São Paulo University

São Carlos School of Engineering

Department of Electrical and Computing Engineering

Undergraduate Thesis

A Photovoltaic Power Forecasting Method using Artificial Neural Networks

Student: Sofia Moreira de Andrade Lopes (9022133)

Adviser: Elmer Pablo Tito Cari

São Carlos

2018

Sofia Moreira de Andrade Lopes

A Photovoltaic Power Forecasting Method using Artificial Neural Networks

Monograph presented to the Electrical Engineering Course of the São Carlos School of Engineering of São Paulo University as part of the requirements for obtaining the title of Electrical Engineer.

Adviser: Elmer Pablo Tito Cari

Student: Sofia Moreira de Andrade Lopes (9022133)

São Carlos

2018

Ficha catalográfica elaborada pela Biblioteca Prof. Dr. Sérgio Rodrigues Fontes da EESC/USP com os dados inseridos pelo(a) autor(a).

Lopes, Sofia Moreira de Andrade
L864 A Photovoltaic Power Forecasting Method using Artificial Neural Networks / Sofia Moreira de Andrade Lopes; orientador Elmer Pablo Tito Cari. São Carlos, 2018.

Monografia (Graduação em Engenharia Elétrica com ênfase em Sistemas de Energia e Automação) -- Escola de Engenharia de São Carlos da Universidade de São Paulo, 2018.

1. Redes neurais artificiais. 2. Métodos de previsão. 3. Sistemas fotovoltaicos. I. Título.

Eduardo Graziosi Silva - CRB - 8/8907

FOLHA DE APROVAÇÃO

Nome: Sofia Moreira de Andrade Lopes

Título: "A Photovoltaic Power Forecasting Method using Artificial Neural Networks"

Trabalho de Conclusão de Curso defendido e aprovado em 22 / 06 / 2018 ,

com NOTA 10 (dez , zero), **pela Comissão Julgadora:**

Prof. Dr. Elmer Pablo Tito Cari - Orientador - SEL/EESC/USP

Prof. Associado Rogério Andrade Flauzino - SEL/EESC/USP

Mestre Alexandre Prodossimo Sohn - Doutorando - SEL/EESC/USP

Coordenador da CoC-Engenharia Elétrica - EESC/USP:
Prof. Associado Rogério Andrade Flauzino

RESUMO

LOPES, S. M. A. **A Photovoltaic Power Forecasting Method using Artificial Neural Networks**. 2018. 90p. Monografia (Trabalho de Conclusão de Curso) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2018.

O crescimento do número de instalações de sistemas fotovoltaicos no Brasil e no mundo tem sido notável nos últimos anos. Todavia, a inserção destes sistemas na matriz energética não traz apenas vantagens. Juntamente com o aumento do número de instalações, surgem alguns impactos negativos que estas causam à rede de distribuição. Tais impactos afetam tanto a estrutura da rede, quanto seu planejamento e operação. Para contornar tais problemas, este trabalho propõe um método para realizar a previsão da potência fotovoltaica em um sistema elétrico de forma que este método possa posteriormente ser utilizado por operadores do sistema de distribuição.

O método utilizado para realizar a previsão utiliza técnicas de redes neurais artificiais. O estudo foi desenvolvido a partir do teste de três estruturas diferentes de redes. Os dados utilizados para treinar tais redes foram obtidos a partir de medições feitas no sistema fotovoltaico instalado no prédio do Departamento de Engenharia Elétrica e de Computação. Os resultados obtidos pelas redes se mostraram satisfatórios e confirmaram a validade do método para esta aplicação. Além disso, foi feita uma análise comparativa entre as estruturas de redes apresentadas, sendo o melhor desempenho obtido pela rede que utilizou apenas dados medidos localmente no sistema.

Este trabalho pode ser posteriormente expandido para um projeto de mestrado, onde poderia ser desenvolvido um software que utiliza o método exposto para realizar a previsão da potência fotovoltaica. Este software poderá ser lançado comercialmente para auxiliar as empresas do setor de redes de distribuição de energia.

Palavras-chave: Sistemas fotovoltaicos, métodos de previsão, redes neurais artificiais.

ABSTRACT

LOPES, S. M. A. **A Photovoltaic Power Forecasting Method using Artificial Neural Networks**. 2018. 90p. Monografia (Trabalho de Conclusão de Curso) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2018.

The number of photovoltaic systems installations in Brazil and in the world has been notable in recent years. However, the insertion of these systems into the energy matrix does not only bring advantages. With the increase in the number of installations, there are some negative impacts that these systems cause to the distribution network. These impacts affect both the structure of the network and its planning and operation. In order to overcome these problems, this work proposes a method to carry out the photovoltaic power forecasting, so that this method can later be used by distribution system operators.

The method used to perform the forecast uses techniques of artificial neural networks. Three different network structures were tested in order to allow a comparison analysis. The data used to train such networks were obtained from measurements made in the photovoltaic system installed in the building of the Department of Electrical and Computer Engineering. The results obtained by the networks were satisfactory and confirmed the validity of the method for this application. In addition, the comparative analysis performed between the network structures showed that the best performance was obtained from the network that only used data measured locally in the system.

This work can be further expanded to a master's project, where a commercial software for power forecast can be developed. In the future, this software can be released to assist companies in the energy distribution network industry.

Key-words: Photovoltaic Systems, Forecasting methods, Artificial neural networks.

Contents

1 Introduction **16**

1.1 Work organization 18

1.2 Definition of the problem 19

1.3 Goals of this work 22

2 Important concepts in photovoltaic systems **23**

2.1 Technology review 23

2.2 Operation of photovoltaic system 25

3 Methods for photovoltaic power forecasting **31**

3.1 Classification of photovoltaic power forecasting methods 31

3.1.1 Indirect approaches or parametric methods 32

3.1.2 Direct approaches or non-parametric methods 33

3.1.3 Hybrid models 33

3.2 State-of-art of forecast methods 34

3.3 Artificial neural networks 38

3.3.1 Basic aspects 38

3.3.2 Multilayer perceptron 47

3.3.3 Multilayer perceptron learning algorithms 49

4 Application of the method **57**

4.1 Description of the photovoltaic system used to obtain the measurements 58

4.1.1 Selection of the inputs and outputs of the model 59

4.1.2 Desired output vector 59

4.1.3 Input vector 61

4.1.4 Training and validation set 62

4.2 Artificial neural network structure 63

4.2.1 Structure A 65

4.2.2 Structure B 66

4.2.3 Structure C 68

4.3 Training algorithms 69

5 Results and Discussion **71**

5.1 Training algorithms performance evaluation 71

5.2 ANN performance evaluation 75

5.2.1 Results of power forecast for structure A 77

5.2.2 Results of power forecast for structure B 79

5.2.3 Results of power forecast for structure C 80

5.2.4 Compilation of power forecast for different structures 82

6 Conclusion **84**

6.1 Conclusions about the study 84

6.2 Future work 85

Appendices **86**

A PV system installation scheme **87**

B List of publications generated by this research **89**

List of Figures

2.1 Photovoltaic cell model 25

2.2 Photovoltaic panel I-V curve 28

2.3 Irradiance influence over the photovoltaic power generation 29

2.4 Temperature influence over the photovoltaic power generation 30

2.5 Photovoltaic generation daily profile of the photovoltaic system installed in the Department of Electrical and Computer Engineering 30

3.1 Classification of photovoltaic power forecast 34

3.2 Artificial neuron model 39

3.3 Activation functions: partially differentiable functions 41

3.4 Activation functions: fully differentiable functions 41

3.5 Generic model of Multilayer Perceptron Network 48

3.6 Multilayer Perceptron Network indexes 49

4.1 Photovoltaic system installed on the roof of the Department of Electrical and Computing Engineering 58

4.2 Inverter and the acquisition system installed on the Department of Electrical and Computing Engineering 58

4.3 Black Box Model 62

4.4 Generic model of structure A 66

4.5 Generic model of structure B 67

4.6 Generic model of structure C 68

5.1 Comparison of the evolution of the errors of the BP and LM algorithms for the best topology ([4 2]) 72

5.2 Comparison of the evolution of the errors of the BP and LM algorithms for topology [5 2] 72

5.3 Comparison of the evolution of the errors of the LM algorithm 73

5.4 Evolution of the errors of the BP algorithm for topology [4 2] for high number of epochs 74

5.5 Overfitting situation for topology [15 10] - training set (Red line - Measured values; Blue line - Forecasted values) 75

5.6 Overfitting situation for topology [15 10] - test set (Red line - Measured values; Blue line - Forecasted values) 75

5.7 Power forecast for structure A 77

5.8 Training error evolution for structure A 78

5.9 Power forecast for structure B 79

5.10 Training error evolution for structure B 80

5.11 Forecast results for structure C 81

5.12 Training error evolution - Structure C 81

List of Tables

4.1 PV module data sheet 59

4.2 Inverter data sheet 59

4.3 Correlation analysis between the climatic variables and the generated power 62

4.4 Input and desired output set 63

4.5 Input and desired output set for structure B 67

4.6 Input and desired output set for structure C 68

5.1 Results of the MLP training made with the LM and the BP algorithm 71

5.2 Value of the variables during the analyzed day (07/03/2018) 76

5.3 Compilation of results 82

Chapter 1

Introduction

In 2015, the United Nations (UN) defined the ”17 Sustainable Development Goals (SDGs)”, in which several nations made a commitment to do real efforts to reach these objectives. The main goals of the UN are to end poverty, inequalities and reduce climate change. Regarding with climate change, the seventh objective (”affordable and clean energy”) deals directly with the role of energy generation in the reality of climate change. It is expected, by 2030, the significant increment of renewable energy sources, such as photovoltaic, wind, hydro, among others, in the world energy matrix [1].

On the other hand, there is also a concern by governments to increase the share of renewable sources in this energy matrix to supply the increasing of energy demand. This concern is also due to the need to reduce dependence on fossil fuels (oil, coal, gas) for energy production, because of their reduced availability and because of the negative impact on the environment that these fuels cause. Due to this fact, renewable energy sources have presented exponential growth in relation to the space occupied in the global power generation matrix. Whether through public or private incentives, the number of facilities of this form of generation has increased significantly in recent years. One example is Germany, which in 2013 documented that 25% of its energy demand was supplied through renewable sources [2].

Among the renewable energies, the one that shows the greatest growth is the photovoltaic energy. For example, the California Solar Initiative (CSI) program, which began in 2007 and plans to increase the photovoltaic installation in the US state. In the period from 2007 to 2016, it was invested 2.167 billion dollars for the installation of 1.940 MW of solar power. Currently the state has an installed capacity of 6.331 MW and continues with incentive programs that help people with reduced rates and aid in the projects of installation of photovoltaic systems [3]. Moreover, a new approved law in the state (2017), requires that from 2020, all new buildings must have photovoltaic

systems installed. [4].

In Brazil, the National Electric Energy Agency (or Agência Nacional de Energia Elétrica (ANEEL), in Portuguese) showed in its Technical Note n^{o} 0056/2017 [5] that the installed capacity of renewable sources in Brazil increased from 0.4 MW in 2012 to 114.7 MW in 2017, characterizing the exponential growth mentioned above. In addition, photovoltaic energy was the one that registered the most outstanding growth, and represents 70% of installed national power, referring to renewable sources [5]. Besides that, the report showed that most of the PV installations correspond to residential consumers, which indicates the increase of micro and mini-generation units in the country.

This type of generation, mini and micro, was also promoted by ANEEL by the Normative Resolution n^{o}482/2012 and updated by Resolution N^{o}687/2018. Those resolutions implemented the credit system for micro and small-scale projects. Thus, consumers who produced energy in their own homes were able to inject into the grid the excess energy that was generated in exchange for credits, which could be used to compensate the cost of the energy bill [6, 7].

However, this exponential increase of photovoltaic installations is not accompanied by the capacity of the distribution system to receive this new source of generation. Although the use of photovoltaic systems for generation is beneficial in sustainable development and environmental terms, large insertion of this generation in the current distribution and transmission networks could affect seriously the quality of the energy and, consequently, the security of the system.

Many studies of the impacts caused by the massive insertion of photovoltaic systems in the distribution network have been developed [8, 9, 10]. Those studies show the necessity of preparing the current distribution system for this new energy scenario. Such measures should be done both at the physical structure of the system and in the planning and operating levels.

In order to lead with problems in the planning and operation of the system, in this research is proposed a tool for photovoltaic power forecasting using an artificial neural network (ANN). The ANN was chosen because of its capacity to deal with hard non-linear problems, as is the case of photovoltaic power. Furthermore, these networks emulate the human brain's ability to solve problems and present a high level of success in many applications. Once the photovoltaic power is forecasted, operators of the distribution system can take correct decisions to diminish the negative impact of photovoltaic installations.

1.1 Work organization

The structure of this works is divided in the next chapters:

- In the chapter ”Introduction” (chapter 1), firstly a discussion will be held on the negative impacts that the large scale insertion of photovoltaic systems generates in the distribution network, in section 1.2. Then, the main goals of the study will be explained in section 1.3;
- In ”Important concepts in photovoltaic systems” (chapter 2), is made an explanation about the functioning and implementation of photovoltaic systems;
- The chapter ”Methods for photovoltaic power forecasting” (chapter 3) presents a literary review that brings together the main works of the present day that deal with the subject of photovoltaic generation forecast. Then, in section 3.3, the theory about artificial neural networks is presented;
- The ”Aplication of the method” chapter (chapter 4) introduces the methods and techniques that were used to carry out the forecast of photovoltaic generation. In this section, an analysis is made on the implementation of ANNs and on obtaining the data used to train these networks;
- In ”Results and discussion” (chapter 5), the forecast results are shown and, based on its results, a critical analysis is made on the validity of the proposed forecast method;
- The last chapter, ”Conclusion” (chapter 6), presents the general and partial conclusions developed during the development of the work. Also, is presented a compilation of scientific productions published on the topic by the research group.

1.2 Definition of the problem

As pointed out in the previous section, photovoltaic systems have received strong incentives for their implementation due to their capacity to generate energy using renewable sources. In addition, other advantageous of such systems are the low maintenance required, the low operating cost and the long service life [11]. In regard to this last feature, a lot of researches have been performed to increase the life and efficiency of the photovoltaic panels. In [12], a study is accomplished to deal with the reduction of losses in the photovoltaic system and to create a cost-benefit relation that is increasingly more advantageous for the consumer. It is supported because each day new panels with greater efficiency and lower cost are launched in the market.

In addition to the environmental and economic advantages aforementioned, some studies show the benefits that the integration of photovoltaic systems can bring to the distribution network. In [13] is shown a case where the insertion of the solar generation in the network improves the oscillatory stability of the system. That research evaluated the distribution network before the integration of solar generation, in this case the system presented unstable modes which characterizes the lack of damping torque. However, after the photovoltaic system was inserted, damping torque to the unstable modes was provided allowing the system to recover its stability.

Even for high load simulations, that situation was still repeated. The study also concluded that the benefits created by solar generation are stronger when this resource is accomplished in a distributed way. That is, instead of large photovoltaic systems (such as a solar farm), several small distributed systems installed in a large area are better [13].

However, even with all the advantages aforementioned, the massive insertion of photovoltaic systems can cause negative impacts that must be studied and analyzed in order to create solutions or forms of mitigation for these problems. In [2], some impacts are cited:

- Reverse power flow in distribution systems;
- Reduction of energy quality levels: unacceptable levels (fluctuations) of voltage and frequency;
- Reduced service life of certain elements of the distribution system, such as voltage regulators;
- Increase in losses in distribution system lines;
- Difficulties in planning the operation of the distribution system.

The reverse power flow occurs in situations where the power generated by the PV system exceeds the demand of the distribution system. Thus, a bidirectional power flow is verified in the network. This bidirectional flow

affects the system protection because the current protection systems is not prepared for this new reality, that is, many devices do not work in a bidirectional way. In addition, the voltage regulation is also affected, because the increased flow can cause damage to the voltage regulator equipment, reducing its life [14]. Furthermore, the increasing active power available in the network can reach such high levels that the flow can be directed to the transmission lines [14, 2].

In [8] is verified the impacts caused in the distribution network voltage regulation when considering 20% of photovoltaic generation insertion. That study showed that in normal operation, without photovoltaic generation, the grid voltage regulator performed the change of the tap 40 times, for a sampling interval of 1 minute for the voltage variation, considering a period of three days. However, when inserting the photovoltaic generation, the intermittent nature of this triggered an increase in the number of changes of the tap of the voltage regulator, reaching 160 changes in the same time interval (three days). That is, there was an increase of 4 times in relation to the previously registered value. This rise is due to the movement of the clouds over the panel, which influence its generation and consequently cause fluctuations in the voltage levels.

These fluctuations require that the voltage regulator acts to guarantee the stability of the system and, consequently, these fluctuations can reduce the useful life of the voltage regulating equipment due to the wear of its physical parts. Moreover, the study suggested that phenomena such as flicker can be observed in the system due to the insertion of PV generation. However, to substantiate this claim, further studies should be developed considering other sampling rates for the system voltage (with intervals of seconds rather than minutes) [8].

In [10], voltage and frequency fluctuation were observed for high penetration levels of photovoltaic generation. Nevertheless, that it is possible to increase the level of PV generation with small impact in the system when the increment is made through small PV systems distributed over a large area [10, 13]. In addition, [10] showed that it is better to PV systems to be implemented in distribution networks with fast response, because these networks do not present so many variations in their frequency levels.

However, for energy quality, not only voltage and frequency fluctuations must be observed, but also the reduction in energy security that PV systems introduce into the grid. This lack of security is linked to the difficulties of planning the operation of the system. The uncertainty in photovoltaic generation makes it impossible to carry out a correct generation planning in relation to the system load, using traditional methods. This scenario affects the stability of the system and ultimately affects the consumer, who does not have the supply of his demand guaranteed.

Therefore, it is important to create solutions for the technical impacts on the network, such as voltage and frequency fluctuation and damage caused to feeder cables and equipment. For these problems, solutions can be applied, such as the implementation of PV systems through distributed micro systems ([13, 10]), the use of energy

storage systems to mitigate variations in generation levels ([11]) and the updating of certain equipment of the network, that can operate in these new conditions. In addition, it is also extremely important to develop solutions to the impacts caused to network operation planning. In this aim, methods of photovoltaic power forecast became relevant. These methods can be used to predict the power that will be generated by the PV system with a suitable time window, so that the system operator can allocate loads, control the use or not of storage systems, control the rotating reserves, and made all other decisions about the system in a conscious way.

1.3 Goals of this work

The main objective of this work is to carry out the forecast of the generation of a photovoltaic system using techniques of artificial neural networks. This forecast can be used as an aid for the planning and operation sectors of the power distribution system. This main aim can be divided into some specific aims:

- Verify the validity of forecasting methods that use artificial neural networks (ANN) to forecast photovoltaic power;
- Analyze the performance of different ANN approaches applied to the problem in question;
- Observe the effects caused in the forecast due to changes in training algorithms and data sets;
- Propose a forecasting method that can later be transformed into a simple software that can be used by companies of the electric distribution system.

Chapter 2

Important concepts in photovoltaic systems

2.1 Technology review

The interest in photovoltaic solar energy began around the year 1839, when the photovoltaic effect was discovered by Edmond Becquerel. The first solar cell was created in 1876, but for a long time the technology development was restrict to studies on the field. Only in around 1960 this type of technology started to be produced on a industrial scale [15, 16].

There are two landmarks that boosted the development of this technology. The first was based on the search for energy sources that could supply energy to isolated locations. The second one was the ”space race”, that used this type of technology to provide energy to equipments in space [15, 16].

The oil crisis in the 1970s also propelled this development. The search for new sources of energy has made companies from countries like the United States invest in the area. In the early 2000s the Asian market began to invest in the production of photovoltaic modules, achieving leadership in 2009 [15].

According to CRESESB, between 2003 and 2014, the growth of this industry reached the mark of 54.2% per year. This growth occurred mainly due to the incentives for the use of renewable sources rather than those dependent on fossil fuels [15]. In 2016 there were 294 GW of photovoltaic power installed in the world, with China having the highest installed power, followed by the United States and Japan [17].

China has invested heavily in the production and exportation of photovoltaic cells and, currently, Asia is the biggest producer of photovoltaic cells. However, even with increased production, the cost of photovoltaic modules is still a barrier to the deployment of this technology [15]. To overcome this challenge, governments in several

countries have created incentives for research and technological development. In addition, government aid is being provided to increase the installed power levels.

In Brazil, the technological development was more restricted to theoretical studies in research centers and did not evolve to the manufacture of cells and modules. However, despite not producing its own technology, Brazil invested heavily in the installation of photovoltaic systems. The mainly type of system installed in Brazil was to supply energy to areas that were isolated from the National Interconnected System (or Sistema Interligado Nacional, in portuguese (SIN)).

In 2012, Aneel with its Normative Resolution n^{o} 482/2012 set the parameters for the implementation of photovoltaic systems connected to the distribution network. And created a great impulse for the installation of mini and micro producers. [6].

The power installed in Brazil in 2017, related to these systems, was 438.3 MW distributed in 15.7 thousand initiatives [17]. Brazil currently seeks to encourage the growth of these facilities through the application of tax free policies on photovoltaic systems equipments, support plans by government agencies and banks, such as the Brazilian Development Bank (or, in portuguese, Banco Nacional do Desenvolvimento (BNDES)), and incentives for micro and distributed mini-generation systems. With such actions, it is expected that in 2050, solar energy will correspond to 9% of the total national energy supply [17].

One can notice the great interest in this type of generation. The increasing in the efficiency of these systems is related to this interest and also is due to the incentives for research and development. In the 1950s the efficiency of a photovoltaic panel was around 5% and the cost of its energy was US \$1.785/Wp. However, the module efficiency currently reaches values of 15%, with prices around US \$1.20/Wp [17].

In relation to the manufacture of the cells of these systems, silicon is the material most used for this type of application. About 95% of all cells manufactured in the world are made of this material [16, 17]. This is because, in addition to being a cheap material that exists in great quantity, the cells made from it already have a well-known and developed manufacturing process. Commercially there are three major types of photovoltaic cells: monocrystalline silicon, polycrystalline silicon and thin silicon film [16]. The characteristics of each type are shown below:

- Monocrystalline silicon: Monocrystalline photovoltaic cells are made from ultra pure silicon. Currently, this is the technology with the highest cost of manufacturing and also with more efficiency in the market, with values varying between 15 and 18% [16, 18].

- Polycrystalline silicon: This type of cell has a cheaper manufacturing process than that of monocrystalline

silicon because it has a less stringent control [18]. However, their efficiency levels are lower, varying between 13 and 15% [16].

- Thin films: This is a new type of technology that has very low manufacturing costs. For this type of technology there is no distinction between a cell and a panel. Therefore, the panels of this technology are composed by a large cell. The advantage of this approach is that the system becomes less susceptible to the effects of partial shading. However, the efficiency of these modules is very low compared to other technologies. Also, this panels have accelerated levels of degradation [16]. Among the thin film technologies, amorphous silicon cells can be cited.

2.2 Operation of photovoltaic system

To understand how does a PV cell works and, consequently, to understand the operation of a PV panel, two concepts are important: solar radiation and irradiance. Solar radiation is defined as the solar energy that reaches the earth in the form of electromagnetic radiation. This can be divided into two categories, the direct radiation and the diffuse radiation. The first directly hits a horizontal surface on the planet through a straight path traveled by sunlight. The second corresponds to the light that hits the surface indirectly, that is, after being reflected and diffracted during its course.

The solar irradiance or irradiation is a measure of solar radiation. Its units are $\mathrm{W/m^2}$, that means that the irradiance measures the solar radiation power in relation to the area. The value of $1000\,\mathrm{W/m^2}$ is used as a reference to evaluate PV panels performance.

Figure 2.1 shows the basic model of a photovoltaic cell. This model was built based on those presented by [16] and by [19].

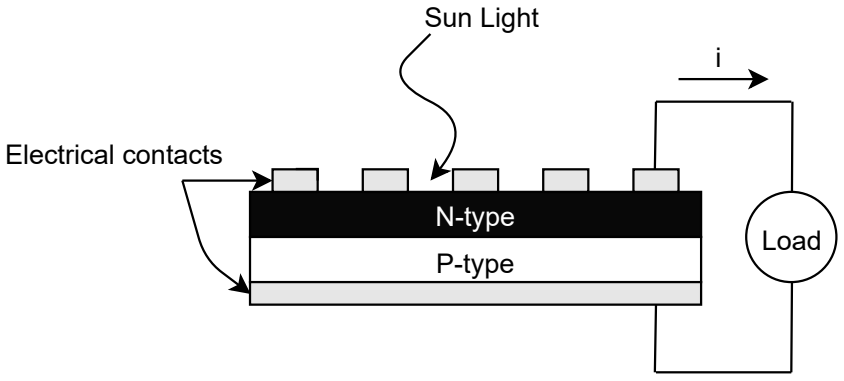


Figure 2.1: Photovoltaic cell model

Analyzing the Figure 2.1, one can notice that the photovoltaic cell is formed by the union of two types of semiconductors materials: the N-type and the P-type. There are also electrical contacts on both sides of the cell, used to close the electrical circuit in which the current flows.

The operation of a photovoltaic cell is based on the photovoltaic principle. This principle describes the process of transforming the incident solar electromagnetic radiation on the surface of the cell into electricity and will be explained in the following paragraphs.

When the light hits the surface of the cristalin silical, some of it is absorbed. If the energy of the incident light is low, it only causes the material to warm up, but does not cause the electrical properties of the material to be affected. However, if the light has enough energy, it can change those properties. In this way, when it reaches the crystal, it excites the electrons of the valence layer allowing them to move through the crystal.

When it detaches itself from the valence shell of its atom, an electron leaves a hole. Like electrons, holes also move through the crystal. This occurs because valence electrons located near the holes are attracted by those, leaving other holes when moving. The higher the radiation energy the crystal receives, the greater the shaking of electrons and holes.

To transform this excitation of electrons and holes into electric current, the potential barrier is needed. This barrier separates the free electrons from the holes, causing one side of the photovoltaic cell to have more electrons and the other to have more holes. In this way, a electric potential difference in the cell that can be used to obtain current is created.

The potential barrier in a photovoltaic cell is created through the junction between a negatively doped silicon shell and a positively doped silicon shell. Pure silicon has 4 electrons in its valence layer. When negatively doped, the material receives atoms of elements with five electrons in this shell. An example of an element that can be used as a dopant is phosphorus. Upon receiving an atom with more electrons in the last shell, there are more free electrons inside the cristalin silical. Thus, its electrical characteristics are altered and it can be called n-type material, with the electrons as the majority carriers and the holes as minority.

For the formation of the potential barrier, it is still necessary to use another material: a shell of silicon positively doped. A positively doped material is obtained when an atom is inserted which, relative to the original element, has one less electron in the valence shell. In this way, the crystal has an excess of free holes and thus can be called a p-type material, having the holes as major carriers and the electrons as minority carriers.

When the n-type and p-type material are brought into contact, the free electrons of the n-type material are attracted to the p-type material, filling in their holes. In addition, the valence electrons of the n-type material are

also drawn into the holes on the other side of the barrier, leaving holes in the n-type material. This process can be seen as if the holes pass from the p-type material to the n-type material. This movement of charges occurs through the junction between the materials. As it occurs, a cluster of positive charges on the side n of the junction begins to form, and a cluster of negative charges on the side p of this junction begins to form.

However, as the charge transfer continues, an electrical barrier in the junction begins to form that prevents the flow of new charges. This is because the holes that migrated to the negative side created a positive potential near the junction, which attenuates the migration of new positive charges, the opposite also happens with the electrons. Thus, over time, the transfer of charges ceases, the equilibrium is established and the potential barrier described above is created.

When the light reaches the surface of the n-type material, it creates a dissociation of electrons. The free electron remains on the n side of the junction. However, the holes are accelerated through the barrier, seeking to recombine with the electrons that form the negative charge of the barrier on the side of the p-type material. When they reach the p-side of the junction, the holes do not find free electrons to recombine, and the p-type material is left with excess holes. The opposite happens when the light hits the p-type material, causing the excess of electrons in the n-type material.

When connecting an electric circuit to the cell, an electric current is created through the electrons leaving the n-type material and passing through the circuit going to the p-type material to recombine with the holes. The amount of current produced is proportional to the energy of the electrons and, therefore, to the light energy received by the cell.

Without this electric circuit, there is no current. However, there is still a voltage between the sides of the cell. This voltage is approximately 0.6 V and it is cause by the potential barrier [16]. An isolated cell produces a low voltage level, therefore, several cells must be connected in series to generate practical voltage levels. A photovoltaic module is formed by a group of electrically connected cells. Depending on its voltage level it can reach up to 60 cells [16]. Regarding the electric current, it depends strongly on the level of incident solar radiation on the surface of the panel. Thus, the area of the cell is a factor that influences the production of current because it determines how much radiation the cell will receive.

The voltage and current of a photovoltaic module are dependent on each other. They can be related through the I-V (current vs. voltage) curve (shown in the image to the left of Figure 2.2). Each point on this curve determines a panel operating point and this point depends on the load connected to the terminals of the panel. Also, the curve has three important points, with values given by the manufacturer: point A - short circuit current; point B - maximum power; point C - open circuit voltage.

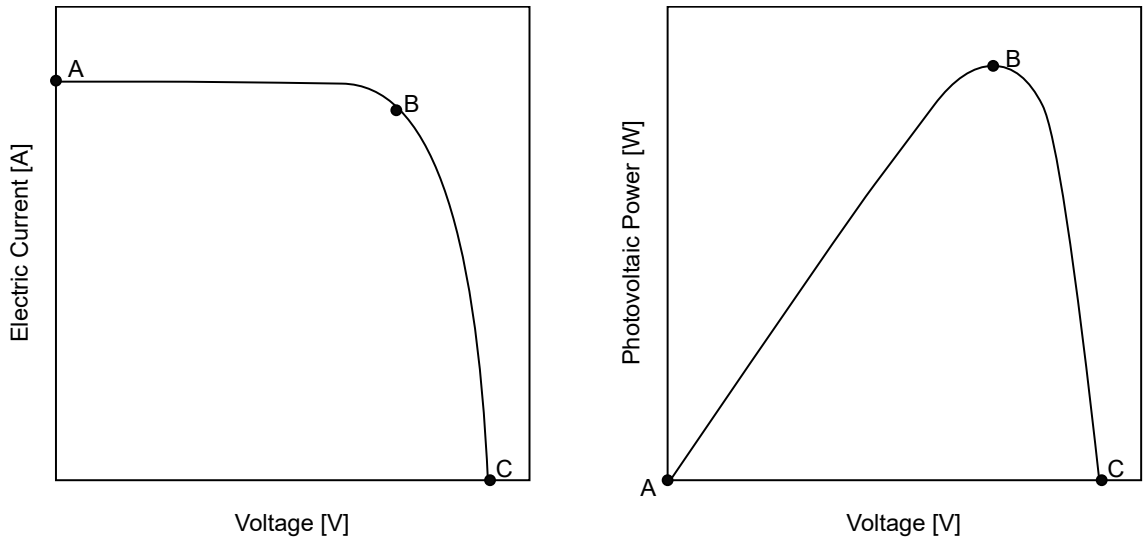


Figure 2.2: Photovoltaic panel I-V curve

The short-circuit current point characterizes the maximum current value that the panel can reach, this occurs when the terminals of the module are short-circuited. At this point, the voltage value is zero. Since the open circuit voltage point characterizes the maximum panel voltage, which occurs for the null current, this voltage is obtained for the panel operating with the open terminals.

The last important point is the one of maximum power, this is the desirable point of operation because it characterizes the greater generation of energy. To ensure that the photovoltaic system always operates at this point, inverters with Maximum Power Point Tracking (MPPT) inputs can be used. This type of device allows the system to produce its maximum power which improves generation levels and also reduces losses [16].

The modules of a grid connected photovoltaic system can be interconnected in series forming a structure known as ”string” [16]. In a system that has an inverter with MPPT input, this string can be connected directly to this input. If you want to increase the efficiency of the system, you can use inverters with more MPPT inputs, so the total set of modules can be divided into more strings and each one can be connected to the MPPT. In this way, each MPPT ensures that its string operates at the maximum power point.

Another curve that characterizes the functioning of these modules is the P-V curve (power x voltage). The three points described above can also be found in this curve. The knowledge of these points is important for the sizing process of photovoltaic systems and also assists in certain studies on these systems.

The I-V curve of a module can undergo through changes due to the influence of the climatic variables on the voltage and current behavior of the panel. The current is sensitive to variations in the irradiance levels on the panel:

the lower is the irradiance values, the lower will be the short circuit current values. This is presented in Figure 2.3.

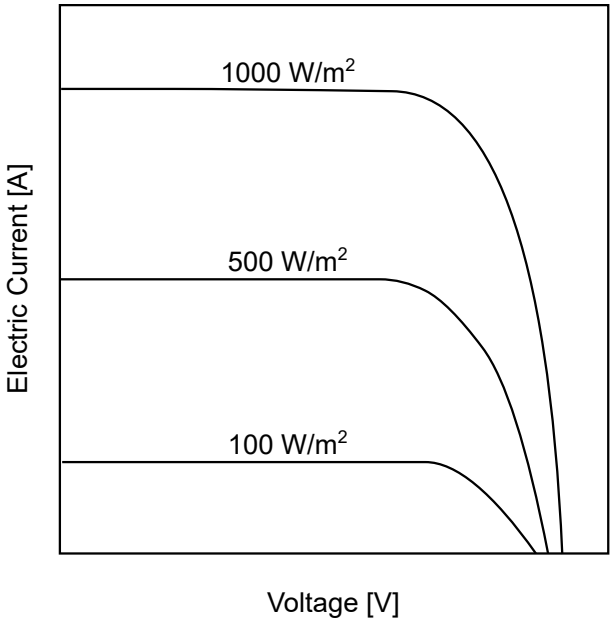


Figure 2.3: Irradiance influence over the photovoltaic power generation

In this way, shading has a strongly negative impact on photovoltaic power generation. The shading characterizes the reduction of the incidence of solar radiation on the system, and can affect the generation even when it occurs partially, that is, when only part of the module is shaded. This occurs because the electrical connection between the cells of the module is done in series, which leads to a cascade effect, which means that if a cell is disconnected from the system, all will be. To solve this problem, manufacturers have implemented by-pass diodes in a set of cells, so if one set is shaded, the others will not have their operation stopping. However, even with this solution, it is a fact that shading negatively affects the generation level.

Another climatic variable that influences photovoltaic generation is temperature. The change in temperature causes changes in the value of the open circuit voltage of a panel. The higher are the temperatures, the lower will be the voltage values, that is, the dependence is inversely proportional, as shown in Figure 2.4.

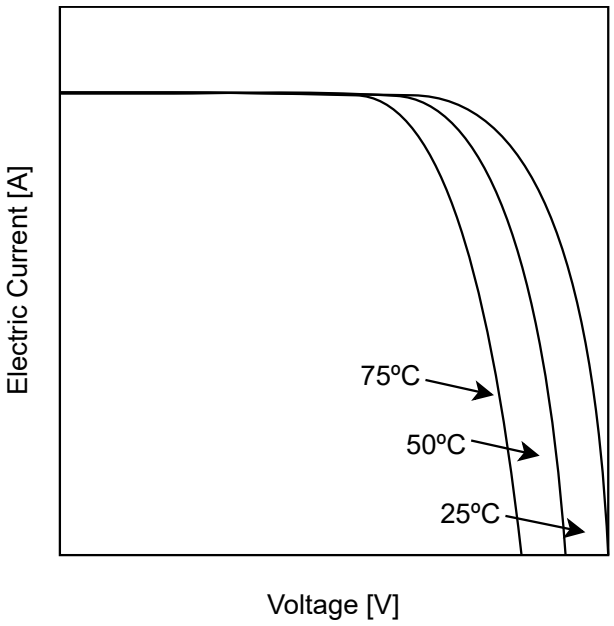


Figure 2.4: Temperature influence over the photovoltaic power generation

Given this, one can see how climatic variables influence the generation levels of a photovoltaic system. Considering that the behavior of such variables is chaotic, it is inferred that the generation of photovoltaic energy also exhibits a volatile behavior. Figure 2.5 shows the one-day generation profile of a real photovoltaic system installed in the Department of Electrical and Computer Engineering. It can be seen the influence of the intermittent behavior of the photovoltaic source and how abrupt variations occur in the profile, due to changes in the climatic variables.

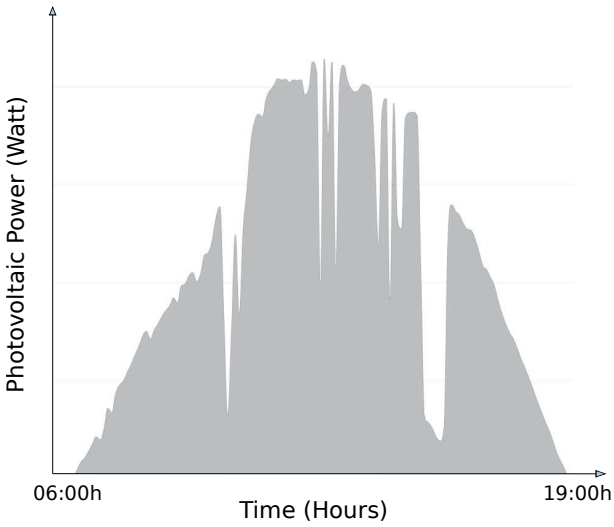


Figure 2.5: Photovoltaic generation daily profile of the photovoltaic system installed in the Department of Electrical and Computer Engineering

Chapter 3

Methods for photovoltaic power forecasting

Due to the necessity of ensuring the safe and planned operation of the distribution system in which the PV system is connected, numerous studies of photovoltaic power forecast have been developed. Most of those studies evaluate which method has the highest accuracy and at the same time is more easily implemented in practical situations.

This chapter presents the main works analyzed during the literature review. Such review was performed to collect information on various forecasting methods and to show examples of application and comparative studies between two or more methods. Therefore, the section is divided in two subjects: initially the classification of the methods proposed by some researchers will be explored; after that, studies dealing with different forecasting techniques will be presented, focusing mainly on research using artificial neural networks, which is the subject of this study.

3.1 Classification of photovoltaic power forecasting methods

In [11], is made a study of advances in the area of photovoltaic forecasting, where is suggested the division of methods into the following main categories: persistence method, physical techniques, statistical techniques and hybrid models. This study still suggests another category that encompasses new forecasting techniques. For Raza et al. (2016), the techniques of numeric weather prediction are into the category of physical techniques, while the use of artificial neural networks and time series for prediction are into statistical techniques.

However, this classification is not absolute. For [20], forecasting methods can be divided into three broad categories: direct methods, indirect methods and hybrid methods. The indirect methods are those that normally

forecast the global irradiance that reaches the photovoltaic plant and not the power generated by this plant, for this reason they receive this name. These methods also receive other names, such as PV performance, physical method, parametric and “white box”. Direct methods encompass statistical techniques and machine learning techniques, generally known as “black box” methods.

Hybrid methods come from mixing or joining two or more different methods. Theoretically, any methods can be combined in order to create a hybrid method. Thus, these can be subdivided according to the methods that were combined to create it, as an example can be mentioned the hybrid-statistical method that joins two or more statistical forecasting techniques in a single method [20]. In [21] such methods are called “grey box”.

Another form of classification found is defined in [22]. In that study, the forecasting methods are classified into four approaches: statistical approach, artificial intelligence approach, physical approach, and hybrid approach. Differently as stated in [11], [22] separates the class of methods that use artificial intelligence techniques from the statistical approach. For [22], this last one is composed by the persistence method and the time series methods. This also differs from what was presented in [11], where a single category is created for persistence methods.

It is worth emphasizing that for this research, the approach presented in [20] will be used. However, it should be noted that even though there are different forms of classification, the techniques to be classified remain the same. In the next section, the most relevant prediction techniques for this work will be discussed following the classification seen in [20].

3.1.1 Indirect approaches or parametric methods

This approach uses a model based on equations that describe the behavior of the system. This approach is also named ”white box”, as aforementioned. In this technique, usually the prediction of the global irradiance is initially obtained at the place where the photovoltaic system is installed. From this information it is possible to predict the generation of energy using a physical model of the photovoltaic system.

This method does not use historical data, instead it uses numeric weather predictor (NWP), and technical data of the photovoltaic system [20, 11]. The NWP consists in a set of equations that model the weather conditions of the PV system installation site [11]. According to Raza et al. (2016), the accuracy of the model is closely related to the stability presented by the climatic conditions. The more the variation of these conditions, the worse the prediction of the method.

Dolara et. al. (2015) states that these methods have advantages when compared with statistical approaches in cases where only fast results for simple systems are required. However, the disadvantage of these methods

lies at the point of the complexity level of the system, in other words, when the system becomes very complex mathematically, the results tend to be less precise [21].

3.1.2 Direct approaches or non-parametric methods

Statistical techniques

This approach uses statistical techniques to make the prediction. The model of this approach is called ”black box” because is not based on equations. Statistical techniques use historical system data to determine the relationship between the input and output information of the model [20, 22]. Such methods are based on learning and error reduction techniques, and are very dependent on the quality of the historical data available [11]. The statistical models can be divided into two subgroups:

a Regressive methods:

They are strongly based on the concepts of ”linear regression”, establishing relations between the dependent variables and the independent variables. Within this class, the main methods are: multiple linear regressions, auto-regressive models, autoregressive moving average, autoregressive integrated moving average, auto-regressive moving average with exogenous variables, non-linear AR-exogenous model, among others [20, 11].

b Artificial Intelligence techniques:

The main method in statistical techniques is artificial neural networks (ANN). Such networks are based on the functioning of the human brain and imitate its capacity for learning and generalization. In this way, they can be applied successfully in solving several problems, including non-linear problems [23]. Thus, ANNs have been successful when applied to solve complex problems. However, the disadvantage of such networks comes mainly from the fact that they require a large historical set of data [21]. Other artificial intelligence methods that can be cited are: k-nearest neighbors, support vector machines, and random forests [20].

3.1.3 Hybrid models

Hybrid methods have been applied successfully in several researches, usually presenting better performance than traditional methods [21]. Hybrid methods are generally considered more efficient than other types of methods because they unite the strengths of the techniques that compose them, while seeking to reduce the negative points of these techniques [22].

The main features of the classification of photovoltaic power forecast can be summing up in the next figure:

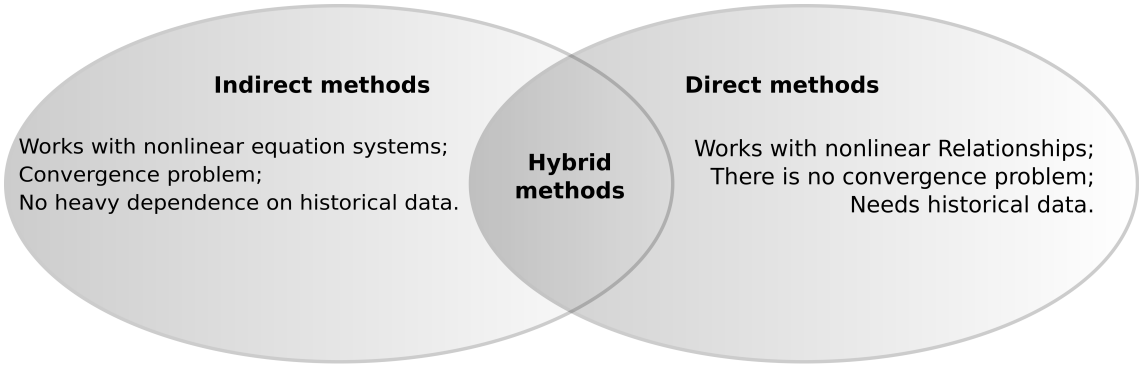


Figure 3.1: Classification of photovoltaic power forecast

3.2 State-of-art of forecast methods

As previously mentioned, several studies were developed with the objective of exploring the methods of forecasting photovoltaic power. In [11, 20, 22] an extensive literature review on such methods was carried out. In those studies each prediction method was explained and some examples were shown. In addition, basic concepts, such as forecast horizon and techniques to analyze the methods were developed.

In [20], an interesting concept was presented: the skill score concept (ss). That index is used to evaluate the performance of the model in relation to persistence models (simpler models, or reference models, used as a basis for comparison) [20].

Besides studies that analyze different forecasting techniques, there are also researches that focus only on the analysis of a single type of method, such as [24], or in the comparison between two or more methods, in order to evaluate the most appropriate for a given task. With respect to the latter type, one may [25] as an example.

In [25] a comparative study is made between the perfomance of the method using artificial neural networks and three physical methods: Osterwald, Araujo-Green and Single Diode Model. The study applied the methods to 3 different photovoltaic systems, using as input, irradiance and temperature data. Thus, the researchers were able to derive as output typical V-I curves of the photovoltaic panel for different input values. The results corresponded to the period of one year.

In all three PV systems analyzed by [25], the ANN method presented smaller prediction errors (errors between 6% and 8%), while the other methods presented errors between 6% and 30%. The justification presented by Almonacid, et. al. (2011) for better performance of ANNs is based on the fact that such networks, when executing the forecast, take into account a greater variety of losses present in the photovoltaic system, such as losses caused

by temperature, irradiation, inclination angle of the panel, among others. Contrary to classical methods, which consider only temperature losses, being more distant from reality [25].

The photovoltaic power generation prediction through the use of ANN techniques has also been reported in [26]. In this, a comparative study was carried out to evaluate the performance of two neural networks applied in this task. The ANNs used were the Multilayer Perceptron and the Elman Neural Network. The Elman Network is a recursive network that contains ”reaction connections” between the neurons of the same layer or of different layers [26].

Both networks used the Backpropagation training algorithm. For the prediction, the input data used was the historical output power values of the PV system. This data was handled as a time series and the adjustment between the input and desired output samples was done using a moving window. In this window, t samples of the series were initially chosen as input to the network, and the sample $(t+1)$ was set as the desired output. After this step, the window moved one position forward and the new range of input samples was built by rejecting the first element of the previous window and including the sample $(t+1)$ [26]. The results were considered satisfactory by the researchers, reaching errors around 0.5%. Also, the researchers inquired that if climatic factors were taken into account in the input variables, the accuracy of the method could be improved [26].

Another comparative study was carried out in [27], where the accuracy of two neural networks applied in the prediction task was compared. The neural networks used in such research are considered dynamic neural networks, which, in contrast to static networks (such as MLP), require less historical data for their full functioning. The networks used were the Focused Time Delay Neural Network and the Distributed Time Delay Neural Network.

The input data used was also based only on the historical data of the time series of the system’s power output power, making use of a moving window, as in [26]. Thus, unlike what normally occurs in studies using ANN, it was not necessary to use historical climatic data, such as irradiance and temperature, from the system installation site.

The results obtained were satisfactory and the argument that this approach is more applicable in reality was raised, because this method does not require a large number of historical data and is not linked to the installation site of the photovoltaic system [27].However, this conclusion contradicts the point raised in [26], which argues that the use of climate data would improve the performance of the forecasting system.

Also, in the application of ANN techniques, an example is given by [28], this work, unlike the previous ones, does not consist of a comparative study. Ding et. al. (2011) forecasts the generation of a photovoltaic system with a forecast window of 24 hours. Using a feed-forward neural network (MLP) and historical power generation data

from the system, it was possible to directly predict the power output [28].

It is worth mentioning two interesting points of such research ([28]). The first one is that the network was trained from the improved backpropagation algorithm. This algorithm consists of an optimized version of the standard backpropagation algorithm and aims to overcome certain difficulties and defects presented by this one. For this, a variable learning rate was implemented, this learning rate was modified according to the proximity between the network and the optimal solution of the problem (minimum prediction error), that is, according to the learning level of the network. This improvement of the algorithm avoids problems such as: convergence to a local minimum of the mean square error function and network instability around the optimal solution [28].

Another interesting method used in [28] is the use of Similar Day Selection Algorithm. This algorithm allows to find historical data of days with similar climatic characteristics with those of the day in which the forecast is wanted, so that this data can be used as input of MLP. This allows improvements in forecasting, especially for days when there is not much solar radiation (non-sunny days) [28].

Another research that explores the use of ANN is [24]. Lo Brano et al. (2014) makes a comparative study between the performance of three different topologies of artificial networks applied in the prediction of a photovoltaic system generation. These topologies are: multilayer perceptron with one hidden layer, recurrent neural network multilayer perceptron and gamma memory artificial neural network. The results presented by Lo Brano et al. (2014), indicate the validity of the use of these networks in the forecast of photovoltaic generation, with prediction errors lower than 1% of the peak power value measured in the real PV system [24].

Also in [24], an interesting approach to input data preprocessing is presented. In this approach a correlation analysis is made between the possible input variables and the desired output variable: the generated photovoltaic power. Therefore, it is possible to verify which input variables have the greatest effect on the output variable. Thus, the entries chosen for ANNs are: air temperature; cell temperature; solar irradiance; wind speed; open circuit voltage and short circuit current [24].

Ronay et. al.(2017) developed an interesting research involving ANNs. This study is interesting because it goes beyond the prediction of the output power generated by the PV system. In this, a solution is proposed for the best operation of a micro-grid composed by a photovoltaic system and by a reversible hydroelectric plant [29]. This solution consists in using the forecasting system to allow the planning of the use of the reversible hydroelectric and consequently of the operation of the whole micro-grid [29].

This approach presented in [29], is interesting because, as explained in [11], a PV system without an energy storage mechanism, that can be used to absorb the oscillations of photovoltaic generation, is detrimental to the

stability of the distribution network. Thus, Ronay et. al.(2017) uses the reversible hydroelectric plant as this storage mechanism. This works as follows: in case of photovoltaic supply failure, the hydroelectric plant starts working, supplying power to the grid. If, on the other hand, the PV system generates more energy than the demand requires, the hydroelectric plant stores this surplus energy in the form of gravitational potential energy. This avoids the lack or excess of active power in the distribution network [29].

The application of a hybrid technique can be observed in Dolara et. al. (2015). In this study it is proposed the union of a parametric technique and a statistic to make the forecast, creating a ”grey-box” method. The parametric or physical technique consists of the Clear Sky Solar Radiation Model (CSRM) method. It performs the modeling of the irradiation incident on the installation site of the photovoltaic system, this irradiation is modeled considering that the sky is clean, that is, without any clouds. With these considerations the CSRM can determine the period during which the photovoltaic system receives irradiation daily and what the maximum value this irradiation [21]. After such parameters are obtained, a neural network is used to predict the energy generation.

Dolara et. al. (2015), still performs the comparison between the performance in the prediction made by the hybrid method and by the use of an ANN (MLP) alone. For both situations the algorithm used for ANN training was the BackPropagation (BP) with the Levenberg-Markquard. The results indicate a better performance of the hybrid method. The author affirms the importance of a method of predicting climatic conditions to improve the efficiency of the photovoltaic generation forecasting [21].

Another interesting approach using hybrid models is presented in [9]. In this, Ogliari et. al. (2013) develops a hybrid prediction method that associates Dynamic Genetic Swarm Optimization (GSO) with ANNs trained with the BP algorithm. This union was made with the objective of increasing the accuracy of the forecast by reducing the defects presented by the EBP algorithm. The results obtained by the hybrid approach were considered satisfactory for a practical application of the forecast. In addition, these results were superior when compared to ANN results acting alone.

After analyzing this set of studies that was presented, some conclusions can be made:

- The treatment of the input data of the forecasting system is extremely important and directly influences the accuracy of the results;
- The inclusion of meteorological data usually helps to improve forecast accuracy. For this reason, techniques like the Similar Day Selection Algorithm are interesting because they include the meteorological factors in the analysis of the problem;
- As previously noted, hybrid methods excel at forecasting tasks, by joining the forces of other methods;

- The choice of input variables for the forecast model that are consistent with the problem is very important. Therefore, correlation analyzes between variables are interesting in the initial stages of the research.

3.3 Artificial neural networks

3.3.1 Basic aspects

According to Jain et. al. (1996), in its history the artificial neural networks (ANN) have gone through three landmarks. The first in 1943 with the work of McCulloch and Pitts, who were the first to address the subject and to develop the initial concept of artificial neurons. Later, in 1960, the Perceptron network won space for Roseblatt's work, but was soon discredited because of its limitations in dealing with problems that are not linearly separable, the disadvantage of the single layer perceptron network was highlighted by Minsky in his work. After this drawback, just a few researches were developed in the area until the 1980s.

However, in that decade, the development of the backpropagation algorithm has rekindled researchers' interest [30, 23]. After that, advances in the area have been notable, new algorithms have been developed and these networks have been used in many different areas, such as economics, health, computing, engineering, among others.

The artificial neural networks, have their basic structure and functioning based on the human brain. Their aim is to emulate the brain's ability to solve problems [23, 31]. Haykin (1999), defines these networks as follows: "A neural network is a massively parallel distributed processor made up of simple processing units, which has a natural propensity for storing experimental knowledge and making it available for use" [31]. These networks have been applied in several tasks, such as: pattern's classification, functions approximation, prediction of parameters, systems optimization, among others [23].

By simulating the behavior of the human brain, these networks have the capacity to learn, to adapt by experience, to adapt to changes in the environment, to generalize solutions, besides being resilient to internal situations of failure [23]. The network learns how to generalize when it can find reasonable outputs to unknown inputs, that means, inputs that were never presented to the network [31].

There are numerous different architectures of neural networks. Each one performs its functions in a characteristic way. However, all of them have their construction based on a basic unit: the artificial neuron.

Artificial Neuron

The artificial neuron is the fundamental processing cell that makes up the ANN. It is based on the biological neuron, so that various analogies can be made regarding the functioning of both. Each ANN neuron receives information on its input channels, performs the weighting of such information and, if certain predetermined requirements have been met, passes them on to the next neurons through its output.

This process is analogous to the one performed by the human brain, in which the biological neurons receive the information through their dendrites and pass this information through synapses to the other neurons, using an output structure called the axon. It is worth mentioning that as in artificial neurons, the information is weighted and its passage to other neurons or to the external environment is directly affected by this weighting. This is the most important aspect for the ANN, since the learning process occurs by adjusting the synaptic weights that weighted the inputs of the neurons [23, 31].

The Figure 3.2 is based on the artificial neuron structure, and will be used to explain in detail the functioning of these cells and the characteristics of each one of its parts. In Figure 3.2 one can see the structure of a **neuron k**. In this figure, it might be noticed the existence of the input signal. This input signal can be represented by a vector X_i. This input vector can have information from the external environment (inputs of the problem) or information from the output of other neurons in the network.

It is worth mentioning that, when such information comes from the external environment, it is interesting to apply a treatment in them, such as the normalization of its values [23]. As stated in chapter 3, several authors stress the importance of this preprocessing of input data for the good performance of the network.

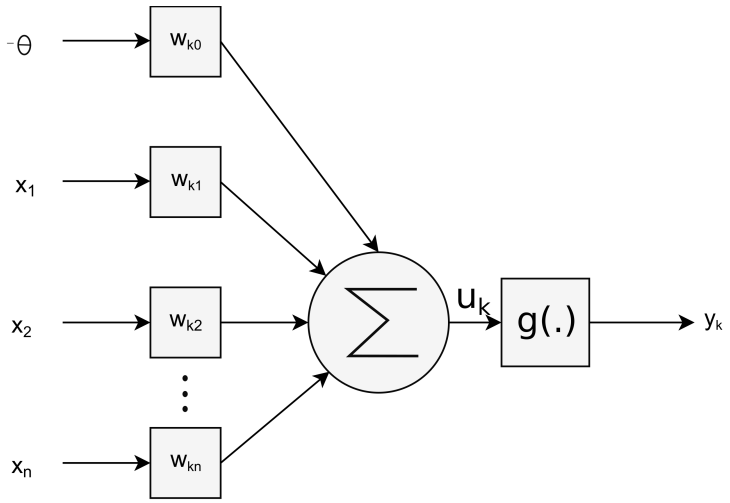


Figure 3.2: Artificial neuron model

After the inputs, are the sets of synaptic weights. Each input has its own synaptic weight, so that all inputs are weighted as them enters the neuron. The synaptic weights impress relevance on the input variables, adjusting the influence of each of them on the functioning of the neuron [23]. It is worth mentioning that the "Activation Threshold" ($-\theta$) was considered as an input to facilitate the notation of the problem. According to theory, the value of θ can be set to -1 and it will be weighted by the synaptic weight w_{k0}, which will represent the value of the activation threshold [23].

After going through the weights, the inputs pass through what Haykin (1999) calls "Summing Junction" and what Silva et. al. (2010) calls "Linear combiner", regardless of the name, this block has the function of performing the sum of the already weighted inputs and relate them in order to generate in its output the "Activation Potential" (u_k) [23, 31]. According to Silva et. al. (2010), the linear combination between the weighted inputs and the activation threshold should result in a value of u_k greater than or equal to zero for the activation potential to be considered excitatory, otherwise it will be inhibitory [23, 30]. The activation potential can be represented by the following relationship:

$$u_k = \sum_{i=0}^{n} w_{ki} \cdot x_i \tag{3.1}$$

The last block that makes up the neuron is the block of the activation function, it will process the data to generate the output (y_k) of the neuron k. This block is represented in Figure 3.2 by $g(.)$. The function implemented in this block takes as argument the activation potential (u_k) and has as output, the parameter y_k. This step is also necessary to limit to an allowable value the amplitude of the output of the neuron, usually this value is in the intervals between [0,1] or [-1,1] [31]. The parameter y_k can be either the final output of the network, ie, the solution of the proposed problem, or an intermediate output, which will be used as input by another network neuron [23, 31]. The relation that makes up the output signal of the network is given by:

$$y = g(u) \tag{3.2}$$

There are several types of functions ($g(.)$) that can be implemented in the neuron. The choice between them depends on the type of application for which the network will be used and the type of network architecture chosen. In Figure 3.3 the main partially differentiable functions can be seen.

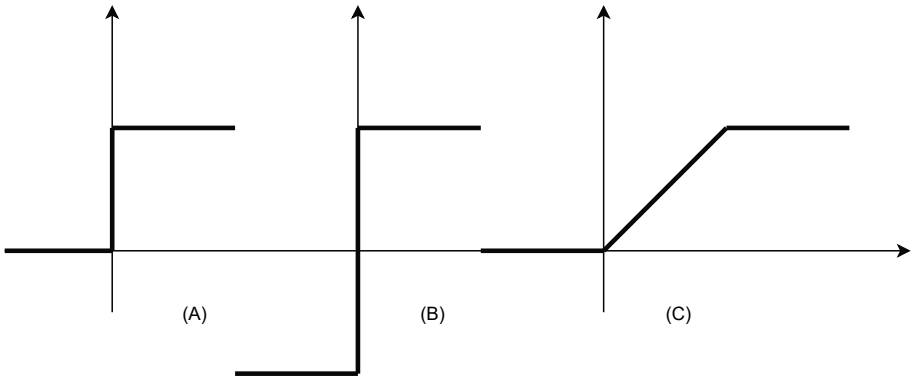


Figure 3.3: Activation functions: partially differentiable functions

The functions step, bipolar step and ramp, represented by figures (A), (B) and (C), respectively, are functions that are partially differentiable, ie, they do not have first-degree derivatives throughout their domain. The logistic, hyperbolic, Gaussian, and linear tangent functions (presented in Figure 3.4 (A), (B), (C) e (D), respectively) are considered to be fully differentiable [23, 30].

The choice of function is based on whether it is totally differentiable or not. Depending on the architecture of the chosen network and the training method used, certain operations require that the function is fully differentiable, which explain this choice. According to Jain et. al. (1996), the type of function most used, due to its properties, is the sigmoid function (logistic function and hyperbolic tangent) [30].

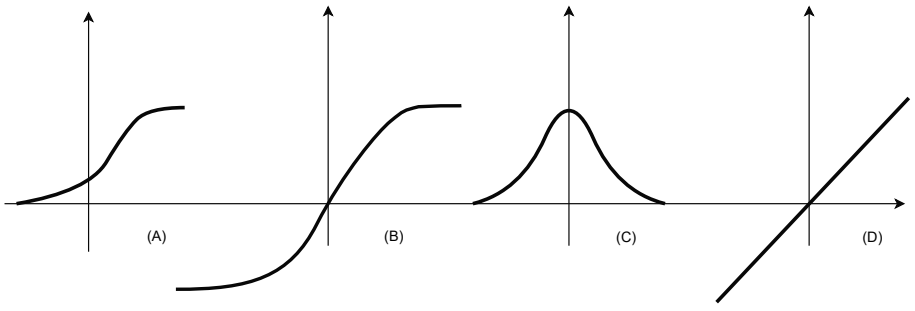


Figure 3.4: Activation functions: fully differentiable functions

This section described in detail the functioning of the artificial neuron. This is the basic cell of a artificial neural network and its understanding is of vital importance for the complete understanding of the next steps of this study. Thus, the functioning of the neuron can be summarized by the following system of equations:

$$u = \sum_{i=0}^{n} w_k i \cdot x_i \tag{3.3}$$

$$y = g(u) \tag{3.4}$$

Neural networks architecture

Silva et. al. (2010), presents the differences of definition between architecture and topology of neural networks. While the first deals essentially with the arrangement of the network neurons and how information flows between these neurons, the second covers more technical details attributed to the architectures, such as the number of neurons, number of layers of the network, type of activation function of neurons, among others factors [23].

A simple classification between types of architectures is proposed by [30]. In this, the architectures are divided into two categories, defined in relation to the form of connection between the neurons of the network. The difference between these categories is that the one of them encompasses networks that do not have feedback connections between their neurons, while the other deals only with networks that have such connections. These categories are [30]:

- Feedforward networks;
- Recurrent networks.

A more detailed classification was proposed by [23]. This classification has taken into account more aspects to define the categories, such as: the direction of the information flow in the network, the types of layers and the form of interconnection between the neurons. Therefore, for [23] the categories are:

- Single layer feedforward architecture, example: Perceptron and Adaline;
- Multi-layer feedforward architecture, example: Multilayer Perceptron;
- Recurring networks, example: Hopfield network;
- Self-organizing map networks, example: Kohonen.

In this paper the classification made by [23] will considered. Regarding the number of layers (single or multiple), ANNs can have three types of layers [23]: the input layer, the hidden layer, and the output layer.

As its name implies, the input layer is the one that receives the preprocessed input data from the external environment (the input data of the problem that is going to be solved, for example). The hidden layers are intermediate layers, their neurons have as inputs the outputs of neurons from the preceding hidden layer. And its outputs are used as input from neurons from another hidden layer. The output layer is the last to receive the network data from previous layers. Its output is the network's output, ie, is the solution to the proposed problem. The number of neurons that make up this layer are determined by the number of outputs required for the problem [23, 31].

Therefore, a single layer network has no hidden layers. This type of network, such as the single layer Perceptron, for example, consists of only the input layer and the output layer, and only the output layer has neurons. Hence, when referring to a single-layer network, only the output layer is considered. The input layer is not considered because no mathematical operations occur in it [23, 31].

In a multi-layer network, in addition to the input and output layers, there are also hidden layers. The neurons of these intermediate layers, act in the construction of the relation between the inputs and outputs of the network. Technically, the greater the number of neurons hidden, the greater the number of connections and, consequently, the greater the generalization capacity of the network [31]. Ergo, a multi layer network can be employed in complex problems that a single layer network would not be able to solve [23]. An example of a multilayer network is the Multilayer Perceptron (MLP).

In relation to the direction of the information flow, feedforward networks have a unidirectional flow, which always follows the path between the input layer and the output layer [23]. These networks are known as "static networks" because they produce only a closed output set, rather than a sequence of output values. Furthermore, because they do not have feedback, the past state of the network does not affect its current state, that is, such networks have no memory [30].

Another type of networks are the recurrent ones, these networks have feedback connections between their layers, so that the output of one neuron can be used as input to another [23]. In order for a network to be considered as recurring, it must have at least one feedback connection [31]. This type of network is also known as dynamic network and is widely used in systems that vary with time, since the past outputs can be used to produce the current outputs [30].

Finally, in the reticulated networks the input information is inserted in all the neurons of the network and each one has its own output. In this type of network there is no division into layers, the spatial arrangement of the neurons is what defines the adjustment of the synaptic weights [23].

The training algorithm of a network is related to its architecture [31]. Many basics concepts about neural

network training are common to different algorithms. Therefore, the following section will present such concepts.

Training the neural network

It is through training that the network learns. Earlier, it was stated that the ability to learn and adapt through experience was characteristic of ANN, this means that the network learns as if there was a process of ”study” to develop knowledge.

For the network learn about the behavior of a system, examples that model the characteristics of this system must be presented to the network. From such examples, the network has its internal parameters changed and starts to respond differently to its inputs [31]. The ability to learn by example is considered one of the advantages of ANNs over classical methods that depends on equations to model the system [30]. It is considered that the network ”learned” when it absorbs the relations given by the examples of the problem and is able to generalize the solutions to this problem, ie, when it returns responses close to those expected [23].

The training process is precisely used to create in the network this capacity of generalization. During the training the synaptic weights and the activation thresholds of the network are altered so that the outputs produced by it approximate the outputs expected for the problem [23, 30].

It is important to note that training is done from a set of samples, examples. Such samples should express the behavior of the treated problem in the best possible way, ie, they should reflect the characteristics of the system that will be studied by the network [23]. It is imperative that the maximum and minimum values of the input variables are part of the training set, since only knowing these values, the network can generalize over the entire domain of the problem. If the network encounters values outside of that domain during its operation phase, its responses will be unpredictable [23].

Another important aspect is that there should be a set of samples for the training, where the network will learn the relationships between the samples, and another set (totally different) for the validation phase of the network. In this last phase it is verified if the network is able to generalize its knowledge, ie, if it returns valid outputs for unknown input samples [23].

Another concept widely used during the training process is the concept of epochs of training. A training epoch occurs after all the input-output pairs of the training set are presented to the network. The number of epochs can be used as a training stop criterion. Another stop criterion that can be used is to check if the tolerance for the error has been reached.

In order to execute the learning process a training algorithm is necessary. There are several algorithms, each one with their characteristics. The basic difference between the training algorithms is the way in which the changes in the synaptic weights of the network are made. Also, the way in which the network receives interacts with the environment also differs between algorithm types. This way of receiving information is called by Haykin (1999) as ”learning paradigm” [31].

The two main learning paradigms are: supervised training and unsupervised training. In supervised training, also known as ”learning with a teacher”, the network outputs for a certain set of inputs are compared to the desired outputs (optimal values) for this same set. Therefore, it is possible to check the error between the desired value and the value obtained by the network. Thus, the training algorithm works by adjusting the internal parameters of the network in order to reduce this error and ensure that the network obtains answers that are closer to the desired ones [23, 31].

Normally the mean square error function is used to evaluate the training. The objective of learning is to ”walk” through the surface of this function in search of its global minimum, which would be the optimal value, since it is desired to reduce the error as much as possible [31]. The training is completed when it is considered that the network is presenting sufficiently satisfactory answers to the inputs that are imposed to it, ie, when it reaches a satisfactory minimum value of the function error.

In the other hand, the unsupervised training does not have the desired output values to be compared to the output values generated by the network. Learning is done through the network self-organization, the network searches for similarities between input samples and tries to create patterns and sets with them. Based on this categorization the network changes its synaptic weights internally [23, 30, 31].

Reinforced training is another type of learning paradigm. This type of training is considered by Haykin (1999) as a sub-category of unsupervised training and by Silva (2010), as a sub-category of supervised training. In any case, training with reinforcement seeks to interact with the external environment, extracting from it all possible information. This information is used to measure network performance, however, it should be noted that this information does not include the desired network outputs. The information only include parameters to indicate whether the network response is satisfactory or not and the network adjusts its weights based on this indication [23, 31].

Topology of the neural network

The performance of a neural network in a given problem depends on many factors besides the choice of its topology, such as data set characteristics and initial values chosen for the weight matrices. But network topology can also greatly influence this performance.

To decide which network topology to use, some methods can be applied. The most common is cross-validation. According to [23], there are three types of cross-validation methods:

- random subsampling cross-validation:

  In this method, the data set of the problem is randomly divided into two other sets: the training set and the test set (or validation set). Usually, in literature [23, 31], this division is made considering from 60% to 90% of the input-output patterns for the training set. And the others 40% to 10% are assigned for validation.

  For each topology the training can be repeated several times, for each time the division of the sets is randomly remade. In the end, the global performance of the topology is obtained based on the performance in each time.

- k-fold cross-validation:

  For this approach the data set is divided into k parts, where one part is used for the validation set and the others compose the training set. For each topology the training-testing process is repeated k times, until all parts are used as test set. As before, the global performance is given in relation to the performance for each training-testing process.

- leave-one-out cross-validation:

  For this method, only one input-output patterns is used as the test set, while all the other patterns stay in the training set. The learning process occurs until all the patterns are used as test set.

Other considerations about artificial neural networks

In literature is stated that more the number of neurons, more the network can generalize and therefore, lower is the training error. This statement is true as long as the network overfitting does not occur. In other words, a network with to few neurons will have trouble in generalizing and its errors during the training phase and the test phase will be very high.

In the other hand, the overfitting is the situation where the network has an unnecessarily large number of neurons and end up decorating the responses during the training phase. Thus, the network has a very low error during this phase. However, this does not mean that the network has a good level of generalization, it just means that it has memorized the data, so when it receives unknown data (as it happens in the test phase) it will present a very high error [23, 31].

Also, the initial guess of the network's weights is very important, because it can highly affect the convergence. During its learning and its operation, the network "walks" through the surface of the mean square error function in search of its minimum value, which is called global minimum, this is the optimal solution and it is the one that the network designer wants to find.

However, the path that the network travels depends on its starting point and this point is defined by the initial values of the synaptic weights of the network. Depending on these values, the network may travel a path that takes it to a local minimum point and it may not be able to get out of this point, considering that this is the optimal solution. To avoid this situation, the network might be trained a few times so that the starting point of operation is changed and different paths are traveled.

The last important topic that should be checked for the implementation of neural networks is the normalization that should be applied to the input data. This normalization process is necessary because it avoids saturation of the network activation functions. With this process, the output values obtained by the network will also be normalized and must go through the opposite process to be used in practical situations.

3.3.2 Multilayer perceptron

To solve the problem proposed in this study, the artificial neural networks technique was chosen. The Multilayer Perceptron Network (MLP) is the most commonly used network within the feedforward architecture [30].

This network was already successfully applied to several problems in different areas and is considered one of the most versatile architectures. Thus, this network was selected to be used in this project to solve the problem of photovoltaic generation forecast. This choice was also based on the existence of an extensive number of researches that have already used such networks to perform this same task.

This section will present basic concepts about this network, discussing its structure, its training algorithm and explaining its characteristics of operation.

Basic concepts

Multilayer Perceptron are networks that have an input layer, an output layer and also hidden layers. This network is part of the feedforward architecture class and, therefore, the information that flows through it follows only one way, the one from the input layer to the output layer, without feedback [23, 31].

In Figure 3.5, an example of a generic MLP is shown. In the figure, the net has as input a vector $\mathbf{X_i}$ with n elements. In addition, there are k hidden layers composing the net, the first one has $n1$ neurons, and the k_{th} layer has nk neurons.

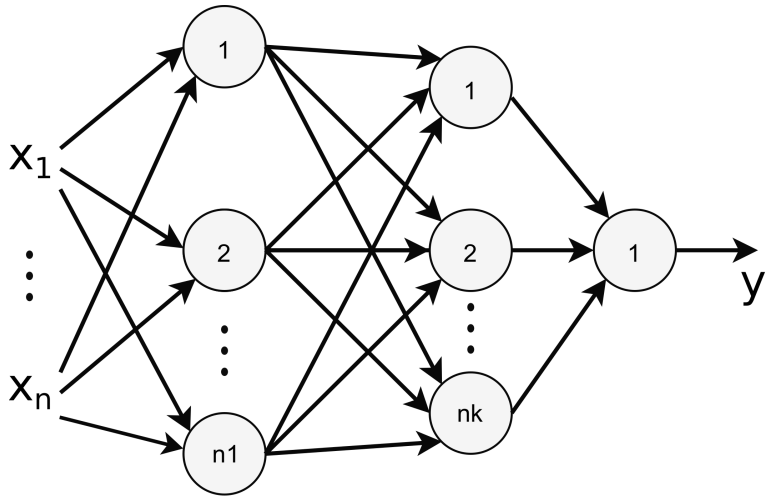


Figure 3.5: Generic model of Multilayer Perceptron Network

The output layer could contain as many neurons as needed to correctly represent a particular problem. In this example it was chosen to use only one neuron. Also, it is worth mentioning that the knowledge of the network is distributed in all its neurons, and the generalization task is performed mainly by the neurons of the intermediate layers [23]. Thanks to these layers the network is able to perform more complex tasks than Single Layer Perceptron networks and to extract more relations between the input-output pairs [31].

Another important feature of these networks is that their neurons have non-linear and totally differentiable activation functions, such as the logistic function. According to Haykin (1999), the non-linearity of the activation function is what differentiates the processing done by the MLP network from the processing done by the Single Layer Perceptron network [31].

The MLP training is done using the learning paradigm of supervised learning. Therefore, all the samples in the data set must have the input values and also the desired output. For the training, the Back-Propagation algorithm, or

some of its variations are used [23, 31]. In the next section the development of these algorithms will be presented.

3.3.3 Multilayer perceptron learning algorithms

In this section will be presented the MLP learning algorithms. First the traditional backpropagation algorithm will be described, then, another method will be introduced, the Levenberg-Maquardt method. To facilitate the calculation development, it is necessary to have a well-defined index set for the network. Therefore, Figure 3.6 shows the network parameters with the indexes that will be used throughout this section.

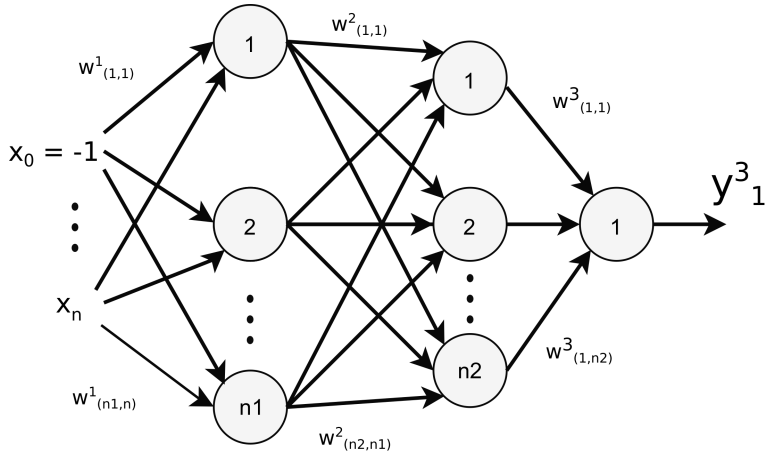


Figure 3.6: Multilayer Perceptron Network indexes

The backpropagation algorithm (BP) is a traditional training method to neural networks. It is used in a lot off application, but it is a know fact that the BP is a algorithm with slow convergence. The slow convergence of this method is due to the step size of the algorithm. The step size, also know as learning rate, determines how much the algorithm will move towards the optimal solution. When the learning rate is large, it can cause oscillation around the optimum solution, without reaching it. However, a small step size leads to a slow convergence, even when the gradient is smooth [32].

The Gauss-Newton method can improve the BP, because in this method the step size for each direction is chosen using second-order derivatives of the error function. But this method only works when the error function can be approximated by a quadratic function [32].

The Levenberg-Maquardt (LM) unites the BP and Gauss-Newton methods. It has the fast convergence of the Gauss-Newton method, but is more robust than this, converging to more complex surfaces. Also, LM has the stability of the BP method, but is much faster. By joining both methods, LM works as follows: When the error surface has a complex curvature, the LM works as the BP, but when the curvature approaches a quadratic function,

the method behaves as the Gauss-Newton algorithm, converging faster to the optimal solution [32].

Backpropagation algorithm

The algorithm is divided into two phases: the forward phase, where the network outputs (function signals) are obtained and the errors signals are calculated; the backward phase, where the weights are updated. It is important to note that the first layer to have its weights updated is the last layer. After this, the other layers' weights are updated in reference to the subsequent layer.

Also, the weights are usually updated on a pattern-pattern basis (online basis). That means that for each pattern from the training set, the weights have their value adjusted in accordance with the errors computed for that specific pattern. After all the patterns in the training set are presented to the network, a epoch is computed. The average error is calculated for each epoch.

Forward phase

For the forward phase, the input (**x**) and the desired output (**d**) matrix are presented to the net, both of these matrices are composed by p vector. Where p is the number of samples in the training set. For the first layer, $a_j^1(p)$ is the input vector of the j_{th} neuron of the first layer, for the p_{th} sample and $w_{j,i}^1(p)$, for: $(i = 1, ..., n)$ is the synaptic weight between the input i, from the input layer, and the neuron j, from the first hidden layer. And $x_i(p)$ is the i_{th} line of the network's input matrix (**x**), for sample p:

$$a_j^1(p) = \sum_{i=0}^{n1} w_{j,i}^1(p) \cdot x_i(p) \tag{3.5}$$

For the other layers:

$$a_j^k(p) = \sum_{i=0}^{nk} w_{j,i}^k(p) \cdot y_i^{k-1}(p) \tag{3.6}$$

One can notice that for $y_0^{k-1} = -1$ and $w_{j,0}^k = \theta$. The output of the j_{th} neuron of layer k is:

$$y_j^k(p) = f^k(a_j^k(p)) \tag{3.7}$$

Considering a M layer network, for a neuron in the output layer, it will be considered Haykin (1999) notation:

$$y_j^M(p) = o_j(p) \tag{3.8}$$

The error signal at the output of neuron j (considering the p_{th} sample) is given by:

$$e_j(p) = (d_j(p) - o_j(p)) \tag{3.9}$$

The quadratic error function will measure the performance of the network, this function is defined, over all neurons in the output layer, as:

$$E(p) = \frac{1}{2}\sum_{j=1}^{nM} e_j^2(p) \tag{3.10}$$

And the average squared error is obtained by:

$$E_{av} = \frac{1}{P}\sum_{m=1}^{P} E(p) \tag{3.11}$$

Where P is the total number of samples in the training set. One can notice that $E(p)$ and E_{av} are functions of the free parameters. These are the synaptic weights and the bias level, that in this study was considered also as a synaptic weight. Since (3.10) and (3.11) measure the error of the network, they also assess the learning performance of this net. Hence, looking at this as a optimization problem, the function given by (3.11) is the cost function of the problem, and the learning algorithm has to minimize this function.

Backward phase

In the backward phase the weights values are updated proportionally to $\frac{\partial E(p)}{\partial w_{j,i}(p)}$. This partial derivative may be expanded using the chain rule:

$$\frac{\partial E(p)}{\partial w_{j,i}(p)} = \frac{\partial E(p)}{\partial e_j(p)} * \frac{\partial e_j(p)}{\partial y_j(p)} * \frac{\partial y_j(p)}{\partial a_j(p)} * \frac{\partial a_j(p)}{\partial w_{j,i}(p)} \tag{3.12}$$

Where:

- Using equation (3.10):

$$\frac{\partial E(p)}{\partial e_j(p)} = e_j(p) \tag{3.13}$$

- Using equation (3.9):

$$\frac{\partial e_j(p)}{\partial y_j(p)} = -1 \tag{3.14}$$

- Using equation (3.7):

$$\frac{\partial y_j(p)}{\partial a_j(p)} = f'(a_j(p)) \tag{3.15}$$

- Using equation (3.6):

$$\frac{\partial a_j(p)}{\partial w_{j,i}(p)} = y_i(p) \tag{3.16}$$

Therefore, the adjustment of the weight is:

$$\Delta w_{j,i}(p) = -\eta \frac{\partial E(p)}{\partial w_{j,i}(p)} \tag{3.17}$$

Where η is the learning rate and $\frac{\partial E(p)}{\partial w_{j,i}(p)}$ is the gradient of the error function. The minus sign represents that the weights' update is made in the direction that minimizes the value of $E(p)$. Replacing in (3.17) the expressions found in (3.13), (3.14), (3.15) and (3.16):

$$\Delta w_{j,i}(p) = \eta \cdot (e_j(p) \cdot f'(a_j(p)) \cdot y_i(p)) \tag{3.18}$$

Defining the "local gradient" (δ) as:

$$\delta_j(p) = -\frac{\partial E(p)}{\partial a_j(n)} \tag{3.19}$$

It can be shown, using (3.13), that:

$$\delta_j(p) = -\frac{\partial E(p)}{\partial a_j(n)} = e_j(p) \cdot f'(a_j(p)) \tag{3.20}$$

Then:

$$\Delta w_{j,i}(p) = \eta \cdot \delta_j(p) \cdot y_i(p) \tag{3.21}$$

Nevertheless, the local gradient is not the same for a neuron j in the last layer and for a neuron j in a hidden layer. That happens because the error signal ($e_j(p)$) is not related in the same way to both neurons. Hence, the weights adjustment has to be divided into two steps: first, the last layer weights are updated; second, the hidden layers, one by one, have their weights updated.

- **Last layer weights update**:

  For a neuron j in the last layer, the error can be obtained in the same way as shown in (3.9). Therefore, considering a net with M layers and using the development presented above, the local gradient is given by:

$$\delta_j^M(p) = e_j(p) \cdot f'(a_j^M(p)) \tag{3.22}$$

  And, equation (3.17) for the last layer is given by:

$$\Delta w_{j,i}^M(p) = \eta \cdot \delta_j^M(p) \cdot y_i^{M-1}(p) \tag{3.23}$$

- **Hidden layers weights update**:

  Using the definition of the local gradient presented in (3.19), this may be obtained to a neuron j in a hidden layer k:

$$\delta_j^k(p) = -\frac{\partial E(p)}{\partial a_j^k(n)} = -\frac{\partial E(p)}{\partial y_j^k(n)} \cdot \frac{\partial y_j^k(n)}{\partial a_j^k(n)} = -\frac{\partial E(p)}{\partial y_j^k(n)} \cdot f'(a_j^k(p)) \tag{3.24}$$

  Since the error is relative to the values of the last layer, the considerations previously used can not be made. Therefore:

$$\frac{\partial E(p)}{\partial y_j^k(p)} = \sum_{m=1}^{nM} \frac{\partial E(p)}{\partial a_m^{k+1}(p)} \cdot \frac{\partial a_m^{k+1}(p)}{\partial y_j^k(p)} \tag{3.25}$$

  Where m is the index of a neuron in the layer $(k+1)$.

  Considering (3.6),

$$\frac{\partial a_m^{k+1}(p)}{\partial y_j^k(p)} = w_{m,j}^{k+1}(p)$$

  And, since:

$$\frac{\partial E(p)}{\partial a_m^{k+1}(p)} = \frac{\partial E(p)}{\partial y_m^{k+1}(p)} \cdot \frac{\partial y_m^{k+1}(p)}{\partial a_m^{k+1}(p)}$$

  That is equal to $-\delta_m^{k+1}(p)$.

Equation (3.25) is given by:

$$\frac{\partial E(p)}{\partial y_j^k(p)} = -\sum_{m=1}^{nM} \delta_m^{k+1}(p) \cdot w_{m,j}^{k+1}(p) \tag{3.26}$$

And following the definition given in (3.19), the "local gradient" of a hidden layer is:

$$\delta_j^k(p) = -\frac{\partial E(p)}{\partial y_j^k(n)} \cdot f'(a_j^k(p)) = \left(\sum_{m=1}^{nM} \delta_m^{k+1}(p) \cdot w_{m,j}^{k+1}(p) \right) \cdot f'(a_j^k(p)) \tag{3.27}$$

Then, (3.17) for the hidden weights is:

$$\Delta w_{j,i}^k(p) = -\eta \cdot \delta_j^k(p) \cdot y_i^{k-1}(p) \tag{3.28}$$

Levenberg-Marquardt algorithm

The Levenberg-Marquardt method is a optimization technique that can be applied to the BP method to speed up its convergence. The LM algorithm is derived from the Newton method. For the Newton method, a function $V(\mathbf{z})$ can be minimized with respect to the parameter vector $\mathbf{z}$ using the following procedure:

$$\Delta \mathbf{z} = -\left(\nabla^2 V(\mathbf{z})\right)^{-1} \cdot \nabla V(\mathbf{z}) \tag{3.29}$$

Where $\nabla^2 V(\mathbf{z})$ is the Hessian matrix and $\nabla V(\mathbf{z})$ is the gradient of $V(\mathbf{z})$.

If $V(\mathbf{z})$ is a sum of squared functions $\left(V(\mathbf{z}) = \sum_{i=1}^{m} e_i^2 \right)$, the following relations are true:

$$\nabla V(\mathbf{z}) = J^T(\mathbf{z}) \cdot \mathbf{e}(\mathbf{z}) \tag{3.30}$$

$$\nabla^2 V(\mathbf{z}) = J^T(\mathbf{z}) \cdot J(\mathbf{z}) + S(\mathbf{z}) \tag{3.31}$$

Where $S(\mathbf{z})$ is given by:

$$S(\mathbf{z}) = \sum_{i=1}^{m} e_i(\mathbf{z}) \cdot \nabla^2 e_i(\mathbf{z})$$

For the Gauss-Newton method, the term $S(\mathbf{z})$ is considered zero. And $J(\mathbf{z})$ is the Jacobian matrix, given by:

$$J(\mathbf{z}) = \begin{bmatrix} \frac{\partial e_1(\mathbf{z})}{\partial z_1} & \frac{\partial e_1(\mathbf{z})}{\partial z_2} & \cdots & \frac{\partial e_1(\mathbf{z})}{\partial z_n} \\ \frac{\partial e_2(\mathbf{z})}{\partial z_1} & \frac{\partial e_2(\mathbf{z})}{\partial z_2} & \cdots & \frac{\partial e_2(\mathbf{z})}{\partial z_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial e_m(\mathbf{z})}{\partial z_1} & \frac{\partial e_m(\mathbf{z})}{\partial z_2} & \cdots & \frac{\partial e_m(\mathbf{z})}{\partial z_n} \end{bmatrix} \tag{3.32}$$

Where m is the number of square functions and n is the number of elements of vector $\mathbf{z}$. Replacing equations (3.30) and (3.31) in equation (3.29), the Gauss-Newton update is:

$$\nabla \mathbf{z} = -\left(J^T(\mathbf{z}) \cdot J(\mathbf{z})\right)^{-1} \cdot J^T(\mathbf{z}) \cdot \mathbf{e}(\mathbf{z}) \tag{3.33}$$

The Gauss-Newton method is already faster than the BP algorithm, but it has a problem: the matrix $J^T(\mathbf{z})J(\mathbf{z})$ may not have a inverse matrix. To solve this problem, Levenberg-Marquardt proposed the following approximation to the Hessian matrix:

$$\Delta^2 V(\mathbf{z}) = J^T(\mathbf{z}) \cdot J(\mathbf{z}) + \mu \cdot \mathbf{I} \tag{3.34}$$

Where μ is called combination coefficient and I is the identity matrix. Therefore, to LM method, equation (3.33) turns into:

$$\Delta \mathbf{z} = -\left(J^T(\mathbf{z}) \cdot J(\mathbf{z}) + \mu \cdot \mathbf{I}\right)^{-1} \cdot J^T(\mathbf{z}) \cdot \mathbf{e}(\mathbf{z}) \tag{3.35}$$

The setting of the μ parameter is what defines the direction followed by the method. If after an iteration, the error of the current output is greater than that obtained previously, the parameter μ is multiplied by a factor β. Thus, the LM method approaches to the steepest descent method. However, if the error decreases between two iterations, the value of μ must be divided by the factor β. With this decrease of the parameter μ, the method approaches the Gauss-Newton method. It is desired that the LM method turns into the Gauss-Newton method as soon as possible, since this method has a faster convergence [33, 32]. In literature, the value of β is usually defined as 10.

To the training of the neural network, $\mathbf{z}$ will be the vector with all the weights of the network:

$$\mathbf{z} = [w^1_{1,1}; w^1_{1,2}; ...; w^2_{1,1}; ...; w^M_{nM,n(M-1)}]^T$$

In this vector, nM is the number of neurons in the last layer and $n(M-1)$ is the number of neurons in the previous layer. Also, each line of the Jacobian matrix will correspond to one pattern of the training set. Consequently, the Jacobian matrix will have P lines and n columns, where n is the number of weights in the network. Thus, the matrix used for training will have the following structure:

$$J(\mathbf{z}) = \begin{bmatrix} \frac{\partial E(1)}{\partial w^1_{1,1}} & \frac{\partial E(1)}{\partial w^1_{1,2}} & \cdots & \frac{\partial E(1)}{\partial w^M_{nM,n(M-1)}} \\ \frac{\partial E(2)}{\partial w^1_{1,1}} & \frac{\partial E(2)}{\partial w^1_{1,2}} & \cdots & \frac{\partial E(2)}{\partial w^M_{nM,n(M-1)}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial E(P)}{\partial w^1_{1,1}} & \frac{\partial E(P)}{\partial w^1_{1,2}} & \cdots & \frac{\partial E(P)}{\partial w^M_{nM,n(M-1)}} \end{bmatrix} \tag{3.36}$$

The LM algorithm can be implemented by making some changes in BP, so that the Jacobian matrix is constructed during the algorithm execution.

Chapter 4

Application of the method

This chapter will present the methods used to solve the problem introduced in section 1.2. Therefore, the implementation of the artificial neural networks techniques will be analyzed, as well as its training algorithms and its input-output parameters characteristics. The theory behind these topics were covered in section 3.3, and this section will only deal with implementation aspects.

It is important to emphasize, that all the implementation of the Artificial Neural Networks and the training algorithms used in this work were built in script code in Matlab software. Despite being possible to use Matlab's toolbox for this purpose, **it was preferred to implement all the routines in Matlab codes**, to have better control In addition, it is expected to adapt the script code to develop a initial version of the software implemented in a program with graphical unit interface, such as Python.

4.1 Description of the photovoltaic system used to obtain the measurements

Figure 4.1: Photovoltaic system installed on the roof of the Department of Electrical and Computing Engineering

The generated power's data was obtained from the photovoltaic system that is installed on the roof of the Department of Electrical and Computing Engineering of São Paulo University - São Carlos Campus. This system is shown in Figure 4.1. This system has a total power of 3.1 kWp and is composed of 12 PV modules of 260 W from Canadian Solar with polycrystalline silicon cells, an Fronius Primo 3.0-1 inverter and a data acquisition system. The data is stored by a control center based on a computer. The acquisition system and the inverter of the photovoltaic installation are shown in Figure 4.2. The multifilar diagram of the PV system can be found in Appendix A.

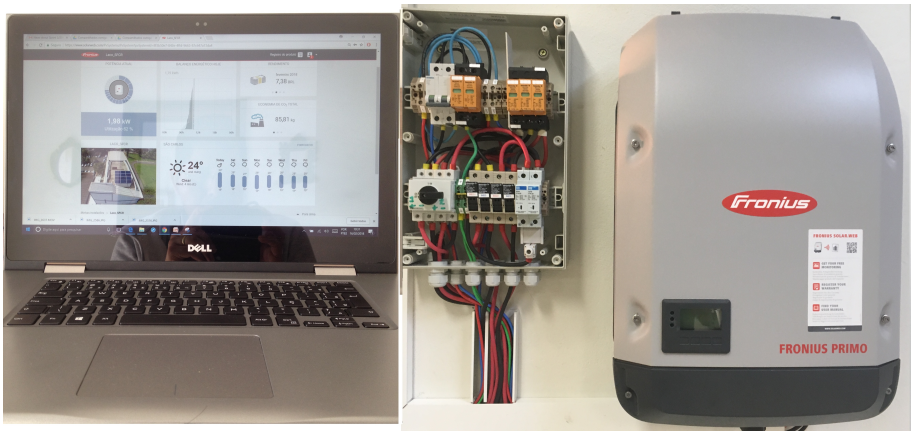


Figure 4.2: Inverter and the acquisition system installed on the Department of Electrical and Computing Engineering

The data sheet information about the panel and the inverter are shown in Table 4.1 and Table 4.2, respectively.

Table 4.1: PV module data sheet

Canadian Solar CS6P-260	
Nominal Max. Power	260 W
Open Circuit Voltage (Voc)	37.5 V
Short Circuit Current (Isc)	9.12 A
Module efficiency	16.16 %

Table 4.2: Inverter data sheet

Fronius Primo $3.0-1$	
Nominal Power	3kW
N^{o} of MPPT	2
Efficiency	98%
Maximum Power of the Photovoltaic generator	4.5kWp

4.1.1 Selection of the inputs and outputs of the model

As mentioned in section 3.3, to train and validate the ANN a set of data is required. For networks with supervised learning, as the MLP, it must contain the inputs of the network and the respective desired output. Since this work aims at analyzing the prediction of power generated by a photovoltaic system, the desired output is the generated power itself and the inputs are chosen based on variables related to this generation. As shown in section 2.1, the climatological variables have strong influence over solar generation. Therefore, those were chosen to form the input vector of the network.

4.1.2 Desired output vector

In this study, the information required is the one obtained by the data acquisition system. This system stores the following info about the panels:

- DC current from MPPT1
- DC current from MPPT2
- DC voltage from MPPT1

- DC voltage from MPPT2
- AC voltage
- AC current

Where, the DC current is the current that leaves the PV panels' terminals. The photovoltaic system is split in two circuits that are connected in two MPPts, as shown in Appendix A. The DC voltage is measured for each MPPT. This voltage is the one on the terminals of each panel set. The AC parameters are not measured in the panels, but in the inverter's output, therefore, these parameters include the inverter's losses. The inverter is the equipment that is directly connected to the grid.

To compute the generated power from the information mentioned above, it is first necessary to choose which power will be used in the study. That's because two types of power can be obtained from the measured data. The first one is generated from the PV modules. This power is calculated using the DC parameters. The second one is the power that the whole PV system (photovoltaic panels and inverter) delivers to the distribution network. This power is related to the AC variables. Since this study wants to forecast only the photovoltaic generation, that is, the work does not take into account any inverter losses information, the first type of power is chosen.

Thus, the PV power (P_{PV}) is used as desired output to the ANN and it is calculated using the following expressions:

$$P_{MPPT1} = I_{DC,MPPT1} \cdot V_{DC,MPPT1} \tag{4.1}$$

$$P_{MPPT2} = I_{DC,MPPT2} \cdot V_{DC,MPPT2} \tag{4.2}$$

$$P_{PV} = P_{MPPT1} + P_{MPPT2} \tag{4.3}$$

Where P_{MPPT1} is the generated power produced by the panels that are connected to MPPT1 (circuit one) , P_{MPPT2} refers to the production of the panels connected to MPPT2 (circuit two). Finally, P_{PV} is the total power generated by all the panels of the system.

To build the desired output vector, a set of measures corresponding to the months of May and April were used. The acquisition system collects information in a five minute interval, therefore, for a full day 288 samples are stored. Since the PV system only works during the day, the number of samples was reduced to 156. This new set covers only the period between 6 o'clock in the morning and 18 o'clock in the afternoon.

Furthermore, since the forecast window size chosen in this study was one hour, the data had to be organized, so that its interval be one hour and not five minutes. For this purpose, the arithmetic mean of the power values during each hour was used. Consequently, the number of samples was reduced to 13 per day, resulting in 403 samples in one month with 31 days, and 390 for a month with 30 days.

Since May and April months were considered, the total number of samples was 793. However, during the month of April, one day had its measures affected due to failures in the distribution network, which resulted in a lack of energy and consequently in the shutdown of the photovoltaic system. Thus, one day of the month of April was excluded and the total number of samples decreased to 780.

4.1.3 Input vector

The input vector is composed from climatological variables. Those were obtained from the Brazilian National Institute of Meteorology [34] (or Instituto Nacional de Meteorologia (INMET), in Portuguese). This institute provides a lot of meteorological variables, such as: atmospheric pressure, temperature, relative humidity, precipitation, solar radiation, direction and wind speed, among others.

To choose the variables for this research, a linear correlation analysis was carried out (similar as accomplished in [24]). The relation between the generated power and each meteorological variable was studied. The correlation coefficient for two variables quantifies the dependency between them. It is mathematically given as:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \hat{x})(y_i - \hat{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \hat{x})^2 \sum_{i=1}^{n}(y_i - \hat{y})^2}} \tag{4.4}$$

Where x_i and y_i are the i_{th} elements of the vectors X and Y, respectively. These are the vectors with the values of the variables. And the parameters $\hat{x}$ and $\hat{y}$ represent the arithmetic mean of the elements in these vectors.

A correlation coefficient near to 1 or -1 characterizes a high dependence. But as this coefficient approaches to zero, the dependency starts to decrease. A null value defines that two variables have no relation to each other. The results are shown in Table 4.3, where only the five highest correlation coefficients were considered.

Table 4.3: Correlation analysis between the climatic variables and the generated power

Variable	**Correlation Coefficient**
Temperature	0.476
Solar Irradiance	0.802
Humidity	-0.425
Wind speed	0.360
Dew point	0.190

One may inferred from the table that the climatic variables with the highest correlation with the solar generation are: solar irradiance (G), temperature (T) and humidity (H). The first two parameters have a positive relation with the photovoltaic power, that means that, as their value increases, the photovoltaic power also increases. In the other hand, the humidity has a negative relation with the PV power. This probably occurs because the greater the humidity, the greater the probability of there being clouds in the sky and, consequently, of shading on the photovoltaic system.

Considering the correlation analysis, these three variables will compose the network input vector. Also, it is important to notice that the parameters were obtained for the months of May and April (corresponding to the interval between $01/03/2018$ and $30/04/2018$, excluding day $21/04/2018$, due to the power outage), with 780 samples for each variable.

In Figure 4.3 is shown the summing of the relations between the input, the output and the desired output of the model for this study.

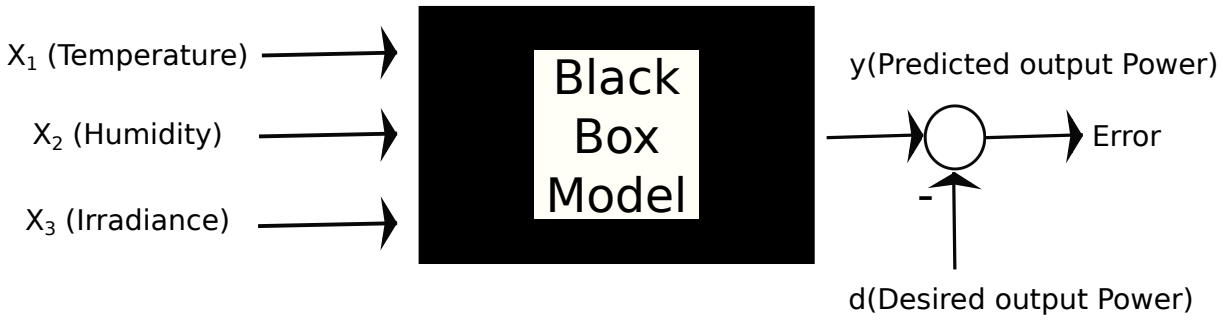


Figure 4.3: Black Box Model

4.1.4 Training and validation set

The data of input and desired output variables were divided into training and validation set.

In this study, the division of the total data set was made considering 70% for the training and 30% for the validation, according to random subsampling cross-validation method, explained in section 3.3. Therefore, the training and the validation set contain 546 and 234 patterns, respectively.

The format of the input-output set is shown in Table 4.4. This set was randomly divided following the considerations made above. For each training-validation process, the division was remade, thus different patterns could be presented to the net. In addition, during the training of the network, the order of the elements in the set was randomly reorganized after each epoch.

Table 4.4: Input and desired output set

X_1 **(Temperature)**	X_2 **(Humidity)**	X_3 **(Irradiance)**	**d (Real Photovoltaic Power)**
19.5	93	0	91.554
21.1	92	163	762.818
23.3	80	1132	1033.559
⋮	⋮	⋮	⋮
22.9	51	50.72	0

4.2 Artificial neural network structure

Three different approaches of ANN were used in this study as a comparison proposed to have better results. All these employed multilayer perceptron networks, but each one considered a distinct structure for the network, as described below:

- **Structure A**: Feedforward multilayer perceptron network, with no feedback connections, performing a non-linear input-output mapping of the problem, using inputs x(t) to obtain outputs y(t).
- **Structure B**: Feedforward multilayer perceptron network, with no feedback connections, performing a non-linear input-output mapping of the problem and also working as a predictor. The output y(t) will be obtained from the input vector $\mathbf{x} = [x^1, x^1(t-1), x^1(t-2), ..., x^1(t-np), x^2, ..., x^k, x^k(t-1), x^k(t-2), ..., x^k(t-np)]$, where x^k is the k_{th} input variable related to the problem and np is the shift degree of x in time.
- **Structure C**: Focused time-lagged feedforward multilayer perceptron network. In this approach, the only variable that will be consider is the generated power. Therefore the measured past values of this parameter will be used to predict its future values.

To implement these network, the input-output set showed in subsection 4.1.4 was used. Thus, for each epoch

of training, 780 input-output patterns were presented to the network. The difference between each structure was the way that this set was introduced to the network.

The index used for comparison is the final error of the network. But other aspects were also taken into account, such as the origin of the network data and the possibility of its practical application. Some considerations are common for all proposed network structures, as the parameters to train the structures, for example. These parameters are:

- Maximum number of epochs (N_{epochs});
- Tolerance (Tol);
- Combination coefficient (μ);
- Parameter β;
- Learning rate (η);
- Activation function used in the hidden layers;
- Maximum number of hidden layers ($Hlay_{\text{max}} = 2$)
- Minimum number of hidden layers ($Hlay_{\text{min}} = 1$)
- Maximum number of neurons ($Neuron_{\text{max}} = 20$)
- Minimum number of neurons ($Neuron_{\text{min}}$ - not defined)

Those parameters are related to the learning algorithm used to perform the training. Before training the network, these variables must be defined. Furthermore, the network input-output data has to be normalized in order to avoid the saturation of the activation function.

To define the network's topology, the cross validation method was used. This method was described in section 3.3 and consists in applying several topologies to the same problem and evaluate the net's response to the training and the test set. For each test-training process the input-output sets was randomly divided, but always keeping the ratio of 30% to 70%. For each structure, twenty topologies were evaluated and each topology was trained and tested five times.

The topologies analyzed were randomly selected. However, even for the random selection, some parameter were defined, as: the maximum number of hidden layers (two), and the maximum number of neurons in each layer

(twenty). Those parameters were important and were based on the literature, because, a large number of neurons or hidden layers can lead the network to a over fitting situation. For this problem, two hidden layers are enough to obtain a satisfactory generalization. Besides that, the number of patterns in the data set is not very large, and if the network had a lot of neurons, it could become badly conditioned.

The minimum number of layers is one, because if the MLP network has no hidden layer it becomes a Single Layer Perceptron network, and these nets cannot solve problems whose variables are not linearly separable. A minimum number of neurons was not defined, but it is know that a network with few neurons has problems in generalizing.

In the cross validation method, the topology that shows the best behavior is chosen. This behavior is analyzed in two steps. First, each topology had its overall performance obtained from the best performance in the submitted five test-training processes. Thus, each topology was trained and tested five times. Because the best performance is given when the network presents the smallest error, the next relation was used:

$$\text{Topology Performance} = \min[0.7 * \text{training error} + 0.3 * \text{test error}] \tag{4.5}$$

Where the index "Topology Performance" in Equation (4.5) was used to determine the topology's best behavior during the five training-testing processes. The training error and the test error were weighted by the percentage of the used samples number.

After defining the global performance for each one of the twenty topologies. They will be analyzed in relation to three characteristics: The number of neurons, the test error and the training error. This analysis is important because, as mentioned in section 3.3, it is necessary to find the balance between the errors values and the number of neurons to avoid the overfitting and the underfitting situation. Therefore, the topology chosen will be the one that has the maximum number of neurons before overfitting, that is, before the test error starts to grow.

The characteristics of each structure are discussed below.

4.2.1 Structure A

In this structure, the inputs of the network will be the temperature (T(t)), the irradiance (G(t)) and the humidity (H(t)). This parameters will be presented to the network, so that the forecast of the power output of the photovoltaic system can be done.

It is important to notice that the input variables of the instant "t" were used to forecast the output variable in

instant ”t”. This is possible because, for this structure, the inputs of the network are results of another forecast method. Companies that deal with weather prediction use methods similar to those presented in chapter 3 to accomplish this task. Thus, it is possible to predict values of temperature, irradiance, and humidity for a specific instant ”t” and use these values to obtain the predicted power output for instant ”t”. A generic model of this arrangement is shown in Figure 4.4. Changing the number of hidden layers and neurons it is possible to obtain different topologies.

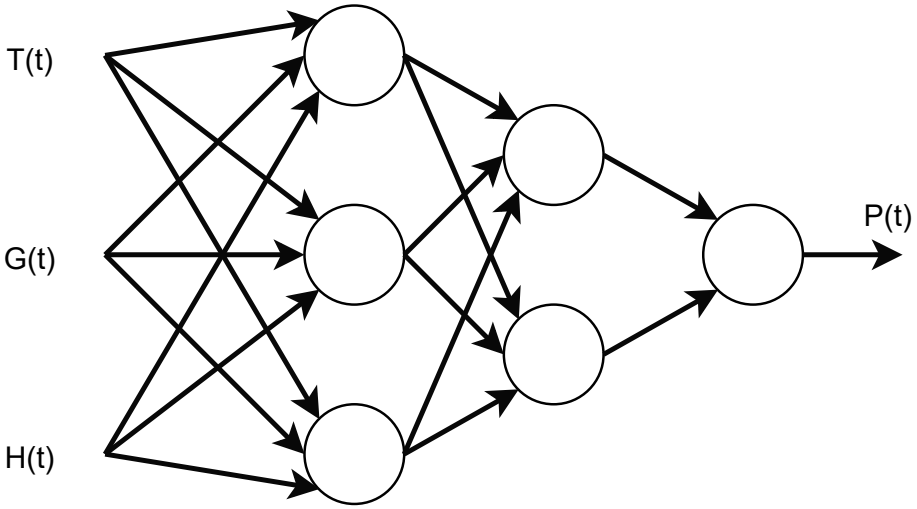


Figure 4.4: Generic model of structure A

4.2.2 Structure B

In this structure, the inputs of the network will also be temperature (T), irradiance (G) and humidity (H). But, differently of structure A, those variables will be considered not only on the instant $(t-1)$, $(t-2)$, until $(t-np)$, where np is the shift of the input parameters in time. In this configuration, the output will be the PV power (P) for the instant ”t”. This means that the temporal window considered for the forecast is one time unit. Since in this study the measures are obtained in a one hour basis, one time unit corresponds to one hour. In other words, the power can be predicted one hour ahead of the actual power.

This approach, was made to evaluate if there is an improvement in the accuracy of the forecast when using more information about the input variables. Since the network acts as an approximation of functions that tries to abstract from the data the type of function that governs the relation between the input and the output, the more information about the input it has, the better this abstraction can be performed.

Therefore, the input vector and the output vector of the network will have the following aspect:

$$\mathbf{x} = [T(t-1), ..., T(t-np), G(t-1), ..., G(t-np), H(t-1), ..., H(t-np)] \tag{4.6}$$

$$\mathbf{y} = P(t) \tag{4.7}$$

Thus, during the training process, the set input and the desired output set will be organized in the following way:

Table 4.5: Input and desired output set for structure B

X_1	...	X_{1+np}	X_2	...	X_{2+np}	X_3	...	X_{3+np}	**d**
$T(t-1)$	⋮	$T(t-np)$	$G(t-1)$	⋮	$G(t-np)$	$H(t-1)$	⋮	$H(t-np)$	P(t)
$T(t-2)$	⋮	$T(t-np+1)$	$G(t-2)$	⋮	$G(t-np+1)$	$H(t-2)$	⋮	$H(t-np+1)$	$P(t+1)$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

The generic model of this approach is presented in Figure 4.5:

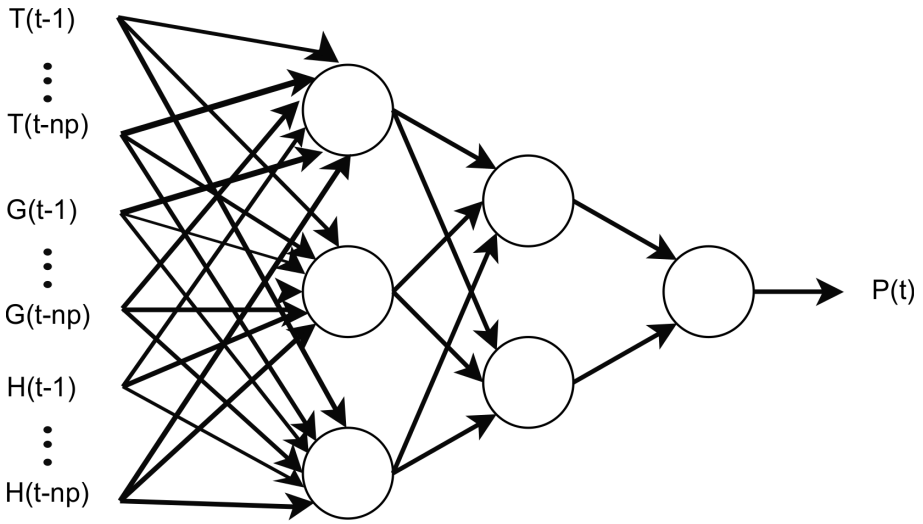


Figure 4.5: Generic model of structure B

In a practical application, this network could be implemented when a measurement system of weather data were available. In this situation, the real weather data of instant $(t-1)$ and the np previous instants, could be used to predict the photovoltaic power of instant "t". After the forecast, the network result might be compared with the measured power for instant "t", to verify the forecast quality.

4.2.3 Structure C

The last MLP proposed is the structure C. This network, differently the other two, only uses power values measured on the photovoltaic system. Thus, the input vector of the network is composed by the past values of the measured power, and they are used to predict the power on instant t. The generic model of this structure is presented in Figure 4.6:

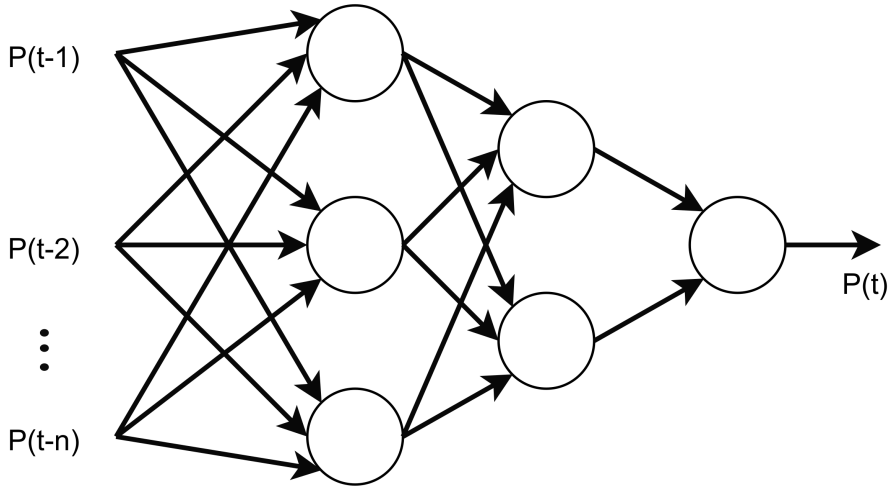


Figure 4.6: Generic model of structure C

It is possible to notice that the input and the output of the network are, respectively, given by:

$$\mathbf{x} = [P(t-1), P(t-2), ..., P(t-np)] \tag{4.8}$$

$$\mathbf{y} = P(t) \tag{4.9}$$

Some considerations can be made about this model. First of all, the parameter np is the shift of the input variables in time. This parameter is defined during the execution of the training algorithm. In this study, the value of np was randomly selected for each training. During the training process, the set will be organized in the following way:

Table 4.6: Input and desired output set for structure C

X_1	X_2	X_3	...	X_{1+np}	**d**
$P(t-1)$	$P(t-2)$	$P(t-3)$	...	$P(t-np)$	$P(t)$
$P(t)$	$P(t-1)$	$P(t-2)$	...	$P(t-np+1)$	$P(t+1)$
⋮	⋮	⋮	⋮	⋮	⋮

It is important to highlight that the whole network training must be done using a real data set, and not data that is

result from another forecast method. Therefore, in a practical situation, this network could be easily implemented, because of the availability of the measures of power to train the network with supervised learning. With the acquisition system, it is possible to acquire new values of power and use those values as inputs to the network to predict the power output.

4.3 Training algorithms

In subsection 3.3.3 the backpropagation and the Levenberg-Marquardt algorithms were presented as methods to train a multilayer perceptron network. In order to choose the best method to accomplish this task, those methods were tested training the same network. This procedure was done in order to chose the best learning method to train the networks for forecasting the photovoltaic power.

They were built based on the deduction presented in subsection 3.3.3. The ***number of interactions required until convergence*** and the ***final error of the training*** were chosen as indicators of performance.

For this analysis the MLP was the one presented in subsection 4.2.1. Both algorithms were tested by training 5 distinct topologies of the network, the performance indicators were analyzed for each case and the decision of the best method was made based on its global performance. The best one was used to train the ANNs that were presented in the previous section.

As explained in section 3.3, to conduct the training it is necessary to define certain network parameters, such as maximum number of epochs, tolerance, learning rate (for the BP algorithm), the combination coefficient and β (for the LM algorithm). In addition, it is necessary to repeat the execution of the training a few times for the same topology in order to avoid convergence of the algorithm to some local minimum in the error function surface. In this analysis, the chosen parameters are:

- N_{epochs}: 100
- Tol: $1e10^{-6}$
- η: 0.1
- μ: 0.1
- β : 10
- Activation function used in the hidden layers: Sigmoid function $\left(\text{given by: f(x)} = \dfrac{1}{1+e^{-\lambda x}}\right)$

The values of η, μ and β were selected based on the scientific review. The topologies tested were randomly chosen. The selected ones were: [3], [4 2], [5 2], [8 6], [15 10], where each column of the vector represents a hidden layer and the value indicates the number of neurons.

For each topology both of the algorithms were executed five times and the results are shown in the following chapter.

Chapter 5

Results and Discussion

5.1 Training algorithms performance evaluation

As described in section 4.3, the LM and the BP algorithms had their performance evaluated in order to decide which one was the best to train the networks. In this section, the results of this evaluation are shown. The chosen topologies used were defined in the previous section. Also, it is important to remember that the maximum number of epochs was set as 100. In Table 5.1 the results for each topology are presented.

Table 5.1: Results of the MLP training made with the LM and the BP algorithm

Topology		**Number of iterations**		**Final Error**		**Training Error**		**Test Error**	
First Hidden Layer	Second Hidden Layer	LM	BP	LM	BP	LM	BP	LM	BP
3	0	3	3	0.0088	0.0099	0.0091	0.0106	0.0082	0.0082
4	2	4	12	0.0081	0.0098	0.0083	0.0108	0.0076	0.0076
5	2	22	12	0.0085	0.0104	0.0080	0.0107	0.0098	0.0098
8	6	3	11	0.0083	0.0121	0.0059	0.0113	0.0139	0.0139
15	10	4	2	0.0118	0.0167	0.0050	0.0119	0.0278	0.0278

As can be inferred from the table, the topology that presented the smallest final error was the one with 4 neurons in the first hidden layer and 2 neurons in the second, and that was trained with the LM algorithm. For the

best topology, the evolution of the error during the execution for both algorithms can be seen in Figure 5.1.

Moreover, it is notable that for all topologies tested the Levenberg-Marquardt algorithm achieved a final error lower than the one found by the backpropagation. **For these reasons, the LM is the selected algorithm to train the networks in the following section.**

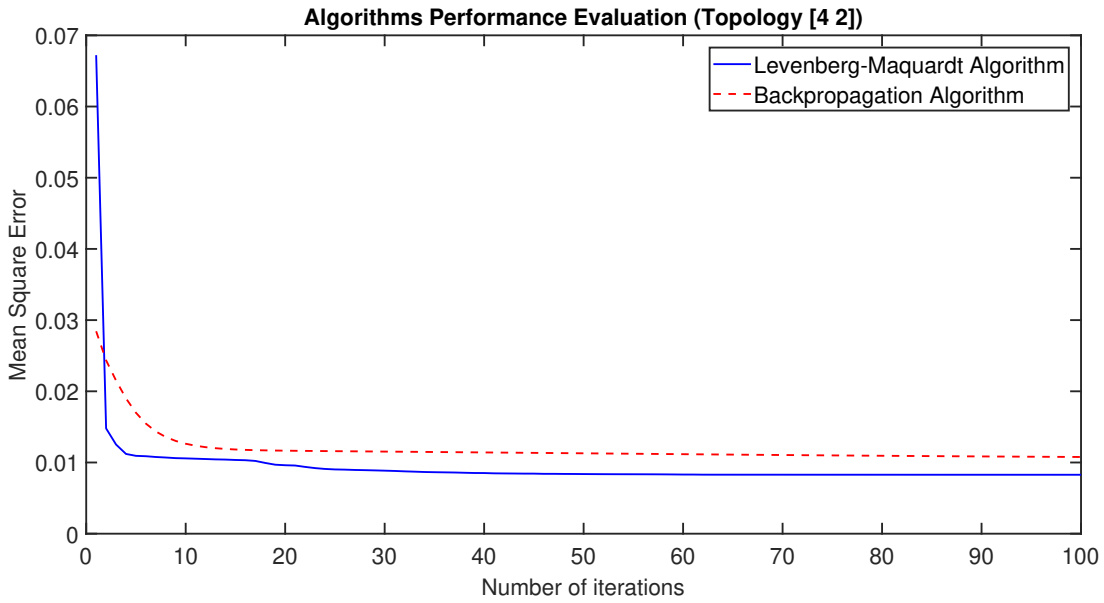


Figure 5.1: Comparison of the evolution of the errors of the BP and LM algorithms for the best topology ([4 2])

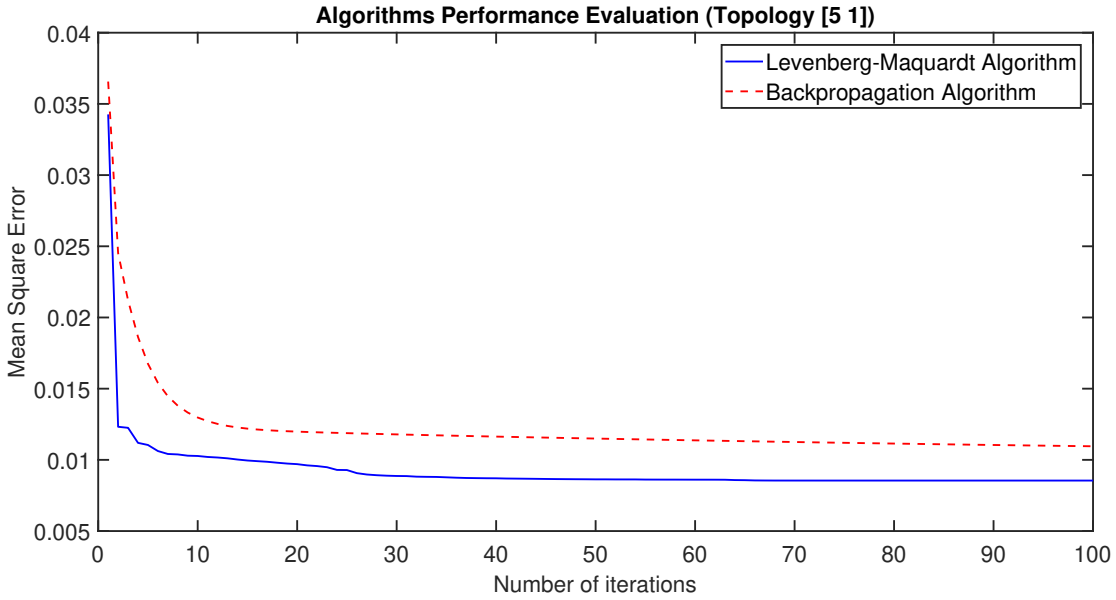


Figure 5.2: Comparison of the evolution of the errors of the BP and LM algorithms for topology [5 2]

With this defined, some considerations can be made regarding the behavior of the algorithms and the training-testing process of the MLP network. First of all, it is clear that the LM method takes less iterations to converge to smaller errors than those found by BP method. And, most of the time, the LM reaches the global minimum faster than the BP reaches a local minimum. Even for topology [5 2], where it takes 22 iterations to the LM to find the global minimum, its error is already lower than the BP's before the fifth iteration, as shown in Figure 5.2.

In literature is stated that more the number of neurons, more the network can generalize and therefore, lower is the training error. This statement is true as long as the network overfitting does not occur, which is noticed when the test error starts to increase even with the training error decreasing. Thus it is important to know how much the network can be increased, so that it improves its level of generalization without getting into an overfitting situation.

The analysis of this network growth limit can be made based on the LM algorithm's training and testing data, shown in Table 5.1. The limit size can be found after creating a graph with the error values of the test and the training as a function of the total value of neurons in the topology as shown in Figure 5.3 (In this graph the total number of neurons is the sum of neurons in all hidden layers).

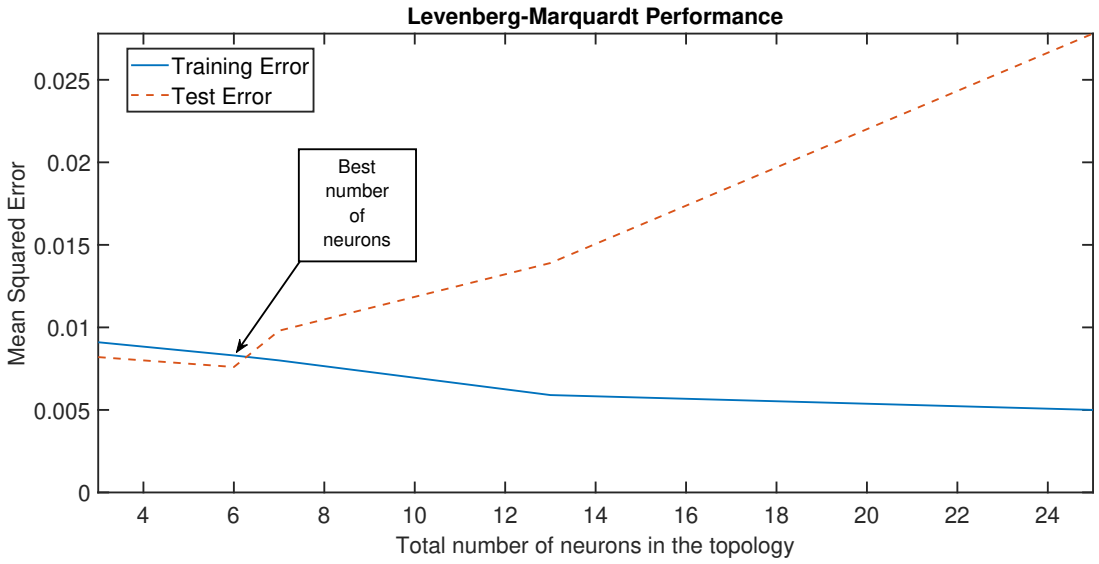


Figure 5.3: Comparison of the evolution of the errors of the LM algorithm

Based on Figure 5.3, the best topology must have a total number os 6 neurons. The topology that matches this condition is [4 2], that has already been chosen for having the least final error. This analysis shows that the previous statement about the number of neurons in a network is true.

However, according to the data in Table 5.1, the BP algorithm does not behave this way, because its training error tends to increase as the number of neurons increases. This happens because for the performance evaluation,

the maximum number of epochs was defined as 100, and the BP algorithm needs much more epochs to converge to a global minimum. Therefore, when the observation time is restricted to 100 epochs, the training error had not yet time to reach its final value.

In Figure 5.4, the BP algorithm was evaluated again for topology [4 2], but now with the maximum number of epochs defined as 10000. It is possible to notice that the algorithm gets stuck in a local minimum for several decades until it reaches a better value around the thousandth iteration.

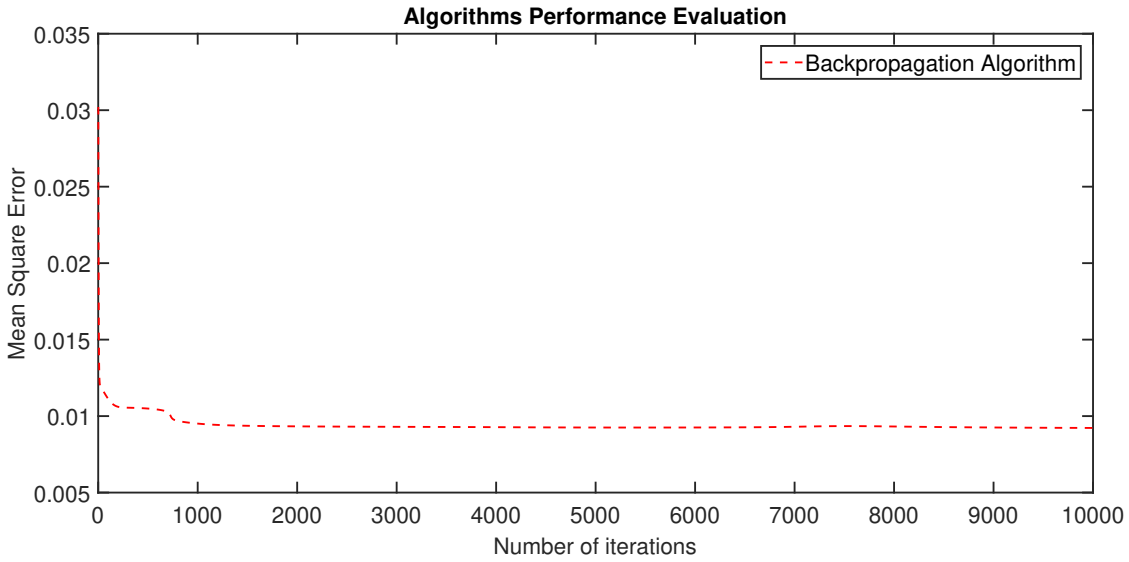


Figure 5.4: Evolution of the errors of the BP algorithm for topology [4 2] for high number of epochs

For this case, the minimum error found by the algorithm was 0.0090, but it is not even the global minimum, because the LM founds a error of 0.0084 with less than fifty iterations. Hence, the BP algorithm would have to be executed for more than ten thousand iterations to be able to find the global minimum or even it might not find it, as discussed in section 4.3.

An example of overfitting phenomenon happens for topology [15 10]. The training and the testing of this topology are presented respectively in Figure 5.5 and Figure 5.6. The overfitting of a neural network, explained in section 3.3, occurs mainly due to an unnecessarily high number of neurons. This phenomenon can be verified when the network presents a low training error, but a very high test error. In Figure 5.5 and Figure 5.6, it can be observed that the topology behaves very well for the training part, presenting minimal error. However, during the test, the error becomes very noticeable. Just looking at the training, the topology may seem a good alternative for the proposed application. However, using the test data (practical situation) to new input-output values, the network shows a very poor performance.

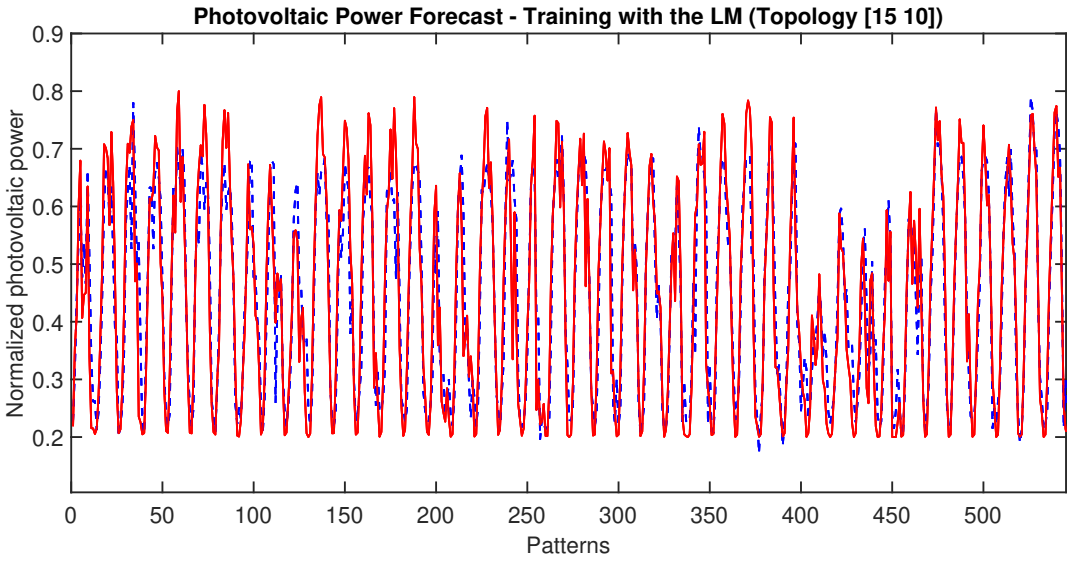


Figure 5.5: Overfitting situation for topology [15 10] - training set (Red line - Measured values; Blue line - Forecasted values)

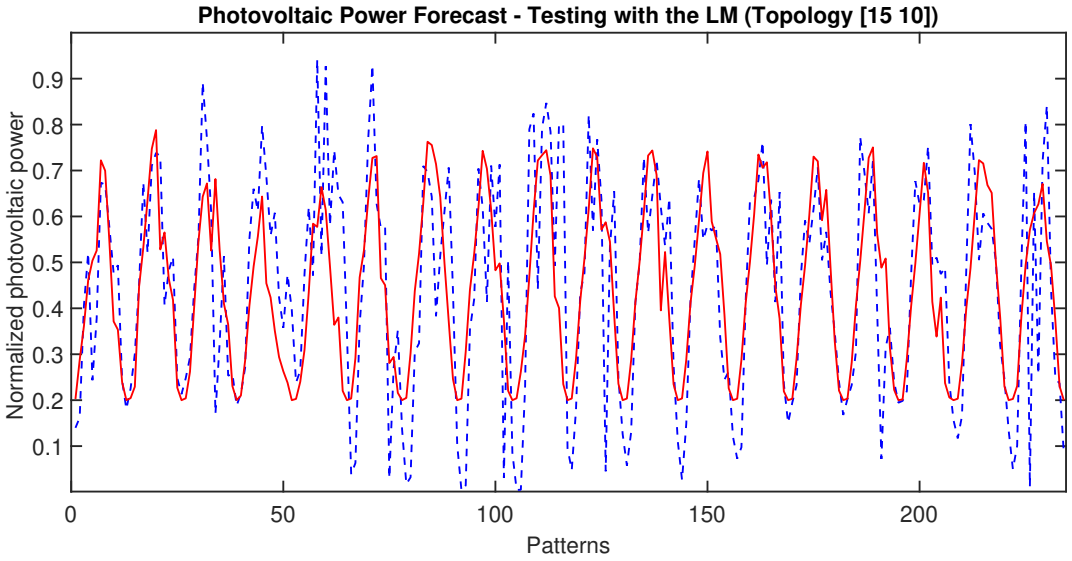


Figure 5.6: Overfitting situation for topology [15 10] - test set (Red line - Measured values; Blue line - Forecasted values)

5.2 ANN performance evaluation

This section shows the results obtained for the photovoltaic power forecast using the ANN. These results were achieved using the methodology described in section 4.2. All the structures were tested using the LM algorithm. The parameters that were defined for these structures are:

- N_{epochs}: 100
- Tol: $1e10^{-6}$
- μ: 0.1
- β : 10
- Activation function used in the hidden layers: Sigmoid function

For comparisons, the forecast data will be related to the same day for all the ANN structures. That day was randomly selected to maintain the impartiality of the study. Hence, for all cases analyzed, the real values of the photovoltaic power will be the same, and only the predicted values will be specific for each approach.

The chosen day for the analysis was 07/03/2018, in average values, it had an temperature of 27.7 °C, an irradiance of 2231 kJ/m^2 and an humidity of 65 %. Besides that, the average solar power generated in this day was 1653.809 W. The behavior of these variables during the day is shown in Table 5.2.

Table 5.2: Value of the variables during the analyzed day (07/03/2018)

Pattern	**Hour**	**Temperature** [°C]	**Humidity** [%]	**Irradiance** [kJ/m^2]	**Photovoltaic Generation** [W]
79	06:00	19.9	93	0	21.017
80	07:00	20.1	94	43.98	474.202
81	08:00	22.8	83	627.1	753.109
82	09:00	23.3	81	1337	1803.423
83	10:00	25.4	73	2233	2513.753
84	11:00	26.7	69	2873	2775.190
85	12:00	28	65	3283	2453.915
86	13:00	28	65	2359	2751.354
87	14:00	30.2	60	3467	2152.878
88	15:00	29.5	55	2645	1653.809
89	16:00	30.1	54	2231	1322.485
90	17:00	29.6	55	1549	518.222
91	18:00	27.7	59	661.7	11.630

An important aspect about the data should be emphasized, it is noted in the first line of the table that there is a low value of power generated for the irradiance equal to zero. This relationship occurs because the generated power

value considered is the average of the values generated during the interval between 06h00 and 06h55, however, the irradiance value used refers only to the instant of 06h00. Thus, as it was not possible to obtain more detailed data on the irradiance, the instantaneous value was used. This explains the relation presented in the table, therefore, it is inferred that the irradiance for other instants in the considered interval is not zero.

In the first column of Table 5.2 is shown the number of the pattern in the total set related to that day. This number may be different depending of the ANN structure used for the forecast. In other words, it will still be the same day, but it might be shifted in relation to the patterns of the set. That happens for structure B and C, because in these, a shift factor (np) is introduced to rearrange the input-output patterns. Therefore, depending on the value of nd, the number of the pattern (that compose the x axis of the following graphs), that refers to this day, will change.

The results of the forecast for each structure will be presented in the following sections.

5.2.1 Results of power forecast for structure A

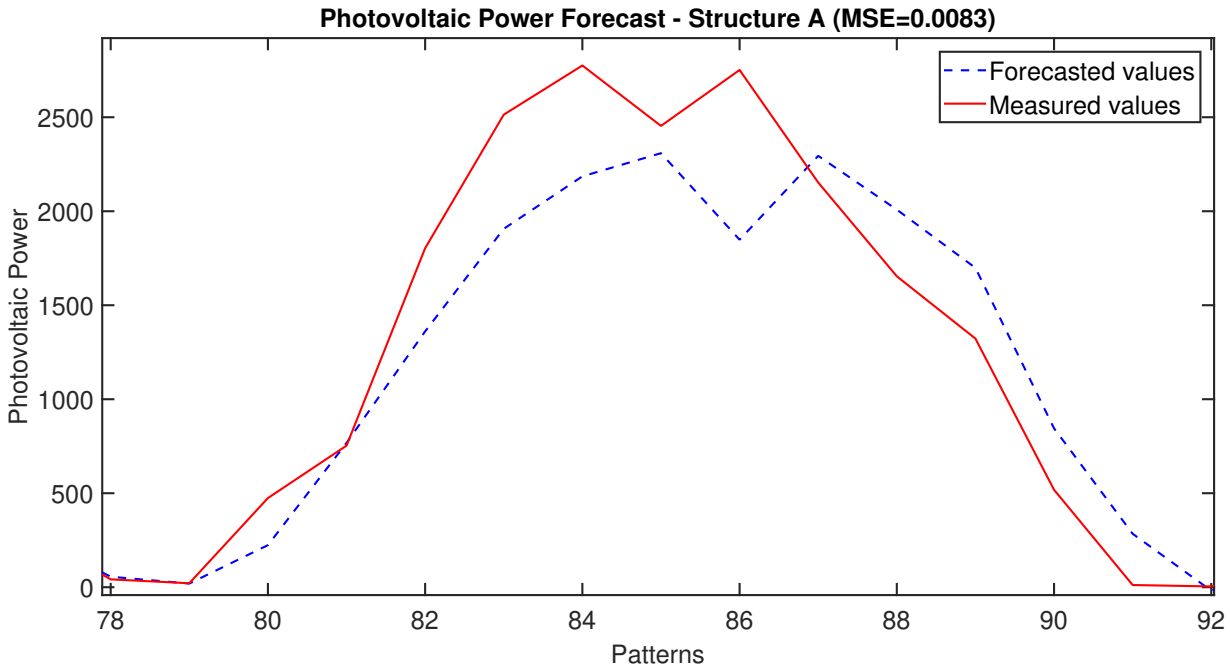


Figure 5.7: Power forecast for structure A

For this structure, the best topology was the one with 3 neurons in the first hidden layer and 6 neurons in the last hidden layer. For this topology the final mean square error was 0.0083. This final error was obtained in relation to the training and test errors, whose values are 0.0089 and 0.0071, respectively.

The forecast power can be compared with the measured power in Figure 5.7. Figure 5.8 shows the change of the error value during the training process. As verified in section 5.1, the LM algorithm presents fast convergence to the final error value. However, it is not possible to affirm if it is a global or local minimum, since it depends on other factors, such as the initial guess of the weights, for example.

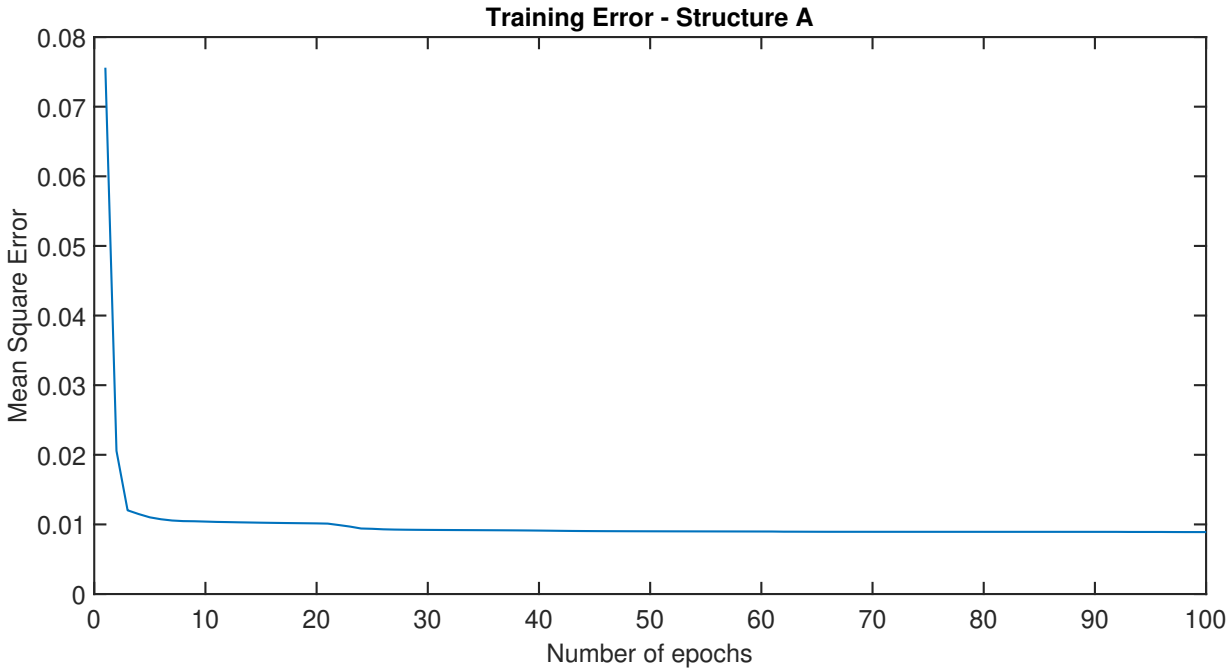


Figure 5.8: Training error evolution for structure A

According to the results, the network demonstrate good performance to carry out the proposed prediction task. However, some of the methods used in weather prediction are not so accurate and, some of the errors generated by them can be propagated trough the process and reflect in the predicted photovoltaic power. This problem restricts the use of this structure in practical applications.

5.2.2 Results of power forecast for structure B

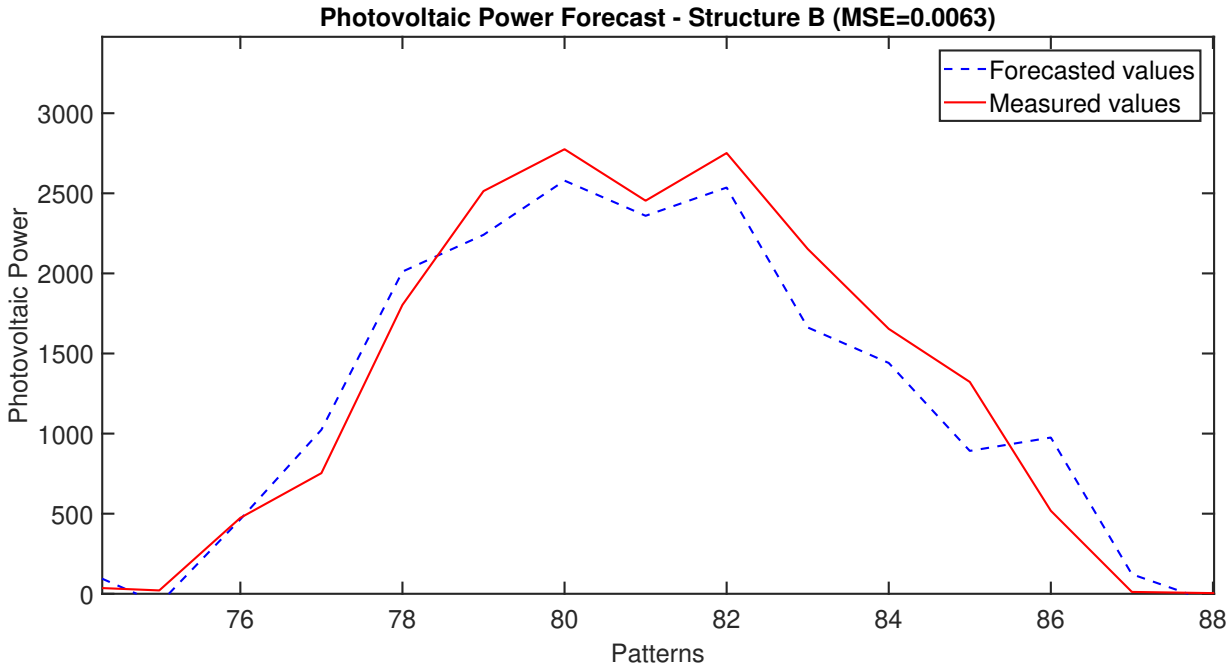


Figure 5.9: Power forecast for structure B

For this approach, the best topology was the one with just one hidden layer with 7 neurons. The final mean square error for this topology was 0.0063, while the training error was 0.0058 and the test error was 0.0075. Unlike structure A, in this approach the network entries were shifted in time by a factor np. For this topology the np selected was 4.

In Figure 5.9, the forecast power can be compared with the measured power. The power values shown in the figure are also relative to 07/03/2018, but the patterns number have changed because of the value of np. The evolution of the error value during the training process can be observed in Figure 5.10. For this structure, the convergence occurs around the tenth pattern.

This structure showed better performance in the prediction than structure A. This can be verified in the smaller value of final error and can also be seen graphically, since the curves from the forecasted values and from the actual values in structure B are closer than in the structure A.

The advantage of this approach is that all the data presented to the network can be obtained from real measures, that is, it may not be a result from another forecast method. In this study, the values of the inputs were obtained from

a external agent, however, they could be obtained from a metering system installed together with the photovoltaic system. Therefore, the weather data could be measured at the same location where the power is being generated, ensuring a better level of accuracy for the forecast. In a practical situation, this would be a good alternative to obtain good forecasting results.

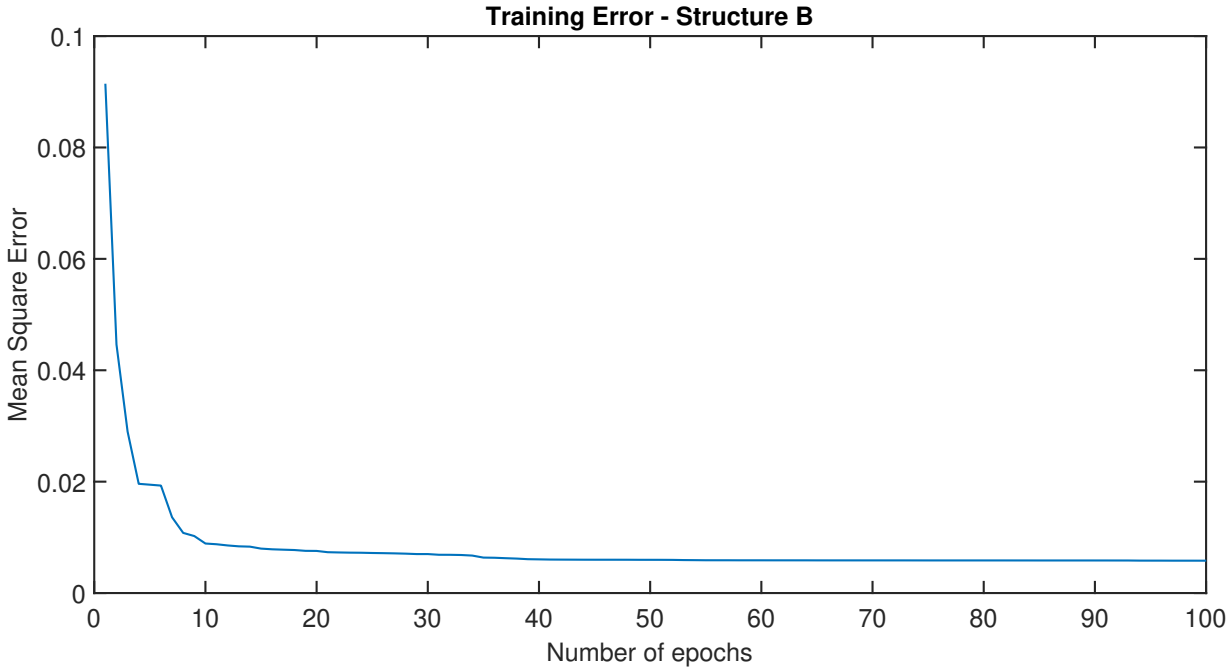


Figure 5.10: Training error evolution for structure B

5.2.3 Results of power forecast for structure C

For this structure, the best topology was the one with 4 neurons in the first hidden layer and 8 neurons in the last hidden layer. The final mean square error was 0.0046, and the training and test errors were 0.0034 and 0.0073, respectively. For this structure the network entries were shifted in time by a factor $np = 9$.

The forecasted power and the real generated power can be compared in Figure 5.11, and the evolution of the error value during the training process can be observed in Figure 5.12.

This structure showed better performance (small value error) in the prediction than structures A and B.

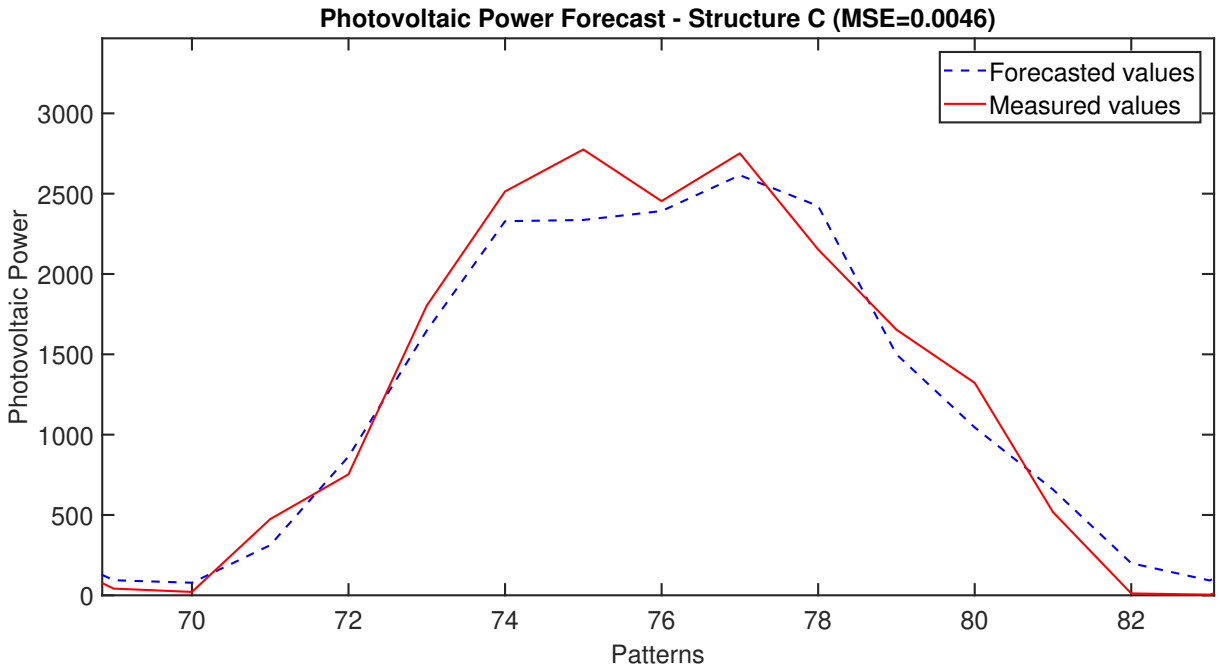


Figure 5.11: Forecast results for structure C

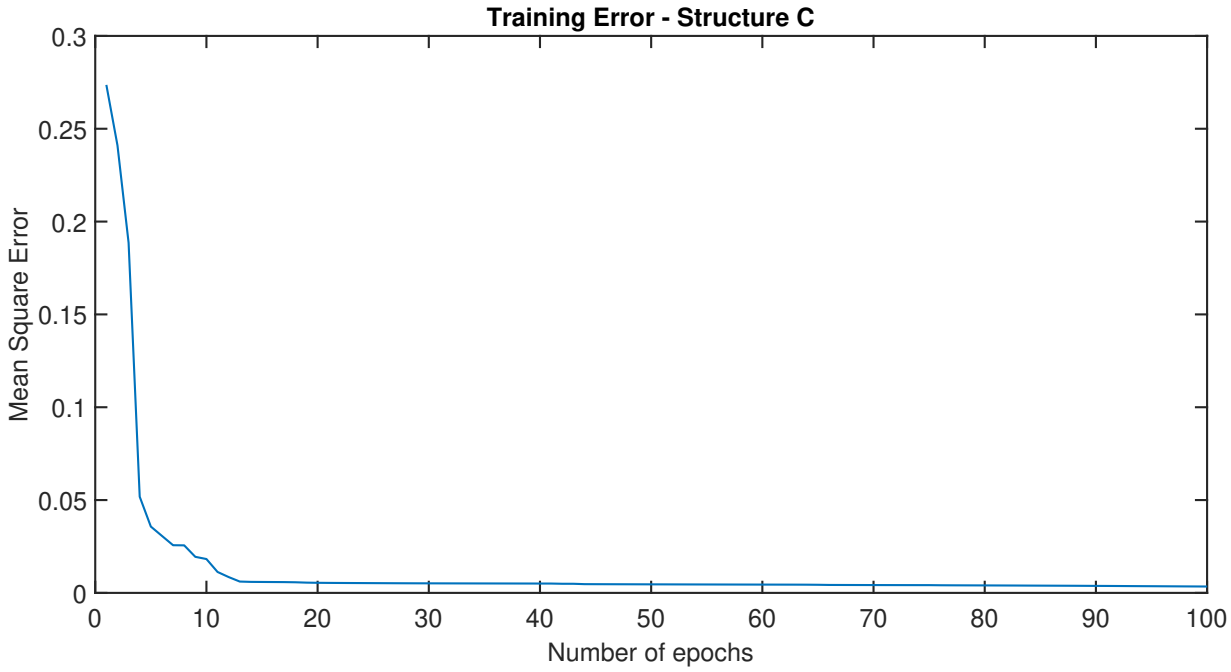


Figure 5.12: Training error evolution - Structure C

In addition, this network could most easily be implemented in a practical situation. First of all, the better results are due to the fact that all the data that was presented to the network was obtained from real measures. Second of all, the system of data acquisition of the generated power is usually installed together with the photovoltaic system. Thus, only one data acquisition system is needed to obtain the necessary data to train the network, not needing another system to measure climate data, as required in structure B. In a practical situation, this might be the best alternative to obtain forecasting results with good levels of accuracy.

5.2.4 Compilation of power forecast for different structures

Table 5.3 compiles the results presented previously. From this table, it can be inferred that the C structure was the one that presented the best precision in the prediction of PV power, which can be noticed due to its low final error.

Table 5.3: Compilation of results

Structure	**Topology**		**MSE**	**Training Error**	**Test Error**	**Number of Epochs**
	First Hidden Layer	**Second Hidden Layer**				
A	3	6	0.0083	0.0089	0.0071	23
B	7	0	0.0063	0.0058	0.0075	10
C	4	8	0.0046	0.0034	0.0073	13

This difference between the networks' results may be due to the fact that only network C has received local and real data. While the others, even receiving real data from climate measurements, did not receive information that was specifically about the location of the photovoltaic system, which may have affected the quality of the forecast. Subsequently, for structure B, this problem can be circumvented by the installation of a weather data measurement system at the power generation site. However, for structure A, the network will always present a similar behavior due to the fact that its data relies on other forecast methods.

Another aspect observed in the results is that networks that have inputs displaced in the time by a factor np presented better prediction results. This may be due to the fact that the np factor allows the network to receive more information about the input set and, therefore, it generalizes better on the relation between this set and the output. However, the value of np can not be very high, because it could cause bad conditioning of the network due to the large number of inputs. Also, np can not be very low, otherwise it would not have the desired effect. To circumvent

this problem of choice of values, several values of np were tested during the learning processes of the network.

Chapter 6

Conclusion

6.1 Conclusions about the study

This study aimed to apply neural network techniques to predict the power generated by a photovoltaic system. According to the results of the study, it can be concluded that such networks have the capacity to successfully carry out this task. Other partial conclusions, that were obtained during the study, are:

- Neural networks are tools that do not work with exact relationships, such as systems of equations. Instead, they operate by relating input values to output values using nonlinear relationships that are internal to the network. Therefore, it is very important that the sets of input and output samples that are presented to the network are well constructed and represent in the best way the system that the network model. In addition, prior to being submitted to the network, such data should be treated with filtering methods, to remove occasional samples that deviate from the normal behavior of the system, and with normalization methods, to avoid the saturation of the internal functions of the network.

- It is also important that the parameters chosen for the network (tolerance, learning rate, initial values for the synaptic weights) are chosen correctly, to allow the network to function in the best possible way. The initial guess of the weights is very important to ensure that the net reaches a good final result and does not get stuck at local minimum points of the error function. Currently, there are several methods to assist in this initial choice of weights, such as those that use genetic algorithms to ensure that the network begins its search for the optimal solution in a region that certainly includes the optimal point.

- An alternative to increase the accuracy of network results for practical applications is to frequently train the network by updating the data set with new measures obtained from the acquisition systems. This ensures that

the network stays updated and that it follows the development of the analyzed system.

- For the validation using real data, it was compared three structures with different topologies. The best result was for the structure using only the real power measures as inputs.

6.2 Future work

In order to improve the result of the power forecast, some actions that could be done are:

- Auxiliary methods can be applied in future studies to improve prediction accuracy. Among these methods can be mentioned the Similar day algorithm, which creates several training networks, one for each type of weather condition (sunny, cloudy and partial cloudy day). This type of method has already been applied successfully in the literature. Its disadvantage is that it requires a large number of historical data and therefore could only be applied in longer studies.

- The use of local data of weather condition (irradiance, temperature, wind) acquired from local measurements. For this goal, in the system installed on the Department of Electrical and Computing Engineering is being installed a climatological kit composed by a pyranometer (to measure irradiance), a temperature sensor (to measure the temperature of the PV modules) and a anemometer (to measure wind features). All the training and testing set of the network used in this work will be reassess with this new set of measurements.

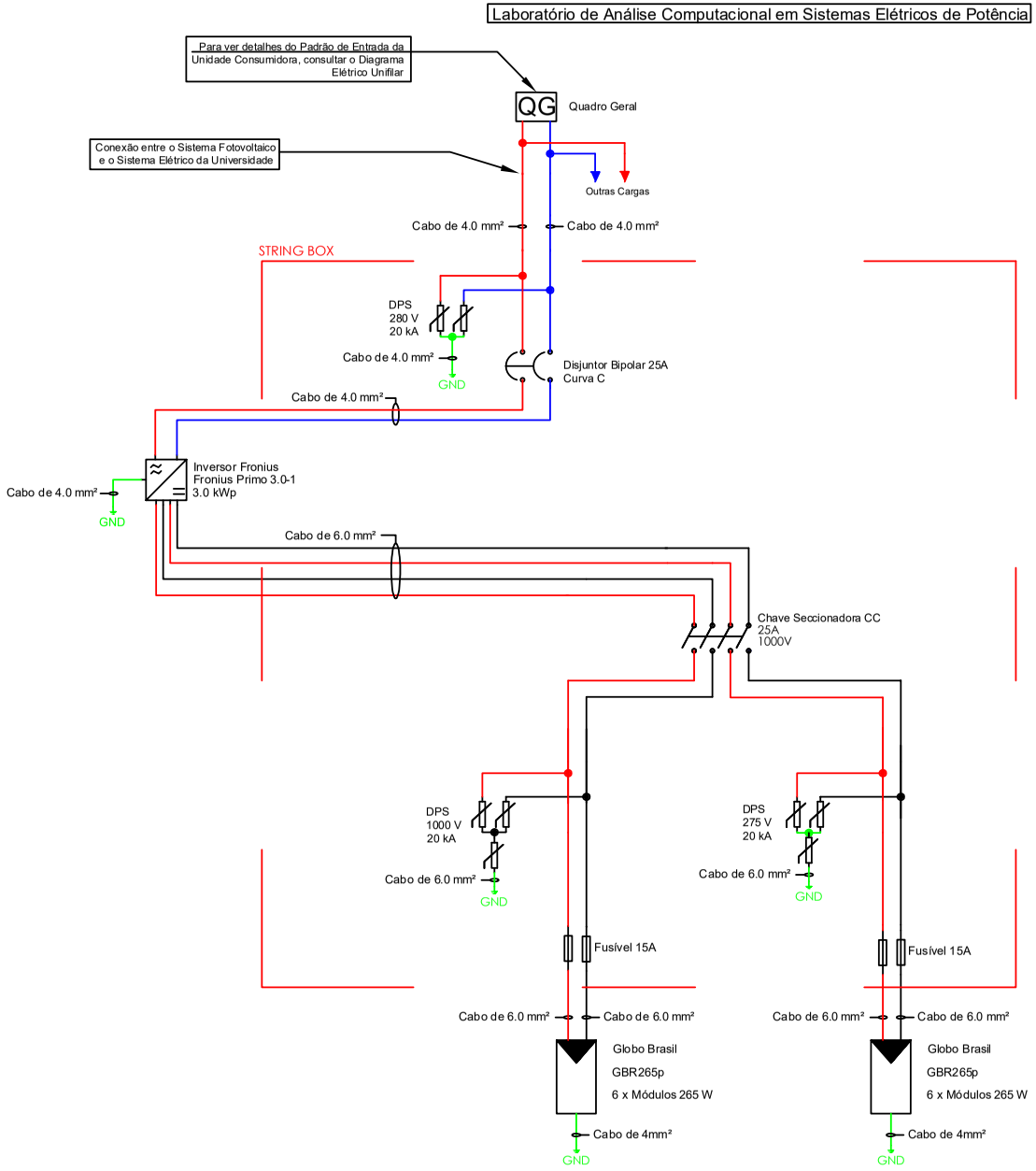
- It is expected to develop a commercial software based on the Matlab script code developed in this research using a graphical unit interface in Python software. This commercial software will support distributed electric companies to forecast the power of photovoltaic installations.

- Finally, it is expected to continue this work by the master study.

Appendices

Appendix A

PV system installation scheme



LEGENDA	
Disjuntor Bipolar 25A Curva C	DISJUNTOR BIPOLAR TERMOMAGNÉTICO PADRÃO IEC -CORRENTE -TIPO
25A 1.000V	CHAVE SECCIONADORA CORRENTE CONTÍNUA -CORRENTE -TENSÃO
Fusível Ultra Rápido Base Fusível 15A	FUSÍVEL TIPO CARTUCHO ULTRA RÁPIDO -TIPO E FABRICANTE -MODELO -CORRENTE
	DISPOSITIVO DE PROTEÇÃO CONTRA SURTO -MÁXIMA TENSÃO DE OPERAÇÃO -CORRENTE MÁXIMA
Inversor Fronius Fronius Primo 3.0-1 3.0 kWp	INVERSOR FOTOVOLTAICO -FABRICANTE -MODELO -POTÊNCIA
Globo Brasil GBR265p 6 X Módulos Fotovoltáicos 265 W	MÓDULO FOTOVOLTÁICO -FABRICANTE -MODELO - QAUNTIDADE E POTÊNCIA

Notas:

1. PARA DETALHES TÉCNICOS COMPLETO DO SISTEMA FOTOVOLTAICO E DA UNIDADE CONSUMIDORA, CONSULTAR DIAGRAMA ELÉTRICO UNIFILAR E O MEMORIAL DESCRITIVO DA INSTALAÇÃO.

REV:	DESCRIÇÃO:	PROJ.:	DATA:
A	EMISSÃO ORIGINAL	Écyo F.	18/10/2017

FASE DO PROJETO: SOLICITAÇÃO DE ACESSO JUNTO À CONCESSIONÁRIA

CLIENTE: Universidade de São Paulo
Colaborador: Prof. Dr. Elmer Pablo Tito Cari

ENGENHEIRO: Écyo Reis C. Farias
CREA - 5069904424
Supervisor: Prof. Dr. Elmer Pablo T. Cari

LOCAL: São Carlos - SP

TÍTULO: SFCR I SEL-USP-SC-Área Norte
Diagrama Elétrico Multifilar

ESCALA:	DATA:	DESENHO:	VISTO/APROV.:
S/ESC	18/10/2017	Écyo F.	Écyo F.

Nº PROJETO:	FOLHA:	REVISÃO:
SFCR-01-2017	02	A

Appendix B

List of publications generated by this research

During the accomplishment of this work, some articles were developed. Those that have been accepted for publication are:

1. Sofia Moreira de Andrade Lopes; Elmer Pablo Tito Cari. Redes Neurais Aplicadas na Previsão de Geração Fotovoltaica. VI Simpósio de Iniciação Científica da Engenharia Elétrica. São Carlos - SP, 2018. Award: Winner of the prize of best work .

2. Francisco Rodrigues Lemes; Sofia Moreira de Andrade Lopes; Elmer Pablo Tito Cari. Photovoltaic Power Forecasting Using a Hybrid Method. Simpósio Brasileiro de Sistemas Elétricos (SBSE). Niterói- RJ, 2018.

3. Vitor Augusto Corrêa; Sofia Moreira de Andrade Lopes; Elmer Pablo Tito Cari. Estimation of photovoltaic systems parameters using the MVMO method. Simpósio Brasileiro de Sistemas Elétricos (SBSE). Niterói- RJ, 2018.

Bibliography

[1] United Nations. *Sustainable Development Goals*. 2016. URL: http://www.un.org/sustainabledevelopment/sustainable-development-goals/.

[2] Jan von Appen et al. "Time in the Sun". In: *IEEE Power & Energy Magazine* February (2013), pp. 55–64. ISSN: 15407977. DOI: 10.1109/MPE.2012.2234407.

[3] California Solar Initiative. *California Public Utilities Commission Program Handbook*. 2017.

[4] *Na Califórnia, energia solar passa a ser obrigatória em imóveis novos*. May 2018. URL: https://noticias.r7.com/internacional/na-california-energia-solar-passa-a-ser-obrigatoria-em-imoveis-novos-10052018.

[5] ANEEL. "Nota Técnica n 0056/2017-SRD/ANEEL: Atualização das projeções de consumidores residenciais e comerciais com microgeração solar fotovoltaicos no horizonte 2017-2024." In: (2017), p. 26.

[6] ANEEL. "Resolução Normativa n 482 de 17 de Abril de 2012". In: *Aneel* D (2012), p. 1.

[7] ANEEL. "Resolução Normativa nº 687 de 2015 da ANEEL". In: *Aneel* (2015), p. 24. ISSN: 1098-6596. DOI: 10.1017/CBO9781107415324.004. arXiv: arXiv:1011.1669v3.

[8] G. K. Ari and Y. Baghzouz. "Impact of high PV penetration on voltage regulation in electrical distribution systems". In: *3rd International Conference on Clean Electrical Power: Renewable Energy Resources Impact, ICCEP 2011*. 2011, pp. 744–748. ISBN: 9781424489282. DOI: 10.1109/ICCEP.2011.6036386.

[9] Emanuele Ogliari et al. "Hybrid predictive models for accurate forecasting in PV systems". In: *Energies* 6.4 (2013), pp. 1918–1929. ISSN: 19961073. DOI: 10.3390/en6041918.

[10] S.A. Pourmousavi, A.S. Cifala, and M.H. Nehrir. "Impact of high penetration of PV generation on frequency and voltage in a distribution feeder". In: *2012 North American Power Symposium (NAPS)*. 2012, pp. 1–8. ISBN: 978-1-4673-2308-6. DOI: 10.1109/NAPS.2012.6336320. URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6336320.

[11] Muhammad Qamar Raza, Mithulananthan Nadarajah, and Chandima Ekanayake. *On recent advances in PV output power forecast*. 2016. DOI: 10.1016/j.solener.2016.06.073.

[12] Yigal Nimni and Doron Shmilovitz. "A returned energy architecture for improved photovoltaic systems efficiency". In: *ISCAS 2010 - 2010 IEEE International Symposium on Circuits and Systems: Nano-Bio Circuit Fabrics and Systems*. 2010, pp. 2191–2194. ISBN: 9781424453085. DOI: 10.1109/ISCAS.2010.5537199.

[13] Rakibuzzaman Shah et al. "Impact of large-scale PV penetration on power system oscillatory stability". In: *IEEE PES General Meeting, PES 2010*. 2010. ISBN: 9781424483570. DOI: 10.1109/PES.2010.5589660.

[14] Juliana Aramizu Paludo. "Avaliação dos Impactos de Elevados Níveis de Penetração da Geração Fotovoltaica no Desempenho de Sistemas de Distribuição de Energia Elétrica em Regime Permanente". In: (2014), p. 186.

[15] CEPEL-CRESESB. *Manual de Engenharia para Sistemas Fotovoltaicos*. Rio de Janeiro, 2014, p. 530.

[16] Marcelo Gradella Villalva and Jonas Rafael Gazoli. *Energia Solar Fotovoltaica: Conceitos e Aplicações*. First. São Paulo: Editora Érica Ltda., 2012, p. 224. ISBN: 978-85-365-0416-2.

[17] Ministério de Minas e Energia MME. *Energia Solar no Brasil e no Mundo*. Tech. rep. 55 61. 2017, pp. 3–6. DOI: 10.1016/j.seta.2013.04.001.

[18] *Energia Solar Fotovoltaica*. June 2008. URL: http://www.cresesb.cepel.br/index.php?section=com_content&lang=pt&cid=321.

[19] U.S. Department of Energy. *Basic Photovoltaic Principles and Methods*. Colorado: Technical Information Office, 1982, p. 71.

[20] J. Antonanzas et al. *Review of photovoltaic power forecasting*. 2016. DOI: 10.1016/j.solener.2016.06.069.

[21] Alberto Dolara et al. "A physical hybrid artificial neural network for short term forecasting of PV plant power output". In: *Energies* 8.2 (2015), pp. 1138–1153. ISSN: 19961073. DOI: 10.3390/en8021138.

[22] C Wan et al. "Photovoltaic and solar power forecasting for smart grid energy management". In: *CSEE Journal of Power and Energy Systems* 1.4 (2015), pp. 38–46. DOI: 10.17775/CSEEJPES.2015.00046.

[23] Ivan Nunes Silva, Danilo Hernane Spatti, and Rogério Andrade Flausino. *Redes neurais artificiais: para engenharia e ciências aplicadas*. Second. São Paulo: Artliber Editora Ltda., 2010, p. 431. ISBN: 978-85-88098-87-9.

[24] Valerio Lo Brano, Giuseppina Ciulla, and Mariavittoria Di Falco. "Artificial neural networks to predict the power output of a PV panel". In: *International Journal of Photoenergy* 2014 (2014), p. 12. ISSN: 1110662X. DOI: 10.1155/2014/193083.

[25] F. Almonacid et al. “Calculation of the energy provided by a PV generator. Comparative study: Conventional methods vs. artificial neural networks”. In: *Energy* 36.1 (2011), pp. 375–384. ISSN: 03605442. DOI: 10.1016/j.energy.2010.10.028. URL: http://dx.doi.org/10.1016/j.energy.2010.10.028.

[26] Cristian-Dragos Dumitru, Adrian Gligor, and Calin Enachescu. “Solar Photovoltaic Energy Production Forecast Using Neural Networks”. In: *Procedia Technology* 22 (2016), pp. 808–815. ISSN: 22120173. DOI: 10.1016/j.protcy.2016.01.053. URL: http://linkinghub.elsevier.com/retrieve/pii/S2212017316000542.

[27] Naji Al-Messabi et al. “Forecasting of photovoltaic power yield using dynamic neural networks”. In: *The 2012 International Joint Conference on Neural Networks (IJCNN)*. IEEE, June 2012, pp. 1–5. ISBN: 978-1-4673-1490-9. DOI: 10.1109/IJCNN.2012.6252406. URL: http://ieeexplore.ieee.org/document/6252406/.

[28] Ming Ding, Lei Wang, and Rui Bi. “An ANN-based approach for forecasting the power output of photovoltaic system”. In: *Procedia Environmental Sciences* 11.PART C (2011), pp. 1308–1315. ISSN: 18780296. DOI: 10.1016/j.proenv.2011.12.196. URL: http://dx.doi.org/10.1016/j.proenv.2011.12.196.

[29] Karoly Ronay, Dorin Bica, and Calin Munteanu. “Micro-grid Development Using Artificial Neural Network for Renewable Energy Forecast and System Control”. In: *Procedia Engineering* 181 (2017), pp. 818–823. ISSN: 18777058. DOI: 10.1016/j.proeng.2017.02.472.

[30] Anil K. Jain, Jianchang Mao, and K. M. Mohiuddin. *Artificial neural networks: A tutorial*. 1996. DOI: 10.1109/2.485891. arXiv: arXiv:1411.3159v1.

[31] Simon Haykin. *Neural networks : a comprehensive foundation*. Delhi: Pearson Education, 1999. ISBN: 81-7808-300-0.

[32] Hao Yu and Bogdan Wilamowski. “Levenberg–Marquardt Training”. In: (2011), pp. 1–16. DOI: 10.1201/b10604-15. URL: http://www.crcnetbase.com/doi/abs/10.1201/b10604-15.

[33] Martin T. Hagan and Mohammad B. Menhaj. “Training Feedforward Networks with the Marquardt Algorithm”. In: *IEEE Transactions on Neural Networks* 5.6 (1994), pp. 989–993. ISSN: 19410093. DOI: 10.1109/72.329697.

[34] *Instituto Nacional de Meteorologia*. URL: http://www.inmet.gov.br/portal/.