Pontifícia Universidade Católica do Rio de Janeiro

Guilherme Fonseca Bassous

Development and validation of a low-cost data acquisition system for very short-term photovoltaic power forecasting

Dissertação de Mestrado

Dissertation presented to the Programa de Pós-Graduação em Metrologia (Área de concentração: Metrologia para Qualidade e Inovação), PUC-Rio as partial fulfillment of the requirements for the degree of Mestre em Metrologia.

Advisor: Prof. Rodrigo Flora Calili
Co-advisor: Prof. Carlos R. Hall Barbosa

Rio de Janeiro
September 2019

Pontifícia Universidade Católica do Rio de Janeiro

Guilherme Fonseca Bassous

Development and validation of a low-cost data acquisition system for very short-term photovoltaic power forecasting

Dissertation presented to the Programa de Pós–Graduação em Metrologia (Área de concentração: Metrologia para Qualidade e Inovação), of PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Metrologia. Approved by the Examination Committee

Prof. Rodrigo Flora Calili
Advisor
Programa de Pós-Graduação em Metrologia – PUC-Rio

Prof. Carlos R. Hall Barbosa
Co-advisor
Programa de Pós-Graduação em Metrologia – PUC-Rio

Prof. Delberis Araújo Lima
Departamento de Engenharia Elétrica – PUC-Rio

Profª. Karla Tereza Figueiredo Leite
Departamento de Engenharia Elétrica – PUC-Rio

Dr. Evandro Luiz Mendes
Operador Nacional do Sistema – ONS

Rio de Janeiro, September 29th, 2019

Guilherme Fonseca Bassous

Bachelor's in Environmental and Sanitary Engineering, Pontifical Catholic University of Rio de Janeiro, 2018.

Bibliography data

Bassous, Guilherme Fonseca

Development and validation of a low-cost data acquisition system for very short-term photovoltaic power forecasting / Guilherme Fonseca Bassous ; advisor: Rodrigo Flora Calili ; co-advisor: Carlos R. Hall Barbosa. – 2019.
87 f. : il. color. ; 30 cm

Dissertação (mestrado)–Pontifícia Universidade Católica do Rio de Janeiro, Centro Técnico Científico, Programa de Pós-Graduação em Metrologia, 2019.
Inclui bibliografia

1. Metrologia – Teses. 2. Metrologia para Qualidade e Inovação – Teses. 3. Metrologia. 4. Fotovoltaica. 5. Previsão. 6. Imagens celestiais. 7. Redes neurais. I. Calili, Rodrigo Flora. II. Barbosa, Carlos R. Hall. III. Pontifícia Universidade Católica do Rio de Janeiro. Centro Técnico Científico. Programa de Pós-Graduação em Metrologia. IV. Título.

CDD: 389.1

Acknowledgements

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Rodrigo Flora Calili, for guiding me for a long time and for believing in my ideas.

Carlos R. Hall Barbosa, for providing endless technical support, no matter how big or simple.

My dear parents, for stimulating me to better myself, for the unwavering support, for all the love you gave me, for being there for the good and bad times, and for being my strongest examples.

Lívia Martins, for all the love and support through the years, for being a warm hearth in the winter and a cool breeze in the summer.

José Chenú, for always extending your hand and never giving up on me.

Filipe Teixeira Silva and João Ricardo Nunes for being brothers in arms during trying times.

Abstract

Bassous, Guilherme Fonseca; Calili, Rodrigo Flora (Advisor). **Development and validation of a low-cost data acquisition system for very short-term photovoltaic power forecasting**. Rio de Janeiro, 2019. 87 p. Dissertação de Mestrado – Programa de Pós-Graduação em Metrologia, Pontifícia Universidade Católica do Rio de Janeiro.

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 PV 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. For a more accurate model, it is also necessary to use deterministic components such as temperature and angle of incidence on the panels in addition to the stochastic effect of clouds. 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. After post-processing, this data was fed into a multilayer perceptron neural network to determine its relevance to the forecasting problem. With a thorough selection of attributes and network structures, the results show very low error with a normalized good fitting 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

Metrology; Photovoltaic; Forecast; Sky images; Neural networks;

Resumo

Bassous, Guilherme Fonseca; Calili, Rodrigo Flora. **Desenvolvimento e validação de um sistema de aquisição de dados de baixo custo para previsão de curtíssimo prazo da potência fotovoltaica.** Rio de Janeiro, 2019. 87 p. Dissertação de Mestrado – Programa de Pós-Graduação em Metrologia, Pontifícia Universidade Católica do Rio de Janeiro.

Dado o recente aumento da adoção de fontes renováveis de energia, é essencial reavaliar os sistemas tradicionais de energia. A intermitência pode causar diversos problemas ligados à qualidade e eficiência energética. O objetivo desta dissertação de mestrado é desenvolver uma ferramenta capaz de subsidiar modelos de previsão solar para aplicações visando a melhoria da operação em tempo real. O atual paradigma de previsão solar sub-horária consiste em usar imagens celestiais para prever a cobertura nebulosa para curtos horizontes temporais. Visando desenvolver um modelo mais exato, é necessária a utilização de componentes determinísticos, como a temperatura e o ângulo de incidência dos raios solares, em conjunto com a modelagem dos efeitos estocásticos das nuvens. Visto que o objetivo da previsão sub-minuto é permitir que se lide com variações de alta frequência, os dados devem possuir informação condizente com estas frequências. Por esse motivo foi feita a coleta de dados por exclusão. O sistema captura dados a cada 1 s e, quando detecta uma mudança suficientemente grande na potência do painel, salva essa informação, 10 s para trás até 4 s à frente da perturbação detectada. Os dados, depois de pré-processados, foram usados para treinar uma rede neural para determinar a relevância dos dados. Com cuidadosa seleção de atributos e arquitetura de rede, o modelo apresentou boa regressão com R^2 maior que 0.93 para ambas variáveis testadas com horizonte de 60 s à frente. Concluindo, portanto, que os dados obtidos são relevantes para previsões de até 60 s à frente.

Palavras-chave

Metrologia; Fotovoltaica; Previsão; Imagens celestiais; Redes neurais;

Table of Contents

List of figures

List of tables

The most exciting phrase to hear in science, the one that heralds the most discoveries, is not "Eureka!" (I found it!) but “That's funny...”

Isaac Asimov

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). In order 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 (UNFCCC[1], 2015; IEA, 2018).

However, each of those 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 (Can Şener et al., 2018; Denholm & Margolis, 2007; Reddy & Painuly, 2004). Most of the variability components are deterministic in nature, which means they can be easily forecasted and addressed, provided it is technically possible.

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 (Dragoon & Schumaker, 2010; Mills & Wiser, 2010). These fast variations in such a short time may cause technical problems in plant and grid operation, such as voltage variations and current harmonics (Bessa et al., 2014; Denholm & Margolis, 2007; Karimi et al., 2016; Liang, 2016; Reddy & Painuly,

[1] United Nations Framework Convention on Climate Change

2004). In order to address these variations, it is necessary to forecast them. In the work done by Mills & Wiser (2010), they stressed the need for power system operators to be able to address generation and load profiles over short time-scales 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 (Diagne et al., 2013). The most widely used physical methods for short-term predictions are sky-image based (Sobri et al., 2018).

A human being is capable of seeing clouds, tracking their movements and visually estimating where they will be in a couple of seconds, perhaps even minutes. Sky-image based forecasting methods follow the same principle, where the camera represents the observer, and the computer represents their brain, capable of detecting, tracking and estimating clouds' positions in a near future. However, unlike a human's prediction, cameras and computer transform all the information from a sky image into numerical data. Computer vision is the field responsible for translating visual observations from the world (i.e. images) into useful numeric data, which can be used for detecting and classifying clouds (Cazorla et al., 2015; Chow et al., 2011; Tingting et al., 2015), estimating their speed and direction of movement (Chow et al., 2015; West et al., 2014; Wood-Bradley et al., 2012; Xu et al., 2015), and deriving new information such as insolation on ground level or PV power (Chow et al., 2011; Sobri et al., 2018).

With this data it is possible to predict PV output in short term and to start addressing solar variability in PV plant operations. This can help reach smoother power curves for plants in partial cloudiness with fast transients and consequently increase solar PV power's reliability, quality, effectiveness and adoption (Diagne et al., 2013; Mills & Wiser, 2010).

The forecasting problem can be divided in two key issues: (i) how to estimate PV power from sky images; and (ii) how to forecast the position of the clouds. By using supervised machine learning methods such as Artificial Neural Networks (ANN or simply NN), it is possible to estimate values based on input data, provided there is a relationship between the input and output data. As for forecasting clouds' positions, there are several methods for detecting and tracking objects, some specifically used for sky-image based applications (Chow et al., 2015; West et al., 2014; Wood-Bradley et al., 2012; Xu et al., 2015).

Stepping from a broad view on the issue to a more focused and local scope, Brazil has seen exponential growth in solar PV energy in the past two decades. Especially in the past decade, going from 57.8 kW in 2009 to 15.4 MW in 2014 and 2.2 GW by August 2019 (ANEEL, 2019), being projected to reach 4 GW by December 2023 (ONS, 2019).

All these non-dispatchable[2] sources have started to concern the National Electric System Operator (ONS) regarding security of the energy supply, especially on the North-eastern region of the country, where most solar farms are situated (ONS, 2018, 2019). Fortunately, regarding both yearly seasonality of renewable sources and daily behaviour of wind vs. solar, all sources complement each other (ONS, 2019). The main issue lies with intra-hour variability, making it difficult to plan ahead and being heavily dependent on local weather and micro-climate (ONS, 2018).

Short-term PV forecast for operational purposes might prove to be the solid foundation on which solar energy may grow to its full potential. It is a complex problem but with the proper tools, i.e., computer vision and machine learning, it should prove to be a feasible task.

During the development of this master's dissertation, two papers were conceived and submitted to international conferences. Their full texts are included in the annexes of this dissertation.

1.1. Research problem

With the growing use of solar PV energy, new technical issues are bound to arise, and energy quality and security are likely to pose new challenges to further PV energy adoption. Having in mind the two key issues mentioned in the previous paragraphs, the main question to be answered by this master's dissertation is:

"How to obtain pertinent and reliable data suitable for intra-hour photovoltaic power forecasting, in order to improve plant and grid operation?"

[2] The dispatchability of an energy source is whether a given generator of that source can be turned on or off as well as increase or decrease power output based on energy demand. The dispatch is made by the electrical system operators.

1.2. Motivation

Due to the increasing share of PV energy on the distribution system, technical constraints such as energy quality and high frequency intermittence in power supply are bound to arise (Bessa et al., 2014; Denholm & Margolis, 2007; Dragoon & Schumaker, 2010; Karimi et al., 2016; Liang, 2016; Mills & Wiser, 2010).

To be able to take the fullest advantage possible from RES, especially PV energy, forecasting techniques must be further advanced to successfully address the intermittency and uncertainty. According to Diagne et al. (2013), several models are able to make solar irradiance predictions within temporal and spatial resolutions high enough to produce useable information for plant operation. However, of these models, only sky image methods are able to use *in situ* physical information pertaining to cloud movement. For computer models, variability due to cloud transients has been inherently stochastic (Mills & Wiser, 2010), but the use of images change that aspect.

Such as a human being is capable of knowing when a cloud will obscure the sun, so are computers, provided they can “see” in the same way as humans do. The field of computer vision is responsible for allowing machines to extract useful information and data from images.

With more forecasting capabilities, PV plants, which have mostly operated passively in response to instant insolation, can employ more active solutions to problems due to high variability. More control capability means more efficiency, less barriers for adoption and more overall competitiveness of PV energy.

Brazil’s energy system operator has already started expressing concern regarding fast expansion of non-dispatchable sources, such as PV plants (ONS, 2018). With solar energy expected to reach 2 % of installed capacity by 2023, and because most solar farms are located on the North-eastern region, ONS fears they might negatively impact grid security (ONS, 2019). Without reliable and affordable means to forecast solar energy variations to meet operational time-horizon constraints, adoption of solar energy will be thwarted by the necessity of providing excessive backup power.

1.3.
Objectives: general and specific

To answer the proposed question, the main objective of this work is to develop a working tool to support PV energy forecast models for real-time operation applications. To achieve the main goal, the following specific objectives must be reached:

- Identify the main detrimental technical factors to intensive PV energy adoption;
- Survey the tools capable of estimating PV conversion in real-time;
- Determine which of those tools is best suited to accurately estimate PV conversion in real-time;
- Develop a data acquisition system capable of providing the required data for high-frequency PV modelling;
- Test the data acquisition system *in situ*;
- Analyse the acquired data to determine its relevance to the issue at hand; and
- Validate the results provided by the data acquisition system using a suitable model for PV conversion forecast.

1.4.
Classifying the research

1.4.1.
Regarding the ends

The research resulting in this master’s dissertation can be considered descriptive, methodological, and applied. First, it is considered descriptive due to the extensive theoretical research and description of the issues and methods addressed. It is also methodological as it assesses the suitability of different data modelling methodologies and proposes its own data acquisition methodology. Finally, it is applied because it applies the proposed data acquisition methodology in a real situation, followed by the chosen data modelling technique aiming to validate it for real world applications.

1.4.2. Regarding the means of investigation

In order to achieve the end results, the research means are bibliographical, experimental, and *ex post facto* in nature. Bibliographical due to the extensive bibliographical survey required to determine the methodology, the variables, and learning models to be used. It is heavily experimental for it revolves around acquiring real-world data and using it to generate relevant information on solar variability. And the final validation of the proposed model and variables for forecasting make it *ex post facto* as well.

1.4.3. Regarding the nature of the research

The research is quantitative in nature, given it will use real-world numeric data.

1.5. Methodology

For visual clarification of the main phases and steps taken throughout the research leading to this dissertation, a concise sketch is presented in Figure 1.

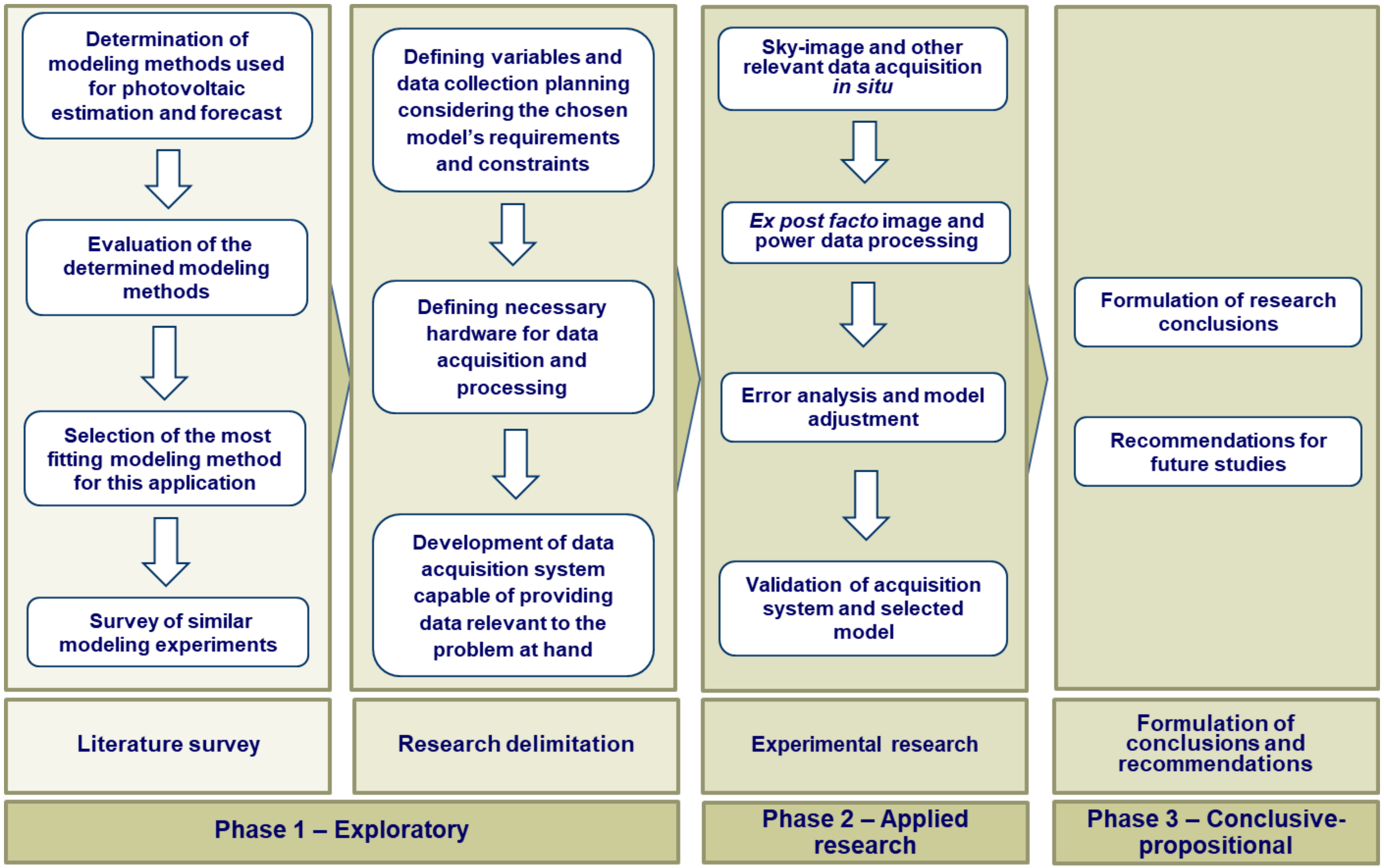

Figure 1 - Research sketch.

On the first phase, or exploratory phase, a survey of the literature pertaining to short-term solar forecast was conducted in order to determine possible models and variables to be used. After setting the direction in which to follow, it was necessary to limit the depth and breadth of the scope while always having in mind the quality of the work to be done and the expected results.

With the experimental roadmap and data acquisition system designed and developed, the next step was to start the measurements. Since the initial data provided was too raw to provide any insight, some post-processing and organizing was necessary to start implementing the selected model and analysing the results in order to validate the choices made on the exploratory phase. With the results, it was then possible to draw conclusions and make recommendations based on what was learned and new questions that had arisen during the development of this work.

Going deeper into the exploratory phase, certain conceptual milestones were required before moving ahead with the research. Figure 2 presents these milestones that laid the foundation of the applied phase. They were split into: base concepts that guided this work; empirical studies from which to draw invaluable insight into testing the system and model; the steps taken in order to choose what data acquisition and modelling tools were most prone to yield solid results; and finally the culmination of the previous steps in the applied phase.

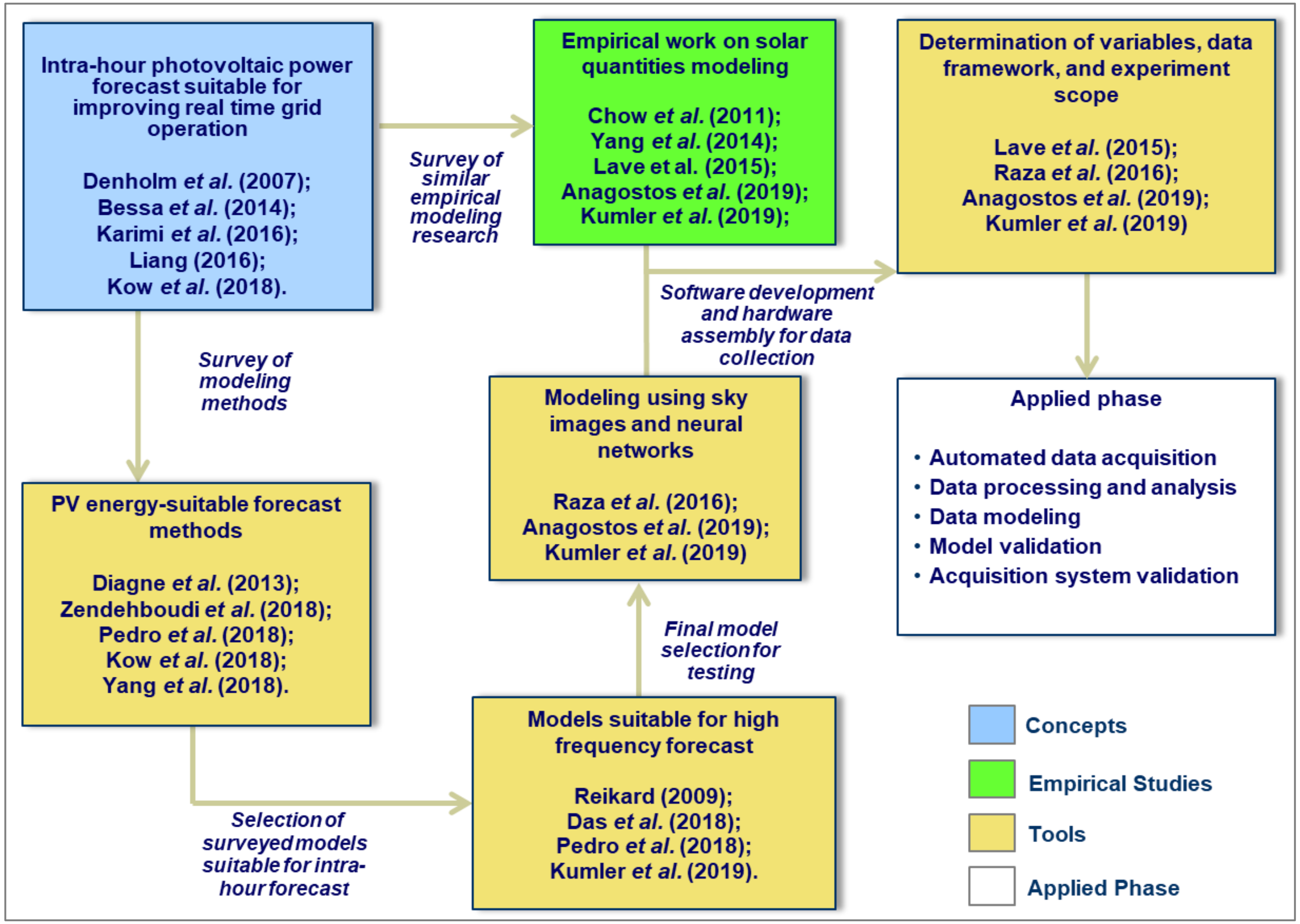

Figure 2 - Conceptual Mapping

1.6.
Structure

This dissertation is divided into five chapters including this introduction and two annexes.

Chapter 2 presents the photovoltaic concepts that cause the limitations addressed in this work. It is divided in five subsections: 1) a brief explanation of PV conversion; 2) enumeration of the short-term variability-originated problems that hinder the adoption of PV energy; 3) role of short-term forecasts in reducing problems caused by high-frequency variability; 4) an introduction to short-term solar forecast models; 5) brief history, and state of the art in solar forecasting using sky-imagery that based the work of this dissertation.

Chapter 3 presents the materials and methods used in this work, the description of the experiment and result analysis. It also contains detailed descriptions of the hardware assembly used in said experiments.

Chapter 4 presents the results obtained by the experiments detailed in the previous chapter. This includes both the raw data obtained and more detailed analysis to provide information on the data acquisition system, experiment premises and overall regarding very short-term solar variability.

Chapter 5 presents the conclusions reached by this master’s dissertation. These conclusions will hopefully help guide future work to further PV forecasting for plant and grid operation purposes.

Annex 1 contains a paper presented in the 2018 International Conference on Energy Engineering and Smart Grids. This paper uses sky images for PV power estimation.

Annex 2 contains a paper presented in the XIII SEMETRO addressing the information presented in Chapter 3 of this dissertation.

2
Photovoltaic energy conversion and short-term forecasting

This Chapter presents the photovoltaic concepts that lead to the limitations addressed in this work, starting with the theory behind PV conversion. Given the properties of PV conversion, in particular the lack of inertia, the following section addresses the limitations caused by high-frequency variability of solar irradiance on the panels' surface. Lastly, the role of accurate short-term forecasting in reducing the impacts of fast cloud transients in the electric system is presented, in order to introduce the following chapter and justify the efforts in researching better ways of short-term forecasting of PV conversion.

This Chapter also addresses the theory behind short-term photovoltaic forecasting and the necessity for high fidelity modelling in achieving accurate forecasting. These are all key components essential to the increase of safe and reliable PV usage, as seen in the previous Chapter.

2.1.
Photovoltaic conversion

Amongst the methods for harvesting solar energy, photovoltaic (PV) conversion is the most used for generating electricity. With an estimated 402 GW installed capacity by 2017 it is also the leading added capacity in the world with the addition of 98 GW in 2017 alone (REN21, 2018).

The energy conversion occurs when light shines on structures that present a junction between different semiconductor materials, the photons being partially absorbed and creating charge carriers. That is, the energy from the photons causes electrons to be excited and, provided there is enough energy, jump from the valence band to the conduction band, thus creating a void in the valence band. This void, called hole, has the behaviour of a positively charged particle and their combination is called electron-hole pair (Smets et al., 2016).

If left alone, the electron-hole pair tends to return to equilibrium and will recombine, with the energy dissipating either through radiative or non-radiative recombination. To harness this energy in the form of work in a circuit, semipermeable membranes must limit the flow of electrons in one direction, forcing them to pass through a circuit in order to recombine with the holes. The most common solar cell designs use n-type and p-type materials to make these membranes (Smets et al., 2016; Würfel, 2005).

Crystalline silicon based solar cells are the most commonly used type, amounting to 90 % of the world market (IEA, 2014). These solar cells are built with an emitter layer, most commonly made of a thin n-type silicon layer, on top of a thicker p-type silicon layer, also called absorber. N-type silicon is silicon doped[3] with electron donors, becoming conductive to electrons but not holes, whereas p-type silicon is doped with electron acceptors, becoming a hole conductor. Metallic contacts on each of the terminals connect them in order to generate electrical current with electrons flowing from the n-type terminal to the p-type, creating useful power in the circuit. These small cells are then connected in series and parallel to form a PV panel with the desired voltage and current (Smets et al., 2016; Würfel, 2005).

PV systems contain at least a module and load on the circuit, varying in size and application, mainly whether the system is connected to the power grid. Off-grid systems are entirely dependent on solar power and may include simple loads such as a water pump or may have more complex design with batteries and charge controllers for storage, and inverters to deliver both alternating and direct current (AC and DC respectively). On the other hand, grid-connected systems convert the DC from the PV panels into AC and feed it into the grid or grid-connected homes and buildings. Grid-connected systems vary in size from residential installations to large PV power plants (Smets et al., 2016).

To comprehend some of the limitations from PV systems, it is crucial to understand the behaviour of solar irradiance and its effects on photovoltaic conversion. The solar radiation that reaches the planet does vary, however it is not within the scope of this work to consider these variations which are minimal if compared with the main sources of variability.

The first source for variation of solar irradiance is due to the translational movement of the Earth around the sun. The yearly cycle of seasons occurs, in part,

[3] Doping is the process in which impurity atoms replace the atoms of a semiconductor. They can be either donors, which have more valence electrons than necessary for chemical bonding in the lattice, or can be acceptors, which have less valence electrons than necessary for chemical bonding.

due to changes in the angle of incidence of solar radiation on the planet. These changes are largely proportional to latitude, the furthest from the Equator line the harshest being the changes in the angle of incidence.

The second source of variation happens because of the rotational movement of the planet. Throughout the day, for a fixed position on the globe, the angle of incidence, as well as the position of the sun on the sky, varies. The variation of the position of the sun is described by two angles, altitude and azimuth. Solar altitude, or elevation, is the angle between the horizontal plane where the observer is standing and the position of the sun. On the other hand, the solar azimuth is the angle between the geographical North and the projection of the sun's position on the horizontal plane (Smets et al., 2016).

Both these sources of variation are deterministic in nature and can be calculated by well-known equations for any given time and place. Now, getting into the scope of this dissertation, the behaviour of solar irradiance begins to change when entering the atmosphere. The atmosphere acts as a filter for the incident light, and its filtering behaviour depends on atmospheric composition, as well as obstructions along the path of travel. Until now only the broad term "irradiance" has been used, but when addressing the change of its behaviour due to the atmospheric filter more specific terms must be used to accurately represent its properties (Myers, 2016; Smets et al., 2016).

Direct Normal Irradiance (DNI) represents the beam of radiation that hits the surface of the planet in a straight line from the sun. The remnant of the difference between the extra-terrestrial radiation and the DNI is what gets scattered or absorbed by the atmospheric components, such as gases, water molecules, particles, and clouds. The changes in density and presence of turbulence along the path of light also affects the beam radiation (Myers, 2016; Smets et al., 2016).

Global Horizontal Irradiance (GHI) is the radiant flux (in watts) reaching a certain horizontal area (thus given in Wm^{-2}). Given that the DNI travels in a fixed direction the beam will not necessarily be perpendicular to a horizontal surface, and according to Lambert's law (Myers, 2016), the radiant flux per unit area (Wm^{-2}) on a given surface is proportional to the cosine of the angle of incidence. In the case of GHI, the DNI component is dependent on the cosine of the solar elevation angle. The missing irradiance component of GHI is the Diffuse Horizontal Radiation (DHI), which occurs due to the scattering of DNI by the atmosphere. These quantities can be measured by a pyranometer for GHI or DHI, if shaded from direct irradiance, and a pyrheliometer for DNI (Smets et al., 2016).

2.2. Limitations of PV systems

Since PV conversion depends on the intensity and angle of incident light on the panel surface, one of the main barriers to its adoption is the variability of insolation. Sunlight has many different components linked to its variability. First, not all light reaches the planet in the same way because, due to the translation movement around the sun, the angle of incidence in which solar rays reach the surface of the planet varies throughout the year, more extremely the furthest away from the equator line. Secondly, due to the rotation movement of the planet, light varies periodically, with small daily changes due to the translation movements, every 24 hours.

These variations can be easily calculated and considered in energy resource management planning. The third variability aspect is due to atmospheric conditions that hinder the passage of light through it, most notably the presence of clouds and their strong impact on available DNI. Fast cloud transients can affect global insolation in a single PV plant by up to 80 % of the theoretical clear-sky insolation in a 1-minute interval (Mills & Wiser, 2010). This information extracted from a 1-minute interval solar database draws attention to the need for higher frequency solar data in order to understand and address these fast changes in insolation due to cloud transients. This subject is well demonstrated in the work by Lave et al. (2015).

The consequences from variations of this intensity and frequency may cause several problems on the power grid. Regardless of solar availability for PV conversion, the load must be met, and to achieve this there must be backup for when PV plants are not able to meet load demands. These backup plants increase operation costs for any RES, especially PV solar, which suffers the most from high frequency and high intensity fluctuations. Conventional generators cannot be expected to meet these ramps, especially in response to cloud transient events, resulting in energy quality issues proportional to the penetration of PV energy in the system and the generator profile, either distributed or concentrated (Karimi et al., 2016; Liang, 2016).

A key concept in understanding the impacts of RES variability in power economics is the capacity factor. It represents the amount of energy generated in proportion to the nominal capacity of that generator for a given period of time. The capacity factor plays an important role in determining the value of an intermittent energy source (Lamont, 2008). Due to several reasons, such as meeting ramp requirements and energy quality standards, intermittent sources such as solar may be curtailed[4] in order to properly fit within grid operation (Bird et al., 2014). Energy curtailment negatively affects the capacity factor, which in turn results in lower value for that generator, reducing its attractiveness as an investment in capacity expansion.

Several different energy quality issues may arise from increased integration of PV energy into the power grid. Among them, some are caused or intensified by the aforementioned high frequency variability, most notably voltage fluctuation and unbalance, frequency fluctuation, islanding operation challenges and stress on distribution transformer (Karimi et al., 2016; Liang, 2016). In order to address these issues and permit a larger share of PV energy, power electronics and energy storage systems (ESS) seem to be the most common solutions that do not involve energy curtailment (Denholm & Hand, 2011; IEC, 2011; Karimi et al., 2016; Liang, 2016; Petinrin & Shaabanb, 2016). However, to further strengthen these systems and generate better and cheaper power, being able to reduce the uncertainty of the solar resource variability is a key step, in other words, accurate forecasting of fast cloud transients is essential (Bessa et al., 2014; Diagne et al., 2013).

As progressively smarter grids take the place of conventional ones, more control strategies will arise, being an example the scheme proposed by Maleki & Varma (2016), Varma & Salehi (2017), and Varma & Siavashi (2018). Smart grids use information and communication to improve grid operation and integration (Hossain et al., 2016). By having accurate short-term forecasting data available for both plant and grid operators, PV energy can be more reliably and cost-effectively adopted at a larger scale (Bessa et al., 2014).

[4] Energy curtailment is an energy management technique consisting of reducing the output of a generator to adequate it to the grid demand. In the case of non-dispatchable sources, power output cannot be simply reduced, so most likely that energy will go to waste.

2.3. Reliability through forecasting

Forecasting is essential in countless applications including energy management and market. As mentioned by Mills & Wiser (2010), system operators need better information about the stochastic behaviour 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 dissertation is on very short-term forecasting to bolster PV plant operation capabilities, reliability, grid integration, and grid operation in a scenario of high penetration. Table 1 presents the terminology regarding forecasting horizons and their applications.

Table 1 – Forecast horizon categories, granularity and applications

Category	Time Horizon	Granularity	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 min to 5 min	Load following Grid operation planning
Medium-term	1 h to 6 h	Hourly	Load following Grid operation planning
Long-term	Day ahead	Hourly	Unit commitment Transmission scheduling Day ahead markets

Source: Diagne et al. (2013); Stefferud, Kleissl & Schoene (2012)

In the work done by Diagne et al. (2013), 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 Diagne et al. (2013) name it, also encompasses regression models (ARMA, ARIMA, CARDS) 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 (Barbieri et al., 2017; Schmidt et al., 2016). Sky-image models keep improving the reliability of very short-term forecasting, shown in the work done by Sobri et al. (2018). This tendency points towards the superiority of using sky-images over what Diagne et al. (2013) refer to as statistical models. In the study conducted by Kow et al. (2018) 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 (Bessa et al., 2014; Denholm & Hand, 2011; IEC, 2011; Karimi et al., 2016; Liang, 2016; Petinrin & Shaabanb, 2016; Varma & Salehi, 2017). The results presented by Kow et al. (2018) depict the beneficial effect that short-term forecasting can have on the operation of PV plants.

Having in mind the information presented in this section, it is safe to assume that very short-term forecasting can have a positive impact in plant level operations in order to increase reliability, but also to include it in longer time-scales grid planning. If the variability component is reduced, solar power can be modelled as a more reliable power source and can even be used as a power management tool.

2.4.
Very short-term photovoltaic 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 behaviour of the studied phenomenon. To that end, many different models are capable of describing or learning the behaviour of PV conversion, some more accurately than others.

Within the statistical category used by Diagne et al. (2013), 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 $\widehat{\boldsymbol{X}}_{\boldsymbol{t+1}}$ to be best described by its value at a previous time $\boldsymbol{X_t}$. In this case, the modelling and prediction are one and the same, it doesn't take into consideration the several variables that affect the behaviour 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. (2013) use historical data either from irradiance or clear-sky index to make predictions. While better than the previous, naïve, predictor 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 minutes to hourly forecasts (Reikard, 2009). In the 5 minutes resolution, results were mixed among the models tested by Reikard (2009), but the 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. (2018).

Switching over to the non-linear models addressed by Diagne et al. (2013), neural network models attempts to simulate the computational and learning process of the human brain (Haykin, 2008). 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 (Haykin, 2008).

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 (Haykin, 2008). They are well suited to model complex problems, especially when involving complex relationships between the variables (Haykin, 2008), such as forecasting energy conversion dependant on cloud passage, location, time and meteorological variables (Das et al., 2018; Raza et al., 2016).

As mentioned previously, neural networks start faring better against other forecast methods at higher temporal resolution (Reikard, 2009), 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 minutes (Kumler et al., 2019; Zendehboudi et al., 2018), which still falls short of the necessary frequency to properly characterize the local solar variability (Lave et al., 2015). Still within the 5 minutes time horizon, sky images can be used to boost forecasting accuracy when coupled with machine learning models and historic irradiance or power data (Pedro et al., 2018).

The conclusion that can be drawn from the consistent number of time-series models limited to the 5 minutes 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. (2011) and Kow et al. (2018), add components of physical and geometrical modelling 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 (Yang et al., 2018).

Throughout the research process that laid the theoretical foundations of this dissertation, 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 modelling, and the materials and methods used in the pursuit of these objectives.

Table 2 – Important works that shaped this master's dissertation.

Work	Objective	Materials and Methods
Chow et al. (2011)	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)	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)	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 & Coimbra (2013)	Forecast of 1-min-average DNI 3 min to 15 min ahead	TSI images every minute + 30 second 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)	Forecast of 1-min-average DNI from 3 min 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)	Forecast of DNI from 0 min to 20 min ahead in 10 s resolution and updated every 10 s	Sky images + 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)	Forecast of 10 min ahead GHI and DNI	Images from 2 sky cameras every 60 s + Irradiance every 30 s Adaptive threshold cloud detection Gridded cloudiness + time lagged irradiance as inputs for NN
Alonso-Montesinos & Battles (2015)	Modelling of GHI, DNI, and DIF	TSI images every 60 s + GHI + DNI every 60 s Correlations of digital image channels to model irradiance

Table 2 – Important works that shaped this master's dissertation (cont.)

Alonso-Montesinos et al. (2015)	Forecast of GHI, DNI, and DIF from 1 min to 180 min at 15 min resolution	Cloud tracking using cloud motion vectors (CMV) Pixel-wise cloud detection Pixel-wise irradiance using correlation of digital channel information
Cazorla et al. (2015)	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
Chu et al. (2015b)	Forecasting of prediction interval for 1-min-average DNI 5, 10, 15, and 20 min ahead	Sky 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. (2015c)	Forecast of PV power 5, 10, 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)	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 auto-regressive forecast model based on the modified covariance method
Pedro & Coimbra (2015)	Forecast of GHI and DNI from 5 to 30 min ahead	5 min averaged irradiance data 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)	Forecast of GHI from 1 min 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

Table 2 – Important works that shaped this master's dissertation (cont.)

Cervantes et al. (2016)	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)	Cloud optical depth modelling	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
Rana et al. (2016)	Forecast of PV power from 5 min 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)	Forecast of 1-min-average clearness index from 1 min to 15 min ahead	GHI, DNI, and DHI measurements every 60 s Modelling of solar zenith-independent clearness index SVR, persistence and autoregressive models of different orders used for forecasting
Schmidt et al. (2016)	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 + GHI every 1 s from 99 pyranometers + CBH measurements averaged over 10 min Area of study of 10 km x 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)	Forecast of PV power and GHI 1, 5, 10, 30, 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 Auto Regressive (AR) estimation Comparison between with and without exogenous inputs
Ai et al. (2017)	Forecast of 30-s-average GHI 1, 2, 3 min ahead	Sky images every 30 s 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)	Forecast of 1-min-average DNI map 15 min ahead with up to 10 m x 10 m spatial resolution	Stereoscopic 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

Table 2 – Important works that shaped this master's dissertation (cont.)

Cheng (2017)	Detection of irradiance ramp down events 5, 10, 15, and 20 min ahead	Sky images every 60 s + 1 min averaged GHI Cloud detection and tracking through feature point clusters
Elsinga & Van Sark (2017)	Forecasts of 1 min average GHI from 1 min 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, 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
Ni et al. (2017)	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)	Forecast of GHI 10 and 15 min ahead	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)	Forecast of PV power 30 s ahead coupled with mitigation system	GHI every 1 s + ambient temperature every 1 s and PV system modelled 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)	Forecast of 1-min-average GHI from 0 to 15 min ahead	Cloud segmentation, detection, and georeferencing using 4 sky cameras 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 & Gueymard (2019)	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)	Forecast of GHI 5, 15, 30, and 60 min ahead	Cloud albedo and fraction modelling 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

2.5. Very short-term photovoltaic forecasting with sky images

The trend in researching sky-based approaches to very short-term solar forecasting began with the work by Chow et al. (2011), despite not being the first to approach the subject (Yang et al., 2018). 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 (Yang et al., 2018). In more recent years, other lower-cost alternatives have been developed for the specific purpose of estimating solar quantities (Richardson et al., 2017; Yang et al., 2018). 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 (Cervantes et al., 2016; Richardson et al., 2017).

Amongst the already mentioned advantages, specifically designed systems have proven to yield superior results to other non-specific sky imaging systems (Gohari et al., 2013; Richardson et al., 2017; Urquhart et al., 2015). Most likely this superior performance is due to the higher data acquisition frequency which provides better insight into local short-term solar variability (Lave et al., 2015). 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.

Given these advantages presented by sky-image based forecasting, the experimental work developed in this dissertation has made use of this framework. More specific details on the sky-imaging system and experimental work are further explained on the next Chapter.

3
Materials and methods

This Chapter presents and explains the materials and methods used for the applied phase. First is the data acquisition system developed as one of the specific objectives of this master's dissertation. Following, the experimental work will be thoroughly explained to provide a better understanding of the resulting data and analysis.

3.1.
Data acquisition system

As pointed out by Richardson et al. (2017), scientific grade sky imagers are far too costly to allow wide geographical dispersion, which is necessary when it comes to solar forecasting in larger operational scales. Therefore, the data acquisition system (DAS) proposed and developed in this dissertation will use affordable, off the shelf components in order to assess the quality of data and information provided.

Following the reasoning laid by the previous Chapter, the system will be centred around sky images and their relationship with PV power. Subsection 3.1.1 presents the information surrounding the hardware for the DAS.

3.1.1.
DAS hardware

Having in mind the goal of keeping the costs low, the system is built around a single board computer (SBC) (Richardson et al., 2017). For this application a Raspberry Pi 3B+ was chosen due to low cost and easy access in Brazil, with ample online documentation and fully integrated and optimized Linux based systems. It boasts a 1.4 GHz quad-core CPU with 1 GB RAM memory, integrated wireless connection, General Purpose Input/Output (GPIO) pins for controlling other sensors that might be necessary and USB ports for connecting and controlling the camera.

Moving on to the sky camera, the chosen model for this application was an ELP-USBFHD01M-L180 camera. It consists of only the printed circuit board (PCB) module with a CMOS OV2710 sensor able to provide images with 1920 pixels by 1080 pixels resolution. This camera comes with a 180° field of view (FOV) C-mount lens.

Ideally, the electrical quantity related to solar energy would be provided by multiple sensors in a PV power plant, however, due to the limited amount of time available to develop this dissertation, no plants capable of providing data at the required frequencies were secured. So, in order to link the sky images with a solar quantity, a 20 cm by 15 cm photovoltaic panel with 6 V and 1 A nominal rating at 25° C was incorporated to the system.

Voltage and current measurements from the solar panel were provided by an Adafruit INA219 DC sensor. This sensor was connected to the GPIO pins on the Raspberry Pi. The measurement circuit and acquisition software will be addressed in the following sections.

When working with solar panels, temperature is a very important variable to PV conversion efficiency (Smets et al., 2016). A Maxim Integrated DS18B20 temperature sensor was placed on the bottom of the solar panel, housed by an aluminium heat exchanger in contact with the panel. This ensemble was enveloped by a dense foam to reduce heat exchange between sensor and atmosphere, working to provide the best possible information on panel temperature. This sensor was also connected to the GPIO pins on the Raspberry Pi.

The Raspberry Pi is supplied by a 5 V/3 A DC power supply connected to a 110 V AC plug. Aside from the measurement components the housing units containing the electronic components are cooled using two computer cooling fans. The basic 3D model for the DAS is presented in Figure 3 to aid in visualizing the layout of the equipment.

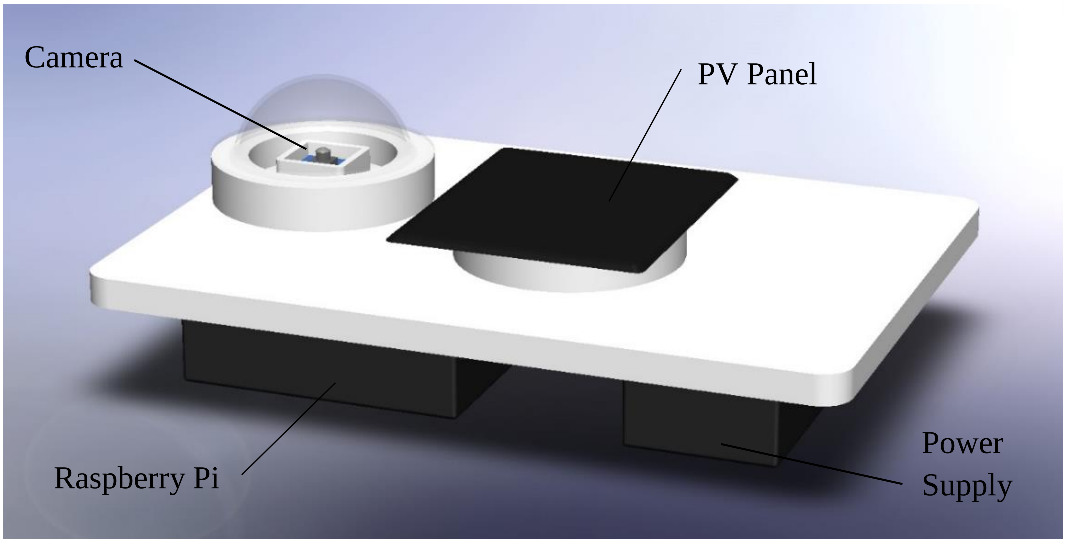

Figure 3 – 3D model of the DAS.

In Figure 3**,** the top side of the support structure holds the sky camera enclosed by an acrylic dome for protection from the elements, and beside it the PV panel. Whereas on the bottom hang two housing units, the right one contains the power supply and DC-DC converter, the left houses the SBC and INA 219 sensor.

3.1.2. Measurement circuit

As mentioned on the previous section, the DAS encompasses a voltage and current sensor as well as a thermometer. This section presents the measurement circuit with the connections between the sensors, the GPIO pins, the solar panel and the circuit load. 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 aluminium 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 (Smets et al., 2016).

As for the INA 219 sensor, it measures both circuit voltage and determines current by measuring voltage across a 0.1 Ω 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 kΩ to stabilize the signal when not communicating with the Pi. Figure 4 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.

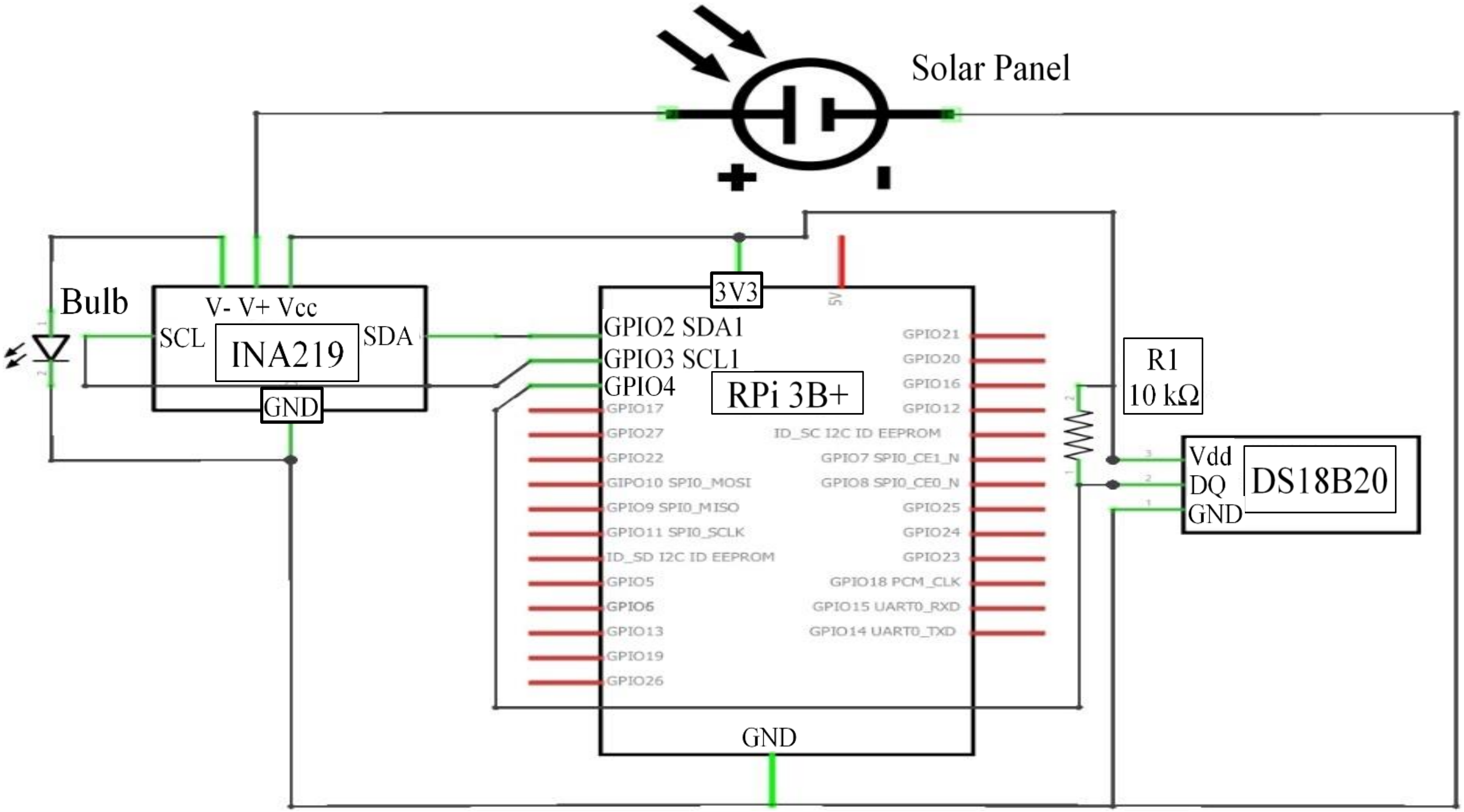

Figure 4 – Measurement circuit schematics.

3.1.3.
DAS software

This section presents information regarding the software used for the measurements from both sensors and the camera. More specific numerical information about thresholds will be presented later, along with other experimental information.

The Pi was running Raspbian Stretch, a Linux based OS specific to it, as well as OpenCV 4.0 and Python 3.7. OpenCV is a highly popular, open source, computer vision library available for several different platforms. Since the OS is highly optimized for use with the Python programming language, the chosen OpenCV 4.0 distribution was for Python, as well as the libraries for both sensors.

There exist several libraries for INA 219 and DS18B20, but amongst the most notorious, and with most documentation available are “pi_ina219” (Borrill et al., 2019) and “w1thermsensor” (Furrer, 2019). These libraries facilitate the access of information from the GPIO pins through high-level programming languages, in this case Python.

Considering the high-speed intermittency caused by fast cloud transients and the approximately point-like dimension of the solar panel used in the DAS, a frequency of 1 Hz was used for data acquisition. This high frequency has been proven to provide highly useful information on very short-term solar variability (Lave et al., 2015). One downside from such an elevated acquisition frequency is the sheer amount of data it is able to generate, and this may be a hindrance when working with images. So, in order to attempt to provide both high-frequency and high-quality data surrounding high-amplitude fast variations in PV power output, the acquired data was screened for possible variation events before being saved.

This strategy is called acquisition by exception, it is based on the principle of: if there is no drastic change to the system, there is no need to record the data. This is not applicable to every situation, but it is most welcome when studying variability. In practice, the acquisition software continuously acquired data during 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 comprised 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} – 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 in order 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. Figure 5 presents a flowchart of the decision process and data flow from the DAS software.

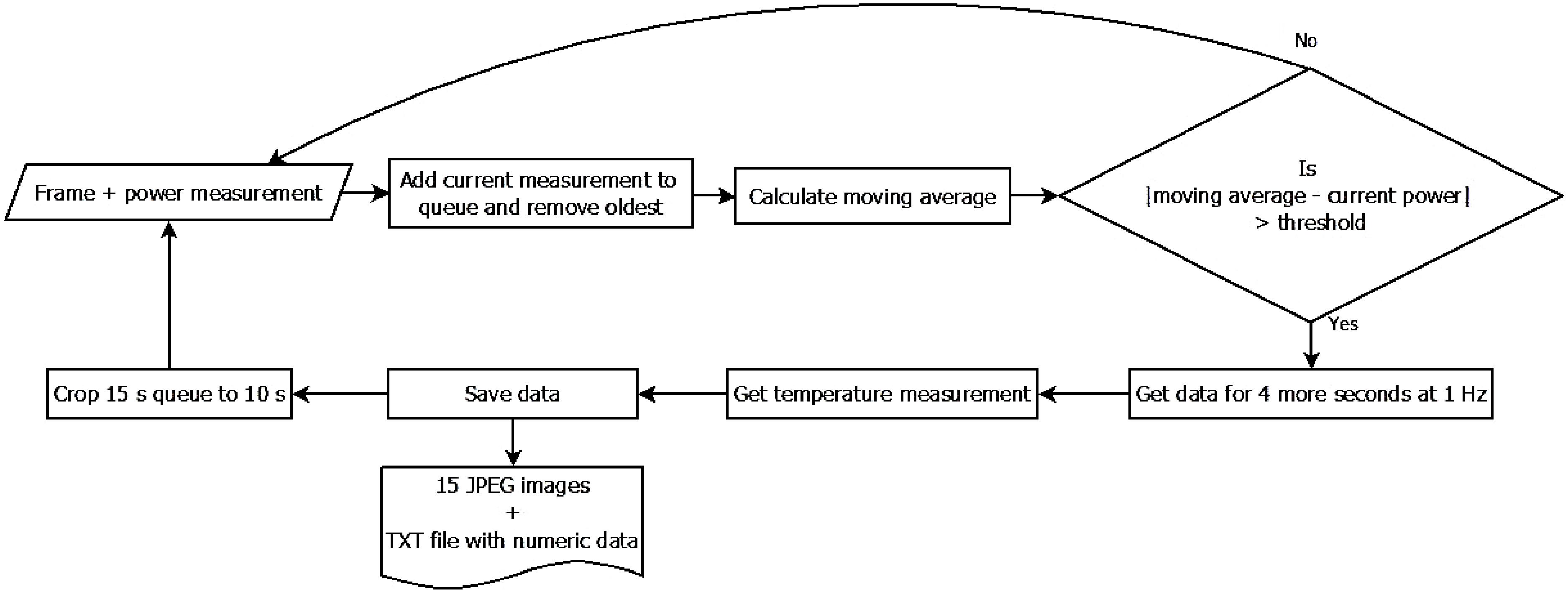

Figure 5 – Flowchart of decision process and data flow within DAS software.

3.2.
Experimental phase

In order to validate the data acquisition system, firstly it would be necessary to put it to use and start acquiring data, which would then be analysed in order to validate the designed system. The equipment was placed on a residential balcony located at 22.9354° S Latitude and 43.1756° W Longitude at an altitude of 46 metres. Coordinates and altitude were obtained using Google Earth. Since the equipment could not be placed on the roof due to safety hazards, the balcony had to suffice, despite having a limited view of the sky, spanning solar position from early morning to about 12 h 30 min.

As the experiment was carried out, some modifications to the program and physical structure of the device had to be made. Starting with the protective acrylic dome, due to repeated exposure to rain in the early testing phase, water managed to infiltrate the 3D printed support designed to hold the camera and dome. This water, when condensed on the dome, made it impossible to visualize the sky, rendering the images useless. The dome had to be removed in order to continue with the experiment, but the DAS could no longer be indefinitely placed outside, so it was removed whenever there was risk of rain.

The official testing phase spanned from February 25th, 2019 at 9 h 21 min 48 s to March 23rd, 2019 at 8 h 2 min 45 s. Despite the long span, records were taken for only 12 of those days non-consecutively due to very unstable or completely clear weather. Another hardware modification was the addition of a neutral density filter in order to darken the images and help with calibration of the camera's exposure parameter. More details on the reasoning behind and results from the addition of the filter will be presented on the next Chapter, containing the results from the experiment and analysis performed on the acquired data with the goal of validating the acquisition system.

4
Results and data analysis

This Chapter presents the data obtained from the DAS testing phase, as well as all the analysis conducted on them. The findings resulting from several steps of analysis are also discussed.

4.1.
Acquired data

Once finished, the experimental phase yielded 500 events, the distribution for each of the successful acquisition data distributed in days is being presented in Figure 6.

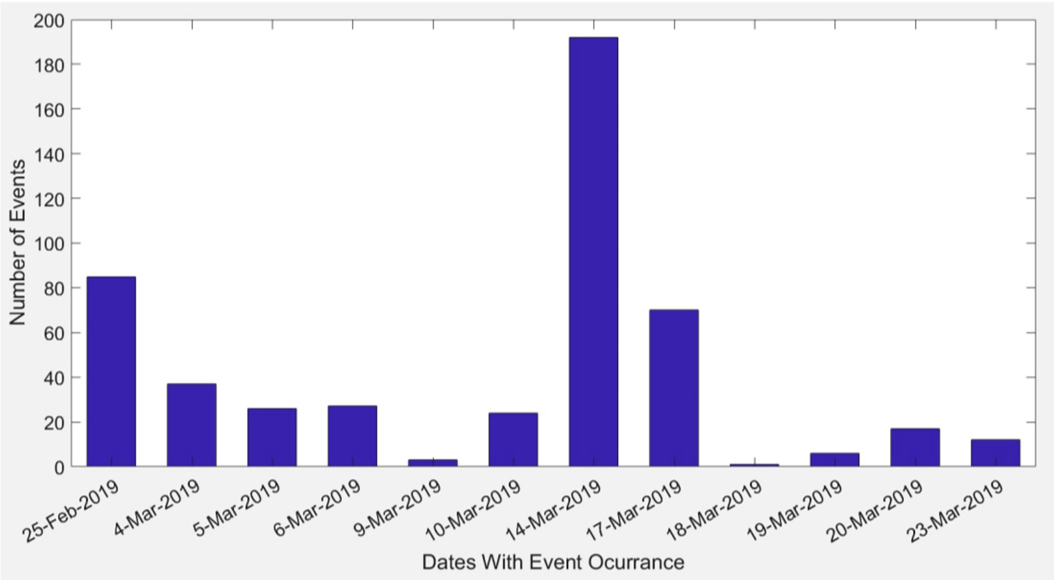

Figure 6 – Distribution of events among successful acquisition days.

Some example images for four different days are presented in Figure 7 to better contextualize the data processing steps.

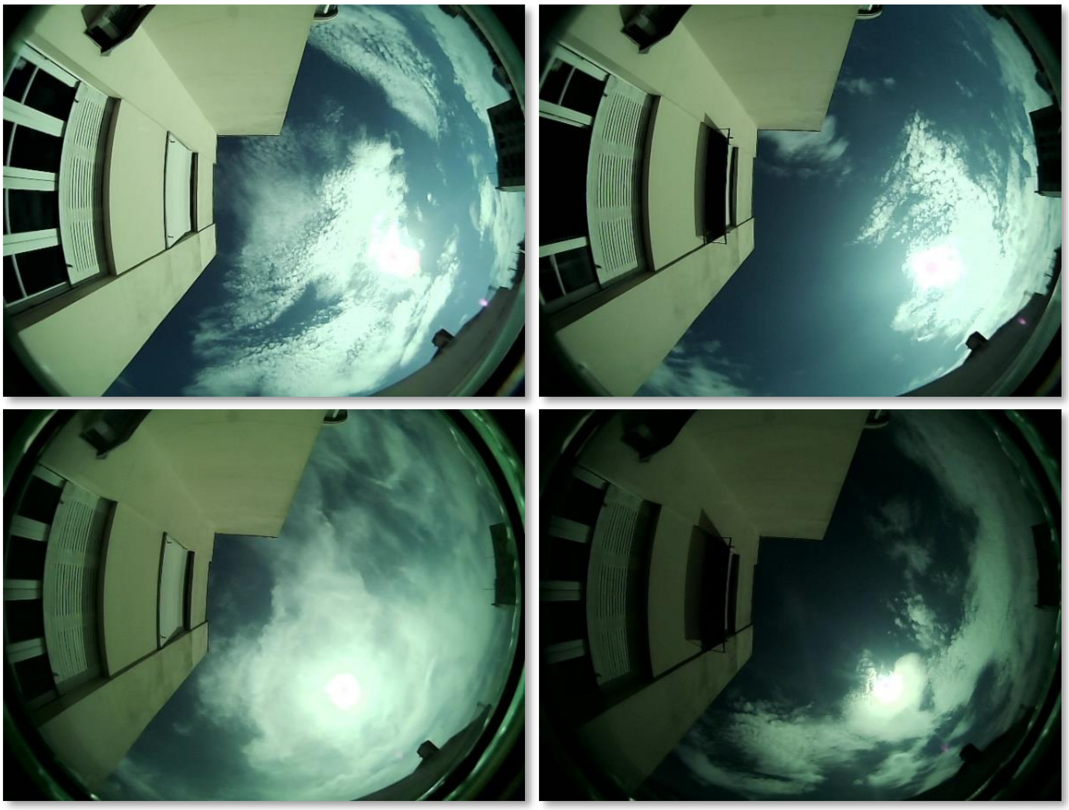

Figure 7 – Raw images obtained on four different days. Starting top-left and moving clockwise: 25/02/2019 – 06/03/2019 – 14/03/2019 – 19/03/2019.

There is a clear difference between the two top and two bottom images in Figure 7 due to the placement of the neutral density filter placed on the 14th of March. It was done for two reasons: i) to reduce the brightness of the images; ii) to protect the camera since the acrylic dome had to be removed. An important impact of its placement is seen on these images: the size of the saturated area around the sun is smaller, enabling to obtain more information closer to the sun.

The first validation of the 1 Hz data is made visible by Figure 8, the plot of power measurements throughout the measurement period of March 14th, the day with most abundant data. As shown in this graph, it is important to perceive that lower resolutions would miss important high-amplitude and high-frequency variations. The graph has non-contiguous lines because of the acquisition by exception resulting in the previously defined events. However, some sections present some longer contiguous lines, that is the result of an event detection occurring while part of the previous event was still in the data queue resulting in overlapping data.

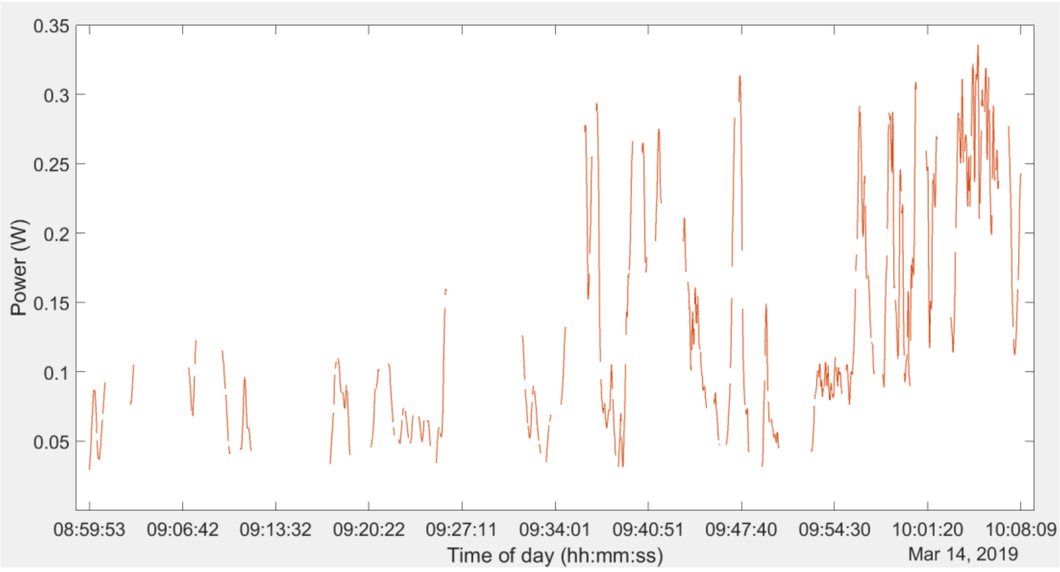

Figure 8 – All power measurements for March 14th.

Next section starts introducing the data analysis processes conducted on the data presented here.

4.2. Data processing and analysis

The data processing and analysis necessary for developing this master's dissertation is presented in the following subsections.

4.2.1. Preliminary visual analysis

In processing the obtained data, the first step was to import it into the Matlab environment from the text files and convert it into proper structured data. Since the presence of buildings in the images would only be detrimental, a simple black mask was applied to the images in order to remove the undesirable parts without cropping the images. The initial goal was to study the data within the events and perform very short-term forecasting based on image subtraction information. By subtracting images, the changes from one frame to another become apparent, and this can be done by using subtracting one from the other in pixel-wise operations per digital channel[5] but capping the subtraction at zero, so the pixel values remain within the

[5] Each digital channel is a matrix corresponding to the red, green, and blue intensities from the image.

0-255 interval. Several event samples from different days were processed using this image subtraction method and visually analysed. The subtraction operation was performed on a grayscale image obtained by converting the red, green and blue (RGB) channels into one “black and white” image. The goal was to be able to visualize the cloud border while it approached the sun, so colour information would not be critical. Some image subtraction examples are presented in Figure 9. In order to ease visualizing the differences between subtracted frames, the resulting matrix was thresholded at an intensity of 5, meaning any values below 5 would be multiplied by zero. After thresholding the contrast was increased for the values to become more apparent to the human eye.

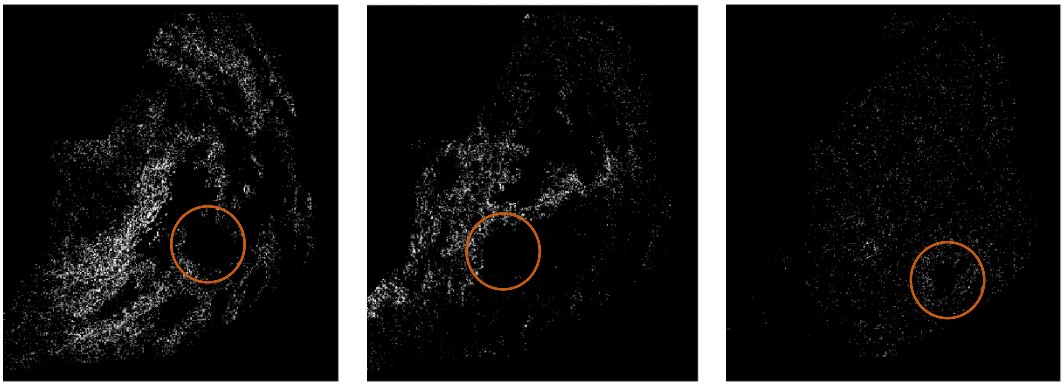

Figure 9 – Image subtraction results for different days with the solar region highlighted.

The orange circles in Figure 9 highlight the solar region. It is possible to tell apart what seems to be some of the edges of the cloud formations, especially in the two leftmost examples, which belong to the same events as the first two images in Figure 8. However, those apparent edges are situated in the middle of the cloud formation.

Another valuable information to be obtained from those samples is the difference in intensity between the two leftmost and the rightmost sample. All three images were subjected to the same processing, thresholding and contrast increase, but the rightmost example is from an image obtained after the neutral density filter was put into place. It may seem to provide less information to the naked eye, but the saturated region within the highlighted area is also smaller, meaning that the camera was able to obtain possibly useful information closer to the sun. For very short-term forecasting this may prove important to enhancing accuracy in intra-minute forecasts.

These preliminary results indicated that visual information from within a single event with 15 s duration was insufficient to provide useful information despite it yielding important insight on power variations. The acquisition by exception is important to reduce the amount of data to be stored but might need to be adjusted in order to optimize the forecasting.

In this sense, the acquired data that was originally structured event-wise was reframed into the calendar days, due to inter-day specificities. Thus, the 86,400 seconds contained in each day were populated (sparsely) by the data contained within the events. The goal in the second phase of analysis, was to determine for which time horizon is it possible to obtain information from images that are correlated with power measurements.

4.2.2.
Correlation analysis

As mentioned in the previous section, the original event structure was put aside, so for the remainder of the data analysis, it was separated only by day due to intra-day specificities. To determine a relevant time horizon for the correlation of images and power data, the first step was to bring the event information together in one data structure to generate interval-based data derived from the raw data. Generating these subsets was done by defining the time steps and then building the datasets for each step. For example: for a time-step of 10 s the values used would be the power difference between $t_{0\text{-}10s}$ and t_0, as well as the actual measurements in both instants and measurement values for panel temperature at t_0. For the image features, energy[6] from a circular region of interest (ROI) around the sun was used. Firstly, image subtraction was performed between the two images corresponding to the time-step, then the energy value was calculated within the ROI of the subtracted image. This was done for each of the three digital image channels, red, green and blue.

Due to the data being sparce, meaning it had empty gaps, some data points did not have a pair one time-step ahead, so these points were disregarded. Since cloud speed impacts cloud position between two different frames and it was an unknown variable, different datasets were made for different ROI raddi. The

---

[6] Energy in image processing is the sum of pixel values from an image or a region of interest.

temperature values missing due to making only one measurement per event were linearly interpolated between two measured values.

In order to determine the solar position in the image, a MATLAB-implemented function was used to determine the solar elevation and azimuth angle for the measurement times and location (Koblick, 2009). With elevation and azimuth angles, the solar position on the image was determined by projecting the spherical coordinates of the sun onto the sensor plane using simple trigonometric functions. Figure 10 presents a draft of the ROI, not to scale.

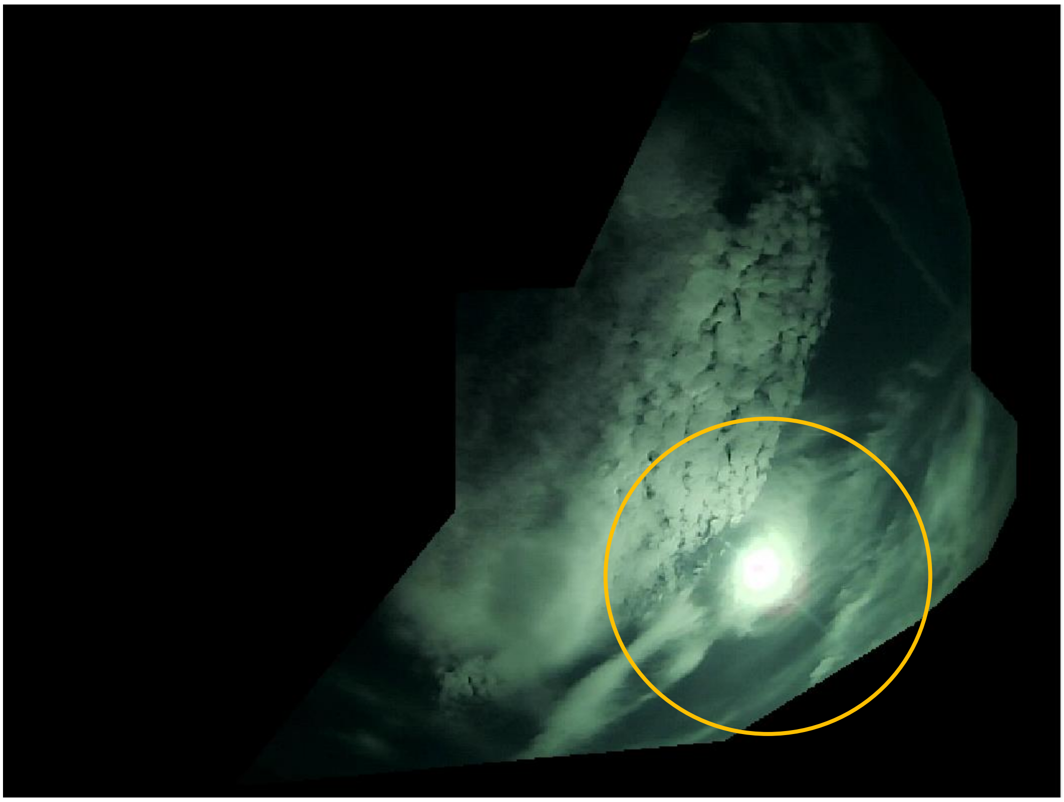

Figure 10 – Example image with draft of the ROI used for calculating energy metrics.

In total, there were 84 combinations of 12 time steps with Δt = {1; 2; 5; 8; 10; 15; 20; 30; 45; 60; 75; 90} seconds and 7 ROI radii r = {25; 50; 75; 100; 150; 200; 250} resulting in 84 datasets (12 x 7). The indexes corresponding to each combination are shown in Table 3.

Table 3 – Indexes corresponding to each combination of Δt and ROI radius.

		Δt											
		1	2	5	8	10	15	20	30	45	60	75	90
ROI RADIUS	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

For each of these datasets, the correlation coefficients were calculated for each combination of the 2 target variables (power at t_0: P_0; and power difference between t_0 and $t_{0-\Delta t}$: ΔP) with the 5 selected input variables (power at $t_{0-\Delta t}$: P_{-1}; temperature at t_0: T_0; and ROI energy difference between t_0 and $t_{0-\Delta t}$ for each digital channel). Correlation coefficients measure linear proportionality in variations between two variables, so are useful for identifying variables with similar behaviours, provided their relationship is linear.

The correlation values between power at t_0 (P_0) and the 5 input variables are shown in Figure 11.

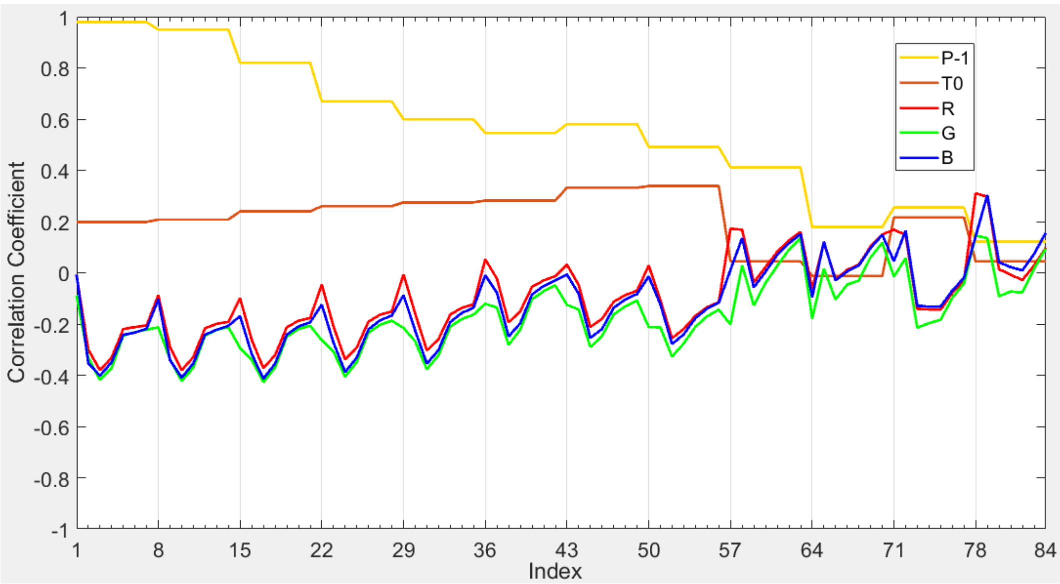

Figure 11 - Correlation between power at t_0 (P_0) and input variables (P_1, T_0, Red, Green, and Blue energy differences).

The horizontal axis represents the index from each combination of Δ*t* and ROI radius aforementioned. The image attributes present negative correlation coefficients, which means that their relationship to power at t_0 (P_0) is inversely proportional, but since the values are lower in module than 0.4, they do not boast good correlation.

Temperature values also are weakly correlated to P_0. Power at $t_{0-\Delta t}$ (P_{-1}) is strongly correlated to P_0 at shorter time intervals, and converges to zero as Δ*t* increases. Another important aspect that can be observed are the peaks for ROI radius = 75 pixels.

The correlation values between ΔP (P_0 – P_{-1}) and the 5 input variables are shown in Figure 12.

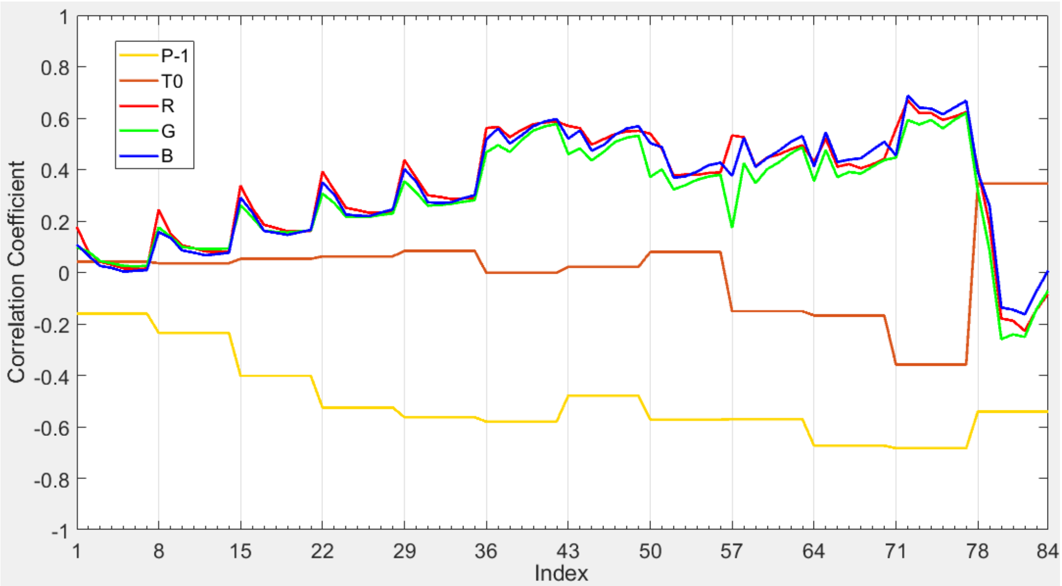

Figure 12 – Correlation between ΔP and input variables (P_{-1}, T_0, Red, Green, and Blue energy differences).

In the case of ΔP the correlation with the input variables increases with Δ*t* with the exception of temperature. For all variables correlation vanishes for Δ*t* = 90 s. For the first five Δ*t* values, the image attributes present a behaviour similar to that of the same variables with P_0, but with an ascending trend. Overall the correlation levels between ΔP and the input variables are higher than for P_0.

The next step in analysing this data is to create regression models to determine the practical viability of the system and data.

4.3. Data validation

The final process was to validate the data acquired by the DAS. First, in order to define a baseline regression performance, the data was used to perform a 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 channel (blue) was used due to a colinearity issue that could adversely affect the model regression.

The coefficient of determination (R^2) is a measurement of how well the model represents the data used for regression. The results of the linear regressions all had low coefficients of determination, even if they had low errors due to extremely low variation rates. This was in line with the findings from the correlation analysis, where for shorter time intervals, P_0 and P_{-1} had high correlation coefficients. This, however, is not sufficient to provide a good regression model. The other variables were statistically insignificant to the model, despite being theoretically relevant. Since the linear model was not able to properly represent the data, given the theoretical relevance of the variables, a non-linear model would probably fare better in representing this data. This led to the following step, training a non-linear model with the same data and comparing it to the linear model. For this step, the chosen model was a regression neural network.

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 (Haykin, 2008). By using larger and more complex network structures it is possible for the network to model more information from the available data, this is the main reason for the use of a neural network for validating the acquired data and selected image features. 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. There are three basic characteristics of the MLP that give its capabilities:

- The individual neurons in the network use a nonlinear activation function that is differentiable;
- The network contains at least one layer that does not directly interact with inputs or outputs, it is hidden; and
- The neurons in the network are heavily interconnected and the importance of these connections depend on the synaptic weights (Haykin, 2008).

The networks used in this phase were all fully connected, meaning that all neurons, or nodes, are connected to all neurons in the previous and next layers. It may seem like overkill, but the learning algorithm gets rid of irrelevant connections by assigning low synaptic weights to them. The selected training algorithm was the feed-forward backpropagation algorithm. Function signals, resulting of the response of the activation function to an input, whether it's the raw input or a forwarded signal from a previous layer, move forward towards the output layer. The final output is compared to a previously known result and an error value is calculated, that error is then backpropagated towards the input layer and the synaptic weights are adjusted to minimize the error. This process may take several iterations depending on the complexity of the model and the network (Haykin, 2008).

This is known as supervised learning, since the output, or target, is known and is used to calculate errors and refine weights. The training process is an iterative task, in which, for each iteration, or epoch, the training error is supposed to converge towards a minimum. However, the overall objective is not to simply minimize the training error, but to minimize error while maintaining generalization ability. This means that even when presented with hitherto unseen data, the network is capable of producing accurate results (Haykin, 2008).

In order to avoid overfitting the model, the opposite of a model that can generalize, several measures can be taken such as: i) ensuring that the training data is of sufficient size and representative of the problem; ii) using a proper architecture and size of network; iii) avoid problems with complexity beyond what the model can solve; iv) stopping the training process before the model is overfitted to the training data (Haykin, 2008).

For the first issue, in the context of this master's dissertation, 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, the camera failed on the last day, most likely due to humidity damage to the circuitry or UV (Ultraviolet) damage to the camera sensor.

Regarding the second issue, the MLP network was tested with several sizes and architectures in order 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 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 couldn’t 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.

4.3.1.
Network training

Network training is a process where the input variables are supplied to the network, the input layer being connected, in this case, to all neurons in the first hidden layer. Each connection has a synaptic weight assigned to it at random in the beginning of training. The input values multiplied by the synaptic weights are supplied to a function on each neuron, and its output is fed forward to the next layer, whether it’s hidden or not. This is repeated for every layer on the network until the output layer, which yields the final result or results.

The next step is calculating the errors and adjusting the synaptic weights. This is done by calculating the error between the expected result and the calculated one. The error is then backpropagated proportionally to the synaptic weights and the gradient for the adjustment of the synaptic weight is calculated for each synaptic connection all the way back to the input layer. These feed-forward signals and error backpropagations are performed multiple times to achieve the most accurate results.

If indefinitely reiterated, the weights will be optimized to only the training data and will fail when presented with new data. For this purpose, an early validation stop is applied when it is determined that the model is being overfitted to the training data through cross-validation. In order to perform cross-validation as described previously, the dataset must be split into two subsets, one for training and one for validation as if it were completely new data. When the error for the validation set ceases to decrease with further iterations, the training is stopped and the network with the best validation error is the final result.

Different networks with different architectures were trained for each combination of Δt and ROI radius in order 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 was 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. 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 (Haykin, 2008).

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 done because these default values yielded solid results and were beyond the machine learning scope of this master’s dissertation.

4.3.2. Validation results

Training was performed for both target variables, P_0 and ΔP, since both had very different linear behaviours and none were successfully represented by linear models. First, the P_0 coefficient of determination is presented in Figure 13 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.

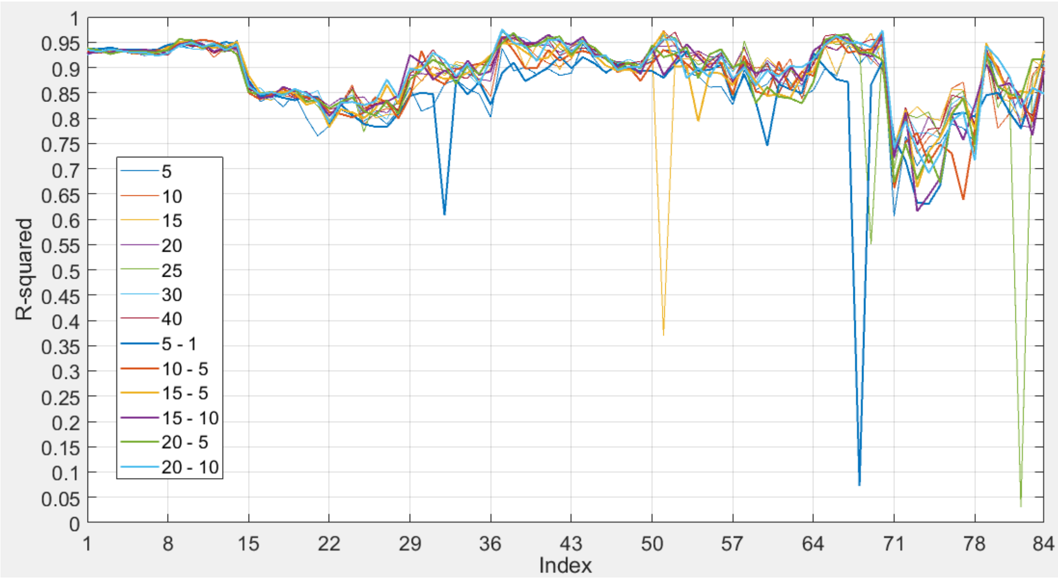

Figure 13 – R^2 values for neural network regression models for P_0.

For the first two Δt values, all 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 more linear Δt range.

Networks with 5 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, this may be due to a less relevant relationship between input and output or due to less training samples availability. This occurs because the data is 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 like 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 14.

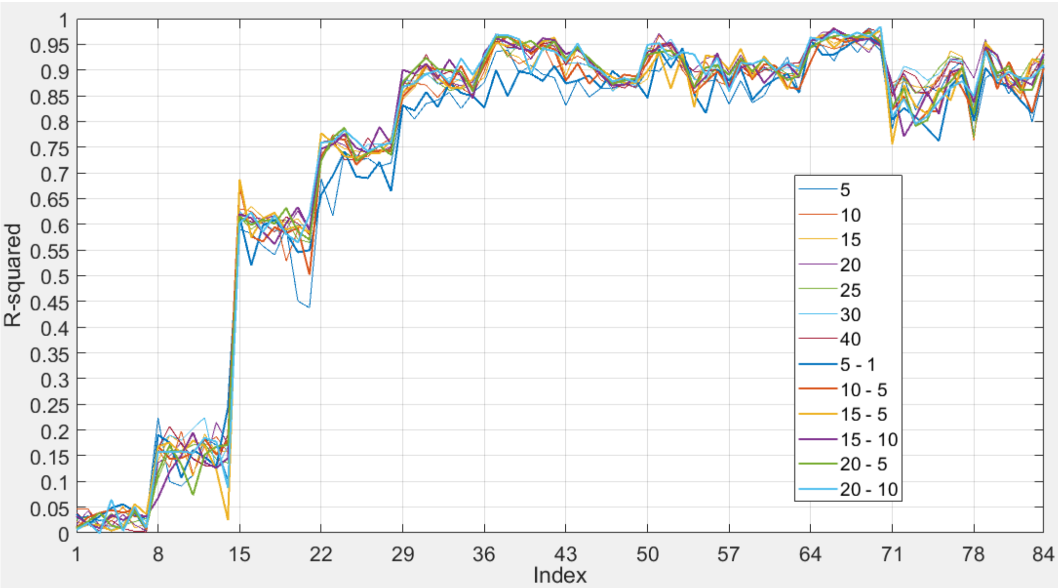

Figure 14 – R^2 values for neural network regression models for ΔP.

This step showed that, for the first four Δ*t* values, neither a linear nor a non-linear method was capable of properly fitting this data. As of the fifth Δ*t* value the neural network model starts presenting good R^2, around 0.9. Similar to the previous plot, it is clear that networks with 5 neurons in the first hidden layer are inferior to the other tested architectures for most data points. After the fifth Δ*t* value the R^2 behaves similarly in both plots, reaching the highest coefficient of determination for Δ*t* = 60 s, closely followed by Δ*t* = 15 s. For both intervals there are small peaks around ROI radius = 75 pixels 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 4. Each row represents one model and each column represents one of the five variables. X's mark when a variable is used.

Table 4 – Variables used in the second step of NN modelling.

		Variables				
		P_{-1}	T_0	R	G	B
MODEL	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 15. Each line in Table 4 represents one line on the plot, and to save some room in the legend, the variables P_{-1} (power at instant 0 - Δt) and T_0 (temperature at instant 0), 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.

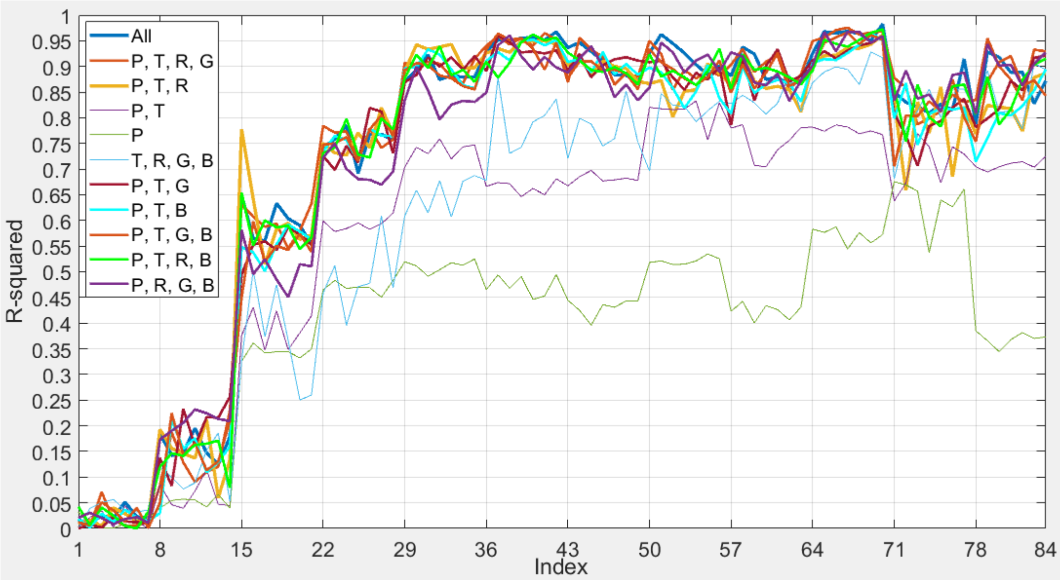

Figure 15 – R^2 values for neural network regression models for ΔP with varying input variables.

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 Δt = 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 are shown in Figure 16.

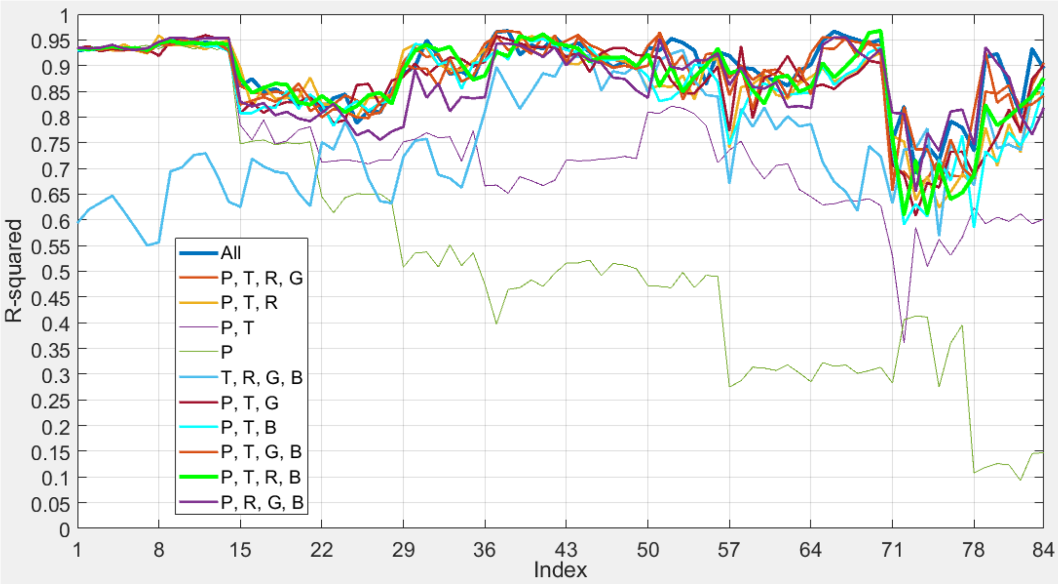

Figure 16 – R^2 values for neural network regression models for P_0 with varying input variables.

Similar to Figure 15, 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. Also, 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 modelled using neural networks, especially compared with linear models. For both cases, previous power, temperature and image attributes from image subtraction proved important to model the targets. This is important because since cloud tracking and modelling is a highly researched matter, as is shown in Table 2, if the tracking and cloud forecast is accurately performed, it is possible to forecast power well up to 60 s head using this data acquisition system.

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 was tested modelling 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 was not used to train the network, it still was capable of estimating the output close to the real measured value.

Figure 17 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.

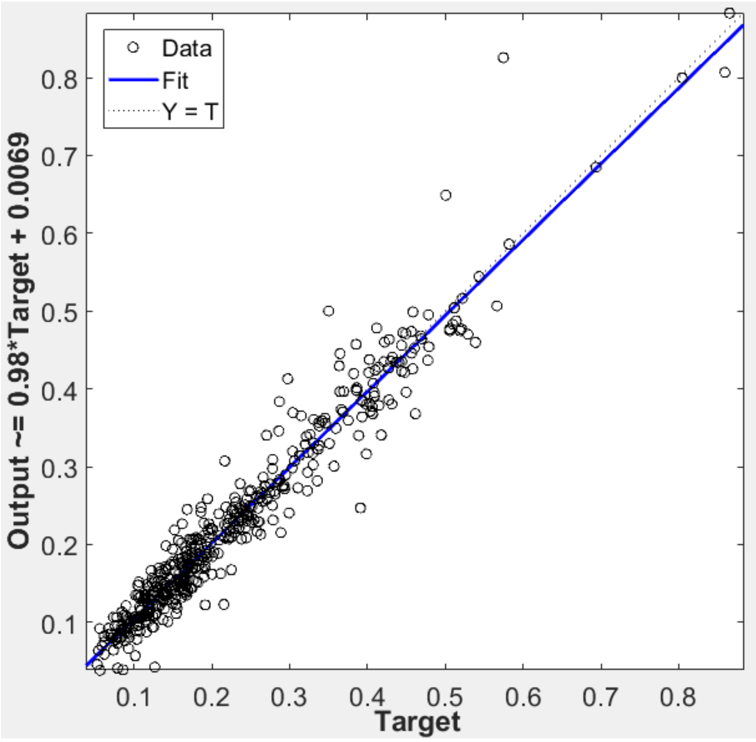

Figure 17 – Regression plot for the validation of a NN model of P_0 with Δt = 60 s step and ROI radius = 250 pixels.

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 modelling ΔP, and the results are presented in Figure 18.

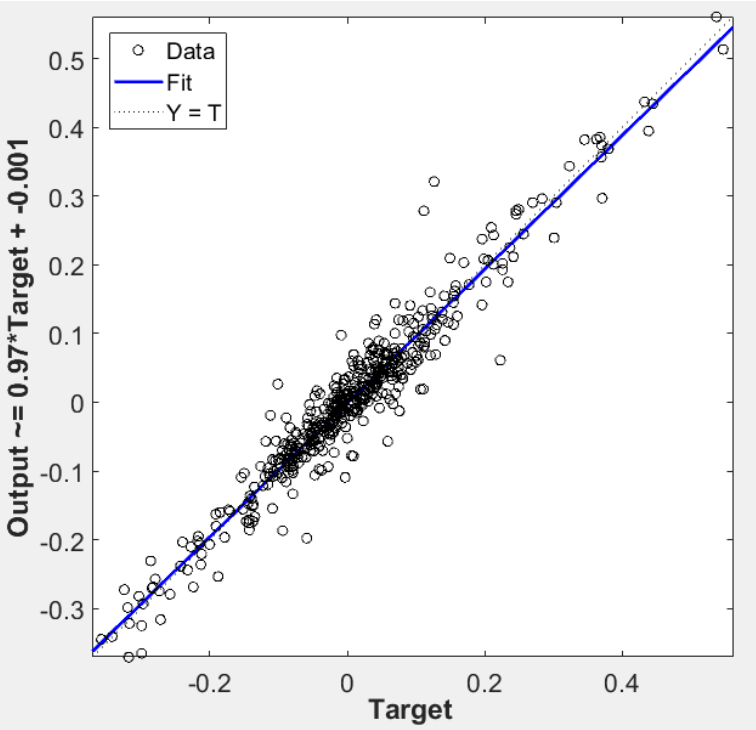

Figure 18 – Regression plot for the validation of a NN model of ΔP with Δ*t* = 60 s step and ROI radius = 250 pixels.

In this case the validation $R^2 = 0.93$, and the regression plot also shows 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.

5 Conclusions

This master's dissertation aimed to develop a working tool to support PV energy forecast models for real-time operation applications. Following the specific objectives listed in Chapter 1, firstly, the most detrimental actor in solar resource availability is cloud occlusion. Due to their mobile nature, the passing of clouds is responsible for the high amplitude and high frequency variations in PV power.

After identifying the main actor, it was necessary to identify the tools with which to deal with these variations. After extensive research, a plethora of different modelling and forecasting tools were surveyed. Of these, considering the very short time scales, sky imaging methods were elected as the most suitable for dealing with this problem due to the physical and geometrical information added by the use of images.

Another critical issue surveyed during the literature research was the need for relevant data in terms of acquisition frequency. Therefore, the data acquisition system developed for this master's dissertation would need to provide sky images at high frequencies, as well as other relevant information such as PV power and panel temperature. This would in theory, based on the extensive survey, provide sufficient information for performing very short-term, more specifically intra-minute, forecasts.

After the acquisition system was finished, it was put to test and produced sufficient information for testing the relevance of all the variables chosen, as well as the acquisition frequency. It was initially pre-processed and analysed in order to determine the ideal forecast horizon and image features to be analysed. This was done through correlation analysis between the variables assumed to be relevant based on the literature survey conducted.

Validation was performed on the data selected during the correlation analysis by using a linear model as 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 itself and of the selected features used for training the neural network have been 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 s 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 master's dissertation are as follows:

- Increase geometrical complexity by using arrays of PV panels, mirroring real-world solar farms;
- Test the system in different seasons, climates and micro-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; and
- Test the developed acquisition system with the NN model with entirely new data.

References

AI, Y.; PENG, Y.; WEI, W. A model of very short-term solar irradiance forecasting based on low-cost sky images. **AIP Conference Proceedings**, v. 1839, n. May, 2017.

ALONSO-MONTESINOS, J.; BATLLES, F. J. The use of a sky camera for solar radiation estimation based on digital image processing. **Energy**, v. 90, p. 377–386, 2015.

ALONSO-MONTESINOS, J.; BATLLES, F. J.; PORTILLO, C. Solar irradiance forecasting at one-minute intervals for different sky conditions using sky camera images. **Energy Conversion and Management**, v. 105, p. 1166–1177, 2015.

ANEEL. **Capacidade de Geração do Brasil**. Disponível em: <http://www2.aneel.gov.br/aplicacoes/capacidadebrasil/capacidadebrasil.cfm>. Acesso em: 19 ago. 2019.

BARBIERI, F.; RAJAKARUNA, S.; GHOSH, A. Very short-term photovoltaic power forecasting with cloud modeling: A review. **Renewable and Sustainable Energy Reviews**, v. 75, n. November 2016, p. 242–263, 2017.

BESSA, R. et al. Handling renewable energy variability and uncertainty in power systems operation. **Wiley Interdisciplinary Reviews: Energy and Environment**, v. 3, n. 2, p. 156–178, 2014.

BIRD, L.; COCHRAN, J.; WANG, X. **Wind and Solar Energy Curtailment: Experience and Practices in the United StatesNational Renewable Energy Laboratory (NREL)**. [s.l: s.n.]. Disponível em: <http://www.osti.gov/servlets/purl/1126842/>.

BLANC, P. et al. Short-term forecasting of high resolution local DNI maps with multiple fish-eye cameras in stereoscopic mode. **AIP Conference Proceedings**, v. 1850, n. June, 2017.

BORRILL, C. et al. **pi_ina219**. Disponível em: <https://github.com/chrisb2/pi_ina219>. Acesso em: 24 jul. 2019.

BOUZGOU, H.; GUEYMARD, C. A. Fast short-term global solar irradiance forecasting with wrapper mutual information. **Renewable Energy**, v. 133, p. 1055–1065, 2019.

CAN ŞENER, Ş. E.; SHARP, J. L.; ANCTIL, A. Factors impacting diverging paths of renewable energy: A review. **Renewable and Sustainable Energy Reviews**, v. 81, n. June 2017, p. 2335–2342, 2018.

CAZORLA, A. et al. Multi-exposure adaptive threshold technique for cloud detection with sky imagers. **Solar Energy**, v. 114, p. 268–277, 2015.

CERVANTES, M. et al. Utilization of Low Cost, Sky-Imaging Technology for Irradiance Forecasting of Distributed Solar Generation. **IEEE Green Technologies Conference**, v. 2016- April, p. 142–146, 2016.

CHAVES-AVILA, J. P.; BANEZ-CHICHARRO, F.; RAMOS, A. Impact of support schemes and market rules on renewable electricity generation and system operation: the Spanish case. **IET Renewable Power Generation**, v. 11, n. 3, p. 238–244, 2017.

CHENG, H. Y. Cloud tracking using clusters of feature points for accurate solar irradiance nowcasting. **Renewable Energy**, v. 104, p. 281–289, 2017.

CHOW, C. W. et al. Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed. **Solar Energy**, v. 85, n. 11, p. 2881–2893, 2011.

CHOW, C. W.; BELONGIE, S.; KLEISSL, J. Cloud motion and stability estimation for intra-hour solar forecasting. **Solar Energy**, v. 115, p. 645–655, 2015.

CHU, Y. et al. Real-time forecasting of solar irradiance ramps with smart image processing. **Solar Energy**, v. 114, 2015a.

CHU, Y. et al. Real-time prediction intervals for intra-hour DNI forecasts. **Renewable Energy**, v. 83, p. 234–244, 2015b.

CHU, Y. et al. Short-term reforecasting of power output from a 48 MWe solar PV plant. **Solar Energy**, v. 112, p. 68–77, 2015c.

CHU, Y.; PEDRO, H. T. C.; COIMBRA, C. F. M. Hybrid intra-hour DNI forecasts with sky image processing enhanced by stochastic learning. **Solar Energy**, v. 98, n. PC, p. 592–603, 2013.

DAS, U. K. et al. Forecasting of photovoltaic power generation and model optimization: A review. **Renewable and Sustainable Energy Reviews**, v. 81, p. 912–928, 2018.

DENHOLM, P.; HAND, M. Grid flexibility and storage required to achieve very high penetration of variable renewable electricity. **Energy Policy**, v. 39, n. 3,

p. 1817–1830, 2011.

DENHOLM, P.; MARGOLIS, R. M. Evaluating the limits of solar photovoltaics (PV) in traditional electric power systems. **Energy Policy**, v. 35, n. 5, p. 2852–2861, 2007.

DIAGNE, M. et al. Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. **Renewable and Sustainable Energy Reviews**, v. 27, p. 65–76, 2013.

DRAGOON, K.; SCHUMAKER, A. **Solar PV Variability and Grid Integration**. Portland, OR: [s.n.]. Disponível em: <http://www.rnp.org/node/rnp-report-solar-pv-variability-and-grid-integration>.

ELSINGA, B.; VAN SARK, W. G. J. H. M. Short-term peer-to-peer solar forecasting in a network of photovoltaic systems. **Applied Energy**, v. 206, n. September, p. 1464–1483, 2017.

FURRER, T. **w1thermsensor**. Disponível em: <https://github.com/timofurrer/w1thermsensor>. Acesso em: 24 jul. 2019.

GOHARI, S. M. I. et al. Comparison of solar power output forecasting performance of the Total Sky Imager and the University of California, San Diego Sky Imager. **Energy Procedia**, v. 49, p. 2340–2350, 2013.

HAYKIN, S. **Neural Networks and Learning Machines: A Comprehensive Foundation**. 3rd ed. ed. Upper Saddle River: Prentice Hall, 2008.

HOSSAIN, M. S. et al. Role of smart grid in renewable energy: An overview. **Renewable and Sustainable Energy Reviews**, v. 60, p. 1168–1184, 2016.

IEA. **Solar Photovoltaic EnergyTechnology Roadmap**. [s.l: s.n.]. Disponível em: <http://www.iea.org/publications/freepublications/publication/TechnologyRoadmapSolarPhotovoltaicEnergy_2014edition.pdf>.

IEA. **2018 World Energy Outlook: Executive Summary**. [s.l: s.n.]. Disponível em: <www.iea.org/t&c/>.

IEC. Electrical Energy Storage - White Paper. **International Electrotechnical Commission**, p. 1–78, 2011.

KARIMI, M. et al. Photovoltaic penetration issues and impacts in distribution network - A review. **Renewable and Sustainable Energy Reviews**, v. 53, p. 594–605, 2016.

KOBLICK, D. C. **SolarAzEl**MATLAB, , 2009.

KOW, K. W. et al. An intelligent real-time power management system with active learning prediction engine for PV grid-tied systems. **Journal of Cleaner Production**, v. 205, p. 252–265, 2018.

KUHN, P. et al. Validation of an all-sky imager–based nowcasting system for industrial PV plants. **Progress in Photovoltaics: Research and Applications**, v. 26, n. 8, p. 608–621, 2018.

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. **Solar Energy**, v. 177, n. September 2018, p. 494–500, 2019.

LAMONT, A. D. Assessing the long-term system value of intermittent electric generation technologies. **Energy Economics**, v. 30, n. 3, p. 1208–1231, 2008.

LAVE, M.; RENO, M. J.; BRODERICK, R. J. Characterizing local high-frequency solar variability and its impact to distribution studies. **Solar Energy**, v. 118, p. 327–337, 2015.

LEE, J. et al. A Neural Network Approach to Cloud Classification. **IEEE Transactions on Geoscience and Remote Sensing**, v. 28, n. 5, p. 846–855, 1990.

LIANG, X. Emerging Power Quality Challenges Due to Integration of Renewable Energy Sources. **IEEE Transactions on Industry Applications**, v. 53, n. 2, p. 855–866, 2016.

LIPPERHEIDE, M.; BOSCH, J. L.; KLEISSL, J. Embedded nowcasting method using cloud speed persistence for a photovoltaic power plant. **Solar Energy**, v. 112, p. 232–238, 2015.

LONG, C. N. et al. Retrieving cloud characteristics from ground-based daytime color all-sky images. **Journal of Atmospheric and Oceanic Technology**, v. 23, n. 5, p. 633–652, 2006.

MALEKI, H.; VARMA, R. K. Coordinated control of PV solar system as STATCOM (PV-STATCOM) and Power System Stabilizers for power oscillation damping. **IEEE Power and Energy Society General Meeting**, v. 2016- Novem, 2016.

MARQUEZ, R.; COIMBRA, C. F. M. Intra-hour DNI forecasting based on cloud tracking image analysis. **Solar Energy**, v. 91, p. 327–336, 2013.

MEJIA, F. A. et al. Coupling sky images with radiative transfer models: A new method to estimate cloud optical depth. **Atmospheric Measurement Techniques**, v. 9, n. 8, p. 4151–4165, 2016.

MILLS, A.; WISER, R. Implications of Wide-Area Geographic Diversity for Short-Term Variability of Solar Power. **Energy**, n. September, p. 1–10, 2010.

MYERS, D. R. Direct Normal Radiation. In: ALGORA, C.; REY-STOLLE, I. (Eds.). . **Handbook of Concentrator Photovoltaic Technology**. [s.l.] John Wiley and Sons Ltd, 2016.

NI, Q. et al. An ensemble prediction intervals approach for short-term PV power forecasting. **Solar Energy**, v. 155, p. 1072–1083, 2017.

ONS. **Plano da Operação Energética 2018 / 2022 Sumário Executivo**. Rio de Janeiro: [s.n.].

ONS. **Plano da Operação Energética 2019 / 2022 Sumário Executivo**. Rio de Janeiro: [s.n.].

PARIDA, B.; INIYAN, S.; GOIC, R. A review of solar photovoltaic technologies. **Renewable and Sustainable Energy Reviews**, v. 15, n. 3, p. 1625–1636, 2011.

PEDRO, H. T. C. et al. Assessment of machine learning techniques for deterministic and probabilistic intra-hour solar forecasts. **Renewable Energy**, v. 123, p. 191–203, 2018.

PEDRO, H. T. C.; COIMBRA, C. F. M. Nearest-neighbor methodology for prediction of intra-hour global horizontal and direct normal irradiances. **Renewable Energy**, v. 80, p. 770–782, 2015.

PETINRIN, J. O.; SHAABANB, M. Impact of renewable generation on voltage control in distribution systems. **Renewable and Sustainable Energy Reviews**, v. 65, p. 770–783, 2016.

QUESADA-RUIZ, S. et al. Cloud-tracking methodology for intra-hour DNI forecasting. **Solar Energy**, v. 102, 2014.

RANA, M.; KOPRINSKA, I.; AGELIDIS, V. G. Univariate and multivariate methods for very short-term solar photovoltaic power forecasting. v. 121, p. 380–390, 2016.

RAZA, M. Q.; NADARAJAH, M.; EKANAYAKE, C. On recent advances in PV output power forecast. **Solar Energy**, v. 136, p. 125–144, 2016.

REDDY, S.; PAINULY, J. P. Diffusion of renewable energy technologies-barriers and stakeholders' perspectives. **Renewable Energy**, v. 29, n. 9, p. 1431–

1447, 2004.

REIKARD, G. Predicting solar radiation at high resolutions : A comparison of time series forecasts. **Solar Energy**, v. 83, p. 342–349, 2009.

REN21. **Renewables 2018 Global Status Report**. [s.l: s.n.].

RICHARDSON, W. et al. A low cost, edge computing, all-sky imager for cloud tracking and intra-hour irradiance forecasting. **Sustainability (Switzerland)**, v. 9, n. 4, p. 1–17, 2017.

SANFILIPPO, A. et al. An adaptive multi-modeling approach to solar nowcasting. **Solar Energy**, v. 125, p. 77–85, 2016.

SCHMIDT, T. et al. Evaluating the spatiooral performance of sky-imager-based solar irradiance analysis and forecasts. **Atmospheric Chemistry and Physics**, v. 16, n. 5, p. 3399–3412, 2016.

SMETS, A. H. et al. **Solar Energy: The physics and engineering of photovoltaic conversion, technologies and systems**. 1. ed. Cambrige: UIT Cambridge Ltd, 2016.

SOBRI, S.; KOOHI-KAMALI, S.; RAHIM, N. A. Solar photovoltaic generation forecasting methods: A review. **Energy Conversion and Management**, v. 156, p. 459–497, 2018.

SOUBDHAN, T. et al. A robust forecasting framework based on the Kalman filtering approach with a twofold parameter tuning procedure: Application to solar and photovoltaic prediction. **Solar Energy**, v. 131, p. 246–259, 2016.

SOVACOOL, B. K. The intermittency of wind, solar, and renewable electricity generators: Technical barrier or rhetorical excuse? **Utilities Policy**, v. 17, n. 3–4, p. 288–296, 2009.

STEFFERUD, K.; KLEISSL, J.; SCHOENE, J. **Solar forecasting and variability analyses using sky camera cloud detection and motion vectors**. 2012

TAPAKIS, R.; CHARALAMBIDES, A. G. Equipment and methodologies for cloud detection and classification: A review. **Solar Energy**, v. 95, p. 392–430, 2013.

TARAVAT, A. et al. Neural networks and support vector machine algorithms for automatic cloud classification of whole-sky ground-based images. **IEEE Geoscience and Remote Sensing Letters**, v. 12, n. 3, p. 666–670, 2015.

TINGTING, Z. H. U. et al. **A Local Threshold Algorithm for Cloud Detection on Ground-based Cloud Images**. Proceedings of the 34th Chinese

Control Conference. **Anais**...2015

UNFCCC. **Paris Agreement**. Paris: [s.n.]. Disponível em: <http://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf>.

URQUHART, B. et al. Development of a sky imaging system for short-term solar power forecasting. **Atmospheric Measurement Techniques**, v. 8, n. 2, p. 875–890, 2015.

VARMA, R. K.; SALEHI, R. SSR Mitigation with a New Control of PV Solar Farm as STATCOM (PV-STATCOM). **IEEE Transactions on Sustainable Energy**, v. 8, n. 4, p. 1473–1483, 2017.

VARMA, R. K.; SIAVASHI, E. M. Enhancement of Solar Farm Connectivity with Smart PV Inverter PV-STATCOM. **IEEE Transactions on Sustainable Energy**, v. PP, n. c, p. 1, 2018.

WEST, S. R. et al. Short-term irradiance forecasting using skycams: Motivation and development. **Solar Energy**, v. 110, p. 188–207, 2014.

WOOD-BRADLEY, P.; ZAPATA, J.; PYE, J. Cloud tracking with optical flow for short-term solar forecasting. **50Th Conference of the Australian Solar Energy Society**, n. November, p. 2–7, 2012.

WÜRFEL, P. **Physics of Solar Cells: From Principles to New Concepts**. 1. ed. Weinheim: Wiley-VCH, 2005.

XU, J. et al. **Solar irradiance forecasting using multi-layer cloud tracking and numerical weather prediction**. Proceedings of the 30th Annual ACM Symposium on Applied Computing - SAC '15. **Anais**...New York, New York, USA: ACM Press, 2015Disponível em: <http://dl.acm.org/citation.cfm?doid=2695664.2695812>. Acesso em: 3 jul. 2016

YANG, D. et al. History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining. **Solar Energy**, v. 168, n. February, p. 60–101, 2018.

YESILBUDAK, M.; COLAK, M.; BAYINDIR, R. **A Review of Data Mining and Solar Power Prediction**. 5th International Conference on Renewable Energy Research and Applications. **Anais**...2016

ZENDEHBOUDI, A.; BASEER, M. A.; SAIDUR, R. Application of support vector machine models for forecasting solar and wind energy resources: A review. **Journal of Cleaner Production**, v. 199, p. 272–285, 2018.

Annex 1

Cloud detection and photovoltaic power estimation by using neural networks and sky images

Guilherme Fonseca Bassous
Pontifical Catholic University of Rio de Janeiro
Brazil
gfbassous@gmail.com

Carlos R. Hall Barbosa
Pontifical Catholic University of Rio de Janeiro
Brazil
hall@puc-rio.br

Marley Maria Bernardes Rebuzzi Vellasco
Pontifical Catholic University of Rio de Janeiro
Brazil
marley@ele.puc-rio.br

Rodrigo Flora Calili
Pontifical Catholic University of Rio de Janeiro
Brazil
calili@puc-rio.br

ABSTRACT

Accurate short-term forecasting is a crucial step in increasing photovoltaic (PV) energy adoption. The first step in deterministically forecasting PV power from sky images is cloud detection, which is mostly done by thresholding or machine learning techniques. The objective of this manuscript is to estimate PV output based on solar occlusion, solar elevation angle and solar panel temperature, using multilayer perceptron networks for image classification and power estimation. The data used consists of power output from a 1 W solar panel, a cloud occlusion metric obtained from sky images, panel temperature at the instant of power measurement and solar elevation angle. The results show that this type of network is well suited for this application, yielding errors of less than 2% of the maximum power measured.

KEYWORDS

Solar power, photovoltaic, sky images, neural networks

1 INTRODUCTION

Forecasting is an important part of energy management, especially when it comes to renewable energy sources (RES). RES short-term variability can cause energy quality problems that limit the real-time operation and insertion of these resources [1, 2]. In particular, cloud cover has a great impact on photovoltaic (PV) energy conversion, since it impacts the direct sunlight reaching the solar panel surface (Parida et al., 2011). Fast cloud transients can also lead to economic problems such as negative prices due to energy curtailment (Chaves-Avila et al., 2017). Therefore, forecasting cloud cover and its impacts on PV panels' performance is paramount to reduce uncertainty and energy quality problems.

To be able to deterministically forecast cloud cover and trajectory, and consequently the behaviour and impacts of clouds on PV energy plants, imaging methods must be used (Diagne et al., 2013). For short-term applications, such as addressing energy quality due to fast cloud transients, only ground-based imagery is suitable. Ground-based imagery can be used for larger time and spatial resolutions than satellite-based imagery, as well as reduced costs and the ability to see small local clouds that satellite cameras can´t detect [5, 6].

The first step is object detection, which is the ability to determine that, for example, a pixel represents a cloud and not sky. There are different ways to detect clouds in an image, the simplest consisting of a fixed threshold based on the ratio of the red and blue channels in an RGB (Red Green Blue colour space) image, as presented by Long et al. (Long et al., 2006). More complex thresholding methods have been proposed, such as local thresholding based on different regions in an image (Tingting et al., 2015), and multi-exposure adaptive thresholding (Cazorla et al., 2015).

Another method for cloud detecting is by using neural networks to classify the image pixels. Neural networks have been used to detect clouds for more than 20 years, one example being the research conducted by Lee et al. (Lee et al., 1990) using satellite imagery. More recently, researchers started using ground-based sky imagery for forecasting. Taravat et al. (Taravat et al., 2015) have achieved reliable results with multilayer perceptron (MLP) neural networks and support vector machine (SVM) deep learning approaches, superior to thresholding for some situations. Machine learning, especially some neural networks with high capabilities for pattern recognition is a powerful and versatile solution to the cloud detection problem due to their capacity of adapting to local specific data and ease to train.

Neural networks have been thoroughly used for power estimation using several different input variables, as shown in the review by Yesilbudak et al. (Yesilbudak et al., 2016). Of those, some used total cloud cover as one of the inputs for the network, however none used a metric for solar occlusion. One of the benefits of using sky images is the possibility to extract spatial information from them. It is possible to derive shadow position from solar angle and cloud position (Stefferud et al., 2012). This is key information in estimating and short-term predicting PV output.

The objective of this paper is to estimate PV output based on solar occlusion, solar elevation angle and solar panel temperature, using MLP networks for image classification and power estimation.

The structure of this paper is as follows: section 2 describes the data used, section 3 describes the methods for processing and analysing the data, section 4 presents the results obtained by the used method and section 5 contains the conclusions of this work and recommendations for future work on this line of research.

2 DATA ACQUISITION

The data acquisition was done by a system comprised of a USB-controlled camera, a 1 W PV panel, an Arduino microcontroller board, a circuit to measure current and voltage of the panel, and an Arduino-compatible thermometer (Figure 1). The system was mounted on a tripod and set so the image sensor from the camera and the panel would be horizontal. The system was mounted on a terrace of a 6-story high building in the main campus of the Pontifical Catholic University of Rio de Janeiro, roughly 22.98° S, 43.23° W and 43 m elevation.

The camera had a 1920 x 1080 pixel resolution and a 180° field of view (FOV). Since the image sensor is rectangular and cropped part of the image, the system was oriented so the largest dimension from the camera was set approximately on the north-south direction. Since the site is in the southern hemisphere and the acquisition was done during winter, the southern half of the image could be cropped without significant loss of information, thus halving the processing and space requirements. All the data was stored in a laptop computer used to control the hardware, for post-processing.

As for power acquisition, the PV panel had a circuit connected to two analog pins in the Arduino board. One pin would measure the voltage of the entire load and the other, the voltage in a small shunt resistance, then inferring the generated current. The board was also responsible for interfacing the thermometer’s digital input through the serial door to a laptop computer.

This computer was running a MATLAB script to acquire an image, voltage and current measurements every 1 second and temperature measurement every 50 seconds for a predefined period. The data were stored in the computer for future processing.

After collecting for different periods in different days, the database had 5089 points consisting of images and power measurements, and 119 temperature measurements. Data were acquired on four days with different cloud cover profiles, from 0 to 100% cloud cover, as for the time of day, acquisition spanned from 11 h to 17 h. Figure 1 shows the acquisition system during operation.

Figure 1: **Data acquisition system.**

3 METHODS

3.1 Image Processing and Classification

The first step was to crop and apply a mask to hide regions with undesired objects such as buildings, trees and hills. This step eases the learning process of the neural network, since the mask is applied by replacing the RGB pixel values on the mask region by (0,0,0), which is an easier pattern for a network to learn. The next step is to mask out the sun, by replacing the RGB values in the solar region by (250,150,0). To do so, since the images come from a 180° FOV camera, to calculate the solar position only the solar azimuth, elevation angle, and a radius value in pixels are required. Both angles can be calculated using the time and coordinates from the site. Figure 2 shows these steps previously mentioned for the same image. On the left, it is shown the original image, on the middle, the cropped image with background mask and on the right, the final image with a solar mask.

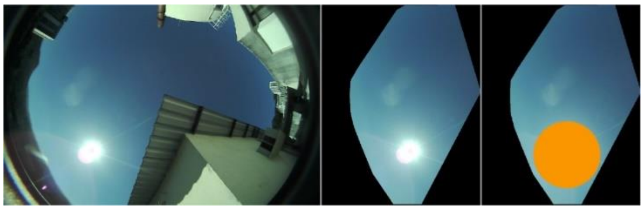

Figure 2: **Crop and masking steps.**

The objective of the solar mask is to cover the solar region so that the neural network won't confuse white cloud pixels with sun pixels. From these final images 8739 points were manually chosen to make up the training vector for the cloud classification network and split randomly 70%/15%/15% into training, validation and test sets, respectively. Table 1 presents the used classes and number of samples used to train the network.

Table 1: **Classes and number of samples per class.**

Class	Samples
Background Mask	284
Cloud	4537
Sky	3485
Sun Mask	433

A two-layer feed-forward network with sigmoid-activated neurons on the hidden layer and softmax in the output layer was used. For this problem and this dataset, a 3-10-4 network was used, yielding 99.9% overall classification accuracy. Still, to determine the presence of clouds on the solar region, another step is required, Tingting et al. (Tingting et al., 2015) used image inpainting to perform the interpolation of the area surrounding the sun and determine if there is occlusion.

By simply inpainting the region of the solar mask, too much error was introduced due to misclassification of pixels around the mask. Instead of using a boolean value as to whether there was occlusion, a red-blue ratio (RBR) similar to the RBR mentioned by Chow et al. (Chow et al., 2011) and Stefferud et al. (Stefferud et al., 2012) was employed. The difference is that the RBR was used as a metric for how cloudy is the solar region. After computing the mean RBR in the solar region and comparing it with the other inputs, it has shown to be correlated, so this was the chosen metric for training the following network along with solar elevation angle and panel temperature. Figure 3 shows all the steps in image processing and classification done in this work. It is shown from left to right: crop+background mask; solar mask; neural network classification; inpainting.

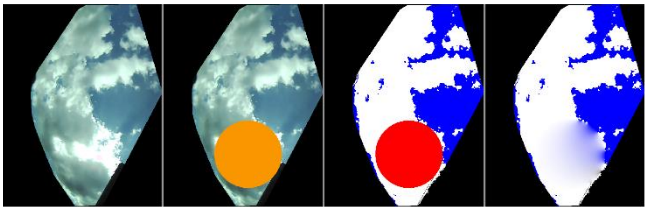

Figure 3: **Image processing and classification steps.**

3.2 Data Fitting

For the data fitting problem, estimating the power output based on 3 input variables, a different network was trained. To find the best possible fit, several topologies and training parameters were tested: number of epochs, number of validation fails before early stopping training/ validation/ test proportions, number of layers, number of neurons in each layer, training algorithm, and transfer functions.

For networks with one hidden layer, the following parameters were tested:

- Number of neurons: between 5 and 60 in 5-neuron increments.
- Transfer functions: hyperbolic tangent sigmoid, log-sigmoid and linear.
- Training algorithms: Levenberg-Marquardt, scaled conjugate gradient, gradient descend with momentum and Bayesian regularization.
- Subsets: between 50-80% for training, 10-25% for validation and 10-25% for test in different combinations of 5% steps.
- Epochs: between 1,000 and 20,000.
- Validation fails: between 20 and 200.

For two hidden layers:

- Neurons on first hidden layer: between 5 and 40 in 5-neuron increments.
- Neurons on second hidden layer: between 5 and 15.
- Transfer functions: hyperbolic tangent sigmoid and log-sigmoid.
- Training algorithms: Levenberg-Marquardt, scaled conjugate gradient, gradient descend with momentum and Bayesian regularization.
- Subsets: between 50-80% for training, 10-25% for validation and 10-25% for test in different combinations of 5% steps.
- Epochs: between 1,000 and 20,000.
- Validation fails: between 100 and 200.

For three hidden layers:

- Neurons on first hidden layer: between 10 and 30 in 5-neuron increments.
- Neurons on second hidden layer: between 5 and 15.
- Neurons on third hidden layer: between 3 and 10.

- Transfer functions: hyperbolic tangent sigmoid.
- Training algorithm: Levenberg-Marquardt.
- Subsets: between 50-80% for training, 10-25% for validation and 10-25% for test in different combinations of 5% steps.
- Epochs: between 1,000 and 2,000 (the maximum number was reduced because the selected training algorithm didn’t require as many iterations as others).
- Validation fails: between 100 and 200.

Not all combinations of the parameters were tested, but the full range of parameters were tested on the three topologies that yielded the best results per number of layers tested. The best results were tested at least 20 times to mitigate the impacts of random weight initialization.

4 RESULTS

After testing numerous topologies with different training parameters, the best results were achieved by a network with two hidden layers, 25 neurons in the first and 10 in the second. Both layers have hyperbolic tangent sigmoid transfer functions. This network was trained using Levenberg-Marquardt backpropagation algorithm, with dataset division ratio of 80%/10%/10%, maximum of 2000 epochs and 100 validation fails.

Figure 4 shows the training, validation, and test performances in mean squared errors (MSE) for each epoch. The training early stopped in epoch 762 due to 100 consecutive validation fails. Using 100 validation checks was enough to escape a local minimum and not prolonging the training without necessity. After this point the network starts to overfit to the training data, losing generalization and its ability to be applied to new data.

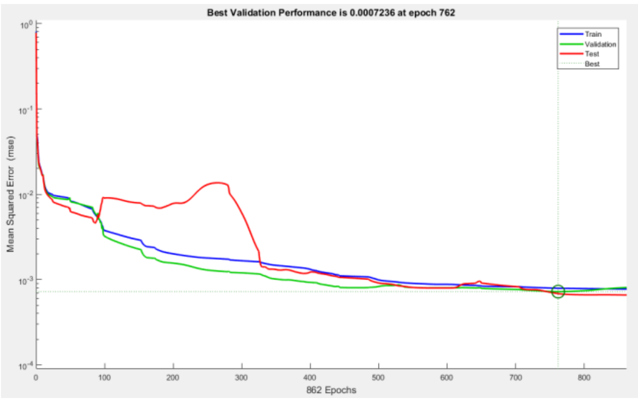

Figure 4: **Validation performance in MSE as function of training epochs.**

Figure 5 is a histogram with all the absolute errors for all steps on epoch 762. The higher concentration of errors closer to zero shows that most errors were small and can possibly be accommodated into a PV forecast system application.

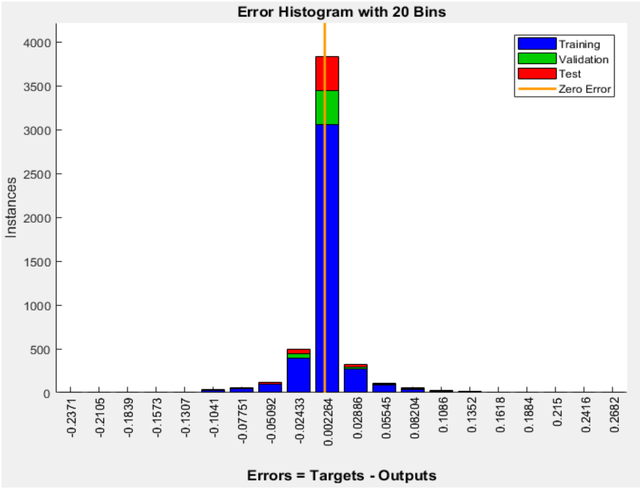

Figure 5: **Histogram with absolute errors from epoch 762.**

Table 2 contains the performance in MSE for the subsets, the whole dataset and the mean absolute error (MAE) for epoch 762.

Table 2: **Error values for each subset.**

Error Metric	Error
Overall MSE	0.000775 W^2
Training MSE	0.000793 W^2
Validation MSE	0.000724 W^2
Test MSE	0.000686 W^2
Mean Absolute Error	0.013458 W

The results show that the multilayer perceptron model is well suited for estimating PV output from image-derived data, temperature and elevation angle. Despite operating in less-than-ideal conditions, the mean absolute error was below 2% of the nominal power of the panel, even lower if the maximum power measured is considered, which was 1.14 W.

5 CONCLUSIONS

The aim of this paper was to estimate PV output based on solar occlusion, solar elevation angle, and solar panel temperature. It can be concluded that this objective was entirely achieved since the model using neural networks yielded a mean absolute error below 2 % of the maximum power observed.

However, it´s important to note that this study has two main limitations: first is the data acquisition system, which should be permanently installed in the test site. And second is the simplification of using a single small PV panel, which doesn't translate into commercial PV plant operation.

With improvements in the data acquisition system, in the image-derived metrics and in the site´s characteristics, far better results should be achieved. As a recommendation for further works, besides improving the training data, the next step would be to expand from a small PV panel to arrays and eventually combining this estimation method with a computer vision system capable of predicting cloud movement.

The main advantage of this paper is to confirm the ability of MLP networks to model the relationships between the chosen variables and PV output power. Future works can embed these models into PV power prediction and plant control systems aiming to reduce the impact of fast cloud transients in energy quality and solar energy adoption.

REFERENCES

[1] S. Reddy and J. P. Painuly, "Diffusion of renewable energy technologies-barriers and stakeholders' perspectives," *Renew. Energy*, vol. 29, no. 9, pp. 1431–1447, 2004.

[2] P. Denholm and R. M. Margolis, "Evaluating the limits of solar photovoltaics (PV) in traditional electric power systems," *Energy Policy*, vol. 35, no. 5, pp. 2852–2861, 2007.

[3] B. Parida, S. Iniyan, and R. Goic, "A review of solar photovoltaic technologies," *Renew. Sustain. Energy Rev.*, vol. 15, no. 3, pp. 1625–1636, 2011.

[4] J. P. Chaves-Avila, F. Banez-Chicharro, and A. Ramos, "Impact of support schemes and market rules on renewable electricity generation and system operation: the Spanish case," *IET Renew. Power Gener.*, vol. 11, no. 3, pp. 238–244, 2017.

[5] M. Diagne, M. David, P. Lauret, J. Boland, and N. Schmutz, "Review of solar irradiance forecasting methods and a proposition for small-scale insular grids," *Renew. Sustain. Energy Rev.*, vol. 27, pp. 65–76, 2013.

[6] R. Tapakis and A. G. Charalambides, "Equipment and methodologies for cloud detection and classification: A review," *Sol. Energy*, vol. 95, pp. 392–430, 2013.

[7] C. N. Long, J. M. Sabburg, J. Calbó, and D. Pagès, "Retrieving cloud characteristics from ground-based daytime color all-sky images," *J. Atmos. Ocean. Technol.*, vol. 23, no. 5, pp. 633–652, 2006.

[8] Z. H. U. Tingting, W. E. I. Haikun, Z. Chi, Z. Kanjian, and L. I. U. Tianhong, "A Local Threshold Algorithm for Cloud Detection on Ground-based Cloud Images," in *Proceedings of the 34th Chinese Control Conference*, 2015, pp. 3702–3706.

[9] A. Cazorla, C. Husillos, M. Antón, and L. Alados-Arboledas, "Multi-exposure adaptive threshold technique for cloud detection with sky imagers," *Sol. Energy*, vol. 114, pp. 268–277, 2015.

[10] J. Lee, R. C. Weger, S. K. Sengupta, and R. M. Welch, "A Neural Network Approach to Cloud Classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 28, no. 5, pp. 846–855, 1990.

[11] A. Taravat, F. Del Frate, C. Cornaro, and S. Vergari, "Neural networks and support vector machine algorithms for automatic cloud classification of whole-sky ground-based images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 3, pp. 666–670, 2015.

[12] M. Yesilbudak, M. Colak, and R. Bayindir, "A Review of Data Mining and Solar Power Prediction," in *5th International Conference on Renewable Energy Research and Applications*, 2016, pp. 1117–1121.

[13] K. Stefferud, J. Kleissl, and J. Schoene, "Solar forecasting and variability analyses using sky camera cloud detection and motion vectors," 2012.

[14] C. W. Chow *et al.*, "Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed," *Sol. Energy*, vol. 85, no. 11, pp. 2881–2893, 2011.

Annex 2

A Low-cost System for High-frequency Solar Imagery and Power Data Acquisition

Guilherme F. Bassous, Rodrigo Flora Calili, Carlos R. Hall Barbosa

Postgraduate Program in Metrology, PUC-Rio, R. Marquês de São Vicente, 225, Gávea, Brazil

Email: <gfbassous@gmail.com>, <calili@puc-rio.br>, <hall@puc-rio.br>

Abstract. With advances in solar energy research and increasingly accurate forecast techniques, intermittency no longer stands as a barrier to the adoption of solar energy. Coupling reliable data and knowledge on the inherent variability of the solar resource with advanced learning and forecast models, renewable energy can take an even bigger role in today's energy paradigm. The objective of this work is to develop and test a low-cost data acquisition system able to provide relevant data for solar energy forecast models. The results yielded from the performed tests indicate high correlation between image derived attributes and power measurements 20 s ahead.

Keywords. Solar energy; Sky-camera; High-frequency data; Photovoltaic; Short-term forecast.

Introduction

For a long time, it's been believed that the intermittency of Renewable Energy Sources (RES) is the main barrier to the massive adoption of greener energy sources, especially solar Photovoltaic (PV). In reality, evidence points towards politics and resistance to change from the traditional energy sector as being the main hindrance (Sovacool, 2009). However, depending on the level of penetration, the inherent intermittence of PV energy may lead to several technical issues and inefficient harnessing of this resource [2,3], especially in systems operation, planning and scheduling.

In order to fully make use of solar PV energy, such problems caused by variability must be addressed. The first step in being able to address variability is to accurately forecast the behaviour of the PV system. The stochastic component of the variability can be split in terms of temporal and spatial resolution and, to each, its own optimal forecasting methods and models can be developed and applied (Diagne et al., 2013).

The main issue preventing predictable PV operation is the presence of clouds casting shadows on the panel surface. Both theory and practice suggest that sky-image based approaches provide significant information for intra-minute and sub-kilometre forecasting [4,5], hence the decision to develop a low-cost system based on sky imagery. Passing clouds can trigger violent albeit short disturbances in global insolation at 1 minute intervals (Mills & Wiser, 2010), but to accurately understand the effect of these disturbances on grid studies, higher frequency data must be used (Lave et al., 2015).

The goal behind having a low-cost system capable of providing data at the required frequencies is increasing the geographical dispersion of data acquisition sites, which can provide more abundant and relevant information on the subject. Thus, this manuscript presents the development and tests of a prototype for a low-cost data acquisition system capable of providing all-sky images with an acquisition frequency up to 1 Hz. Due to the lack of a suitable testbed capable of providing 1 Hz PV conversion data at the time of this study, this prototype also encompasses a 6 W solar panel.

This study is structured as follows: section 2 introduces sky imaging systems, design process and choices for hardware and software of the acquisition system. Section 3 contains information on testing circumstances and results obtained from measurement and data analysis. Finally, section 4 presents the manuscript conclusions and recommendations for furthering this investigation in future works.

Sky-imaging systems

Sky imagers or total sky imagers are devices capable of imaging the whole sky dome with the use of a 180° field of view (FOV) lens. Some have cameras directly pointed towards the sky, while others are pointed down to dome-shaped mirrors reflecting the entire sky. Sky imagers were originally developed for meteorological use, mainly cloud cover analysis. For that reason, many devices have solar-occlusion mechanisms to prevent the high intensity direct solar beam from saturating a portion of the image due to forward scattering (Urquhart et al., 2015). Only in the past decade researchers have begun coupling sky imagers with solar forecasting, driving increased research focus on the area for intra-hour application (Yang et al., 2018).

Unfortunately, being of scientific grade, such equipment represents impeditive costs for the wide geographical dispersion needed for very short-term forecasting in an electrical system scale (Richardson et al., 2017). Even more so if one considers the solar resource availability in developing countries, whose currency inequivalence to the dollar or euro would further the inability to use these imagers in a national operation scale. Hence the need for affordable devices developed specifically for solar forecasting, with all the necessary functionalities cost-effectively wrapped in a small and simple equipment.

Some research centres have developed their own equipment tailored for solar forecasting, and their results show that the increased specificity of the devices yield more relevant information [8,10,11]. Another important feature for PV-specific sky imagers is the necessity of higher frequency acquisition compared to meteorological applications. Image acquisition frequency should be high enough to describe distinguishable events within the desired temporal resolution.

5.1.Hardware

In order to keep the costs as low as possible, the imaging system was built around a single board computer (SBC) (Richardson et al., 2017). The selected SBC was a Raspberry Pi 3B+ due to its easy access, low cost, ample documentation as well as easy access to Linux operating system. Connected to it via USB is an ELP-USBFHD01M-L180 camera consisting of a printed circuit board (PCB) module with a 1920 pixels x 1080 pixels CMOS OV2710 sensor and a 180° FOV lens.

In order to provide an electrical quantity related to solar generation on the same frequency as the image acquisition, a 20 cm x 15 cm solar panel with nominal capacity of 6 V and 1 A at 25° C was used. Current and voltage measurements were done using an Adafruit INA219 DC sensor and, since solar cell temperature has such a strong impact on conversion efficiency, a Maxim Integrated DS18B20 temperature sensor was attached to the bottom of the panel.

Figure 1 shows a basic 3D model of the equipment. The camera and solar panel sit atop the main support structure, while on the bottom two housing units can be seen. One has the power supply components, 110 V AC to 5 V/3 A DC power supply for the SBC and minor cooling fans, plus a 5 V to 12 V DC-DC converter for the main cooling fan. The second unit houses the Raspberry Pi and INA219 DC sensor.

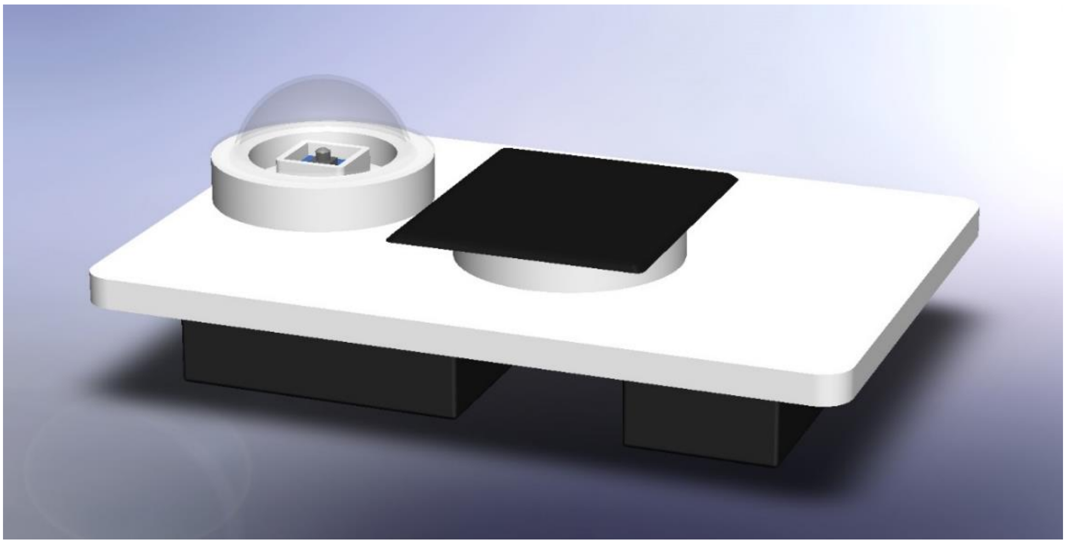

Figure 1: Basic 3D model of data acquisition system, atop sits the camera enclosed by an acrylic dome and the 6 W PV panel used to acquire solar related quantities. On the bottom left, the enclosure with the SBC and DC power sensor INA219.

Measurement circuit

Both sensor boards have embedded analog-to-digital converters and are Raspberry Pi compatible. The temperature sensor was placed on an aluminium heat-exchange pad with the fins bent into contact with the entire sensor area in order to maximize heat transfer from the panel. The pad was then covered with packing foam to reduce heat loss to the atmosphere. As for the electrical quantities' measurement, the INA219 has an internal circuit capable of measuring DC up to 3.2 A by using a 0.1 Ω shunt resistor. A dichroic bulb was used as load on the solar panel, completing the circuit. Finally, the camera was powered and controlled via USB cable. Figure 2 shows a diagram of the measurement circuit.

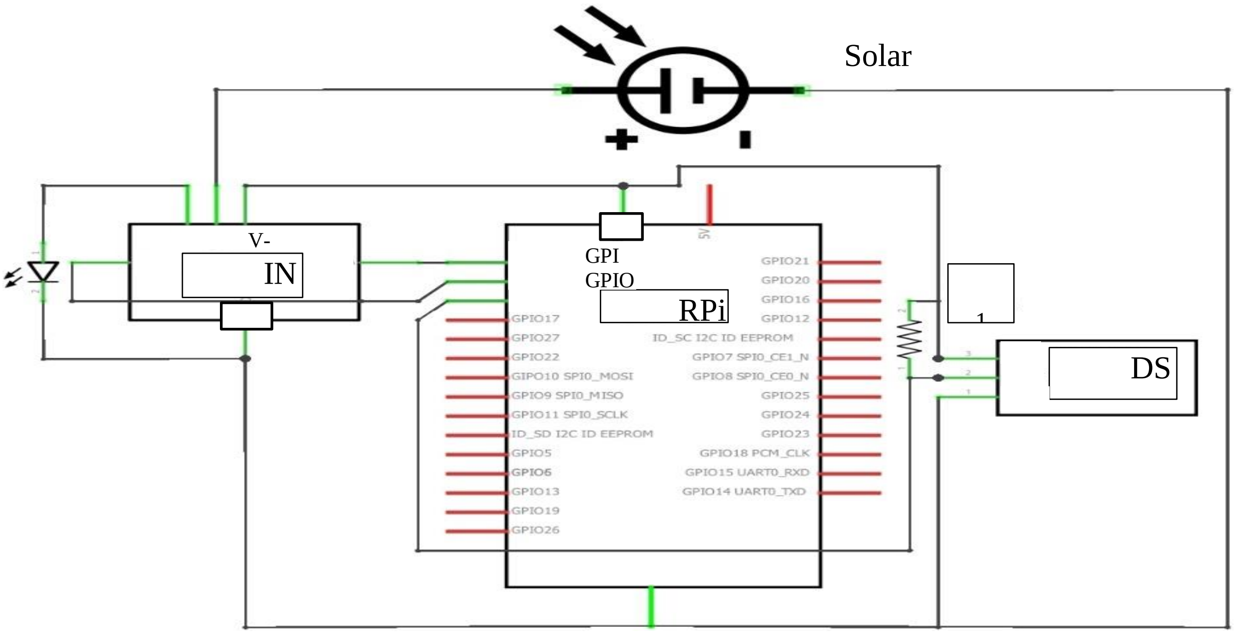

Figure 2: Diagram of measurement circuit.

Software
The Raspberry Pi was running Raspbian Stretch, a Linux OS made for the SBC, with OpenCV 4.0 Python distribution and Python 3.7. OpenCV is an open source image processing and computer vision library available for several platforms. It was exclusively used for controlling the camera and saving the images. All post-processing was performed in Matlab 2017a.There are several Python libraries available for the INA219 and DS18B20 for use with the Raspberry Pi's GPIO pins, the most notorious, and with largest online documentation being "pi_ina219" (Borrill et al., 2019) and "w1thermsensor" (Furrer, 2019). The final program integrated these libraries to record images, power and temperature data on the required frequency for studying cloud induced intermittency on PV power.

As for the acquisition frequency, 1 Hz has proven to provide very useful information about the problem addressed (Lave et al., 2015). However, one image per second for several days yields too much data, with part of it not being useable at all. Hence the decision to go with the principle of acquisition by exception. As the main goal is studying high-amplitude fast variations of power, the system keeps getting images and power measurements every second, but it does not record all the data. Only when triggered by a variation larger than a certain predefined threshold, does the system record the measured data.

In this case, the system keeps the previous ten measurements – t_{-10s} … t_{-1s} – and calculates a moving average of the past 3 values, so if the absolute difference between the present measurement and the calculated average is larger than a certain threshold the data recording process begins. Aside from the past 10 s, it also records the present measurement (t_0) and keeps recording for 4 more seconds – t_{+1s} … t_{+4s}. Those 15 values with their corresponding images and time stamps, plus a temperature measurement, are then saved and the program goes back to listening for intermittence triggers. This data structure is called an "event" throughout this work. The reason for only one temperature measurement per event is that the sensor's ADC often takes over 1 s to convert it, and this makes it impossible to achieve 1 Hz, but fortunately the temperature variations are minimal in only a couple seconds span due to the thermal inertia from the panel. Figure 3 presents the decision process and data flow from the acquisition software.

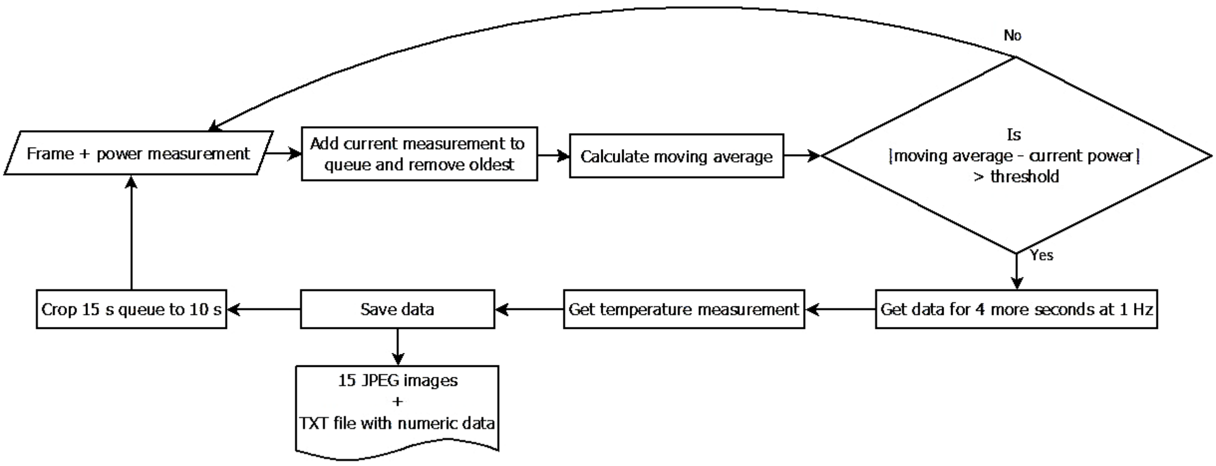

Figure 3: Decision process and data flow from the acquisition software.

Testing and results
The Data Acquisition System was placed on a balcony located at Lat 22.9354° S / Long 43.1756° W, at an altitude of 46 metres. Coordinates and altitude were obtained using Google Earth. Unfortunately, due to building obstruction, only the morning solar path was available, up to roughly 12 h 30 min.

Initially, the camera was protected by an acrylic dome; however, due to exposure to rain and consequently infiltration of water into the dome, too much condensation rendered possible sky images unusable. A decision was made to remove the dome and to cover or expose the equipment depending on the forecast.

5.2.Testing

The first successful capture occurred on February 25th, 2019 at 09 h 21 min 48 s, with the acquisition period spanning until March 23rd, 2019 at 8 h 2 min 45 s. Records were taken for 12 non-straight days due to very clear or cloudy/rainy days in this period. Midway through the period, after analysing the first images, an attempt was made at darkening the images due to oversaturation of the solar region and a large area around it.

A neutral density filter was placed over the lens on March 14th, 2019 and, after adjusting the camera exposure parameter, the images appeared to have a smaller saturated area around the solar region. It also served as a shield protecting the camera from rain, so after its placement the equipment was no longer removed when it rained.

5.3.Results and data analysis

A total of 500 events were recorded over the twelve days. The first processing step was image subtraction to investigate the visible difference between each point in an event. For this, firstly the images were converted to grayscale, then the intensity matrices subtracted. The results depicted mostly noise, but on the scarce ones where more significant border information was available, the solar region was blacked out due to sensor saturation. This pointed towards an inadequacy of the event structure in yielding valuable information for a forecast model. Figure 4 shows image subtraction results for different sky conditions with a highlighted area tracing the solar region. Since the pixel differences are very small and therefore barely visible, the image contrast was increased for differences larger than 5.

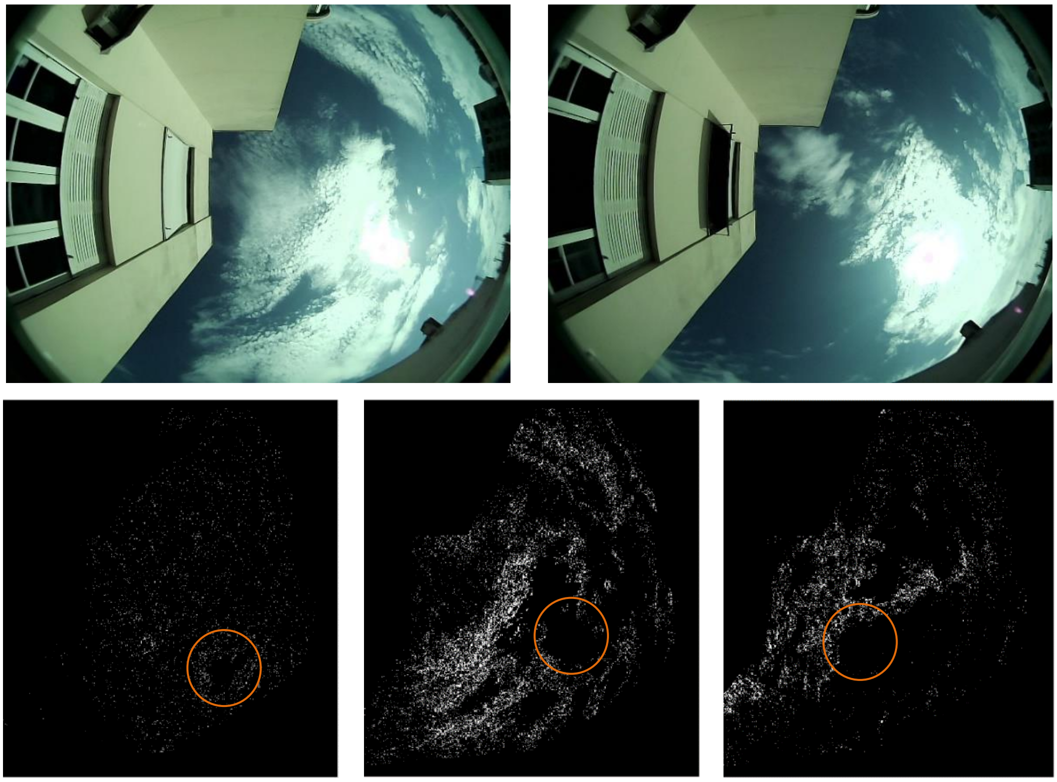

Figure 4: Two examples of raw images from different cloudy days and three cropped image subtraction results for diverse sky conditions with highlighted solar regions.

Next step was to analyse the data as a continuous time series, only separating it by day due to the major differences between each one. The time series plots very clearly showed the fast high-amplitude variations as can be seen in figure 5.

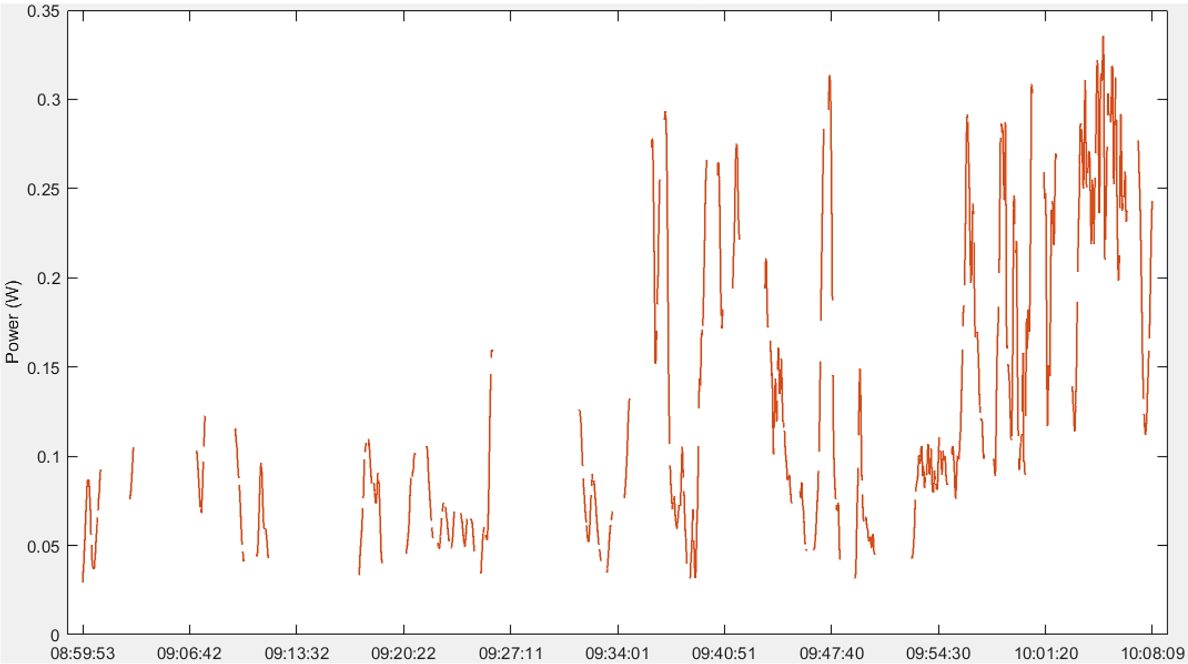

Figure 5: Time series plot of power measurements for March 14th, 2019.

The next goal was to find some relationship between the images and power data. The chosen attribute to be analyzed first was the energy of the pixels in a region of interest (ROI) around the sun. To achieve that, a derived dataset was put together with the power measurements, temperature measurements and ROI energy. Several different ROI radii were used along with differential values as a function of a time interval Δt. The differential values used were for power, temperature and 6 different energy values: absolute difference for each image channel and image subtraction for each channel. The image channels represent red, green and blue components of the colored images and the distinction between absolute difference and image subtraction is, whilst the former allows negative numbers, the latter does not.

The final group of variables were: power at t_0 and $t_{0+\Delta t}$; power difference between those two measurements; temperature at t_0 and $t_{0+\Delta t}$ (for some instances temperature values had to be filled in using a linear method); temperature difference between those two measurements, absolute difference between the ROI energies at t_0 and $t_{0+\Delta t}$; energy of the ROI from the image subtraction at t_0 and $t_{0+\Delta t}$.

In total, 70 combinations of Δt and ROI radius pairs were computed for each of the 12 days, with $\Delta t = \{1; 2; 5; 8; 10; 15; 20; 30; 45; 60\}$ seconds and ROI radius = {25; 50; 75; 100; 150; 200; 250} pixels. Then, the correlation coeficients were calculated for each combination for each day, resulting in 840 matrices of 12x12 correlation coefficients between the aforementioned variables. In order to more easily evaluate the results, the Identity matrix was subtracted from the correlation matrices, then the first 6 columns were binarized between larger and smaller than 0.8, then summated columnwise. This step produced 840 vectors of size 6 with the amount of correlation coefficients larger than 0.8 in each of the first 6 columns.

Through these compiled metrics, it was possible to determine the combination with the highest count of higher (> 0.8) correlation coefficients between the first 6 variables and all twelve. The best results happened with Δt = 20 s and ROI radius = 150 pixels. The compilation of the 840 vectors is presented in Table 1.

Table 1: Compilation of vector coefficients for each combination of Δt and ROI radius.

		ΔT								
		2	5	8	10	15	20	30	45	60
ROI RADIUS	25	0	2	9	12	17	19	8	4	19
	50	0	2	9	11	17	20	8	8	25
	75	0	2	9	12	18	21	8	7	25
	100	0	2	9	12	18	20	9	7	24
	150	0	2	9	12	18	25	9	4	22
	200	0	3	9	12	16	23	7	4	17
	250	0	3	11	12	15	19	7	4	16

Figure 6 presents power measurements 20 seconds ahead from the reference point and ROI energy differences with Δt = 20 s for ROIs with 150-pixel radii. Values were normalized to help visualization.

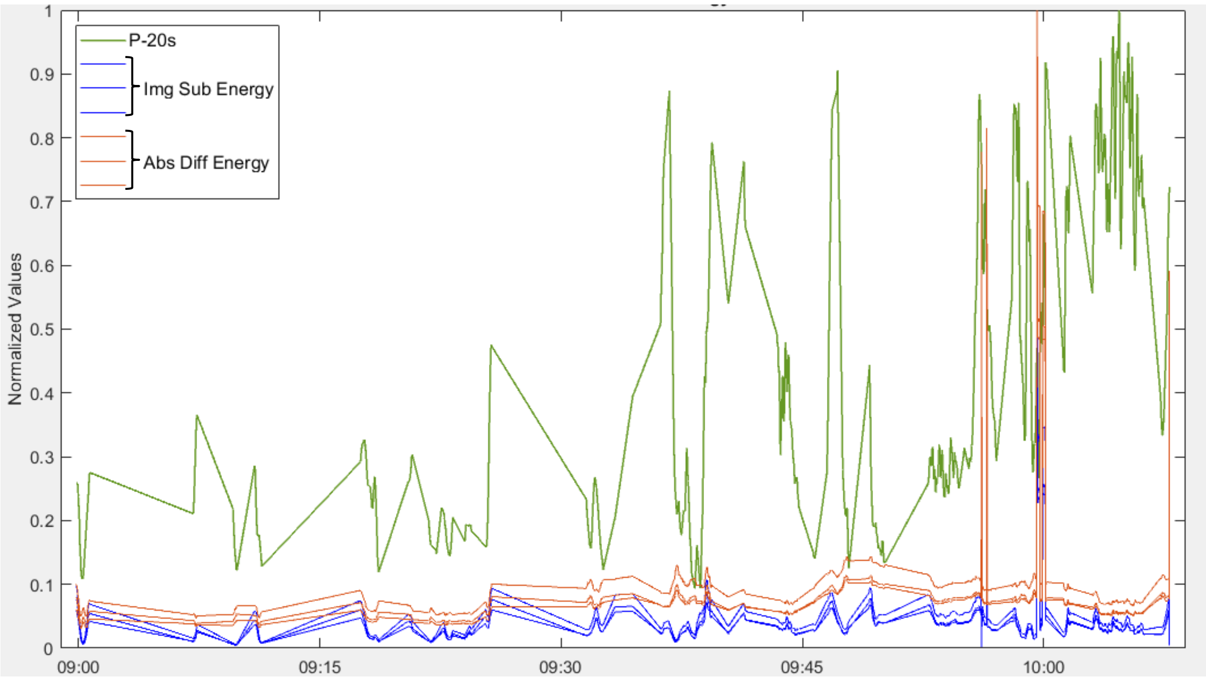

Figure 6: Time series plot of normalized power measurement 20 seconds ahead and different image energy calculations for March 14th, 2019 In blue, energy data from image subtraction and in orange, energy from absolute image difference.

Figure 6 makes visually apparent the correlation between the attributes derived from image processing, – in blue and orange – and power measurements taken 20 s after the reference point – in green. Both curves vary proportionately on the same points save for some apparent outliers near 10:00. This was numerically determined in the previous step in data processing. Temperature measurements also boasted a good correlation coefficient but due to a low range of values they were not visually apparent and therefore kept off the plot to ensure clear visualization of the most critical relationships.

Conclusion and Future works

This work presented the development and testing process of a low-cost data acquisition system aimed at providing data for solar energy forecast models. The initial assumptions using an event-based approach did not yield relevant information, but upon further data analysis, high correlation was found between images and power differences at a horizon of 20 s. Future work should focus on three aspects:

- Increasing ruggedness, reliability and control of the data acquisition system;
- Validating the acquired data by using a nonlinear regression model such as a neural network; and
- Increasing the geometrical complexity of the problem by using power and temperature data from an array of PV panels.

Acknowledgements

The authors thank for the financial support provided by the Brazilian funding agencies CNPq and CAPES.

References

[1] B. K. Sovacool, "The intermittency of wind, solar, and renewable electricity generators: Technical barrier or rhetorical excuse?" , Util. Policy, vol. 17, no. 3–4, pp. 288–296, 2009.

[2] M. Karimi, H. Mokhlis, K. Naidu, S. Uddin, and A. H. A. Bakar, "Photovoltaic penetration issues and impacts in distribution network - A review" , Renew. Sustain. Energy Rev., vol. 53, pp. 594–605, 2016.

[3] P. Denholm and M. Hand, "Grid flexibility and storage required to achieve very high penetration of variable renewable electricity" , Energy Policy, vol. 39, no. 3, pp. 1817–1830, 2011.

[4] M. Diagne, M. David, P. Lauret, J. Boland, and N. Schmutz, "Review of solar irradiance forecasting methods and a proposition for small-scale insular grids" , Renew. Sustain. Energy Rev., vol. 27, pp. 65–76, 2013.

[5] S. Sobri, S. Koohi-Kamali, and N. A. Rahim, "Solar photovoltaic generation forecasting methods: A review" , Energy Convers. Manag., vol. 156, pp. 459–497, 2018.

[6] A. Mills and R. Wiser, "Implications of Wide-Area Geographic Diversity for Short-Term Variability of Solar Power" , Energy, no. September, pp. 1–10, 2010.

[7] M. Lave, M. J. Reno, and R. J. Broderick, "Characterizing local high-frequency solar variability and its impact to distribution studies" , Sol. Energy, vol. 118, pp. 327–337, 2015.

[8] B. Urquhart, B. Kurtz, E. Dahlin, M. Ghonima, J. E. Shields, and J. Kleissl, "Development of a sky imaging system for short-term solar power forecasting" , Atmos. Meas. Tech., vol. 8, no. 2, pp. 875–890, 2015.

[9] D. Yang, J. Kleissl, C. A. Gueymard, H. T. C. Pedro, and C. F. M. Coimbra, "History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining" , Sol. Energy, vol. 168, no. February, pp. 60–101, 2018.

[10] W. Richardson, H. Krishnaswami, R. Vega, and M. Cervantes, "A low cost, edge computing, all-sky imager for cloud tracking and intra-hour irradiance forecasting" , Sustain., vol. 9, no. 4, pp. 1–17, 2017.

[11] S. M. I. Gohari et al., "Comparison of solar power output forecasting performance of the Total Sky Imager and the University of California, San Diego Sky Imager" , Energy Procedia, vol. 49, pp. 2340–2350, 2013.

[12] C. Borrill, T. Timmons, T. van Staveren, T. Kluyver, and S. Bauer, "pi_ina219" , 2019. [Online]. Available: https://github.com/chrisb2/pi_ina219. [Accessed: 24-Jul-2019].

[13] T. Furrer, "w1thermsensor" , 2019. [Online]. Available: https://github.com/timofurrer/w1thermsensor. [Accessed: 24-Jul-2019].