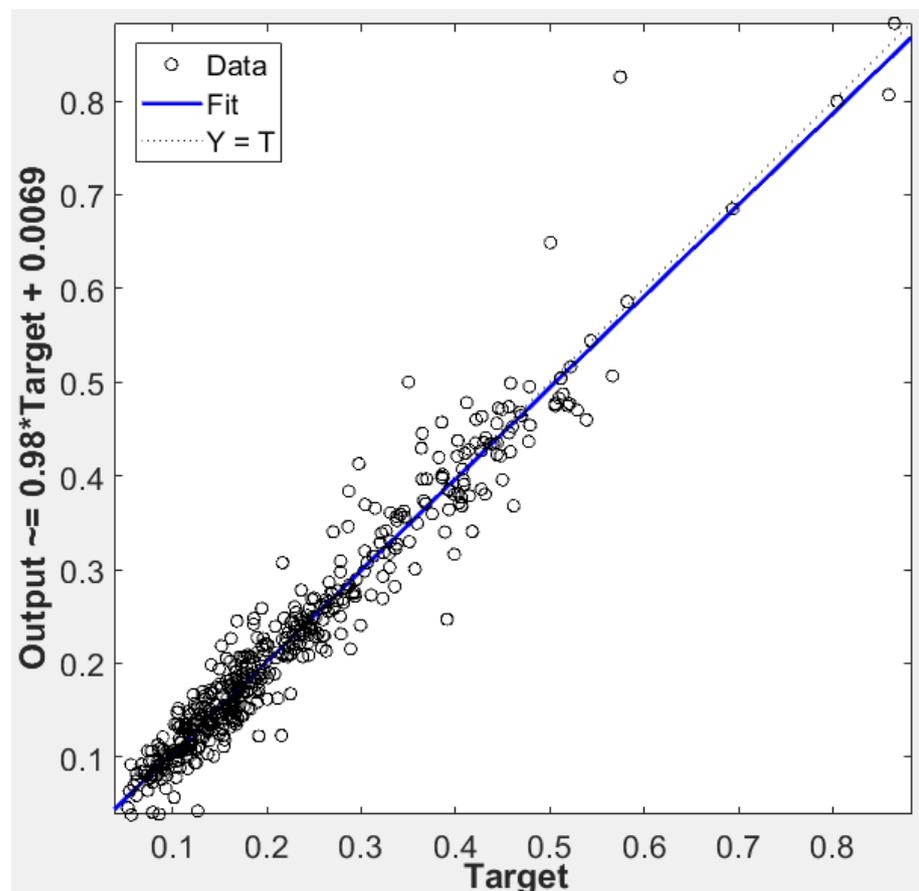


Figure 15. Regression plot for the validation of a NN model of P_0 with $\Delta t = 60$ s step and ROI radius = 250 pixels.

In a perfect model, all points would stand on the line, but this result shows a very close representation of the relationship between input variables and the target variable. Next, the same process was applied to modeling ΔP , and the results are presented in Figure 16.

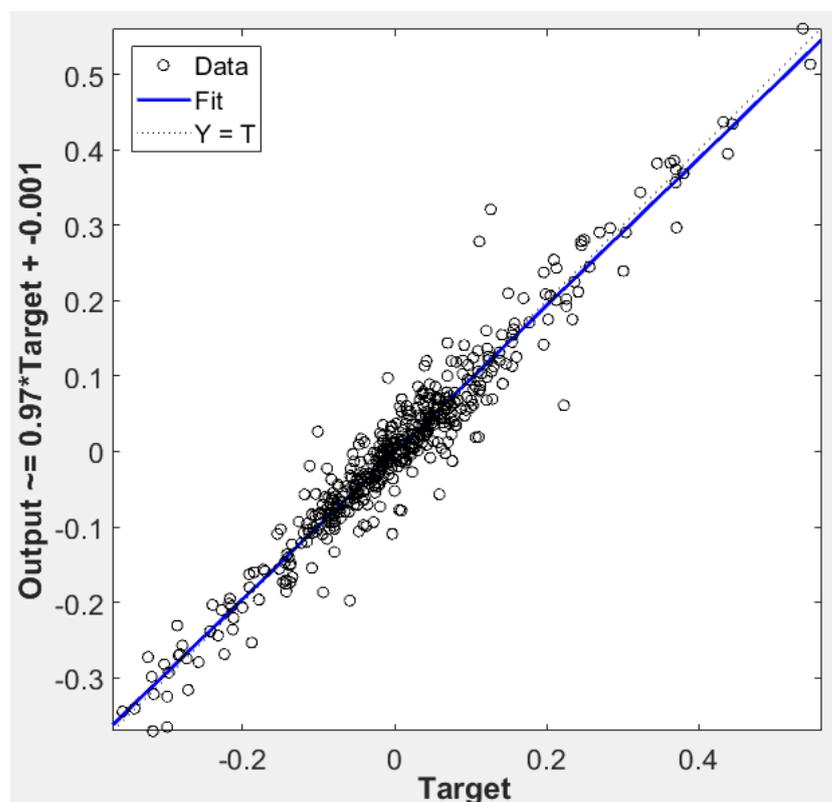


Figure 16. Regression plot for the validation of a NN model of ΔP with $\Delta t = 60$ s step and ROI radius = 250 pixels.

In this case, the validation $R^2 = 0.93$ and the regression plot also show how well the model represents the relationship between input and target variable. It is safe to claim that neural networks are well suited to model this type of data.

4.4. Discussion

When evaluating the obtained results relative to the literature herein presented, the first important point to consider is that the higher frequency of acquisition provided by the current system does offer more, and useful, information for modeling PV generation at a plant level [8,35]. This better resolution, coupled with the acquisition strategy, serves to confirm what other works have shown, that higher resolutions are important for PV forecasting [19,35,51,52,59,60].

The results also show the suitability of neural networks for modeling the relationship between image data and PV power. Its use in applications such as determining cloud albedo [33] or optical depth [55] should aid in forecasting efforts. Due to the low cost of the equipment used and the key information it was able to provide, approaches that employ multiple imagers [51,55,60] should be more easily employed. The combination of high-fidelity regression, high-frequency data and low cost permits not only a higher accessibility for developing countries to endeavor in PV energy research, but also for a more complex and in-depth look into the PV forecast problem.

5. Conclusions

Validation was performed on the data selected during the correlation analysis by using a linear model as the baseline and a neural network regression as a nonlinear model. It was

possible to model power variations with up to 60 s intervals based on the data acquired by the developed system. Both the characteristics of the data themselves and of the selected features used for training the neural network were proven relevant to the intra-minute solar forecast problem.

Given the high-accuracy results, the data frequency and chosen variables were deemed relevant for intra-minute forecasting. The acquisition by exception proved to yield data rich in information surrounding solar variability; however, the event structure should be redefined in order to more accurately translate the reality. Since, through the data analysis, a 15 to 60 s horizon was deemed ideal given the available data, and that assumption was validated by the neural network model, an event structure capable of fully encompassing this horizon is recommended. Based on the information provided by this experimental research, an event structure with 90 s prior to the point of detection and 30 s after it should be enough to provide a clearer view on the subject of study.

Through forecasting, renewable energy sources will become more reliable and help steer the energy paradigm into a less fossil-reliant reality. With the coupling of multi-horizon forecasting, power electronics and energy storage systems, RES can lead to a new and clean energy era. To make this happen, more research into forecasting of the solar resource in different temporal and spatial scales is required, as well as the combination of forecasting with energy storage. The recommendations to improve upon the foundation laid by this work are as follows:

- Increase geometrical complexity by using arrays of PV panels, mirroring real-world solar farms;
- Test the system in different seasons and climates;
- Couple the model with a cloud tracking and forecast algorithm to provide power forecasts with the system;
- Model the impact of 60 s ahead forecasts for energy-storage management and PV variability mitigation;
- Test the developed acquisition system with the NN model with entirely new data.

Author Contributions: Conceptualization, G.F.B., R.F.C. and C.H.B.; methodology, G.F.B., R.F.C. and C.H.B.; software, G.F.B.; validation, G.F.B., R.F.C. and C.H.B.; formal analysis, G.F.B., R.F.C. and C.H.B.; investigation, G.F.B.; resources, G.F.B., R.F.C. and C.H.B.; data curation, G.F.B.; writing—original draft preparation, G.F.B.; writing—review and editing, G.F.B., R.F.C. and C.H.B.; visualization, G.F.B., R.F.C. and C.H.B.; supervision, R.F.C. and C.H.B.; project administration, R.F.C. and C.H.B.; funding acquisition, R.F.C. and C.H.B. All authors have read and agreed to the published version of the manuscript.

Funding: The authors thank for the financial support provided by the Brazilian funding agencies CNPq, CAPES, FINEP and FAPERJ. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available in Mendeley Data at 10.17632/r83r6g5y6t.1, reference number [72].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. UNFCCC; Paris Agreement: Paris, France, 2015.
2. IEA. *World Energy Outlook: Executive Summary*; IEA: Paris, France, 2018.
3. United Nations. *Transforming Our World: The 2030 Agenda for Sustainable Development*; United Nations: New York, NY, USA, 2015.
4. Can Şener, Ş.E.; Sharp, J.L.; Anttil, A. Factors impacting diverging paths of renewable energy: A review. *Renew. Sustain. Energy Rev.* **2018**, *81*, 2335–2342. [[CrossRef](#)]
5. Denholm, P.; Margolis, R.M. Evaluating the limits of solar photovoltaics (PV) in traditional electric power systems. *Energy Policy* **2007**, *35*, 2852–2861. [[CrossRef](#)]

6. Reddy, S.; Painuly, J.P. Diffusion of renewable energy technologies—barriers and stakeholders' perspectives. *Renew. Energy* **2004**, *29*, 1431–1447. [[CrossRef](#)]
7. Dragoon, K.; Schumaker, A. *Solar PV Variability and Grid Integration*; Renewable Northwest Project: Portland, OR, USA, 2010.
8. Mills, A.; Wisser, R. *Implications of Wide-Area Geographic Diversity for Short-Term Variability of Solar Power*; Lawrence Berkeley National Laboratory: Berkeley, MI, USA, 2010.
9. Bessa, R.; Moreira, C.; Silva, B.; Matos, M. Handling renewable energy variability and uncertainty in power systems operation. *Wiley Interdiscip. Rev. Energy Environ.* **2014**, *3*, 156–178. [[CrossRef](#)]
10. Karimi, M.; Mokhlis, H.; Naidu, K.; Uddin, S.; Bakar, A.H.A. Photovoltaic penetration issues and impacts in distribution network—A review. *Renew. Sustain. Energy Rev.* **2016**, *53*, 594–605. [[CrossRef](#)]
11. Liang, X. Emerging Power Quality Challenges Due to Integration of Renewable Energy Sources. *IEEE Trans. Ind. Appl.* **2016**, *53*, 855–866. [[CrossRef](#)]
12. Diagne, M.; David, M.; Lauret, P.; Boland, J.; Schmutz, N. Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. *Renew. Sustain. Energy Rev.* **2013**, *27*, 65–76. [[CrossRef](#)]
13. Sobri, S.; Koochi-Kamali, S.; Rahim, N.A. Solar photovoltaic generation forecasting methods: A review. *Energy Convers. Manag.* **2018**, *156*, 459–497. [[CrossRef](#)]
14. Solargis Methodology—Solar Radiation Modeling. Available online: <https://solargis.com/docs/methodology/solar-radiation-modeling> (accessed on 16 August 2021).
15. Natural Earth Countries. Available online: http://www.naturalearthdata.com/download/10m/cultural/ne_10m_admin_0_countries.zip (accessed on 4 June 2020).
16. Solargis. Longterm yearly average of global irradiation at optimum tilt. *Global Solar Atlas* **2019**.
17. United Nations. *The Sustainable Development Goals Report*; United Nations: New York, NY, USA, 2021.
18. Barbieri, F.; Rajakaruna, S.; Ghosh, A. Very short-term photovoltaic power forecasting with cloud modeling: A review. *Renew. Sustain. Energy Rev.* **2017**, *75*, 242–263. [[CrossRef](#)]
19. Schmidt, T.; Kalisch, J.; Lorenz, E.; Heinemann, D. Evaluating the spatiooral performance of sky-imager-based solar irradiance analysis and forecasts. *Atmos. Chem. Phys.* **2016**, *16*, 3399–3412. [[CrossRef](#)]
20. Kow, K.W.; Wong, Y.W.; Rajkumar, R.; Isa, D. An intelligent real-time power management system with active learning prediction engine for PV grid-tied systems. *J. Clean. Prod.* **2018**, *205*, 252–265. [[CrossRef](#)]
21. Denholm, P.; Hand, M. Grid flexibility and storage required to achieve very high penetration of variable renewable electricity. *Energy Policy* **2011**, *39*, 1817–1830. [[CrossRef](#)]
22. IEC. Electrical Energy Storage—White Paper. *Int. Electrotech. Comm.* **2011**, 1–78. [[CrossRef](#)]
23. Petinrin, J.O.; Shaaban, M. Impact of renewable generation on voltage control in distribution systems. *Renew. Sustain. Energy Rev.* **2016**, *65*, 770–783. [[CrossRef](#)]
24. Varma, R.K.; Salehi, R. SSR Mitigation with a New Control of PV Solar Farm as STATCOM (PV-STATCOM). *IEEE Trans. Sustain. Energy* **2017**, *8*, 1473–1483. [[CrossRef](#)]
25. Richardson, W.; Krishnaswami, H.; Vega, R.; Cervantes, M. A low cost, edge computing, all-sky imager for cloud tracking and intra-hour irradiance forecasting. *Sustainability* **2017**, *9*, 482. [[CrossRef](#)]
26. Oliveira, A.; Calili, R.; Almeida, M.F.; Sousa, M. A Systemic and Contextual Framework to Define a Country's 2030 Agenda from a Foresight Perspective. *Sustainability* **2019**, *11*, 6360. [[CrossRef](#)]
27. Bassous, G.F. Development and Validation of a Low-Cost Data Acquisition System for Very Short-Term Photovoltaic Power Forecasting. *PUC-Rio* **2019**. Available online: <https://www.maxwell.vrac.puc-rio.br/47953/47953.PDF> (accessed on 17 September 2021).
28. Stefferud, K.; Kleissl, J.; Schoene, J. Solar forecasting and variability analyses using sky camera cloud detection and motion vectors. In Proceedings of the 2012 IEEE Power and Energy Society General Meeting, San Diego, CA, USA, 22–26 July 2012. [[CrossRef](#)]
29. Reikard, G. Predicting solar radiation at high resolutions: A comparison of time series forecasts. *Sol. Energy* **2009**, *83*, 342–349. [[CrossRef](#)]
30. Haykin, S. *Neural Networks and Learning Machines: A Comprehensive Foundation*, 3rd ed.; Prentice Hall: Upper Saddle River, NJ, USA, 2008; ISBN 9780131471399.
31. Raza, M.Q.; Nadarajah, M.; Ekanayake, C. On recent advances in PV output power forecast. *Sol. Energy* **2016**, *136*, 125–144. [[CrossRef](#)]
32. Das, U.K.; Tey, K.S.; Seyedmahmoudian, M.; Mekhilef, S.; Idris, M.Y.I.; Van Deventer, W.; Horan, B.; Stojcevski, A. Forecasting of photovoltaic power generation and model optimization: A review. *Renew. Sustain. Energy Rev.* **2018**, *81*, 912–928. [[CrossRef](#)]
33. Kumler, A.; Xie, Y.; Zhang, Y. A Physics-based Smart Persistence model for Intra-hour forecasting of solar radiation (PSPI) using GHI measurements and a cloud retrieval technique. *Sol. Energy* **2019**, *177*, 494–500. [[CrossRef](#)]
34. Zendejboudi, A.; Baseer, M.A.; Saidur, R. Application of support vector machine models for forecasting solar and wind energy resources: A review. *J. Clean. Prod.* **2018**, *199*, 272–285. [[CrossRef](#)]
35. Lave, M.; Reno, M.J.; Broderick, R.J. Characterizing local high-frequency solar variability and its impact to distribution studies. *Sol. Energy* **2015**, *118*, 327–337. [[CrossRef](#)]

36. Pedro, H.T.C.; Coimbra, C.F.M.; David, M.; Lauret, P. Assessment of machine learning techniques for deterministic and probabilistic intra-hour solar forecasts. *Renew. Energy* **2018**, *123*, 191–203. [[CrossRef](#)]
37. Chow, C.W.; Urquhart, B.; Lave, M.; Dominguez, A.; Kleissl, J.; Shields, J.; Washom, B. Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed. *Sol. Energy* **2011**, *85*, 2881–2893. [[CrossRef](#)]
38. Yang, D.; Kleissl, J.; Gueymard, C.A.; Pedro, H.T.C.; Coimbra, C.F.M. History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining. *Sol. Energy* **2018**, *168*, 60–101. [[CrossRef](#)]
39. Cervantes, M.; Krishnaswami, H.; Richardson, W.; Vega, R. Utilization of Low Cost, Sky-Imaging Technology for Irradiance Forecasting of Distributed Solar Generation. In Proceedings of the 2016 IEEE Green Technologies Conference (GreenTech), Kansas City, MO, USA, 6–8 April 2016; pp. 142–146. [[CrossRef](#)]
40. Urquhart, B.; Kurtz, B.; Dahlin, E.; Ghonima, M.; Shields, J.E.; Kleissl, J. Development of a sky imaging system for short-term solar power forecasting. *Atmos. Meas. Tech.* **2015**, *8*, 875–890. [[CrossRef](#)]
41. Gohari, S.M.I.; Urquhart, B.; Yang, H.; Kurtz, B.; Nguyen, D.; Chow, C.W.; Ghonima, M.; Kleissl, J. Comparison of solar power output forecasting performance of the Total Sky Imager and the University of California, San Diego Sky Imager. *Energy Procedia* **2013**, *49*, 2340–2350. [[CrossRef](#)]
42. Chu, Y.; Pedro, H.T.C.; Coimbra, C.F.M. Hybrid intra-hour DNI forecasts with sky image processing enhanced by stochastic learning. *Sol. Energy* **2013**, *98*, 592–603. [[CrossRef](#)]
43. Marquez, R.; Coimbra, C.F.M. Intra-hour DNI forecasting based on cloud tracking image analysis. *Sol. Energy* **2013**, *91*, 327–336. [[CrossRef](#)]
44. Quesada-Ruiz, S.; Chu, Y.; Tovar-Pescador, J.; Pedro, H.T.C.; Coimbra, C.F.M. Cloud-tracking methodology for intra-hour DNI forecasting. *Sol. Energy* **2014**, *102*, 267–275. [[CrossRef](#)]
45. West, S.R.; Rowe, D.; Sayeef, S.; Berry, A. Short-term irradiance forecasting using skycams: Motivation and development. *Sol. Energy* **2014**, *110*, 188–207. [[CrossRef](#)]
46. Chu, Y.; Pedro, H.T.C.; Li, M.; Coimbra, C.F.M. Real-time forecasting of solar irradiance ramps with smart image processing. *Sol. Energy* **2015**, *114*, 91–104. [[CrossRef](#)]
47. Alonso-Montesinos, J.; Batlles, F.J. The use of a sky camera for solar radiation estimation based on digital image processing. *Energy* **2015**, *90*, 377–386. [[CrossRef](#)]
48. Alonso-Montesinos, J.; Batlles, F.J.; Portillo, C. Solar irradiance forecasting at one-minute intervals for different sky conditions using sky camera images. *Energy Convers. Manag.* **2015**, *105*, 1166–1177. [[CrossRef](#)]
49. Cazorla, A.; Husillos, C.; Antón, M.; Alados-Arboledas, L. Multi-exposure adaptive threshold technique for cloud detection with sky imagers. *Sol. Energy* **2015**, *114*, 268–277. [[CrossRef](#)]
50. Chu, Y.; Li, M.; Pedro, H.T.C.; Coimbra, C.F.M. Real-time prediction intervals for intra-hour DNI forecasts. *Renew. Energy* **2015**, *83*, 234–244. [[CrossRef](#)]
51. Chu, Y.; Urquhart, B.; Gohari, S.M.I.; Pedro, H.T.C.; Kleissl, J.; Coimbra, C.F.M. Short-term reforecasting of power output from a 48 MWe solar PV plant. *Sol. Energy* **2015**, *112*, 68–77. [[CrossRef](#)]
52. Lipperheide, M.; Bosch, J.L.; Kleissl, J. Embedded nowcasting method using cloud speed persistence for a photovoltaic power plant. *Sol. Energy* **2015**, *112*, 232–238. [[CrossRef](#)]
53. Pedro, H.T.C.; Coimbra, C.F.M. Nearest-neighbor methodology for prediction of intra-hour global horizontal and direct normal irradiances. *Renew. Energy* **2015**, *80*, 770–782. [[CrossRef](#)]
54. Xu, J.; Yoo, S.; Yu, D.; Huang, D.; Heiser, J.; Kalb, P. Solar irradiance forecasting using multi-layer cloud tracking and numerical weather prediction. In Proceedings of the 30th Annual ACM Symposium on Applied Computing—SAC’15, Salamanca, Spain, 13–17 April 2015; ACM Press: New York, NY, USA, 2015; pp. 2225–2230.
55. Mejia, F.A.; Kurtz, B.; Murray, K.; Hinkelman, L.M.; Sengupta, M.; Xie, Y.; Kleissl, J. Coupling sky images with radiative transfer models: A new method to estimate cloud optical depth. *Atmos. Meas. Tech.* **2016**, *9*, 4151–4165. [[CrossRef](#)]
56. Rana, M.; Koprinska, I.; Agelidis, V.G. Univariate and multivariate methods for very short-term solar photovoltaic power forecasting. *Energy Convers. Manag.* **2016**, *121*, 380–390. [[CrossRef](#)]
57. Sanfilippo, A.; Martin-Pomares, L.; Mohandes, N.; Perez-Astudillo, D.; Bachour, D. An adaptive multi-modeling approach to solar nowcasting. *Sol. Energy* **2016**, *125*, 77–85. [[CrossRef](#)]
58. Soubdhan, T.; Ndong, J.; Ould-Baba, H.; Do, M.-T. A robust forecasting framework based on the Kalman filtering approach with a twofold parameter tuning procedure: Application to solar and photovoltaic prediction. *Sol. Energy* **2016**, *131*, 246–259. [[CrossRef](#)]
59. Ai, Y.; Peng, Y.; Wei, W. A model of very short-term solar irradiance forecasting based on low-cost sky images. *AIP Conf. Proc.* **2017**, *1839*, 020022. [[CrossRef](#)]
60. Blanc, P.; Massip, P.; Kazantzidis, A.; Tzoumanikas, P.; Kuhn, P.; Wilbert, S.; Schüler, D.; Prah, C. Short-term forecasting of high resolution local DNI maps with multiple fish-eye cameras in stereoscopic mode. *AIP Conf. Proc.* **2017**, *1850*, 140004. [[CrossRef](#)]
61. Cheng, H.Y. Cloud tracking using clusters of feature points for accurate solar irradiance nowcasting. *Renew. Energy* **2017**, *104*, 281–289. [[CrossRef](#)]
62. Elsinga, B.; van Sark, W.G.J.H.M. Short-term peer-to-peer solar forecasting in a network of photovoltaic systems. *Appl. Energy* **2017**, *206*, 1464–1483. [[CrossRef](#)]
63. Ni, Q.; Zhuang, S.; Sheng, H.; Kang, G.; Xiao, J. An ensemble prediction intervals approach for short-term PV power forecasting. *Sol. Energy* **2017**, *155*, 1072–1083. [[CrossRef](#)]

64. Kuhn, P.; Nouri, B.; Wilbert, S.; Prah, C.; Kozonek, N.; Schmidt, T.; Yasser, Z.; Ramirez, L.; Zarzalejo, L.; Meyer, A.; et al. Validation of an all-sky imager-based nowcasting system for industrial PV plants. *Prog. Photovolt. Res. Appl.* **2018**, *26*, 608–621. [CrossRef]
65. Bouzgou, H.; Gueymard, C.A. Fast short-term global solar irradiance forecasting with wrapper mutual information. *Renew. Energy* **2019**, *133*, 1055–1065. [CrossRef]
66. Chow, C.W.; Belongie, S.; Kleissl, J. Cloud motion and stability estimation for intra-hour solar forecasting. *Sol. Energy* **2015**, *115*, 645–655. [CrossRef]
67. Wood-Bradley, P.; Zapata, J.; Pye, J. Cloud tracking with optical flow for short-term solar forecasting. In Proceedings of the 50th Conference of the Australian Solar Energy Society, Melbourne, VIC, Australia, 21–22 August 2012; pp. 2–7.
68. Borrill, C.; Timmons, T.; van Staveren, T.; Kluyver, T.; Bauer, S. pi_ina219. Available online: https://github.com/chrisb2/pi_ina219 (accessed on 24 July 2019).
69. Furrer, T. w1thermsensor. Available online: <https://github.com/timofurrer/w1thermsensor> (accessed on 24 July 2019).
70. Smets, A.H.; Jäger, K.; Isabella, O.; van Swaaij, R.A.; Zeman, M. *Solar Energy: The Physics and Engineering of Photovoltaic Conversion, Technologies and Systems*, 1st ed.; UIT Cambridge Ltd.: Cambridge, UK, 2016; ISBN 978-1906860325.
71. Amelink, H.; Hoffmann, A.G. Current trends in control centre design. *Int. J. Electr. Power Energy Syst.* **1983**, *5*, 205–211. [CrossRef]
72. Bassous, G.F.; Hall, C.; Calili, R. Sky Images and PV Measurements. Mendeley Data. 2021. Available online: <https://data.mendeley.com/datasets/r83r6g5y6t/1> (accessed on 17 September 2021). [CrossRef]