Universidade do Algarve

Faculdade de Ciências e Tecnologia

Implementation of an intelligent sensor for measurement and prediction of solar radiation and atmospheric temperature

João Mealha Gomes

Masters in Electronics and Telecommunications Engineering

2011

Universidade do Algarve

Faculdade de Ciências e Tecnologia

Implementation of an intelligent sensor for measurement and prediction of solar radiation and atmospheric temperature

João Mealha Gomes

Supervisors

António Eduardo de Barros Ruano
Pedro Miguel Frazão Fernandes Ferreira

Masters in Electronics and Telecommunications Engineering

2011

Acknowledgments

Financial support for this study was provided by the project "PTDC/ENR/73345/2006 - Intelligent use of energy in public buildings" through the Centro de Sistemas Inteligentes (CSI) of the Universidade do Algarve, Portugal.

I would like to thank my parents for additional (and vital) financial support!

Many thanks also to:

- my supervisors, Prof. Dr. António Ruano and Dr. Pedro Ferreira for their patience and support;
- José Neves,Paulo Luz, Américo Carneiro, Sérgio Miguel and João Lopes from *Publirádio*®, Rodrigo, Bimbo, Ivo, Amin, Coquet, Sérgio, Igor and Aldric for their technical support;
- NEIUAlg;
- Mahaut for the inspiration and constructive reviews.

Abstract

The aim of this study was to develop an intelligent sensor for aquiring temperature, solar radiation and cloudiness index data, and use these measured values to predict temperature and solar radiation in a near future. The prototype produced could ultimately be used in systems related to thermal comfort in buildings and the efficient and intelligent use of energy resources.

So to incorporate these functionalities, a small and portable prototype was developed, which consisted in: a CCTV camera with a *fish-eye* lense for sky images aquisition; a computer of format *mini-itx* with a Linux operating system, for data aquisition and processing; a GPS to enable automatic use independent of the system's geographical position; a pyranometer for regular measurements of solar radiation; a temperature probe, for regular measurements of outdoor temperature; a shadow band to eliminate the sun's flare effect on sky images; Arduino, an *open source* electronics prototyping platform that aquires data from the temperature and solar radiation sensors, as well as processing the data provided by the GPS and controlling the shadow band; neural networks of the type NARX (not developed in this present study), which use the aquired data to forecast the cloudiness index, solar radiation and temperature, in the next four hours.

The system was programmed to aquire data, both from the sensors and the camera, every five minutes.

Keywords: intelligent sensing, temperature, solar radiation, cloudiness indexes, neural networks

Resumo

Este estudo prende-se com a necessidade de desenvolver um protótipo que para além de aquisição de dados de temperatura, radiação solar e índices de nebulosidade, faça também uma previsão da temperatura e radiação solar num futuro próximo. Protótipo esse que possa ser utilizado em sistemas visem o conforto térmico, bem como uma utilização inteligente de energia.

Por forma a incorporar estas funcionalidades, foi desenvolvido um protótipo de dimensões reduzidas que incorpora uma câmara de CCTV com uma lente *fish-eye*, utilizada para a captura de imagens do céu; um computador de formato *mini-itx* com um sistema operativo Linux, que trata da aquisição e processamento de dados; um GPS, para permitir uma utilização automática e independente do posicionamento do sistema; um piranómetro, para realizar medições regulares de radiação solar; um sensor de temperatura, para efectuar medições regulares da temperatura exterior; uma banda solar, utilizada para eliminar os efeitos do brilho do sol nas imagens capturadas; uma plataforma de prototipagem electrónica *open source* chamada Arduino, que realiza a aquisição dos dados dos sensores de radiação solar e temperatura, bem como o tratamento dos dados que provêm do GPS e o controlo do movimento da banda solar; redes neuronais NARX (não desenvolvidas neste estudo), que utilizam os dados adquiridos para fazer a previsão do índice de nebulosidade, radiação solar e temperatura num horizonte de 4 horas. O sistema efectua as medições dos sensores e a captura de imagens de cinco em cinco minutos.

Palavras chave: instrumentação inteligente, temperatura, radiação solar, índices de nebolusidade, redes neuronais

Contents

List of Figures

1. Introduction

1.1. Motivation

In the field of meteorological data acquisition/estimation, more specifically cloudiness indexes, solar radiation an atmospheric temperature, there is room for futher developments. First, there are not a lot of different systems which incorporate, as a package, sky pictures and sensor data. No commercial existing solutions that take photographs for cloudiness studies are portable in a way that they do not allow an easy assemble on the spot. Furthermore, and more importantly, no portable commercial solution incorporates the possibility of forecasting the estimated of measured variables.

This study follows a previous work done by Eduardo M. Crispim et al. (2008), while the estimation is an evolution of A.E. Ruano et al. (2005) and Pedro M. Ferreira and A. E. Ruano (2007). The starting point was the TSI440 from Yankee Environmental Systems, Inc., as it only takes photographs of the sky but does not incorporate any type of sensor, thereby lacking in environmental information.

This study arises from the need to have a device that centralizes a set of tasks - estimates the cloudiness index using photographs of the sky, measures the global solar radiation, measures the atmospheric temperature and predicts the evolution of these values within an horizon of four hours. The previous work done by Eduardo M. Crispim et al. (2008) focuses in the prediction of solar radiation for 30 minutes in future, using multi-objective genetic algorithms for the off-line design of radial basis function neural networks.

This study aims to develop the future in ground to sky imaging systems.

1.2. Goals and Contributions

The main goal of this study is to implement a portable system to measure and predict solar radiation, cloudiness indexes and atmospheric temperature in a four hours horizon,

based on neural networks prediction models. For this, a prototype had to be developed to take photographs of the sky, measure solar radiation and atmospheric temperature, every minute. Using this data, three Radial Basis Function neural networks will predict the evolution of these variables, over the chosen horizon. As it will be seen during this thesis, several steps were performed towards this goal.

In the end, the prototype was built and validated, constituting therefore the contribution of this work.

1.3. Thesis Structure

This thesis is divided into eight different chapters. Chapter 1 contains the introduction of this thesis, as well its motivation and goals. Chapter 2 introduces the reader to the *state of art* in acquiring and processing cloudiness, solar and thermal data. Chapter 3 presents the system that became the reference to this study, as well as its main features and drawbacks. Chapter 4 introduces and explains in greater detail the construction of the prototype developed during this study, the considered hypothesis in software and hardware, as well as the implementation details and the experimental conditions and results of the system. Chapter 5 presents the validation results. These are compared with the ones from the system presented in Chapter 3. In Chapter 6 the results presented in the previous chapters are summarized, and future work possibilities within this study are also presented/discussed. This Thesis includes also a Reference Chapter and Appendix. The former includes all the used references in the development of the thesis; the latter has all the code, annotations, circuits and specifications of the equipment and software that were not included in any other chapter.

2. Literature Review

Nowadays there is a global need for decreasing our energy consumption, to ensure that what we get from our natural resources is efficiently used, and that we live in balance with the environment. Increasingly so, efforts in R&D are directed towards this goal. A major field consists in developing intelligent systems capable of integrating environmental data for a more efficient use of resources and sustainable functioning of man-made utilities. In this study, the main focus is on global solar radiation, as it influences the majority of living beings in many different ways. Thus an accurate prediction of its evolution in time is important for several different areas of application, such as the field of renewable energy (see Kalogirou S. A (2001) for an extensive review), people's comfort in buildings, where possible applications are luminescence control and thermal comfort as Eduardo M. Crispim et al. (2008), among others, proved to be relevant. In the study of global solar radiation, relevant data to be acquired consists in cloudiness indexes, solar radiation and temperature. Not only are these parameters intricately linked, as global solar radiation is usually presented as a time series, making non-parametric methods, such as artificial neural networks good tools for this type of study.

Though no neural network model was developed in the present study, the ones developed by Igor António Calé Martins (2010) were used. Therefore, in this chapter, special attention is given to the possibilities and different types of data processing when using neural networks, particularly when processing and estimating cloudiness indexes, solar radiation and atmospheric temperature.

2.1. Cloudiness Indexes

The existing research in cloudiness indexes differs in their goal, in how data is obtained and how it is processed. Some are concerned with the type of clouds being observed, others only with the amount of radiation filtered. As for data acquisition, the most used technologies are: satellites, ground all-sky imaging systems and spectroradiometers.

One of the most used geostationary meteorological satellites is the *Meteosat*, that has been operational since the 1970's. Many researchers still use this technology, such as Arnout Feijt et al. (2000), to detect clouds, measure temperature and test new models, among other applications. This satellite allows a temporal resolution of 30 minutes and a moderate spatial resolution, which provides an easy way to monitor fast atmospheric processes. Arnout Feijt et al. (2000) introduced a cloud detection algorithm Meteosat Cloud Detection and Characterization KNMI (Metclock) scheme, which analyses the infra red and visual channels measurements over an area from about 25^oW to 25^oE and from 35^o to 70^oN, covering Europe and a small part of northern Africa. This area was divided in 20 segments for analysis, that were then classified into four different categories: sea, coastal, land and mountainous. In this scheme, each pixel was then classified as cloudy or cloud free by two algorithms, one based on the threshold comparison of the apparent brightness temperature (infra red test) and the other based on the comparison of the threshold of reflectivity (VIS test). Over the sea, the scheme presented quite good results, since the spatial variability was relatively low. However over land, because of high temperature gradients that occur due the variability of surfaces, the results can be misleading. The Metclock differs from other cloud detection scheme's based on infra-red analysis, in how the temperature threshold is obtained from the infra-red data. One of the most used methods, is the frequency analysis of the spatial distribution of temperatures measured from the satellite. These provide very good results for over sea analysis due to low spatial variability, nevertheless over land analysis the spatial variability is quite high depending on the global zone. As referred in the study, the characterization part of Metclock, used to retrieve cloud cover fraction, optical thickness, emissivity, and correct cloud-top temperature from the reflected sunlight as measured from the VIS channel, is still under development.

Another type of satellite is the moderate resolution imaging spectroradiometer (MODIS), as used by Xiaoning Song et al. (2004) to detect clouds and analyse their images. The study aims to improve cloud removing and precision in remote sensing image recognition. MODIS has thirty six bands with a spectral range from 0.4 to $14\mu m$, twenty visible/near-infra red bands and sixteen thermal infra red bands; but only seven of them were used, the ones sensitive to clouds, aerosol, temperature and water vapour. The area

used for this study had several singularities, such as vegetation, sand and water, and since these have typical spectrums, it was easy to recognize clouds. The cloud pixel information was obtained through the low temperature on the cloud top using thermal infra red bands, but the thin clouds were not detectable. When temperatures from land surface and top thin clouds were alike, it was necessary to statistically analyse the eigenvalues of each land cover in the image to know the distribution of values. The thresholds used for automatic cloud detection, were related to time, region and sensor which translates data into a spatial texture, that altogether with a neural network provided an automatic detection method. As for the results of the method, the precision was up to 95%, since it is said that it works quite well in those cloudy pixels that are uneasy to distinguish in naked eye.

Alternatively to satellite observation are the all-sky imaging systems, which promote a more accessible way of cloud monitoring. In the study made by G. Pfister et al. (2003), that aims to monitor the impact of the clouds in the radiation field, without relating it to the temperature, two different all-sky imaging systems were used to measure the total opaque and thin cloud fraction, indicating as well whether the sun was or not covered by clouds. The used all-sky imaging systems are:

Allsky1: developed by the National Institute of Water and Atmospheric Research (NIWA);

Allsky2: also known as Total Sky Imager (TSI) 880 built by Yankee Environmental System, Inc.;

The two systems cover the same field of view, and were placed 300 m from each other, allowing the possibility of using the images for stereographic analysis of cloud height. Both systems use a wide-angle lens, a hemispheric mirror and a shadow band to prevent sun's reflections. The analysis of the pictures was based on red-to-blue ratio (the ratio of red to blue light in a pixel), since white clouds have a higher red-to-blue ratio than blue sky. A red-to-blue threshold was then used to determine whether a pixel is cloud or clear sky. There was also an estimated fraction cloud cover, that was given by the number of cloudy pixels divided by the total number of pixels in the field of view. Besides the comparison of the two systems, which was done through the comparison of $\pm$ 67000 images, where in about 80% of the images the absolute difference in the total and opaque was less than 10%, there was an effort to relate the cloud effect over the solar

irradiation. Despite the fact that the cloud fraction is an important element in the amount of solar radiation that reaches the surface, it was not possible to relate the cloud coverage alone to the solar radiation, even when the sun was obscured. Cloudiness indexes were not sufficient to explain the radiation field. Nevertheless it was shown that this type of technology proved to be a good way of making long term measurements, and also to continue studying the cloud coverage with high temporal resolution. It is also important to mention that this study was an important starting point for the use of this type of systems.

A different type of on ground all-sky imaging, is the one used by A. Cazorla et al. (2007), which aims for aerosol characterization using neural networks. The study is based on the principle that aerosols affect Earth's radiation and the temperature field by changing the energy balance and solar distribution in the atmosphere. The instruments used in this work were: an all-sky imager developed by Atmospheric Physics Group (GFAT), to provide images of the whole sky dome in daytime for cloud characterization; a Peltier system to control the temperature of the system; a solar shadow system, a 2AP Sun Tracker (by Kipp & Zonen), that must cover the camera lens every moment. To obtain a characterization of the aerosol optical depth (AOD) they used a sun-photometer CE-318-4 (by CIMEL Electronic). Two radial basis network were used, one for estimating the AOD, and another for the α estimation (characterizes the spectral features of the aerosol, Ångström (1964), and is related to the size of the particles, Shifrin and K.S. (1995)). The results were compared with the CIMEL CE318 sun photometer, and for the AOD the results are in the limits of nominal precision of the CIMEL CE 318; for the α parameter about 80% of the cases have deviations of below 0.1, which is also between the limits of nominal precision. But there is an uncertainty to the precision of the system, and the developed neural networks were quite sensitive to the quality of the data used for the training.

In a more recent work by Francisco J. Olmo et al. (2008), the retrieval of optical depth using a CCD camera was studied, using the same system as A. Cazorla et al. (2007). The study uses a real-time iterative procedure. The first method did a cloud classification of pixels derived by a neural network (using a multilayer perceptron model), reconstructing spectral sky radiance at zenith by a linear pseudo-inverse spectral estimation algorithm, and finally, deriving the optical depth at $550nm$ using radiative transfer code. In this

method the clouds were classified based on three parameters, mean of the pixel and its neighbours in the red and blue channels, and the variance of the pixel and its neighbours in the red channel. The information obtained was the percentage of thin clouds, *oktas*[1] for opaque and thin clouds, the percentage of opaque and thin clouds in every octant, and the sun position in the octant. The second method was used to obtain real-time sky radiances (spectral power distribution) at zenith, from 380 to 780 nm, by means of the pseudo-inverse spectral estimation algorithm, which provided a spectral radiance for each pixel 1280×1024 image in real time and required a training set. The training set was based on grouping the similar spectral radiances, according to their distances measured by the colorimetric and spectral combined metric, defined by M. A. López-Álvarez et al. (2005). The third method showed that it is possible to recover zenithal sky radiance spectra, with a reduced number of broadband spectral filters used in the CCD camera. The authors of this study claim that the methodology presented solves the strong forward scattering problems associated with the large particle load in neural networks. As the only training set concerns the spectral sky radiance reconstructions, the absolute radiance calibration was not necessary, and the training set does not depend on the location.

Based on radiometric data the work of H.F. Assunção et al. (2007), used 5 minutes averaged values of global and indirect solar irradiance and daily values of sunshine duration, considering four types of cloud reflection, clear sky, sun reflecting in clouds, sun partially covered and sun totally covered. With the developed algorithm, it was possible to estimate the daily and hourly conditions of the sky. Previous work had been done on this subject, such as J. Calbó et al. (2001), D.H.W. Li and J.C. Lam (2001) and A. Orsini et al. (2001), which catalogue the clouds in different forms and with different time estimations. The method used in this study to characterize the sky was based on solar irradiation values that allows the estimation of the cloudiness indexes for the past five minutes. H.F. Assunção et al. (2007) chose this time interval based on H. Suehrcke and P.G. McCormick (1988) work, that observed that the frequency distribution for 5 minutes-averaged solar irradiance values is similar to the frequency distribution of instantaneous values and concluded that hourly and daily values of solar irradiance do not represent appropriately

[1]Unit used to define the clearness of the sky. Each *okta* represents one eight of the sky covered by clouds. For further information see http://worldweather.wmo.int/ .

the statistical properties of instantaneous values. The data acquisition for this study was made in Botucatu in the state of São Paulo, Brazil; the global solar irradiation was made with a pyranometer of spectral precision (PSP), and direct solar irradiation was measured by a pyrheliometer coupled to a sun tracker model ST-3, manufactured by Eppley Inc. The measurement period was six years, from 1996 to 2001. The proposed algorithm was based on the principle that short term variations of cloudiness were due to changes of the relative optical mass of air. Different frequencies were used for clear and cloudy sky, but it was also necessary to set boundaries to the different types of categories, that according to M. Jurado et al. (1995), can be mathematically justified and are well defined. The algorithm consists in a matrix of four columns and the number of lines corresponds to the number of observed indexes during a day, made to obtain the real duration of the day. The collected datum was compared with a recording device from Campbell-Stocks, with few errors. The authors also suggest that this algorithm is to be used only in places where there are solar irradiation records, with at least five years.

The study made by Calbó and Sabburg (2008), show a method that analyses and recognize clouds according to their patterns. In order to achieve that purpose, pictures from two different devices were used:

- the TSI (*Total Sky Imager*), has a field view of 160^o, and the photos are 352 $\times$288 pixels in JPEG format. After capture, the photos were transformed to 256 $\times$ 256, through a bilinear interpolation using the MATLab® software.
- the WSI (*Whole Sky Imager*), is a colour digital camera with a fish-eye lens, providing a field of view of 180^o , that aims to the zenith. The captured photos are 768$\times$576 in BMP format, although it is referred that the JPEG format with a compression of 80% was better for the image digital processing.

Besides their own experience in the field, the authors also based this study on previous works of Buch et al. (1995), defining 8 conditions to be used in cloud classification. These are not based on the classical classification of clouds, but on the visual aspect of the cloud instead. The photos were characterized through their texture, analysing the histogram properties. These process can be applied to gray-scaled pictures, that is pictures whose pixels information is represented by a single value. Since the captured pictures had

three components, red, green and blue (RGB), they had to be transformed to gray-scale pictures. Two transformed pictures were used, one had the red component transformed to blue and the other only had the value of the intensity, given by $(R + G + B) \div 3$. Cloud pattern was analysed in frequency using a Fourier transform, based on González et al. (2004) work. It is important to mention that the initial values are real, and that the spectral power is symmetric relatively to the origin. Since the power spectral functions from different types of sky are different between them, it was necessary to identify simple characteristics in spectral power, to identify each type of cloud. These characteristics are:

Correlation with clear (CC) sky: quantifies and compares the spectral power of the analysed image, with a reference (spectral power of image with clear sky). The CC value is the linear correlation coefficient between the logarithms of the two power spectral functions, between 0 and 1. The bigger this value is, the more homogeneous is the sky.

Spectral intensity (SI): depending on the presented cloud patterns, there will be more or less spectral power.

Since the previous methods only analysed the image pixels, and since there was not a distinction between cloud pixels and clear pixels, other characteristics were added allowing a better recognition of each type of sky. This distinction was made using a *threshold* in the image R/B, and its value was calculated using a MATLab® script, value which was based on the the picture histogram. According to this study, it is possible to recognize sky patterns subsequent cloud classification. It is mentioned, that during the research of the study, there were explored two different characteristics: Fourier transform, one of most promising approaches to cloud-pattern recognition, became limited when tried to summarize a limited number of values, the information contained in the FFT; pixel distinction, could have led to better results if the value of threshold was changing according to the position of the pixel.

In the study made by Atsu S. S. Dorvlo et al. (2002), a set of neural networks is presented that estimated the cloudiness indexes using latitude, longitude, altitude, light intensity and month of the year. They used radial basis function neural network ($RBFNN$), the learning method involved the input/output data to exploit the next pa-

rameters: the network inputs, the bandwidth of the Gaussian function, and the weights between networks neurons and output layer. One of the techniques used to get the described parameters was based on fixed *RBFNN*. This *RBFNN* has random selection for the centres in the set and the width of Gaussian radial basis function was expressed in terms of the maximum distance between the chosen centre and the number of centres. To train this neural network the author used a back propagation algorithm implemented in the neural networks toolbox of MATLab®.

Another neural network used was the multi layer perceptron (*MLP*) network, which consisted in one input layer, and one or more hidden layers. The architecture used in this model had five neurons in the first layer, one hidden layer with sigmoid activation function, and a output layer with a linear activation function. This network training was made using a back propagation algorithm belonging to the neural network toolbox (*trainbr*). In the end, the final network was implemented with one, two and three hidden layers, and the best results were selected using the mean square error.

The data sets were collected from eight meteorological stations during ten years, and were used to train three *MLP* and one *RBFNN* neural networks. The output parameter was the cloudiness index. To get the solar radiation, the author used this index and multiplied it by the maximum radiation index.

It is said in the study that both network types had reasonable results, but the *RBFNN* model was preferred due to its best results and smaller computational time.

In this thesis, the implemented sky classification method was developed on the study of Pedro M. Ferreira et al. (2010) and the ANN's used to predict cloudiness was developed on the study of Igor A.C. Martins et al. (2010).

The sky classification method was implemented using a new approach based on ANN's alternatively to the common threshold method. This implementation was also compared with three different methods: fixed threshold, *RCT* method, *Otsu's* method and a *Neural Network* method. A common approach in all methods, is the calculation of the threshold value on a given pixel intensity scale. The testing of this approach involved several intensity scales. On the original RGB image the red channel were considered, and were also considered the grayscale images after a transformation being performed.

This transformation consists on the conversion of the original RGB image into the Hue-Saturation-Value (HSV) colour model, setting the value channel to 1 for all pixels, and converting the transformed image back to RGB. The transformation applied to the value channel, has an equalization effect on the pixels of the luminosity, since the colours on the HSV colour model become more distinct. The converted RGB images have an improved contrast within the red channel, which allows a clearer distinction between the clear sky and the cloud pixels. The fixed threshold method is based on a histogram analyses in order to identify the best single threshold value to be applied to every image of the used set. The *RTC* method consists on an histogram iteration form, presented by Trussel (1979), of an iterative thresholding algorithm proposed by T. Ridler and S. Calvard (1978). This method evaluates a unique threshold from each image, assuming a bimodal histogram, being the threshold $T = (\mu_0 + \mu_1)/2$. The μ_0 and μ_1 are the means of each of the two components of the histogram separated by the threshold. The *Otsu's* method, proposed by Otsu (1979) , consists on searching the pixel intensity scale for the threshold that maximises the inter-class variance of each component of the histogram. In the Neural Network method an attempt was made to identify a Radial Basis Function (RBF) Neural Network (NN) image thresholder using a Multi-Objective Genetic Algorithm (MOGA) developed by C. Fonseca and P. Fleming (1998). The NN's were trained by the Levenberg-Marquardt algorithm using a modiffed training criterion exploiting the nonlinear-linear RBF NN topology (P. M. Ferreira and A. E. Ruano (2000); Ferreira et al. (2002)). In this method the set of 410 images was broken into three sub-sets: the training set (290 images) for NN training; the testing set (60 images) for generalisation testing; the validation set (60 images) to evaluate the NNs after the MOGA execution. From all the images and from the performed transformations, a total of 69 features were extracted. First, from the original RGB image the HSV and Hue-Saturation-Lightness (HSL) images were obtained. Secondly, on the HSV and HSL images the V and L channels were set to 1 and 0.5, respectively, and these transformed images were converted back to RGB mode generating two additional RGB images. Finally, from each RGB mode image a gray intensity image was generated. This resulted in a total of 7 different images and 17 distinct intensity channels. From each of these, the sample mean, standard deviation, and skewness were extracted. Additionally, from the red and gray intensity channels (6 in total) histogram, the most frequent, first non-zero, and last non-zero intensity levels were also extracted.

The NN structure (set of input features and the number of neurons) was selected using the MOGA in a similar way to that presented in P. M. Ferreira et al. (2003). Two objectives were setup for minimisation: the training and generalisation NN output (Root Mean Square) RMS errors. The NN revealed an improvement in accuracy of approximately 50% when compared to the best results obtained by the remaining methods.

The cloudiness prediction methodology used the RBF NN image thresholder selected in the identification experiment described in the previous study, cloudiness time-series were estimated by feeding the NN with the features extracted from consecutively acquired (1 minute sampling interval) images, and applying equations 1 to 4 for each image and corresponding estimated threshold. Which was done for a three months periods: October, 2004; February, 2005; and May, 2007. The correlation of measured global solar radiation to the cloudiness was also inspected. These time-series were afterwards employed in the identification of RBF NN cloudiness predictive models. As these were meant to be employed in a global solar radiation prediction scheme which operates on the basis of a 5 minute sampling interval, the cloudiness time-series were averaged over 5 minute consecutive periods, thus generating smoothed time-series synchronised with the solar radiation data. The experiments for the identification of RBF NN cloudiness predictive models by means of a MOGA, followed the same principles outlined in the previous section for the NN image thresholder. In this case the NN cloudiness predictive models were identified as autoregressive One-Step-Ahead (OSA) predictive models which, for a given instant k, were iterated in a multi-step fashion to generate predictions from instants $k+1$ to $k+ph$, where ph is the prediction horizon, in this case set to 48 steps corresponding to 4 hours. The MOGA was employed to evolve a preferable set of models whose number of neurons and selected input terms optimise a number of pre-specified goals and objectives. In this case the MOGA chromosome, the cloudiness auto-regressive predictive model, are a collection of delayed output signals, together with the number of neurons. The input terms considered were selected from the delays within the period of the past 24 hours.

To assess the models performance three objectives were used: the RMS, the OSA error and the third objective was computed on the basis of the long-term model prediction error taken from the multi-step model simulation over the prediction horizon. The best

model structure has 6 neurons and employed 13 input terms, and also corresponds to the best one in the preferable set in terms of the long term prediction objective, the minimization of the summed RMS prediction error.

Thus for cloud detection, satellites tend to be more used, as shown by Arnout Feijt et al. (2000),Xiaoning Song et al. (2004), Patrick Minnis et al. (2008) and B. H. Kahn et al. (2008). Nevertheless, ground to sky imagers tend to be a cheap alternative allowing images with greater resolutions, the frequency between each new image being higher than the satellites and also providing more focused results, as shown by Buch et al. (1995), G. Pfister et al. (2003), A. Cazorla et al. (2007), H.F. Assunção et al. (2007), Calbó and Sabburg (2008) and Francisco J. Olmo et al. (2008).

2.2. Solar Radiation

Solar radiation estimation is not a trivial subject, specially if zones with a complex topology are considered, such as mountain cases, where it is difficult to support economically and maintain the equipment. Another problem in this type of terrains is the number of variables involved; these require a denser neural network. There are a few studies in this area, such as J.L. Bosch et al. (2008), who performed a study in which the performance of active and passive energy systems are analysed and J.C. Winslow et al. (2001), who developed a model of daily solar irradiance estimated from air temperature and precipitation data. In this last approach, the best results were obtained through extrapolation/interpolation of the acquired data in a station near of the study location, but even there it was difficult to take problems such as terrain complexity, into account.

Other studies also explore the benefit of using neural network instead of an empirical model, such as the study performed by Jérôme G. Fortin et al. (2008), which considers also two classical approaches to simulate the propagation of solar radiation on earth surface developed by Tymvios F.S. et al. (2005). Besides the classical methods, the solar radiation could be estimated by radiation measurements interpolation, as studied by Y. Xia et al. (2000), or by satellite images analysis, as studied by Lèfevre M. et al. (2007).

The study developed by Jérôme G. Fortin et al. (2008), used two different methods:

the *HS* model developed by Hargreaves G.H. and Samani Z.A. (1982), which is a simple empirical model to simulate daily solar radiation received in an horizontal surface to estimate the evapo-transpiration; and the *MH* model developed by Mahmood R. and Hubbard K.G. (2002) that uses a different approach in the empirical model, in which the transmissivity is a function of the day of the year, and the daily solar radiation received by the surface is correlated with temperature amplitude and extraterrestrial radiation. The latter is identical to the previous, the main difference being the unbiased characteristic. These models were optimized using a non-linear standard unrestricted scheme.

The neural network method used was the *MLP*, trained with the optimized *Levenberg-Marquardt Back-Propagation Optimization* with a Bayesian regulation. The selected inputs were the extraterrestrial radiation, the day of the year and maximum and minimum air temperature. The best performance was achieved with eight neurons in the hidden layer. The collected data was provided by eleven meteorological stations, and consists of temperature and solar radiation, where the data from six of them were used for trainning and optimizing the empirical method. The other 5 stations were only used for testing. Each station gave about 225 data sets of solar radiation, in the years 2004 and 2005. Each variable had 2522 observations, 1385 of them to train the network and 1137 to validate it. The empirical model that obtained the best results was the *HS*, and even after the error optimization, the errors obtained by this model were greater than the results obtained by the second best neural network. These results were almost unbiased. The MLP formulation had enough flexibility to explore a wide range of multivariate regressions in a single step. This advantage prevails over empirical models like HS or MH where the formulation structure was fixed beforehand. Furthermore, the non-linear property of the MLP activation function was particularly valuable for reproducing the behaviour of a natural phenomenon like surface incoming solar radiation and its relation with common meteorological variables.

The model used in the study of J.L. Bosch et al. (2008) used artificial neural networks, since these methods allow more accurate results in the solar radiation forecast and estimation than other models. The data used was collected by 12 stations localized near the Sierra Nevada, in the city of Huéneja in Granada, Spain. These stations collected solar

radiation using LICOR 200-SZ pyranometers between 2003 and 2005, with an interval of 2.5 minutes. In the study, the data of one of the 12 stations were chosen be the reference of the artificial neural networks, and the data of the remaining stations were used to generate comparison values. The artificial neural network used was a MLP with one hidden layer, due to its simplicity; the activation function was the hyperbolic tangent. In this study a method is presented, Automatically Relevant Determination (ARD), based on the Bayesian structure (see G. López et al. (2005)) for determining the structure of artificial neural networks. This method limits the complexity of the networks by identifying the variables' relevance, and avoids the data separation in subsets, therefore allowing the use of the whole data for training. The results were demonstrative and proved that the day of the year and cloudiness index are very important to solar radiation forecast and estimation in these conditions, since these variables proved to be vital for the results accuracy.

The study made by Eduardo M. Crispim et al. (2008), also aims for the solar radiation forecast, based on sky image acquisition and solar radiation measurements. The artificial neural network types used were RBF composed by three layers. The input layer was composed by the sensor inputs, i.e., by the data from the sensors. The second was an hidden layer composed by several neurons chosen by a MOGA, and the third and last layer had a neuron that corresponded to the output. The artificial neural networks' learning was performed in two parts: the first, in which an unsupervised procedure, OAKM (optimal adaptive k-means clustering) was used to compute the initial position of the centres, its spreads being computed by a simple algorithm. The second part consisted in the training method, through the LM algorithm. To verify the influence of the cloudiness index on the solar radiation forecast, three methods were compared, one without cloudiness index and two others with a different cloudiness index as exogenous input. One of the NARX models had in account all sky and the other just an ellipse area near the sun. The device used to acquire sky images is a Total Sky Imager (TSI), through a CCD camera. This device has a black strip to protect the CCD from the reflected sun on a concave mirror. After collecting the images, these are processed to get the cloudiness index (in percentage). The sky was classified as clean sky, clouds with low density level and high density level. The TSI system has an algorithm that classifies the sky, but results were unsatisfactory, especially when the sky had a different hue; for example, when the sky is completely covered with

clouds, the algorithm presents a result of 0% of clouds, i.e., clean sky. To determine the cloudiness index, two NARX models with different type of cloudiness index were used. These models were later compared with a NAR model. To build a realistic model of solar radiation three different models were used: a NARSR model; a NARX model, with the PCS values as exogenous input; and a NAR model. The data used to train the networks had 10080 points extracted from six days, and the validation set used data from four days, with a total of 7200 points. The solar radiation values were measured with a pyranometer. The lag terms were chosen by the MOGA, employing also an heuristic relation which took into account the solar radiation variation through time, and the relation between the solar radiation and the cloudiness indexes history. The model evaluation was achieved using the *Root-Mean-Square* of the one-step-ahead training errors (OSATE), of the generalized set (OSAGE) and the RMSE long time forecast. According to the results in the study of Eduardo M. Crispim et al. (2008), the cloudiness index is a relevant parameter in the solar radiation forecast; however the capture methods are not optimized and result in large errors in the cloudiness index determination.

In this thesis, the solar radiation forecast follows precisely the used methodology in the study of Igor António Calé Martins (2010). Similarly to the previous mentioned studies, MOGA was employed in order to identify the ANN's input-output structure. The solar radiation data was acquired using a 1 minute sample rate and then re-sampled applying a 5 minute average over the entire data set. The available inputs in MOGA, the lagged solar radiation values, were taken from the past 24 hours. The considered inputs were, all the lagged terms from the most recent 8 hours, inputs taken at 10 minute intervals in the following 8 hours and the input terms taken at 15 minute intervals for the last 8 hours. Such choice was made, since the radiation value near the forecasted instant is more important than the values taken in the previous 24 hours. As for the NARX model, the exogenous input terms considered were extracted from the previous 4 hours to the prediction instant. The number of neurons of the models was restricted to range from 2 to 16 and the number of input terms was restricted to lay in a range between 7 and 32. The ANN training method used was the LM algorithm using a modified criterion exploiting the non-linear/linear topology of the RBFNN, and the stopping criteria used was early stopping. The NARX model and the exogenous inputs considered used a recursive multi-step prediction, an approach usually called the Minimum Error Prediction

(MEP). As for the results, the long term forecast was considered to compare the NAR and NARX approaches. The best result for the NAR system was achieved with a NN with 22 input terms and 6 neurons in the hidden layer. This structure achieved an RMS error of about $80W/m^2$ at the four hour ahead prediction. For the NARX system, the best RMS error at the four hour horizon was around $50W/m^2$. This represents an error decrease of approximately 30%, relatively to the NAR model.

2.3. Temperature

The study of Kostas I. Chronopoulos et al. (2008) has the purpose of obtaining ANN's models to estimate meteorological data values in areas with sparse meteorological stations. This study relies on the need of having available meteorological data in the required spatial resolution, which might no be available due to the complex terrain of the studied area, or due to the sparse network of meteorological stations in the major area of the study. The goal of this study was to develop and apply ANN's models to estimate atmospheric temperature as a function of the corresponding values of one or more reference stations. The application site is the canyon of Samaria, located on the south-west Crete Island, Greece, covering a total distance of 18 km. The data set used consists of measured mean hourly temperature data from 4 meteorological stations, which were HOBO® type of Onset Computer Corporation. The sensors were protected with radiation shields and were placed on trees about $1.5m$ above ground. All measurements were taken every 10 minutes and then were averaged to hourly values. A statistical model (multiple regression model), was used as comparison to test the performance and evaluate feasibility of the developed ANN's. The data set used consists on a total of 1550 hourly observations. The data was divided into two separate overall data sets, the first data set was used for training, while the second was used for testing and consisted on 550 hourly observations. The variables are the mean hourly temperature data from the four stations and the mean hourly relative humidity data from two of the stations. An ANN model was developed, a multilayer preceptron (MLP) and a statistical model, multiple linear regression (MLR). The MLP was trained with a back propagation algorithm, and the network consists on 1 hidden layer and 3 nodes. The method uses the final weight as initial weights of the validation networks to take into account the problem that a network

trained with all of the known data set may not be identical to networks trained with subsets of the data. On the MLR, the regression coefficients represent the independent contributions of each independent variable to the prediction of the dependent variable. The evaluation of the performance of the results obtained using the MLP and MLR, was done through two different criteria: the determination coefficient between the observed and estimated values (R^2) and the Mean Absolute Error (MAE). The evaluation of the preliminary results of an ANN model used in the present study to estimate air temperature in areas where meteorological stations were sparse, showed that fair to very good air temperature estimations may be achieved depending on the number of the meteorological stations used as reference stations. The comparison of the ANN model results with the results obtained from the MLR model, showed the better performance of the ANN model. In addition, it is mentioned that the ability of ANN models to take into account any non-linear relationships among meteorological parameters, may provide advantages over traditional methods.

The control of the environment inside a greenhouse, is also an import topic when studying the prediction of atmospheric temperature, such as the study preformed by Irineo L. López-Cruz et al. (2007), in which a dynamic linear model of auto-regression with exogenous variables was developed to predict the air temperature inside the greenhouse. And also the study performed by P.M. Ferreira et al. (2005), where a radius basis function (RBF) model was developed in order to predict the greenhouse temperature and humidity.

The study developed by Irineo L. López-Cruz et al. (2007), aimed for an evaluation on the methodology used to generate ARX models in order to predict the air temperature of a Mexican greenhouse with natural ventilation. The greenhouse has $1080m^2$ and is located at an altitude of $2244m$ in the experimental field Tlapeaxco of the Universidad Autónoma Chapingo in México. The meteorological and input variables were temperature, relative humidity, solar radiation and wind speed. They were measured during 3 months, using two Davis® Weather automatic meteorological stations. The output variable was the air temperature inside the greenhouse. Samples were taken of the input and output variables of the model every 5 minutes during a crop cycle. To determine the structure of the best model, 100 000 ARX models were evaluated using the information criteria and final prediction error of Akaike (1974). The adjustment between the simulated and

observed temperatures, and the residual analysis, indicated that ARX models of second degree or above, adequately predict the behaviour of the temperature inside the greenhouse. It is said, that the obtained ARX models had a fit of more than 70% over the group of data of estimation and validation, and also that the models also satisfied the residual analysis.

The study developed by P.M. Ferreira et al. (2005), applied a MOGA to identify the radial basis function neural network coupled models of humidity and temperature in a greenhouse. This study arise from a similar need as the study of Irineo L. López-Cruz et al. (2007), since the greenhouse temperature and humidity are the two most important environmental variables, that affect the crop production. The data used in study was acquired with a sampling rate of 1 minute in a plastic covered greenhouse, and was collected during 12 days. The variables involved were the greenhouse relative humidity and temperature, the outside temperature and solar radiation. The wind speed and direction was not considered, since the authors did not have any information on the greenhouse's opening state, and it is mentioned that in previous studies the MOGA discarded these variables from the model. The number of points used was reduced by applying a 5 minute average over the entire data set, and afterwards, due to the different scales in the variables considered all were scaled to the an interval between 0 and 1. The resulting data was split into three different data sets, model training, generalisation testing and validation. The RBF NN's were trained as one-step-ahead (OSA) predictors by an algorithm based on unconstrained deterministic optimisation employing a Levenberg-Marquardt method. Early stopping was used to terminate training. In the MOGA each individual was a pair of humidity-temperature models. Both models structure were a Non-linear auto-regressive model with exogenous inputs (NARX). The quality of each trained NN was assessed by its prediction performance over an horizon of 3 hours (36 steps) using the generalized testing data set. To reduce the computational time, the authors computed the prediction starting at one hour intervals, which resulted in 96 prediction horizons. The respective error of each prediction was measured using the *root mean square error* (RMSE). The obtained networks had 6 neurons and 11 inputs for the humidity model, and 2 neurons and 29 inputs for the temperature model. The study presents good fittings for the prediction in 1, 18 and 36 steps ahead. The mean and maximum absolute error values obtained for humidity and temperature were 3.0383%, 24.5791%, 0.8434^oC, and 6.1402^oC.

In this thesis, the temperature forecast follows precisely the used methodology in the study of Igor António Calé Martins (2010). Similarly to the previous mentioned studies, MOGA was employed in order to identify the ANN's input-output structure. The sampling rate used in data acquisition was the same as for the solar radiation model mentioned in the previous section, as well as the temperature input terms selection. In this model the exogenous input is the solar radiation, whose lagged terms were extracted from the previous 6 hours to the forecasting instant. The number of neurons for the models was restricted to the range from 2 to 20. The training methodology, the stopping criterion and network type were the same mentioned in the previous section. The MOGA population size was set to 100 individuals, with the introduction of 10 new individuals randomly generated. As for the results, the best long term temperature forecast produced a mean-squared error of approximately 1.5^oC on the four-hours horizon, having an asymptotic behaviour right above this value. The NN model has 19 input terms and 4 neurons in the hidden layer. As a final remark, the author mentions that the obtained result will deteriorate due to error propagation.

3. Background

3.1. Summary

This project took as a reference an industrial total sky imager used in the *Centro de Sistemas Inteligentes* (CSI) laboratory in UALg, the $TSI440A$ from Yankee Environmental Systems, Inc. (see Figure 3.1).

Figure 3.1: TSI 440A (Yan (2007))

For a brief explanation of the system it will be separated in three different parts: its structure, the logic module and the $Axis200+$ ip camera. Below is a description of the

main characteristics of the system (see Yan (2007)):

Imager resolution: 328×288, 24-bit colour;

Sampling rate: variable, with a maximum of one image every 10 seconds;

Operating temperature: $-30^oC \Leftrightarrow 34^oC$;

Weight: approximately 23kg;

Power: 110/220 VAC;

Data storage: disk on local computer or remote computer over a full-time TCP/IP connection

3.2. *TSI* Structure

The *TSI* structure has two parts (see Figure 3.3(a)), a chrome-plated steel mirror on which the whole sky is reflected on, and an arm that holds an $Axis200+$ ip camera that captures the reflection of the sky in the mirror (see Figure 3.2). The acquisition of the images is done by accessing the ip camera's web server.

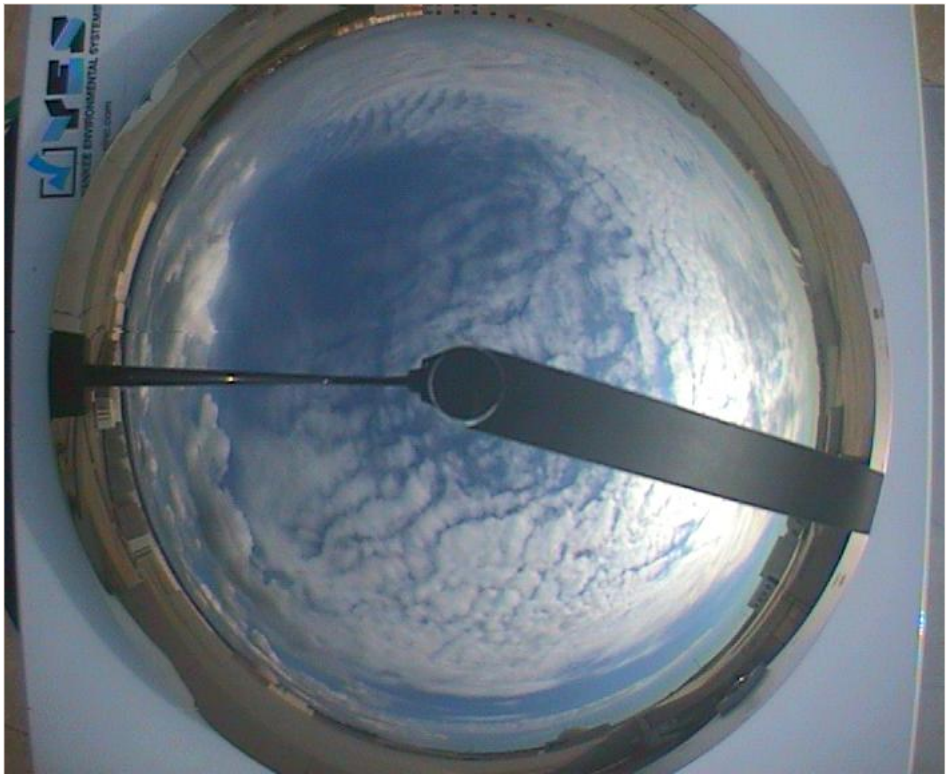
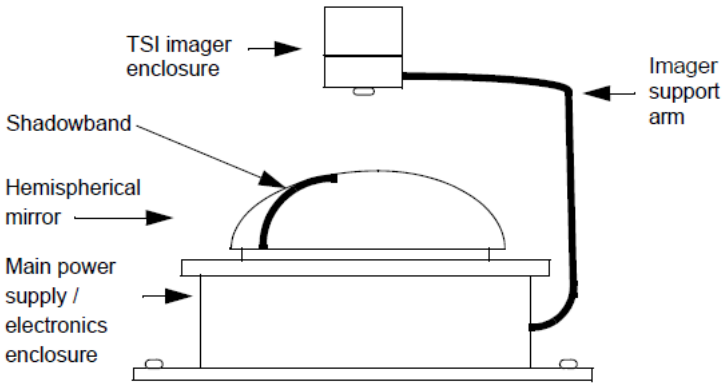


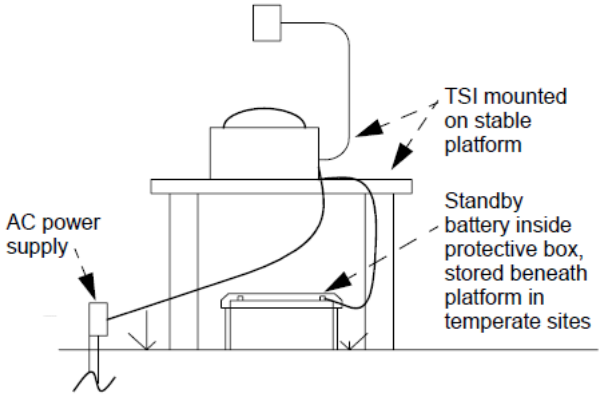
Figure 3.2: TSI 440A sky image

The assembly of the TSI involves, besides AC power, an external battery to be used

in case of power failure. The Figure 3.3(b) shows the setup used.



(a) General view



(b) Focused view

Figure 3.3: TSI components and power scheme (Yan (2007))

In order to achieve the desired results, the system must not have any visual interference. It must be placed in a high place, and even there, the responsible for its assembly must check if it suffers or not from visual interference. If it does, a software mask must be applied in order to not compromise the results.

3.3. *TSI* Logic Module

The logic module of the *TSI* was replaced by a custom made module (see Figure 3.4), due to the failure of the original one. The custom made module did not add any functionality besides the existing ones. The module is responsible for the calculation of the azimuth in degrees for the sun's position, and depending on its value, the chrome-plated

steel mirror rotates until it gets to the azimuth position. The azimuth value depends on the geographical coordinates, these must be hard-coded in its calculations.

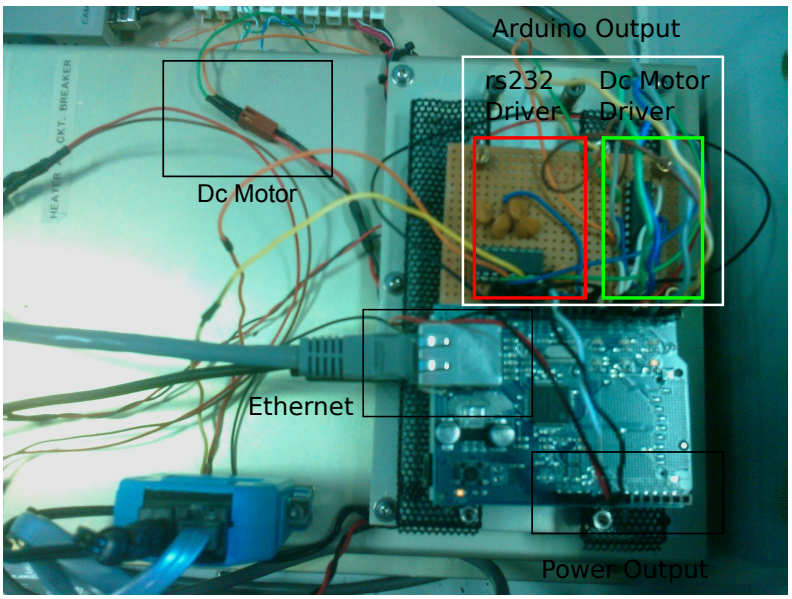


Figure 3.4: TSI custom made logic module

This system uses an encoder from US Digital to retrieve the position, and a dc motor to move the chrome-plated steel mirror. The following equipment was used to rebuild the logic module:

Arduino Duemilanove: core of the logic module;

Arduino Ethernet Shield: retrieve date and time from an NTP server;

MAX3232CPE: interface serial communications between Arduino and the US Digital encoder;

LM293E: interface the dc motor that controls the shadow band;

The logic module consists in two *C* programs: one used only for **setup**, in which the *TSI* origin is calibrated; and another for **production** purposes, where the azimuth, time/date and communications are performed. The azimuth calculations follow the study done by Carruthers D et al. (1990).

The module depends on the UTC (*Coordinated Universal Time*), which is provided by two services. The module attempts to connect the NTP server inside the CSI laboratory, if the connection fails, the module will access the camera through telnet service with the command *date*, to get the date and time. The ip camera is also synchronized by the previously mentioned NTP server.

3.4. *Axis*200+ ip camera

As already mentioned in the beginning of the chapter, the imager resolution is 328×288, 24-bit colour; and the sampling rate is variable, with a maximum of one image every 10 seconds. The ip camera has an internal server which allows the user to choose between a whole set of options related to the image (see Figure 3.5), such as its format, size, brightness, contrast and saturation. The camera has the option of choosing larger formats than its resolution, however these are software transformed.

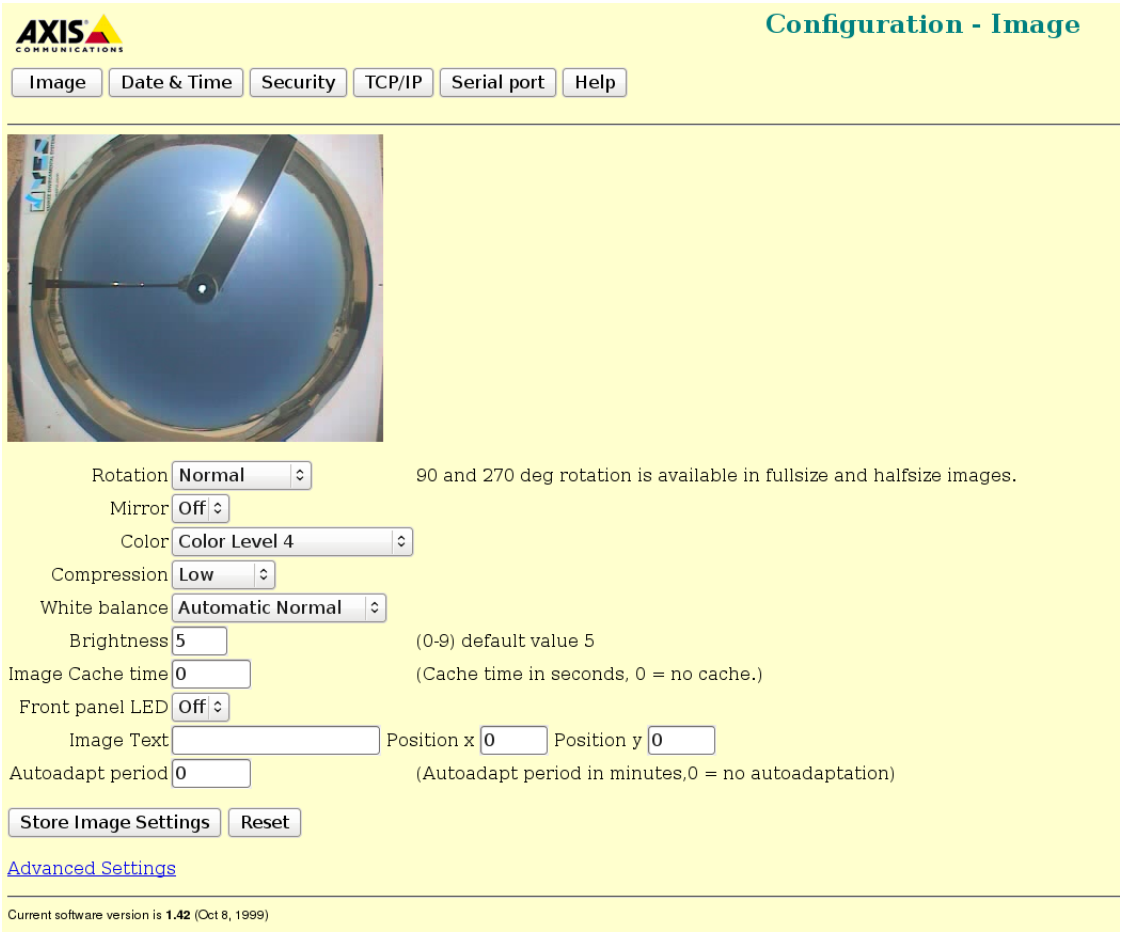


Figure 3.5: Axis 200+ configuration menu

4. Prototype

4.1. Summary

A prototype was developed to perform measurements of global solar radiation and atmospheric temperature and to take photographs of the sky, and then feed the data collected to a set of neural networks (developed by Igor António Calé Martins (2010)), in order to predict the atmospheric temperature and solar radiation in a four hours horizon.

4.2. Requirements

In order to develop a prototype small enough to carry, easy to assemble and that starts functioning automatically once powered up, some construction features had to be taken into account. As already mentioned in Chapter 2, the most relevant variables to make an estimation of cloudiness indexes, atmospheric temperature and solar radiation using ANNs, are the variables themselves.

In order to measure cloudiness indexes (Section 4.4), it is necessary to use a camera that can photograph the largest possible area of the sky. In order to be compact, the prototype was designed to take the photographs by aiming the camera at the zenith, since this option would require a smaller amount of space and does not require a special structure to fix it.

A gps (Section 4.4.2.1) was also included, since the system must be portable and must also function automatically independently of its location, so that the user does not need to insert the geographical coordinates every time the system location is changed.

To measure the atmospheric temperature (Section 4.5) and solar radiation (Section 4.6), two sensors had to be considered, that could be used for long time intervals, but still provide reliable results.

Regarding the box to contain the system itself, it had to be light enough to be carried around, have the necessary space to house the equipment, and the most important of all,

it must be resistant to all weather conditions.

And finally, for the whole system to work, a computer had to be used, powerful enough to collect the data from the sensors, process the ANNs in less than a minute and not take a huge amount of space inside the system's structure. The computer runs Linux®, which was chosen due to its portability, stability and performance.

4.3. Logic Module

The logic module is used to make the connection between the sensors and the computer which is going to implement the ANNs, but also to control the shadow band and process the data retrieved from the GPS. The device used for these tasks, is an open-source electronics prototyping platform called Arduino®.

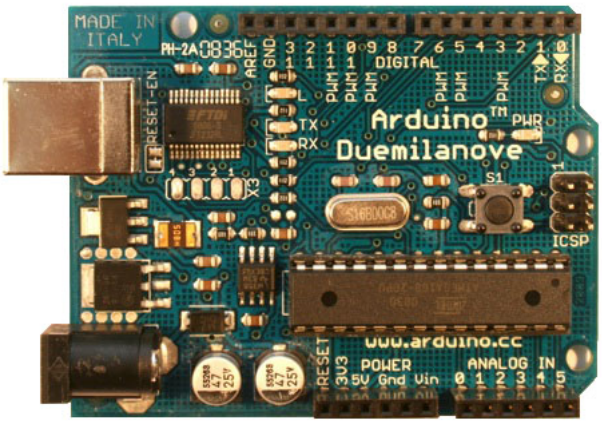


Figure 4.1: Arduino Duemilanove (image taken from http://arduino.cc)

The device was chosen since it is easy to integrate with almost any system, and also because it has the necessary characteristics to perform the control and data acquisition required by the prototype. These characteristics are: size, $6.9cm \times 5.3cm$; 14 digital input/output pins, of which 6 can be used as *Pulse Width Modulation* outputs; 6 analogue inputs with a 10 bit ADC; USB connection with serial communication between the device and its host; power pins, $5V$, $3V3$, V_{in} and GND. Figure 4.2 shows the operational flowchart of the logic module from the moment it is powered up. The main tasks are:

Serial Initialization: the serial communication is started at 9600 baud rate, 8 bit, no parity and 1 stop bit;

Set ADC Internal Reference to $1.1V$**:** configures the reference voltage used for analogue input to $1.1V$, by default it is $5V$;

Shadow Band Initialization: this is where the shadow band starts moving from its current position towards bottom, and stops once it activates the dead end switch. Finally the shadow band is moved towards to the 0^o position that corresponds to the 0^o of sun altitude (see Figure 4.9);

GPS: this routine initializes the GPS module, and retrieves the location coordinates, date and time in GMT;

Compute Sun Altitude and Azimuth: the sun altitude and the azimuth are computed using the GPS data and a sun path algorithm (see subsection 4.4.2.3 for further details);

Position: this is a position check, if the position has changed since the last check then the shadow band moves to the new current position, otherwise it does nothing.

These routines are going to be explained with more detail in the next sections, for further electrical details see Figure A.1.6.

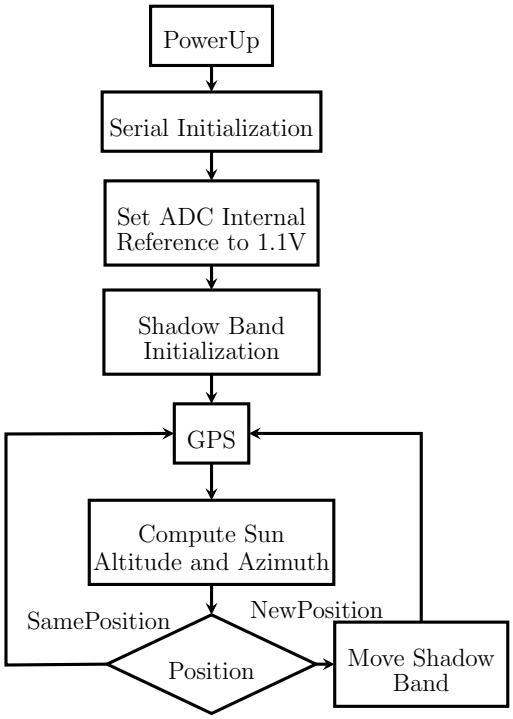


Figure 4.2: Logic module state diagram

4.4. Cloudiness Indexes

Since cloudiness indexes are based on sky images, the first approach was to take photographs aiming directly at the zenith. However this type of photography can be affected by the extreme brightness of the sun. Therefore, to test the feasibility of this option, a Baxall CDSP9372WF CCTV camera was used (see Appendix A.1.2.1), using a lens without a wide field of view.

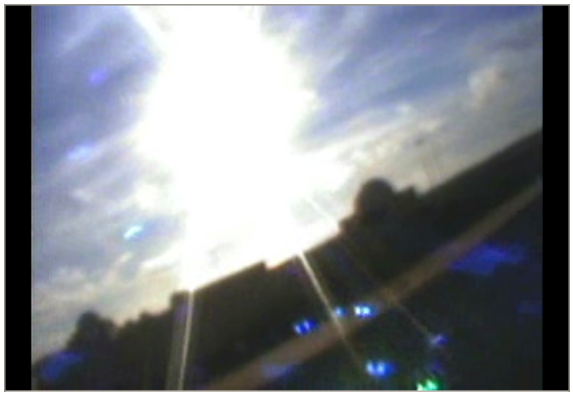


Figure 4.3: Shutter speed: 1/10,000; AGC off

Figure 4.4: Shutter speed: 1/10,000; AGC on

Figure 4.5: Auto-iris; AGC on

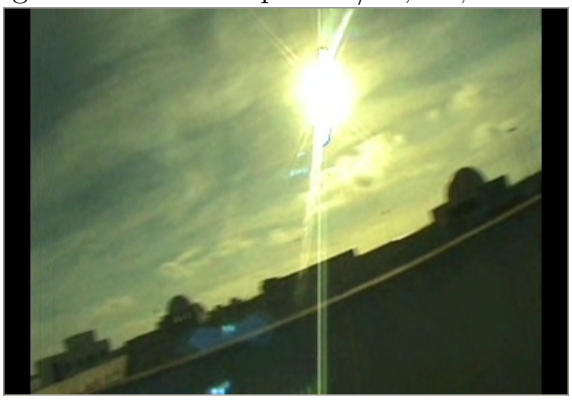


Figure 4.6: Auto-iris; AGC off

Changing between the different configuration options, Figures 4.3, 4.4, 4.5 and 4.6 were obtained. Every figure show that despite the glow of the sun, it is possible to have a good perception of what is cloud and what is not. Figures 4.4 and 4.5 are darker than the remaining, which might lead to a sky type classification mismatch in terms of computational analysis. As for the photographs in Figures 4.3 and 4.6, it is possible to see a clear and improved contrast between colours, which eases the image processing determining what is cloud and what is clear sky.

For this study, the computer system described in Appendix A was used, and two types of implementations were made: one without a shadow band and another using a shadow band that eliminates the reflection on the lens originated by the sun's flare.

4.4.1 Implementation A: without a shadow band

This implementation used the computer setup described in Appendix A. These sky images are taken automatically each minute and one mask is applied to them.

Figure 4.7 illustrates the system operation from booting onwards:

Snapshot: the routine that takes the snapshot of the sky was declared on *crontab*[1] to run every minute during the entire year;

Apply Mask: applies an external mask which covers buildings or any other visual interference that affects the components being studied;

Store Image: stores the processed image in a predefined directory, for further analysis;

Timer: triggered every minute, using linux *crontab* service;

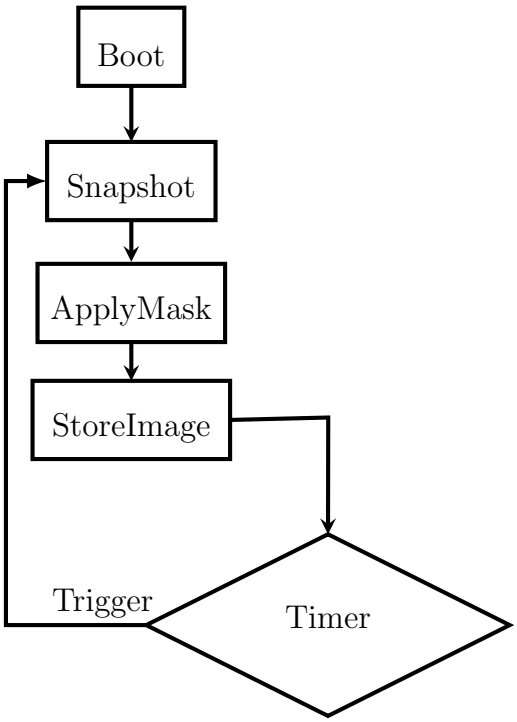


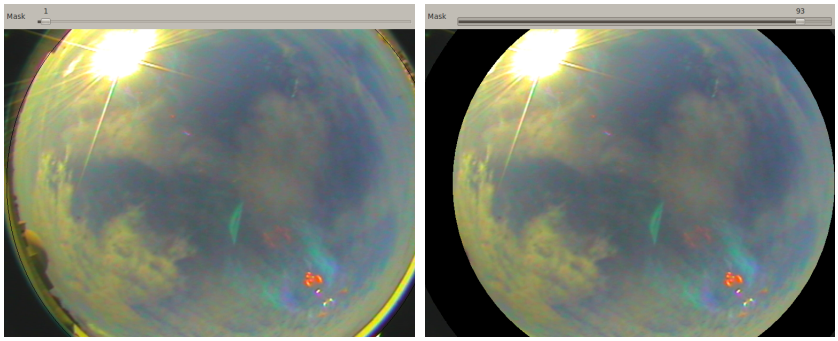
Figure 4.7: Flow chart for implementation A

The use of an external mask, has a special relevance to the system. Ideally the system

[1]Time-based job scheduler, for further information see http://crontab.org

should be positioned in a high location, without taller buildings or other structures nearby, but in urban areas this is quite impossible. This mask allows the user to define the outline for the outer mask of the image, blackening whatever might appear on the image that is not supposed to be classified as clear sky or cloud.

The mask is applied using a software application, where a circle is drawn with a thickness specified by the user (see Figure 4.8). This is done during the setup stage of the system. The user selects the thickness of the circle by moving a slider in the application's window, which is saved in a file when the application is closed, and used in the sun's mask algorithm in a production environment.



(a) Before mask definition (b) After mask definition

Figure 4.8: External mask

4.4.1.1 System testing

The implementation was tested mainly on sunny days at noon, to test its behavior when exposed to bright light conditions. These conditions were considered as most relevant, since they are one of the types of conditions that lead to many cloud misclassification. Despite the fact that the auto-iris made a difference by reducing the sun's glow, the lens caused a reflection that can be seen on Figure 4.8. This reflection was not acceptable, since it leads to a misclassification of that part of the image (i.e. clear sky to a cloud). To solve this problem, a shadow band was considered and implemented.

4.4.2 Implementation B: with a shadow band

This second implementation aims to improve the results obtained using the approach described on section 4.4.1, by adding a shadow band (see Figure 4.9) that follows the sun path position during the day, in order to eliminate the problems caused by the reflection previously mentioned.

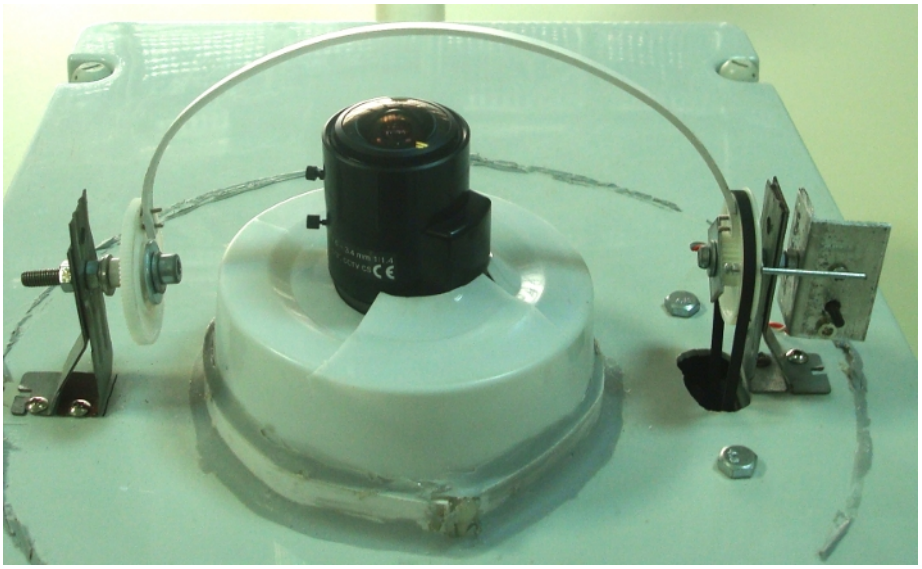


Figure 4.9: General view of the shadow band

4.4.2.1 Shadow band

To control the shadow band and its positioning, the following equipment was used: a unipolar stepper motor (see Figure 4.10), $NMBPM55L - 048 - HP69$; an *Arduino Duemilanove*; a PVC strip to make the shadow band; a CD-ROM belt, to control the shadow band positioning; two plastic discs with $4cm$ each to couple the shadow band and obtain the desired resolution; and a $GlobalSatEM - 406A$ GPS module (see Figure 4.11) coupled with the *Arduino* to obtain the geographic location of the system.

The stepper motor used has a full step resolution of $7.5^o/step$, but using a half step sequence instead, the resolution is of $3.8^o/step$. Since the cdrom belt had no teeth, and the teeth of both stepper motor hub and the $4cm$ disc did not match, its resolution was

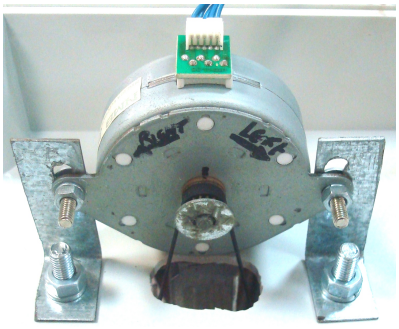


Figure 4.10: Stepper motor

determined using equation 4.1.

$$\text{Stepper motor resolution} = \frac{360^o}{\text{number of steps in a turn}} = \frac{360^o}{326 steps} \simeq 1.1^o \qquad (4.1)$$

As for the control of the motor and initialization process of the shadow band, it was done using the *Arduino*. The first action after the system boot is the shadow band initialization. Once the system has power, the band starts moving from north to south until it pushes an end switch. The switch is used to control the initialization process of the band, determining a deference point from where it will start moving. Then the band moves a predetermined number of steps until it reaches de 0^o position, that is perfectly aligned with the zenith, becoming the origin for the upcoming steps.

4.4.2.2 Geographic positioning

The next step of the system initialization depends on the GPS. This is one of the parts that can take more time to initialize, because it has to wait for reliable data from the satellites. Therefore it might take a while depending on the amount of satellites available for connection. The amount of time it takes to connect and retrieve its geographical position and UTC time, depends also on the systems location. It is important to point out, that during tests on our location this procedure was actually very fast.

Once the device is connected and has a *fix*, the system continues its normal flow. The scheme below, shows the string retrieved from the GPS, and a brief explanation of its fields. The geographic location and UTC time and date, will then be used as inputs to the sun path algorithm.

Figure 4.11: GPS

GPRMC,	225446,	A,	4916.45,	N,	12311.12,	W,	0.5,	54.7,	191194,	20.3,	E	*68
	1	2	3	4	5	6	7	8	9	10	11	12

1 : UTC time of fix

2 : Data status (A=Valid position, V=navigation receiver warning)

3 : Latitude of fix

4 : N or S of longitude

5 : Longitude of fix

6 : E or W of longitude

7 : Speed over ground in knots

8 : True course

9 : UTC date of fix

10: Magnetic variation degrees (Easterly variation subtracts from true course)

11: E or W of magnetic variation

12: Checksum

4.4.2.3 Sun path calibration

The sun path calibration was computed using an algorithm based on the study of Carruthers D et al. (1990), which takes as input: time, date, latitude and longitude; and outputs the azimuth, solar declination, solar altitude (in degrees), sunrise time and sunset time. With the calculated data, it is possible to make a stereographic projection of the sun's path (see Figure 4.12). By plotting this projection on top of a ground sky image with the coordinates perfectly aligned, it becomes possible to see the sun's path in that specific image.

Analysing Figure 4.12, the vertical axis gives the altitude of the sun. It's altitude starts on 0^o and ends on 90^o. The circle represents the angle of the azimuth, its value vary from 0^o to 360^o.

Contrary to the system described in chapter 3, the solution to the positioning problem of the prototype depends on both the altitude and the azimuth of the sun.

Figure 4.13 represents the base model used to calibrate the shadow band, it is important to consider that the shadow band is displaced along the *yy's* axis, and only moves along the *xx's* axis. Figure 4.13 has a representation of the shadow band on the point where the sun is intercepted. On the model, ϕ represents the altitude of the sun, θ represents the azimuth of the sun,r represents the radius of the shadow band and the black dot represents the point where the sun intercepts the shadow band.

Using the model in Figure 4.13, a conversion from polar to Cartesian coordinates was made, which is represented in Figure 4.14. This represents the triangulation of the three points ABC, which results on the angle (β) that the shadow band has to move to intercept the sun. By computing this angle , the number of steps is computed using equations 4.1 and 4.2

$$\text{Number of steps} = \beta/1.1 \tag{4.2}$$

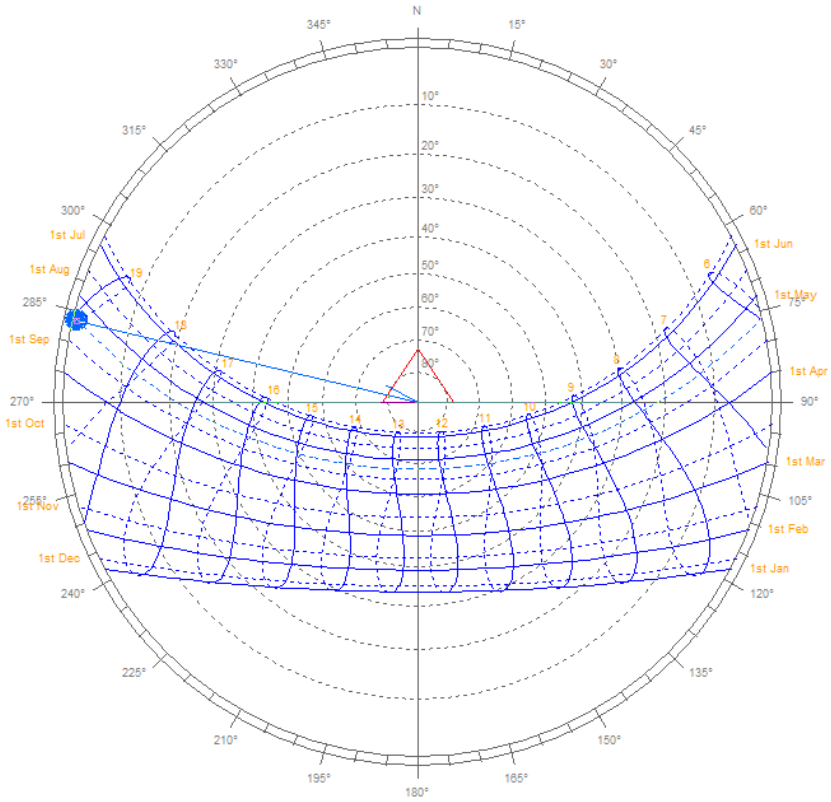


Figure 4.12: Stereographic projection of the sun path for Faro(lat. 37^o,lon. -7.6^o), obtained using a trial version of *Solar Tool 2011, Autodesk Inc. 2010*

Figure 4.15 has the result of the triangulation obtained in Figure 4.14. The equations 4.3 where used to convert the coordinates from polar to Cartesian, and finally to determine the angle β that the shadow band has to move.

$$\begin{aligned} radius &= 8\ cm \\ x &= radius \cdot \cos(\theta) \cdot \sin(\phi) \\ y &= radius \cdot \sin(\theta) \cdot \sin(\phi) \\ z &= radius \cdot \cos(\phi) \\ \beta &= \arctan(x/z) \end{aligned} \tag{4.3}$$

4.4.2.4 System testing

The shadow band was tested in the same conditions as those described in subsection 4.4.1.1. As it may be seen in figure 4.16, the reflex that appeared before was eliminated. As already mentioned there are four different masks applied to each image. Figure 4.16(a)

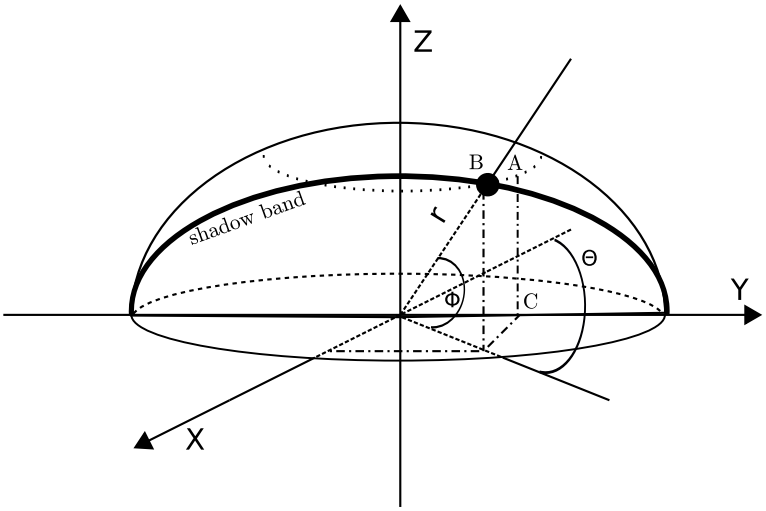


Figure 4.13: Spherical representation of point where the sun intercepts the shadow band

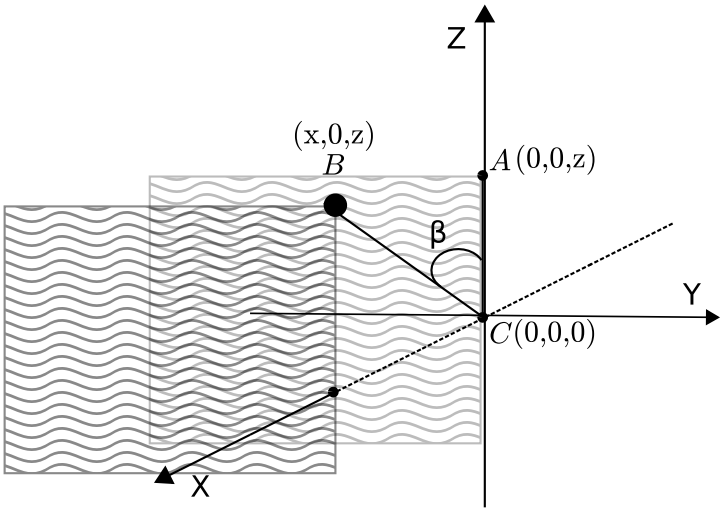


Figure 4.14: Cartesian coordinates representation of Figure 4.13

shows the photo as it is taken, without any mask. There are two types of external masks applied: the first, represented in Figure 4.16(b), was used to mark the unused part of the photo; and an optional external mask used to eliminate the visual interference in the photo, represented in Figure 4.16(d). The shadow band mask is applied dynamically depending on the altitude of the sun in each photo, see Figure 4.16(c).

Figure 4.16(d) shows the complete mask once all masks are applied. The black part is discarded while the remaining pixels of the image will be analysed. The masks were developed using the python library of OpenCV®[2].

[2]**Open Source Computer Vision**, for further information see http://opencv.willowgarage.com/wiki/

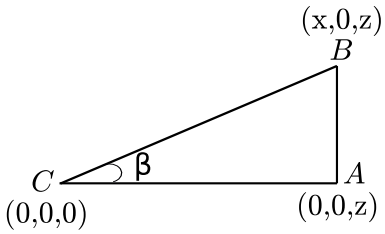


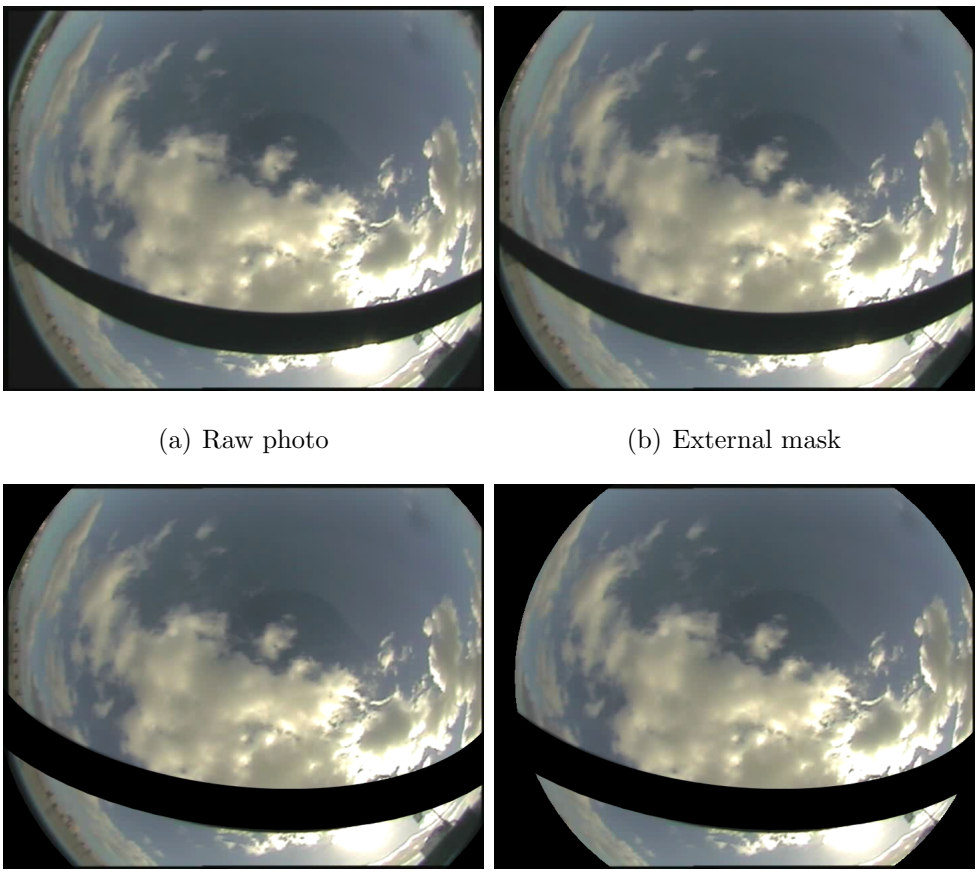
Figure 4.15: Representative triangle of the shadow's band angle

The software shadow band was computed using the same principle described in equations 4.3 and on Figure 4.12. First it is necessary to obtain β with equations 4.3. As previously explained for Figure 4.12, the value for the sun altitude is between 0^o and 90^o. The centre of the image corresponds to the 0^o, while the top and bottom part of the image corresponds to the 90^o of sun altitude. Once the the value of β is computed, it has to be converted from degrees to pixels. In order to do so, it is necessary to employ the height of the image and the vertical field of view of the lens (126^o). The multiplication of the angle of the shadow band by half of the image height, and division of the result by half of the lens field of view, gives the position of the shadow band in pixels. The reason of using half of the image height and half of the lens field of view, is due to the fact that the sun only appears on the lower half part of the image and lens.

At the end one final correction is made if necessary. The lower position of the shadow band is always 260 pixels, so if by any change the value of the altitude is bigger than 260 pixels the altitude is corrected.

Listing 4.1: Conversion of β to pixels

```
def alt2px(altitude,azimuth,imageHeight):
  radius = 8
  x = radius * cos(azimuth*pi/180)*sin(altitude*pi/180)
  y = radius * sin(azimuth*pi/180)*sin(altitude*pi/180)
  z = radius * cos(altitude*pi/180)
  altitude = arctan(x/z)*180/pi
  altitude = (90 - altitude)
  altitude = (altitude*(imageHeight/2))/63
  if altitude > 260:
    altitude = 260
  return int(altitude)
```



(a) Raw photo

(b) External mask

(c) Shadow band mask

(d) Final photo with optional external mask

Figure 4.16: Used masks

4.5. Temperature

The atmospheric temperature is acquired using an LM35DZ sensor, which is a precision centigrade temperature sensor that linearly translates temperature into millivolts. The LM35 was designed for routine measurement of temperature and has the following characteristics:

- Calibrated directly in Celsius (Centigrade)
- $+2^oC$ to $+150^oC$
- Linear + 10.0 $mV/^oC$ scale factor
- 0.5^oC accuracy guaranteeable (at $+25^oC$)
- Suitable for remote applications
- Low cost due to wafer-level trimming
- Operates from 4 to 30 volts

- Less than 60 μA current drain
- Low self-heating, 0.08^oC in still air
- Nonlinearity only $\pm 1/4^oC$ typical
- Low impedance output, 0.1Ω for 1 mA load

The *Arduino* has a 10-bit analogue to digital converter, which by default, maps the input voltage between 0 and $5V$, into integer values between 0 and 1023, allowing a resolution of $\frac{5V}{1024} \backsimeq 4.9mV$. This resolution was improved, by setting an $1.1V$ *Internal Reference* in *Arduino*, to $\frac{1.1V}{1024} \backsimeq 1.074mV$. Figure A.1.6 shows how the sensor was connected, and equation 4.4 shows the volts to Celsius conversion.

$$\text{Temperature in Celsius} = \frac{\text{ADCvalue} \times 1.1}{1023 \times 0.01} + 2.0 \tag{4.4}$$

4.6. Solar radiation

The solar radiation will be acquired using a pyranometer from Kipp & Zonen®, the SP Lite2 (Figure 4.17), that was designed for routine measurement of solar radiation, and has the following characteristics:

Spectral range: 400 to 1100 nm

Sensitivity: $71\mu V/W/m^2$

Response time: $<< 1s$

Directional error (Up to 80^o with 1000 W/m^2 beam): $\pm 50W/m^2$

Temperature dependence: $\pm 0.1\%/K$

Operating temperature range: $-30^oCto + 70^oC$

Maximum solar irradiance: 2000 W/m^2

Field of view: 180^o

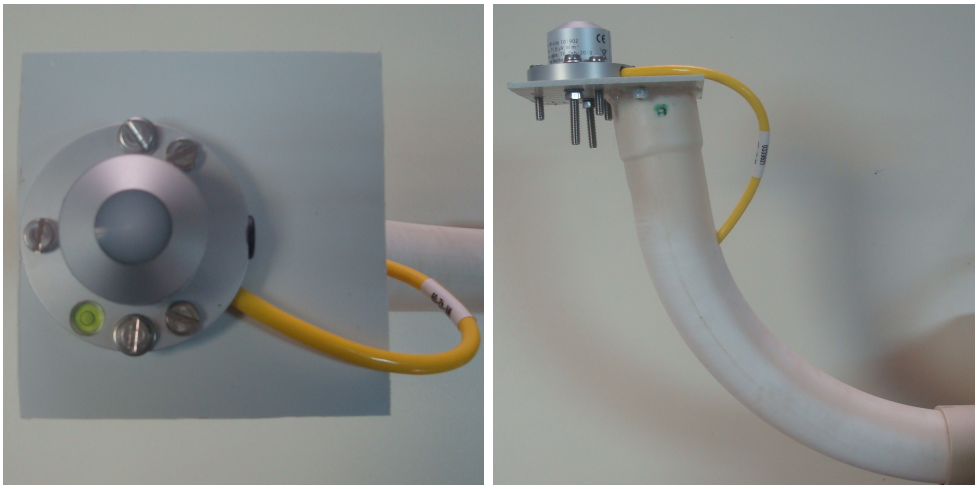
The output range of the pyranometer does not allow a direct connection to the Arduino®, since its output is between $0 - 15mV$, and its sensitivity is $71\mu V/W/m^2$. The 10 bit ADC has a default input range of $0 - 1.1V$ and a sensitivity of $1.074mV$, considering a reference of $1.1V$. Comparing the output of pyranometer with the input of the ADC, it is possible to conclude that the 10 bit ADC will only detect changes if they are bigger

that $1.074mV$, which means that the ADC will only read approximately 211 different values. If the default reference of $5V$ was used, the ADC would read only approximately 14 different values.

In order to overcome the lack of sensitivity of the ADC, an instrumentation amplifier was used. The gain of the device is set using an external resistor and can be from 1 to 1000, in this case a gain of 10 was used ny means of a resistor of $10k\Omega$. The amplification of the pyranometer output signal, sets the μV variations to mV, which can be read by the ADC. Considering the output range of the pyranometer, the expected output of the instrumentation amplifier range from 0 to $150mV$. To adjust the output range of the pyranometer in order to be fully readable by the Arduino® ADC, a voltage divider was added (see Figure 4.6). This adjustment also corrects the incoming range of the amplifier to be between $0V$ and $1V$, since the Arduino® was set with a reference of $1.1V$ it cannot read values above this.

Figure A.1 shows the connections diagram between the instrumentation amplifier and the Arduino®. As for the solar irradiation, it will be computed using equation 4.5, that converts the ADC value into V.

$$\text{Irradiance} = \frac{\text{ADC value} \times 1.1V}{1023 \times 71\mu V} \tag{4.5}$$



(a) Top View (b) Side View

Figure 4.17: Pyranometer

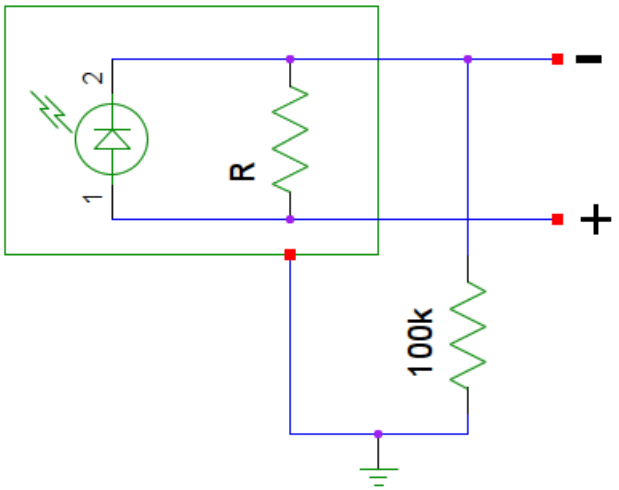


Figure 4.18: Pyranometer impedance adjustment circuit

4.7. Neural networks

The neural network set was identified using the multi-objective genetic algorithm (C.M. Fonseca (1995)) to optimize their structure in order to develop a non-linear autoregressive with exogenous input model type of neural network (NARX). The networks were trained using a retro-propagation algorithm based on Levenberg-Marquardt's optimisation algorithm (P. M. Ferreira and A. E. Ruano (2000);Ferreira et al. (2002)) with early stopping as a stopping criteria.

Figure 4.19 shows the conceptual model of the neural networks prediction scheme. The images are processed by the *ANN Cloudiness Estimator* to estimate the amount of cloudiness of each image. The cloudiness along with the temperature and solar radiation measurements, are fed into the respective neural network models in order to make a prediction up to 48 steps ahead. Since each step is five minutes, 48 steps correspond to a four hours horizon.

The complete cascade of models consist of: one NAR cloudiness predictive model; one NARX solar radiation predictive model, using cloudiness as the exogenous input; and a NARX temperature model, using solar radiation as the exogenous input.

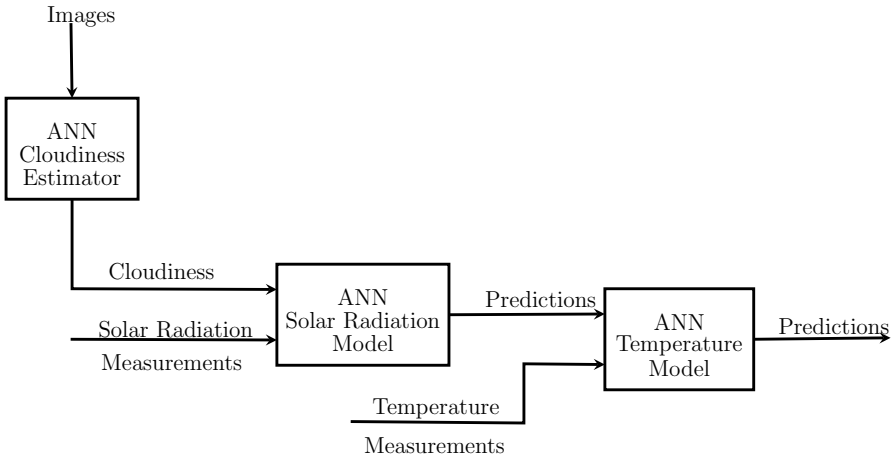


Figure 4.19: Neural network model

Perhaps the most significant drawback of the use of ANNs, is the fact that the use of this technology at different locations requires a previous training of the network for that specific location.

4.8. System's routines

The system's routines are divided in three different scripts, as shown in Figure 4.20. Initially these scripts were meant to run at each minute, but since the script ***IACE*** takes almost a minute to run due to the camera capture device driver, the frequency of the scripts had to be changed. In the initial implementation the images were taken every minute, and five minute averages were made for further computation of the predictions. In order to overcome the camera device driver problem and ensure that all scripts finish their execution before the next image is taken, the frequency was set to 5 minutes. This problem was detected when the second implementation was tested.

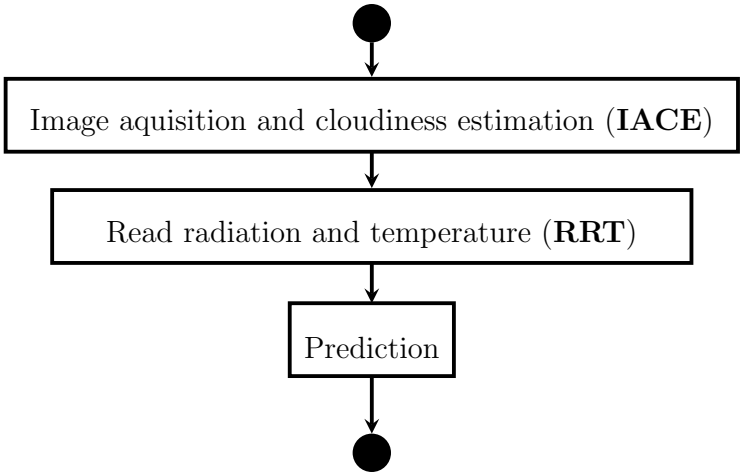


Figure 4.20: Main script

The ***IACE*** module, has the tasks of acquiring sky photos and estimate the percentage of clear sky associated with each photo. The algorithm was set to work only during daylight hours, for the remaining hours of the day (during night), it was established that the clear sky value is zero.

The ***RRT*** module, loads the retrieved data from the temperature and radiation sensor, and performs a validation check to the values. Contrary to what is done with the images, these values are collected during the whole day.

The ***Prediction*** module, makes the predictions of cloudiness, radiation and temperature. Initially the module loads the neural networks structure, and the required previously acquired values of clear sky, radiation and temperature. The NAR network is triggered once there are 24 hours of datum of cloudiness. The $NARX$ networks are triggered under

the same conditions as the *NAR* network, but it also require the datum of temperature and solar radiation. The *NARX* makes two predictions for temperature: one where the exogenous variable is the radiation provided by a NAR model, and another where the exogenous variable is the radiation provided by a NARX model.

4.9. Housing

The structure of the system is composed by two parts (see Figure 4.21): the system itself ($70cm \times 30cm \times 30cm$) and an external structure ($60cm \times 50cm \times 20cm$). The external structure was used to ease the fix of the system to any surface, but also to shadow the system to prevent overheating from direct sunlight exposure in hotter days.

Figure 4.21: System

5. Results

In this chapter, results are presented for system validation, which was done in three different ways:

- the first validation consisted on comparing the results of the prototype system with those obtained on a CSI Laboratory computer, using the same data. This aims to validate the prototype implementation regarding the models.
- on the second validation the results obtained were compared to those obtained by the reference system. The cloudiness was compared with the system described in Chapter 3, and its validation results are presented in Section 5.1. Temperature and solar radiation results were compared with those acquired by a weather station, incorporating a pyranometer (BF3 from $\Delta - T^{®}$) and a temperature sensor (RHT2 from $\Delta - T^{®}$). The solar radiation validation results are presented in Section 5.2, and the validation results regarding temperature are presented in Section 5.3. This aims to validate the chosen sensors and the acquisition hardware.
- the third validation is a comparison between the predictions obtained by means of images acquired by the prototype, and the predictions obtained using images acquired by the *TSI* (see Chapter 3).

The data used in the validation was acquired between the 28^{th} and the 30^{th} of January of 2011. The first day was used to initialize both NAR and NARX models, while the two remaining days where used for comparison of predictions.

In the following sections the developed prototype will be referred to as *CloudSpotter*.

5.1. Cloudiness Indexes

The first validation was performed by analysing the set of images acquired by the *CloudSpotter*, for the period of analysis, in a computer of the CSI laboratory, and by measuring the maximum absolute error between the the estimation of cloudiness computed in the laboratory computer and the *CloudSpotter*. As for the result, a maximum absolute error of approximately $5.0e^{-14}$ was obtained. This validation was also applied to the NAR model used for prediction of cloudiness, using data acquired by the *CloudSpotter*. The validation consisted on the comparison of the maximum absolute errors obtained when comparing the predictions computed by the *CloudSpotter* model implementation, to those computed by the laboratory computer, considering the 1 step ahead prediction. The maximum error obtained was $7.1e^{-15}$.

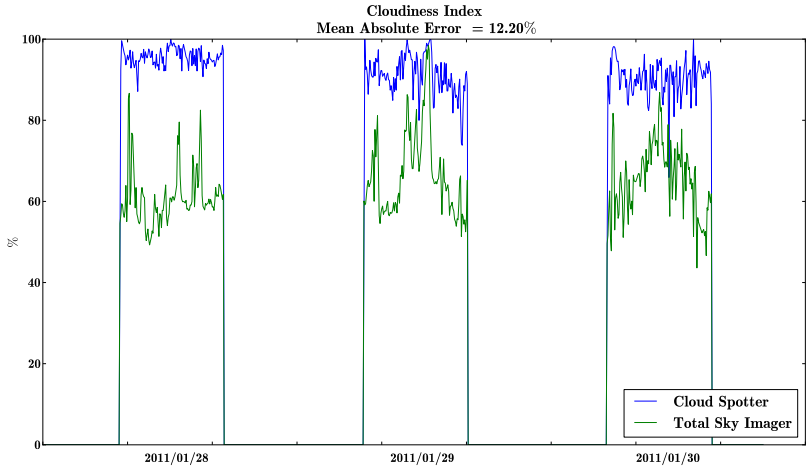


Figure 5.1: Estimated cloudiness indexes

The second validation consisted on the comparison of the cloudiness estimated by means of *CloudSpotter* images to that estimated by images acquired using the *TSI*, for the same period of time (see Figure 5.1). The mean absolute error between the two systems is 12.20% with a variance of 15.39%.The difference observed between the two systems, is due to the fact that the neural network employed was trained to provide a threshold using images from the *TSI*. When this threshold is applied to images with a different tonality, as those acquired by the *CloudSpotter*, it induces an error in the cloudiness estimation (see Figures 5.2, 5.3, 5.4 and 5.5).

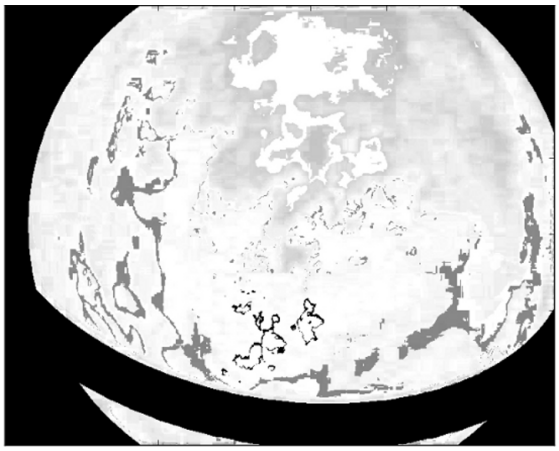


Figure 5.2: *CloudSpotter* with threshold

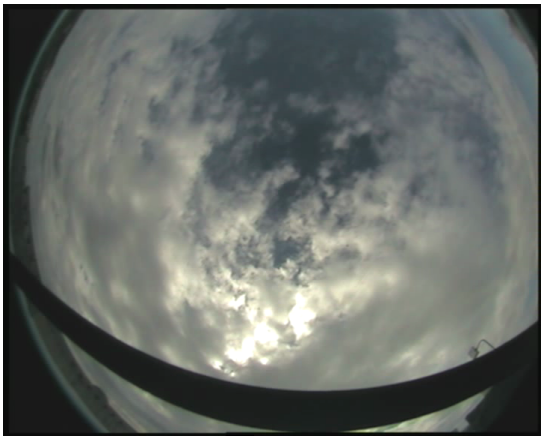


Figure 5.3: *CloudSpotter* original image

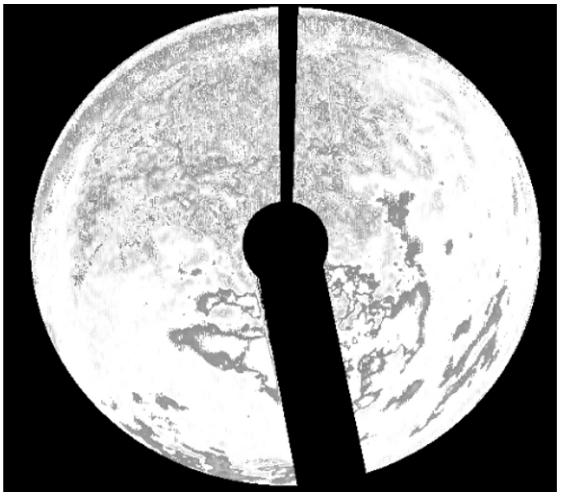


Figure 5.4: *TSI* with threshold

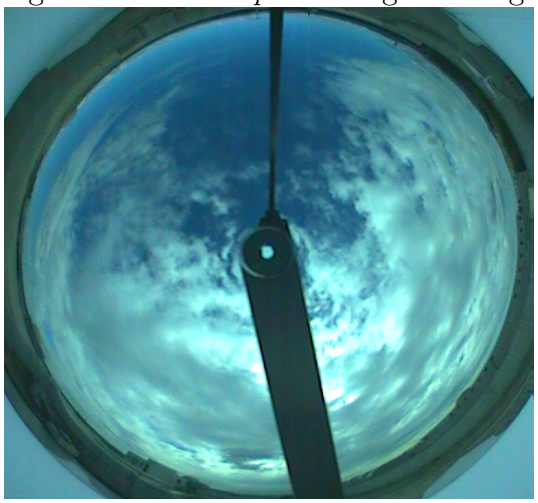


Figure 5.5: *TSI* original image

In Figures 5.2, 5.3, 5.4 and 5.5, it is possible to observe the difference in applying the same threshold to images acquired by both systems. As already mentioned, the *CloudSpotter* images have a different tonality than those of the *TSI*, which translated into a cloud misclassification when the image is processed. Comparing Figures 5.2 and 5.3, is possible to see an area of clear sky on the top of de original image that was translated into cloud after the threshold was applied, as well as other areas that are cloud on the original image and were translated into clear sky. This situation does not happen with the images of the *TSI*, because the neural network was trained with its images.

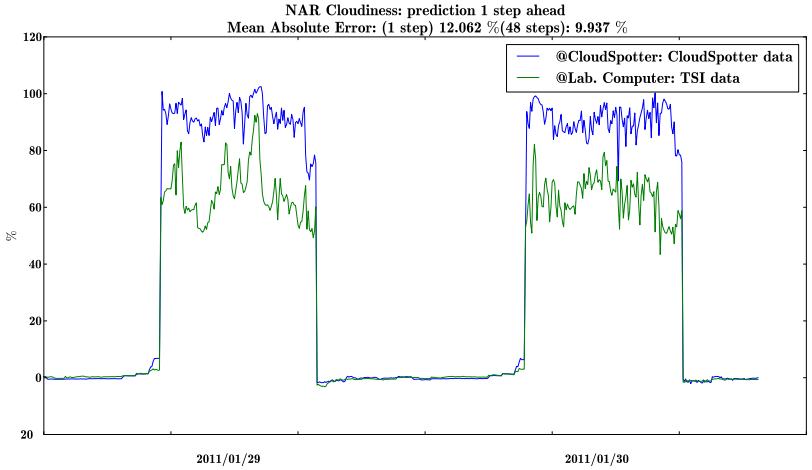


Figure 5.6: NAR cloudiness indexes prediction

The results of the third validation are shown in Figure 5.6, which presents a comparison of the 1 step-ahead (5 minutes) predictions obtained by the cloudiness NAR predictive model, when using data from the *TSI* and from the *CloudSpotter*. The mean absolute error between the two systems is, for 1 step ahead 12.062% with a variance of 14.156% and for 48 steps ahead 9.937% with a variance of 10.65%, which are in accordance with the error obtained between the cloudiness estimations(see Figure 5.1).

5.2. Solar radiation

The first validation was performed by comparing the maximum absolute errors for the NAR and NARX (cloudiness is the exogenous variable) prediction models for solar radiation. The comparison was made between the the 1 step-ahead predictions computed in the *CloudSpotter* and corresponding predictions computed in a laboratory computer, for the previously mentioned time period. This validation was made with data acquired by the SPLite2 from *Kipp & Zonnen*®, that was employed in the *CloudSpotter*. For the NAR prediction the maximum absolute error was $4.5e^{-13}$, and for the NARX prediction the maximum absolute error was $8.9e^{-16}$.

The second validation was performed by comparing the measurement from the two sensors, the SPLite2 and the BF-3 from $\Delta - T$®. The results of this validation are shown in Figure 5.7. The mean absolute error between the two sensors is $16.525 W/m^2$ with a variance of $49.182 W/m^2$. Another important observation, is the fact that the dynamics of the two sensors are very alike, as expected.

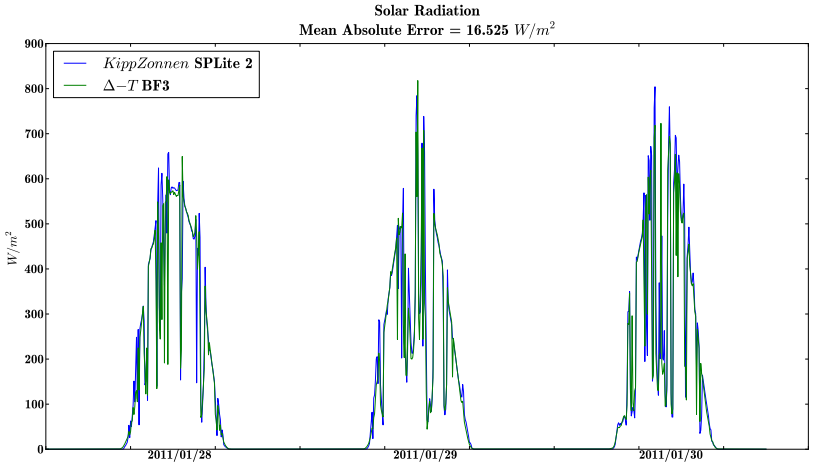


Figure 5.7: Measured solar radiation

The results of the third validation are presented in Figures 5.8 and 5.9. The first shows the results of the NAR predictive model of the solar radiation NAR predictive model. This was performed using the datum of solar radiation acquired and computed by the *CloudSpotter*, in comparison to the datum acquired by the BF-3 and computed in a laboratory computer. The mean absolute error between the two systems is not significant and decreases when the number of steps is increased. For 1 step-ahead it is $15.874 W/m^2$

with a variance of $35.515W/m^2$, and for 48 steps-ahead it is $8.816W/m^2$ with a variance of $6.111W/m^2$.

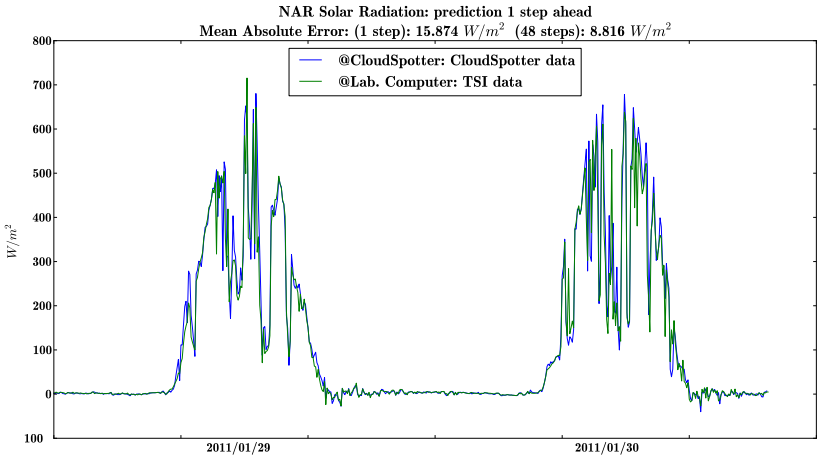


Figure 5.8: NAR solar radiation prediction

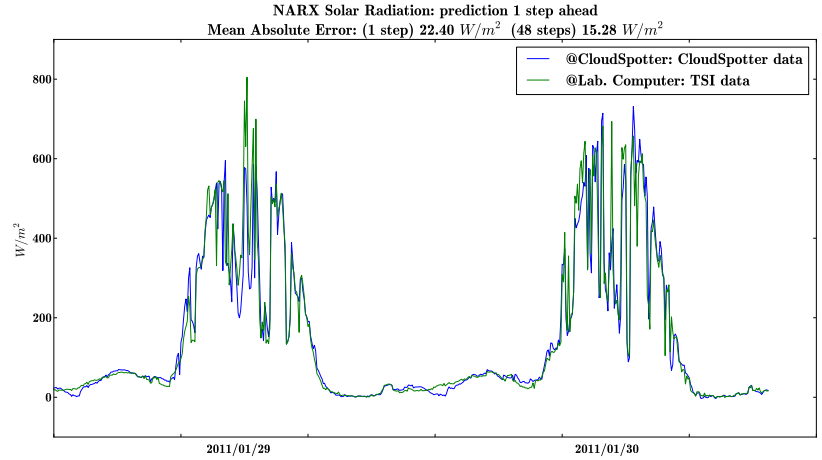


Figure 5.9: NARX solar radiation prediction

Figure 5.9 shows the predictions made by solar radiation NARX prediction model. As before, the validation was made by the prediction computed using the datum of solar radiation acquired and processed by the *CloudSpotter* in comparison to the prediction computed using the datum acquired by the BF-3 and processed in a laboratory computer. The mean absolute error between the two systems increased when compared with the NAR predictions. For 1 step-ahead and 48 steps-ahead, $22.40W/m^2$ with a variance of $42.48W/m^2$ and $15.28W/m^2$ with a variance of $19.40W/m^2$, were achieved respectively.

Most certainly, these errors are partially induced by the error of the exogenous variable predictions, that are employed in the NARX approach.

5.3. Temperature

The first validation was performed by comparing the maximum absolute errors for the two NARX temperature predictive models, one where the radiation was provided by a NAR model and the other, where the radiation was provided by a NARX model, cloudiness being the exogenous variable. The comparison was between the model predictions computed in the *CloudSpotter*, and the corresponding predictions computed in the laboratory computer. This validation as performed using the the data acquired by the LM35DZ temperature sensor from *National Semiconductor*®, that was employed in the *CloudSpotter*, considering the 1 step-ahead prediction. For the case where the radiation was provided by a NAR model the maximum absolute error was $3.2e^{-15}$. For the the other case, where the cloudiness was the exogenous variable, the maximum absolute error was $1.8e^{-15}$.

The second validation was made by comparing the measurements of two sensors employed. The results are presented in Figure 5.10.

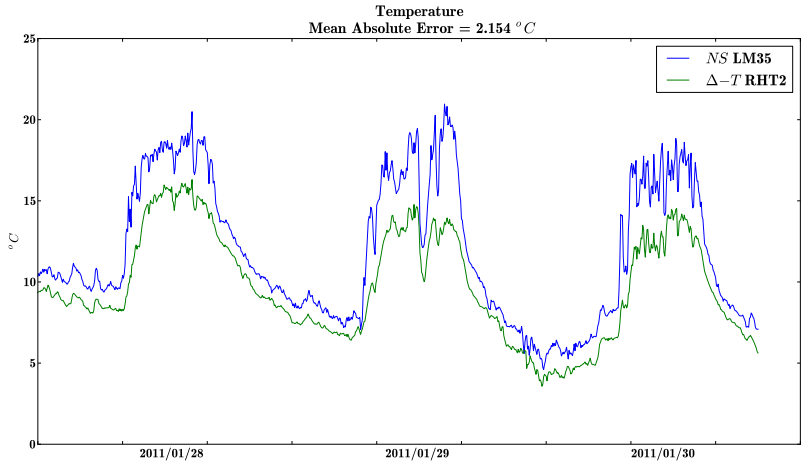


Figure 5.10: Measured temperature

As it can be observed, there is an offset between the measurements, which translates into a mean absolute error of 2.154^oC and a variance of 1.371^oC. The observed offset between the sensors, might be explained by inappropriate housing of the LM35DZ sensor in the prototype, which was too small and probably providing insufficient air flow. Part of this offset, would probably be removed if the two sensors would have been calibrated at the same time using the same references. Nevertheless, the observed patterns are very

alike suggesting that the LM35DZ can be used in future designs provided that proper calibration is performed and adequate housing is built.

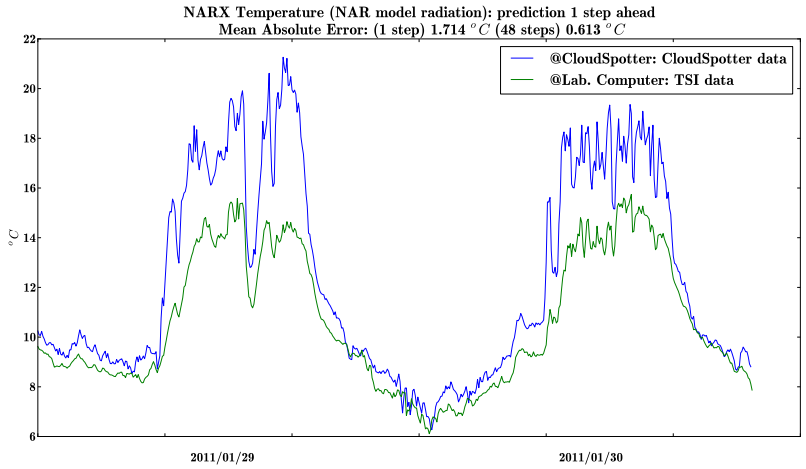


Figure 5.11: NAR temperature prediction

The third validation is presented in Figures 5.11 and 5.12. The first shows the prediction computed by the NARX model of temperature where the radiation was provided by a NAR model. The plot shows the model prediction computed using the datum of temperature acquired and processed by the *CloudSpotter*, in comparison to the prediction computed by means of the datum acquired by the RHT2 and processed in a laboratory computer. The mean absolute error between the two systems, considering 1 step-ahead is 1.714^oC with a variance of 1.468^oC and for 48 steps-ahead is 0.613^oC with a variance of 0.109^oC.

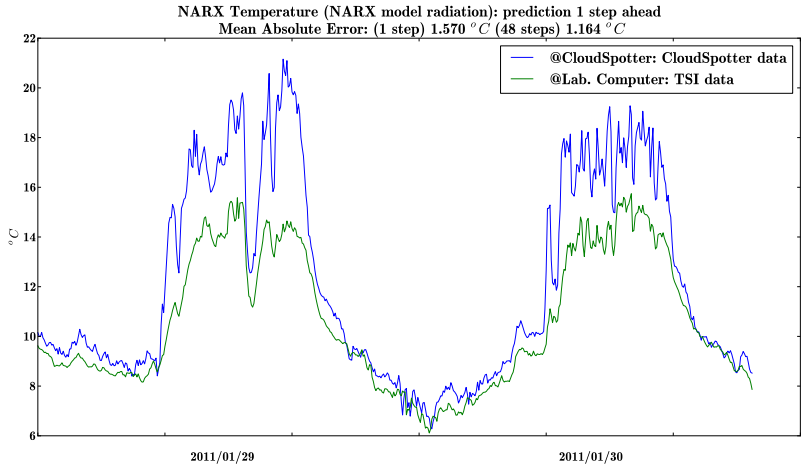


Figure 5.12: NARX temperature prediction

Figure 5.12 shows the prediction computed by the temperature model where solar radiation is provided by means of an exogenous variable, the cloudiness. As before, the prediction computed using the datum of temperature acquired and processed in the *CloudSpotter*, is compared to the prediction computed using the datum acquired by the RHT2 and processed in a laboratory computer. The mean absolute error between the two systems, considering 1 step-ahead is 1.570^oC with a variance of 1.536^oC and for 48 steps-ahead is 1.164^oC with a variance of 0.935^oC. These larger error values, as for the solar radiation case, are probably due to the inclusion of the NAR cloudiness predictive model as an exogenous input to the solar radiation model, which in turn was employed as an exogenous input in one of the temperature predictive models.

6. Conclusion

Apart from the estimation of cloudiness, which still needs improvements, the other functionalities of the implemented prototype were successfully completed, as it may be concluded by analysing the results obtained. The cloudiness estimation neural network was trained with images from the *TSI*, and since the cameras of the *TSI* and the prototype have distinct characteristics, the difference of results obtained was expected. This result influences the prediction of the radiation by means of NARX models, because cloudiness indexes are the exogenous variable on these models. In addiction to improvements in the estimation of cloudiness, the quality of the images acquired by the prototype may also be improved. This should be accomplished by changing the camera protective dome material in order to decrease undesired reflections.

Regarding the measurement and prediction of air temperature, the results obtained were highly satisfactory. To this respect, the prototype may only be improved by changing the temperature sensor housing to a more appropriate one. This should eliminate the measured offset that was obtained.

The measurement of solar radiation had excellent results, when comparing the two sensors used. The mean absolute error between the two is small enough to be explained by the simple fact that the sensors are different, and therefore have different operation characteristics.

Importantly, as a final remark, the implemented system is portable, easy to assemble, easy to setup and operates properly under all weather conditions.

6.1. Future Work

Although the prototype implemented operates satisfactory, it may easily be improved regarding some of its aspects:

Camera protection dome: the dome must be replaced by one that does not reflect the inside part of the camera housing;

Cloudiness estimation: regarding this aspect two actions may be considered. The first consists in fine tuning the camera, to change the colour setup so it resembles the one of the *TSI*. The second option consists in training one neural network with images captured by the prototype.

Shadow band system: although the shadow band worked perfectly, its durability is not long enough due to the materials employed. It could be replaced by one that does not depend on rubber bands, like a gear system or servo motors directly coupled to the shadow band.

Temperature sensor housing: the housing should be changed to a larger one, also providing increased air flow. This should eliminate the observed offset.

Power supply: this was not investigated in this study. A battery and solar panel system should be designed.

System cooling: the system might need increased cooling, if it is going to be used in extremely hot environments. The cooling system must still adequately seal the system from the external environment.

A great improvement in the sky imaging acquisition, would be the use of a magnetic fields to control the position of a metallic dot on a surface, whose location on that surface would shade the sun in the acquired image. Using the same or a similar algorithm to calculate the position of the sun, azimuth and altitude, would be simple enough to know the exact position of the sun in the used reference system. This improvement would increase the usable area of the image to its possible maximum, by excluding only the sun.

7. References

A. Cazorla, F.J. Olmo, and L. Alados-Arboledas. Using a Sky Imager for Aerosol Characterization. *Atmospheric Environment*, 2007.

A. Orsini, C. Tomasi, and F. Calzolari. Cloud Cover Classification Through Simultaneous Ground Based Measurements Of Solar And Infrared Radiation. *Atmospheric Research*, 2001.

A.E. Ruano, E.M. Crispim, E.Z.E. Conceição, and M.M.J.R. Lúcio. Prediction Of Building's Temperature Using Neural Networks Models. *Energy and Buildings*, 2005.

Hirotugu Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Contro*, 19(6):716–723, 1974.

A. Ångström. The parameters of atmospheric turbidity. *Tellus*, 1964.

Arnout Feijt, Paul De Valk, and Sibbo Van Der Veen. Cloud Detection Using Meteosat Imagery and Numerical Weather Prediction Model Data. *American Meteorological Society*, 2000.

Atsu S. S. Dorvlo, Joseph A. Jervase, and Ali Al-Lawati. Solar Radiation Estimation Using Artificial Neural Networks. *Applied Energy*, 2002.

B. H. Kahn, M. T. Chahine, G. L. Stephens, G. G. Mace, R. T. Marchand, Z. Wang, C. D. Barnet, A. Eldering, R. E. Holz, R. E. Kuehn, and D. G. Vane. Cloud Type Comparisons of AIRS, CloudSat, and CALIPSO Cloud Height and Amount. *Atmospheric Chemistry and Physics*, 2008.

Buch, K.A., C.H. Sun, and L. R. Thorne. Cloud Classification Using Whole Sky Imager Data. *Proc. Fifth Atmospheric Radiation Measurement Science Team Meeting, San Diego, CA, U.S. Dept. of Energy, 35 39,*, 1995.

C. Fonseca and P. Fleming. Multiobjective optimization and multiple constraint handling with evolutionary algorithms i: A unified formulation. *IEEE Transactions on Systems, Man and Cybernetics - Part A: Systems and Humans*, 28(1):26–37, 1998.

J. Calbó and J. Sabburg. Feature extraction from whole-sky groundbased images for cloud-type recognition. *Journal of Atmospheric and Oceanic Technology*, 25, 2008.

Carruthers D, Uloth C, and Roy G. G. An Evaluation of Formulae for Solar Declination and the Equation of Time. *Research Report No RR17*, 1990.

C.M. Fonseca. *Multiobjective genetic algorithms with application to control engineering problems*. PhD thesis, Department of Automatic Control and Systems Engineering, University of Sheffield, UK, 1995.

D.H.W. Li and J.C. Lam. An Analysis Of Climatic Parameters And Sky Condition Classification. *Build Environment 36, 435 445,*, 2001.

Eduardo M. Crispim, Pedro M. Ferreira, and Antonio. E. Ruano. Prediction of The Solar Radiation Evolution using Computational Intelligence Techniques and Cloudiness Indices. *International Journal of Innovative Computing, Information and Control*, 4, 2008.

P. M. Ferreira, E. Faria, and A. E. Ruano. Neural network models in greenhouse air temperature prediction. *Neurocomputing*, 43(1-4):51–75, 2002.

Francisco J. Olmo, Alberto Cazorla, Lucas Alados-Arboledas, Miguel A. López-Álvarez, Javier Hernández-Andrés, and Javier Romero. Retrieval of the Optical Depth Using an All-sky Camera. *Applied Optics*, 47(34):H182–H189, 2008.

G. López, F.J. Batlles, and J. Tovar Pescador. Selection Of Input Parameters To Model Direct Solar Irradiance By Using Artificial Neural Networks. *Energy*, 30(9):1675–1684, 2005.

G. Pfister, R. L. Mckenzie, J. B. Liley, A. Thomas, B. W. Forgan, and C. N. Long. Cloud Coverage Based on All-Sky Imaging and Its Impact on Surface Solar Irradiance. *American Meteorological Society*, 2003.

González, R. C., R. E. Woods, and S. L. Eddins. *Digital Image Processing Using MATLAB*. Prentice Hall, 2004.

H. Suehrcke and P.G. McCormick. The Frequency Distribution Of Instantaneous Insolation Values. *Solar Energy 40, 413 42,*, 1988.

Hargreaves G.H. and Samani Z.A. Estimating Potential Evapotranspiration. *Journal Irrig. Drain. Eng.*, 108:225–230, 1982.

H.F. Assunção, J.F. Escobedo, and A.P. Oliveira. A New Algorithm to Estimate Sky Condition Based on 5 Minutes-averaged Values of Clearness Index and Relative Optical Air Mass. *Theor. Appl. Climatol.*, 2007.

Igor A.C. Martins, Pedro M. Ferreira, and António E. Ruano. Estimation and prediction of cloudiness from ground-based all-sky hemispherical digital images. *Conference on Control Methodologies and Technology for Energy Effciency*, 2010.

Igor António Calé Martins. Neural models design for solar radiation and atmospheric temperature forecast. Master's thesis, Universidade Do Algarve - Faculdade De Ciências E Tecnologia, 2010.

Irineo L. López-Cruz, Abraham Rojano-Aguilar, Waldo Ojeda-Bustamante, and Raquel Salazar-Moreno. Arx models for predicting greenhouse air temperature: A methodology. *Agrociencia*, 41:181 192, 2007.

J. Calbó, J. A. González, and D.P. A Method For Skycondition Classification From Ground Based Solar Radiation Measurements. . *Journal Appl. Meteorology*, 2001.

J.C. Winslow, E.R. Hunt, and S.C. Piper. A Globally Applicable Model Of Daily Solar Irradiance Estimated From Air Temperature And Precipitation Data. *Ecol Model*, 2001.

Jérôme G. Fortin, François Anctil, Léon-Étienne Parent, and Martin A. Bolinder. Comparison Of Empirical Daily Surface Incoming Solar Radiation Models. *Agriculture And Forest Metereology*, 2008.

J.L. Bosch, G. López, and F.J. Batlles. Daily Solar Irradiation Estimation Over A Mountainous Area Using Artificial Neural Networks. *Renewable Energy*, 33(7):1622–1628, July 2008.

Kalogirou S. A. Artificial neural networks in renewable systems applications: a review. *Renewable and Sustainable Energy Reviews*, 5:373–401, 2001.

Kostas I. Chronopoulos, Ioannis X. Tsiros, Ioannis F. Dimopoulos, and Nikolaos Alvertos. An application of artificial neural network models to estimate air temperature data in areas with sparse network of meteorological stations. *Journal of Environmental Science and Health*, A(43):1752–1757, 2008.

Lèfevre M., L. Diabaté, and Wald L. Using Reduced Data Sets ISCCP-B2 From The Meteosat Satellites To Assess Surface Solar Irradiance. *Solar Energy*, 81:240–253, 2007.

M. A. López-Álvarez, J. Hernández-Andrés, J. Romero, and Jr. R. L. Lee. Designing a Practical System for Spectral Imaging of Skylight. *Applied Optics*, 2005.

M. Jurado, J.M. Caridad, and V. Ruiz. Statistical Distribution Of The Clearness Index With Radiation Data Integrated Over Five Minute Intervals. *Solar Energy 55, 469 473.*, 1995.

Mahmood R. and Hubbard K.G. Effect Of Time Of Temperature Observation And Estimation Of Daily Solar Radiation For The Northern Great Plains. *USA. Agron. Journal*, 94(4):723–733, 2002.

N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1):62–66, 1979.

P. M. Ferreira and A. E. Ruano. Exploiting the separability of linear and non-linear parameters in radial basis function neural networks. *IEEE Symposium 2000: Adaptive Systems for Signal Processing, Communications and Control*, pages 321–326, 2000.

P. M. Ferreira, A. E. Ruano, and C. Fonseca. Genetic assisted selection of rbf model structures for greenhouse inside air temperature prediction. *IEEE Conference on Control Applications*, pages 576–581, 2003.

Patrick Minnis, Qing Z. Trepte, Szedung Sun-Mack, Yan Chen, David R. Doelling, David F. Young, Douglas A. Spangenberg, Walter F. Miller, Bruce A. Wielicki, Ricky R. Brown, Sharon C. Gibson, and Erika B. Geier. Cloud Detection in Nonpolar Regions for CERES using TRMM VIRS and Terra and Aqua MODIS Data. *IEEE Transactions on Geoscience and Remote Sensing*, 46(11):3857–3884, 2008.

Pedro M. Ferreira and A. E. Ruano. Application of Computational Intelligence Methods to Greenhouse Environmental Control. Part II: Model Predictive Control. *Computers and Electronics in Agriculture*, 2007.

Pedro M. Ferreira, Igor A.C. Martins, and António E. Ruano. Cloud and clear sky pixel classification in ground-based all-sky hemispherical digital images. *Conference on Control Methodologies and Technology for Energy Efficiency*, 2010.

P.M. Ferreira, A.E. Ruano, and C.M. Fonseca. Evolutionary multiobjective design of radial basis function networks for greenhouse environmental control. *Proceedings of the 16th IFAC World Congress*, 16(1), 2005.

Shifrin and K.S. Simple Relationships for the ångström Parameter of Disperse Systems. *Applied Optics*, 34(21):4480–4485, 1995.

T. Ridler and S. Calvard. Picture thresholding using an iterative selection method. *IEEE Transactions on Systems, Man and Cybernetics*, 8(8):630–632, 1978.

H. Trussel. Comments on "Picture thresholding using an iterative selection method". *IEEE Transactions on Systems, Man and Cybernetics*, 9(5):311, 1979.

Tymvios F.S., Jacovide C.P., Michaelides S.C., and Scouteli C. Comparative Study Of Angstroms And Artifcial Neural Networks Methodologies In Estimating Global Solar Radiation. *Solar Energy*, 78:752–762, 2005.

Xiaoning Song, Yingshi Zhao, and Zhenhua Liu. Cloud Detection and Analysis of MODIS Image. *Geoscience and Remote Sensing Symposium*, 2004.

Y. Xia, M. Winterhalter, and P. Fabian. Interpolation Of Daily Global Solar Radiation With Thin Plate Smoothing Splines. *Theor. Appl. Climatol*, 66:109–115, 2000.

Total Sky Imager Installation and User Guide Model 440A Rev.C. Yankee Environmental Systems, Inc., 2007.

A. Appendix A

A.1. System setup

A.1.1 Computer Hardware

- System

  CPU: VIA Eden ULV 500MHz

  BIOS: AMI BIOS

  System Chipset: VIA CX700M

  I/O Chip: Winbond W83697

  System Memory: Onboard 512MB DDR2 RAM

- I/O

  1 x EIDE (UltraDMA 133)

  1 x Type I/II CF Slot

  1 x PS/2 K/B & Mouse

  2 x RS-232

  1 x Mini PCI sockets

  USB 3 x USB 2.0 Ports

- Ethernet

  Chipset: Realtek RTL8100B 10/100 Base-T Ethernet

  Remote Boot ROM: Built-in boot ROM function

- Mechanical & Environment

  Power Requirement: 5V @ 3A

  Operating Temp.: $0 - 60^oC(32 - 140^oF)$

  Operating Humidity: $0\% - 90\%$ relative humidity, non-condensing

  Size (W $\times$ H $\times$ D): $115 \times 115 \times 35$ mm

  Weight: 505g

A.1.2 CCTV camera

A.1.2.1 Baxall CDSP9372WF CCTV camera

Sensor: 1/3" Sony Hyper HADTM CCD with micro-lens over each pixel

Effective pixels: $752(H) \times 582(V)$

Sensitivity: 1.5 lux colour/0.7 lux mono. 40IRE at F1.2 with AGC (80% scene reflectance)

Resolution: 480TVL

Shutter speeds: 1/50, 1/125, 1/250, 1/500, 1/1000, 1/2000, 1/4000, 1/10,000 & flickerless via 3 DIP switches

Video output: 1V p-p composite video via 75Ω BNC connector on the rear of the camera Video Output

- PAL: 625 lines 50 fields per second 2:1 interlace
- NTSC: 525 lines 60 fields per second 2:1 interlace

Signal to noise ratio: Better than 48dB

White balance: Auto Tracing, Indoor, Outdoor, Fluorescent selectable via 2 DIP switches

AGC: Built in 28dB selectable On/Off via a DIP switch

Electronic iris: 1/50s - 1/100,000s. Selectable On/Off via a DIP switch

Backlight compensation: BLC selectable On/Off via DIP switch. Weighted to centre third of scene

Gamma correction: User-selectable 1.0 (linear) or 0.45 (normal) via a DIP switch

Synchronization system: Line-lock. The option to switch to internal synchronization is provided via a DIP switch. In line-lock the setting of the V phase can be set to fixed or adjustable

Adjustable V phase: The line lock vertical phase may be adjusted by $\pm 120^o$

A.1.2.2 Genie CCTV camera

Power source: DC12V - 160mA

CCD :

Total pixels: NTSC - $811(H) \times 508(V)$; PAL - $795(H) \times 596(V)$

Effective pixels: NTSC - $768(H) \times 494(V)$; PAL - $752(H) \times 582(V)$

Size: 1/3”

Scanning system: 2:1 interlace

Synchronization: INT/LL selectable; 0 359^o (adjustable)

Electrical :

Resolution: 540 TV lines

S/N: 50dB (AGC off)

Min. illumination: 0.3Lux/F1.2(color), 0.002Lux(sens-up)

Color: On/AUTO

Sens-Up: Off/AUTO

Gain control: OFF/LOW/MIDDLE/HIGH

Digital noise reduction: OFF/LOW/MIDDLE/HIGH

Iris control: DC/video

Operating temperature/humidity: -10^o $+ 50^oC/30$ $90\%RH$

A.1.3 Camera lens

Focal length: $1.6 \sim 3.4mm$

Max. aperture ratio: 1 : 1.4

Operation :

Iris: $F1.4 \sim close$

Focus: Manual

Zoom: Manual

Angle of view :

Diagonal: $183.9^o \sim 98^o$

Horizontal: $172.2^o \sim 78.8^o$

Vertical: $126.2^o \sim 58.1^o$

Distortion: $-104 \sim 30.3\%$

Operation temperature: $-10^o \sim 50^oC$

A.1.4 Operating system

Linux FaCtOrY 2.6.31-16-generic #53-Ubuntu SMP Tue Dec 8 04:01:29 UTC 2009 i686 GNU/Linux

A.1.5 Python modules

matplotlib; pylab; numpy;mx.DateTime; opencv

A.1.6 Electrical Scheme

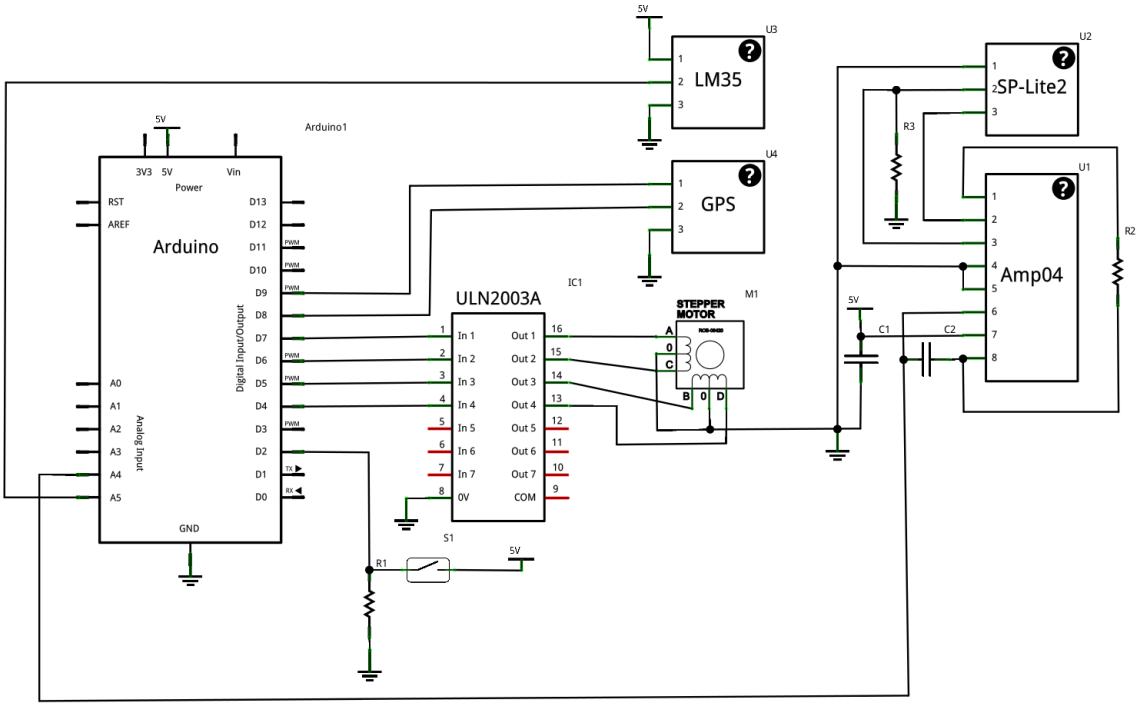


Figure A.1: Logic Module Wiring: $C1$ $0.1\mu F$; $C2$ $220nF$; $R1$$10k\Omega$; $R2$$10k\Omega$; $R3$$100k\Omega$. The connection with a computer is made through a USB cable.