INPE ePrint: sid.inpe.br/mtc-m18@80/2008/07.22.19.39 v1 2008-07-23

Title Page

Institute: Center for Weather Forecasts and Climate Studies/Brazilian Institute for Space Research (CPTEC/INPE)

Authors:

Ricardo André Guarnieri
Fernando Ramos Martins
Enio Bueno Pereira
Sin Chan Chou

Number of figures: 7 (seven)

Address for galley proofs:

Dr. Fernando Ramos Martins

Climate and Environment Division (DMA)
Center for Weather Forecast and Climate Studies (CPTEC)
Brazilian Institute for Space Research (INPE)
P.O. Box 515. São José dos Campos. São Paulo.
Zip Code: 12245-970. Brazil.
Email: fernando@dge.inpe.br

SOLAR RADIATION FORECAST USING ARTIFICIAL NEURAL NETWORKS

R. A. Guarnieri, F. R. Martins, E. B. Pereira, Sin Chan Chou

Center for Weather Forecast and Climate Studies - Brazilian Institute for Space Research

CPTEC-INPE. P. O. Box 515. São José dos Campos, Brazil. 12245-970

Summary

The fast increase in importance of the solar energy resource as viable and promising source of renewable energy has boosted research in methods to evaluate the short-term forecasts of the solar energy resource. There is an increasing demand from the energy sector for accurate forecasts of solar energy resources in order to support the planning and management of the electricity generation and distribution systems. The Eta model that runs operationally in the Brazilian Center of Weather Forecast and Climate Studies (CPTEC/INPE) has outputs for solar radiation flux at the surface. However, these radiation forecasts are greatly overestimated. In order to achieve more reliable short-term estimates for the downward solar radiation flux at the surface, forecasts for the future atmospheric conditions provided by Eta/CPTEC model were used as inputs in Artificial Neural Networks (ANNs). Ground data of downward solar radiation flux acquired in two SONDA sites located in south Brazil (Florianópolis and São Martinho da Serra) were used for ANNs training and for forecasts evaluation. The forecasts from ANNs have presented higher correlation coefficients and lower deviations than the model Eta/CPTEC output. The bias observed in solar radiation forecasts provided by the Eta/CPTEC model was removed by the ANNs. The improvement in RMSE obtained with ANNs over the model Eta/CPTEC was higher than 30%.

1. Introduction

The study of the incident solar radiation is of chief importance in the meteorological climate research. In addition, from the applied point of view, it has several implications in the agriculture, natural and artificial illumination, heating/refrigeration of buildings and residences, and in the search for new renewable energy resources.. Currently, the demand for new and more accurate information on the solar radiation resources has increased mostly because of the public awareness on the environmental impacts of fossil fuel consumption.

Developed countries are responsible for the environmental damages caused by years of fossil fuels consumption, increasing the carbon monoxide and other greenhouse gases concentrations in the atmospheric and triggering global climate changes. However, developing economies such as Brazil, India, China, and Russia are increasingly sharing this responsibility owing to their growing demands for energy to support their development. The commitment to reduce the emissions of carbon dioxide (and other greenhouse gases) by the countries that ratified Kyoto Protocol and the perspectives of oil depletion in next decades (Bentley, 2002; Geller, 2003) are other factors capable to boost new markets for the solar energy technology.

Furthermore, instabilities in hydroelectricity generation during dry seasons and political or economical crisis call for complementary energy resources in the energy matrixes of most countries. In Brazil the solar energy is one of the most promising option since most of its territory is located in the inter-tropical region where solar energy resources are accessible all the year round (Pereira et al.,2006). The main obstacles to the commercial exploitation of solar energy resources are its higher cost when compared to the conventional electricity generation technologies, and its strong dependence on the weather and climate conditions and to the day cycle. The generation costs are expected to fall in next decades due to technological advances and market demands. This leads to an increasing worldwide demand for more reliable information on the solar resources, including its spatial and temporal variability in the short and long terms. There is an increasing demand from the electricity energy sector for accurate forecasts of solar energy resources in order to manage hybrid systems. In addition, an accurate short-term forecast of solar radiation is important information for the management of energy dispatches in transmission lines, since the solar radiation influences the heat dissipation by the cables.

Mesoscale weather forecast models usually have radiation parameterization codes, since solar radiation is the main energy source for atmospheric processes. Nevertheless, forecasting solar irradiation, even one day in advance, is a complicated task. Part of the difficulties arises from the solar radiation dependence on clouds and meteorological conditions, which intrinsically involves non-linear physical processes. Several studies are been developed in order to reduce the inaccuracy associated to the modeling approaches adopted in numerical weather models to simulate the complex non-linear atmospheric processes (Hinkelman, 1999).

The model Eta/CPTEC that runs operationally in the Brazilian Center of Weather Forecast and Climate Studies (CPTEC/INPE) has outputs for many meteorological variables, including solar radiation incidence on ground. However, these radiation forecasts are greatly overestimated. As an attempt to get better predictability for the solar energy resources using model Eta/CPTEC, Artificial Neural Networks (ANNs) were used as statistical post-processing model. This study describes the development of an operational method to forecast the solar radiation flux at the surface. The goal is to help stakeholders in the management and optimization of several energy generation and distribution processes. Therefore, it constitutes an important application of the meteorology science to the energy segment of the society.

2. Methodology

The solar irradiance at top of the Earth's atmosphere (*top of atmosphere* will be referred as *TOA* hereafter) , also known as *solar constant*, is about 1368 W·m^{-2}, despite variations around ± 0.6 W·m^{-2} along the 11-year solar-activity cycle and ± 3.4% along the year, due to the eccentricity of the Earth's orbit around the Sun (Kidder e Vonder Haar, 1995). Some geometrical factors should be considered to compute the TOA, since the solar zenith angles depends on latitude, declination (variable along the year) and time of day.

The atmosphere modifies the solar radiation flux until it reaches the Earth's surface. Absorption and scattering are the main processes that affect the solar radiation transmittance through the atmosphere. Clouds are the main factor that controls the solar radiation at the surface. (For more details about solar radiation and atmospheric influences see Kidder e Vonder Haar, 1995; Liou, 1980; Robinson, 1966; or Iqbal, 1983). Therefore, the atmospheric optical properties should be known in order to correctly evaluate the solar irradiation at any specific site and time. These properties depend on clouds, aerosols, humidity and other factors. Forecasting solar irradiation depends on prospecting of future atmospheric conditions. Despite the intrinsic uncertainties, numerical weather prediction (NWP) models provide routinely information about many meteorological variables for several future times, including solar radiation data and atmospheric optical properties. Earlier studies have demonstrated that solar radiation data provided by NWP presents a larger bias than the required to manage several solar energy applications such as grid connected or hybrid photovoltaic (PV) systems (Hinkelman *et al*., 1999; Chou *et al*., 2002).

This paper describes a statistical method for improving the confidence level and reliability of solar irradiation forecasts provided by the NWP model used at the Center for Weather Forecast and Climate Studies (CPTEC) of the Brazilian Institute for Space Research (INPE), known as model Eta/CPTEC. The method consists of using the atmospheric data outputs from the mesoscale NWP model for future atmospheric conditions to feed Artificial Neural Networks (ANNs). Further, the calculated solar radiation on the *top of Atmosphere* (TOA) was also supplied to the ANNs,. The main goal is to obtain a solar radiation forecast with error deviations lower than the forecasts provided directly by the mesoscale model through its radiative code, for a given site of investigation. The solar data acquired in two SONDA ground stations located in the Southern region of Brazil were used as reference for training the ANNs and for evaluation of the solar radiation forecasts provided by the method.

2.1. Model Eta/CPTEC

Detailed descriptions about model Eta can be found in the following specific literature: Mesinger *et al.* (1988), Janjić (1994), Black (1994) and Ničhović *et al.* (1998). Finite Difference schemes are applied to solve the system of equations in space and time. One of the features of this model is the vertical coordinate "Eta", η, defined as (Mesinger, 1984):

$$\eta = \frac{p - p_t}{p_{sfc} - p_t} \frac{p_{ref}(z_{sfc}) - p_t}{p_{ref}(0) - p_t} \tag{1}$$

where p_t is the pressure at the top of the model atmosphere, p_{sfc} and z_{sfc} are the pressure and height of the lower boundary (surface), and p_{ref} is a reference pressure to the vertical profile. The lower surface heights can take only discrete values since the orography is represented by step-like functions and the tops of model mountains coincide with the η-coordinate surfaces (Ničhović *et al.*, 1998). The constant η-surfaces are relatively horizontal, so that the errors associated with the determination of the pressure gradient along a steeply sloped coordinate surface are minimized. The model Eta was adapted and optimized to the South America conditions (model Eta/CPTEC) and runs operationally at Brazilian Center of Weather Forecast and Climate Studies (CPTEC/INPE) since 1996. The model Eta/CPTEC runs routinely for an area that covers the most of South America continent and neighboring oceans: latitudes between 50.2ºS and 12.2ºN, and longitudes between 83ºW and 25.8ºW. It is configured to 40 km of horizontal resolution and 38 vertical layers. The discretization of the space domain is through the Semi-Staggered Arakawa E-grid in the horizontal and the Lorenz grid in the vertical. The radiation modeling uses the schemes of Lacis and Hansen (1974) for shortwave radiation, and Fels and Schwarztkopf (1975) for long wave radiation. Chou *et al*. (2002) showed that model Eta/CPTEC systematically overestimates the solar radiation incidence, as well as the surface fluxes of sensible and latent heat.

The model Eta/CPTEC runs twice a day, with initial conditions at 00UT and 12UT. The initial conditions are the NCEP analyses. The lateral boundary conditions are taken from the CPTEC Atmospheric Global Circulation Model (MCGA - www.cptec.inpe.br/prevnum/exp_global.shtml) and updated every 6 hours. The model Eta/CPTEC provides two sets of data (00UT and 12UT) comprising forecasts 6-hourly spaced for next 7 days, matching with the synoptic times (6, 12, 18 and 24UT).

The model Eta/CPTEC provided forecast values for each 58 atmospheric variables: 49 of them are surface variables and present a single value representing the atmospheric column total (*single variables*). The other nine variables are *profile variables* and values for 19 atmospheric pressure levels are supplied. The model Eta/CPTEC delivers two kinds of data in its output file: instantaneous forecast values for some atmospheric variables in each synoptic time; and averages, integrals or cumulative quantities related to the last 6-hour period before the synoptic time. For the current study, a set of 33 *single variables* was selected. In this work, only data provided by model Eta/CPTEC for the grid-points nearby from the two ground stations (described later) were used.

2.2. SONDA network

SONDA (*Sistema de Organização Nacional de Dados Ambientais para o Setor de Energia* – National Organization System of Environmental Data for the Energy Sector) is a network of ground measurement sites, set up and managed by CPTEC/INPE, that aims at acquiring reliable surface solar radiation and wind data all over Brazil to develop, improve and validate a variety numerical models for environmental research and for the assessment of renewable energy resources. Besides that, the SONDA database will provide valuable information to be used by energy sector in planning and deployment of energy resources for electricity generation and distribution in Brazil. In this work, measurements of global solar radiation performed by SONDA ground stations using *Kipp & Zonen CM-21* pyranometers (Kipp & Zonen, 2006) were used. The data were taken from two sites located in the Southern region of Brazil:

– Florianópolis (SC): *FLN site* (*Lat.: 27.60ºS; Long: 48.52ºW*)
– São Martinho da Serra (RS): *SMS site* (*Lat.: 29.44ºS; Long: 53.82ºW*)

The Figure 1 shows the location of all SONDA sites featuring FLN and SMS sites. The ground data are available daily as mean irradiances for each 1-minute, along the 24 hours. The data used in this paper comprise periods from January/2001 to October/2005 for FLN, and from July/2004 to October/2005 for SMS. The data provided by model Eta/CPTEC for each location were taken for these same periods.

2.3. *Data Management*

The solar and meteorological data used in this work comprises forecasts provided by the model Eta/CPTEC and the calculated variables and ground data described earlier. Table 1 presents the complete list of 33 variables from model Eta/CPTEC with a brief description for each along with their units. In addition, other three variables were calculated in order to supply ancillary information to the ANN: solar radiation flux at TOA (rtoa), mean air mass (airm), and mean solar zenith angle (szam). Altogether, 36 variables were managed to be used as predictors for the ANNs.

The SONDA ground data were used for the ANNs training and validation as described in the next topic. Besides that, ground data were used to compare the results between the two methods: model Eta/CPTEC and ANNs.

The database was subdivided into three sets: the *training set* (575 days for FLN and 236 days for SMS), the validation set (288 days for FLN and 118 days for SMS) and the investigation *set* (287 days for FLN and 118 days for SMS). The *training set* is used for the ANNs' training and the *validation set* is used for the real-time evaluation of the training process and to establish an end for this process. The *investigation set* is used for simulations and for the comparative evaluation of the solar radiation forecasts calculated by the ANNs and by the model Eta/CPTEC.

The variable *ocis* (Table 1) from model Eta/CPTEC stands for forecasts of solar irradiance. The *ocis* variable comprises the averages of solar irradiances along the 6-hour periods preceding the synoptic times. The forecasts provided by the model are available for each synoptic time and represents the mean solar irradiance along the former 6-hour interval.

2.3.1 Solar Data.

In order to achieve the same time-scale presented by solar irradiance forecasts provided by model Eta/CPTEC, the ground data of solar irradiation acquired in FLN and SMS sites with 1-minute temporal resolution were averaged in 6-hour intervals and it was attributed to the same synoptic times adopted by the model. Both, ground data and Eta/CPTEC forecasts, were converted to energy integrals, expressed in $MJ \cdot m^{-2}$ (mega joules per squared meter). After data-quality verification, 1150 valid days for FLN and 472 valid days for SMS were taken for the analyses.

Solar radiation flux at TOA was calculated according Iqbal (1983), with 1-minute resolution, for the both locations, FLN and SMS sites. As the ground data, the TOA solar radiation flux were averaged in 6-hour intervals and stored in units of $MJ \cdot m^{-2}$. In addition, the mean solar zenith angle and the mean air mass were also obtained and stored for the same 6-hour intervals, to be used as additional inputs in ANNs.

2.3.2 Atmospheric data provided by model Eta/CPTEC.

All instantaneous variables provided by model Eta/CPTEC were averaged by using two consecutive values. The averages were attributed to the final time of 6-hour interval in order to set up all data with values that better represent the variability of the atmospheric and meteorological variables in such time interval.

Thus, all data used to feed ANNs have now the same temporal resolution and represent the same time-intervals as *ocis variable*. Variables with low temporal variability or not straight linked to the solar radiation transmittance in atmosphere were discarded.

2.4. ***Artificial Neural Networks (ANNs)***

Artificial Neural Networks (ANNs) are computing systems, which attempt to simulate the structure and function of biological neurons. The networks generally consist of a number of interconnected processing elements, called *neurons*. Figure 2 presents an artificial neuron. Synapses connect each neuron with the neurons of neighboring layers. The input values (x_i) are weighted by values associated with each synapse (w_{ij}), called *synaptic weights*. The *activity level* of the neuron ($\boldsymbol{v}_j$) is determined by summing up all weighted values together with a value called bias (b_j). The neuron output is a result from an *activation function* ($\varphi(\boldsymbol{v}_j)$). Generally, the activation function is a linear or hyperbolic-tangent function. The use of non-linear function allows ANNs to learn non-linearity behaviors and complex patterns.

There are several structures of ANNs, but *feedforward* is the most used. In *feedforward networks*, the neurons are disposed in layers. Signals flow from the input layer through to the output layer via unidirectional connections, called *synapses* (Figure 3). Feedforward ANNs with multiple layers of neurons are commonly called *multilayer perceptrons* (Haykin, 1994).

Preliminary analyses revealed that, for the purposes of this study, better ANNs' performances were achieved using two hidden layers of neurons. Table 2 shows the best neurons distributions verified for each ANN-model. The number of neurons of input layers is equivalent to the number of variable in input dataset. Only one neuron is present in the output layer once only one output is expected: solar radiation flux at surface. The schematic diagram of a feed forward ANN model used at this study is illustrated in Figure 3.

The training step is the first procedure to set up the ANN model. In this work we have trained multilayer perceptrons using as inputs the meteorological data generated by the model Eta/CPTEC, and three other theoretically calculated values: the solar zenith angle and the solar radiation flux at TOA and air mass (total of 36 predictors). The training algorithm uses a training set of data to adjust the network parameters (weights and bias), in order to reduce the errors in output. For each iteration, the output produced by using a set of inputs is compared with the *target* (or the expected value, in this case, solar radiation ground data), and incremental corrections are calculated for each network parameter – wij and b_j – aiming to reduce the error in output.

A second data set, called validation set, is used to verify the performance of the ANN with an independent data sample, not directly used in learning process. This verification allows to check the generalization capacity along the training and to find out the appropriate moment to stop the training, avoiding *overlearning*. The most widespread training algorithm used for multilayer perceptrons are the *Back propagation* algorithm (Rumellhart *et al.*, 1986). In this work, we use a modified version of Back propagation, called *Resilient Back propagation* or *Rprop* (Riedmiller and Braun, 1993). After training, the weights and bias are fixed, and the ANN is ready to be used in simulations, using the investigation dataset as input data.

3. Results and Discussion

Since the 36 predictors and the ground data (measured solar radiation) are disposed in 6-hour variables, each variable has values for four times in each day: 6:00, 12:00, 18:00 and 24:00UT; each of them representing the intervals 0:00-6:00UT, 6:00-12:00UT, 12:00-18:00, and 18:00-24:00UT, respectively. This paper presents the results for the time interval between 12:00 and 18:00UT, hereafter referred as *Rad18UT*, once it is the interval presenting the highest fraction (63% – 80%) of daily total amount of solar radiation flux at the surface throughout year in both ground sites (Guarnieri, 2006). In this work, it discussed only the results obtained by using the meteorological average conditions predicted for the period Rad18UT of a particular day by the model Eta/CPTEC in the beginning of the same day (i.e., Eta/CPTEC running with initial conditions at 00UT). The ANN outputs are referred as *P00UT*.

The investigation dataset was used for evaluation of solar radiation flux estimates provided by both methods: model Eta/CPTEC and ANN. The forecasted values (*forecasts – F*) were compared with measured values (*observations – O*), and deviations between them (*F - O*) were calculated. The performance of the models was checked with two statistical indices: mean error (ME) or bias, and root mean squared error (RMSE). ME provide information about the systematic deviations of the methods, indicating the amount of overestimation or underestimation in the forecasted values. RMSE provides an estimative of the mean absolute deviations between forecasts and observations.

Both indices were normalized and expressed as percentages of the mean measured global solar radiation value as shown in eq. (2) and (3).

$$ME\% = 100 \cdot \frac{\sum_{i=1}^{N}(F_i - O_i)}{\sum_{i=1}^{N}(O_i)}\% \tag{2}$$

$$RMSE\% = 100 \cdot \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(F_i - O_i)^2}}{\frac{1}{N}\sum_{i=1}^{N}(O_i)}\% \tag{3}$$

where N is the number of data pairs (forecast and observation) used in the evaluation (it is equivalent to the number of days in the investigation dataset once there is one pair of solar radiation flux at Rad18UT for each day).

In addition, it was also calculated the Person's correlation coefficient (R), as described in eq. (4), and the determination coefficient (R^2) taking the square of R.

$$R = \frac{\sum_{i=1}^{N}(F_i - \overline{F})(O_i - \overline{O})}{\sqrt{\sum_{i=1}^{N}(F_i - \overline{F})^2} \cdot \sqrt{\sum_{i=1}^{N}(O_i - \overline{O})^2}} \tag{4}$$

In order to evaluate the improvement brought by one method over another, it was defined the skill-score index as follows:

$$Skill \quad (Score, ref) = \frac{Score - Score_{ref}}{Score_{perf} - Score_{ref}} \tag{5}$$

where *Score* can be the ME% or the RMSE% values calculated for the method in evaluation, $Score_{ref}$ is the score calculated for a reference method and $Score_{perf}$ is the score value expected for perfect-forecast (zero, for ME% or RMSE%).

Initially the model Eta/CPTEC forecasts and ground data for solar radiation flux were compared. As demonstrated in previous studies (Chou *et al.*, 2002; Hinkelman *et al.*, 1999), it was observed a significant positive bias (overestimation) in solar radiation flux provided by Eta/CPTEC. Table 3 shows the performance indices obtained for Eta/CPTEC estimates (P00UT-Rad18UT) taking all data in account (1150 days for FLN and 472 days for SMS) and taking only investigation dataset (N = 287 for FLN; N = 118 for SMS). Similar R, R^2, ME% and RMSE% were presented by model Eta/CPTEC for both cases denoting that the investigation dataset are representative of the complete dataset. Since ANN performance can be evaluated using the investigation dataset, only the Eta/CPTEC performance indices obtained with this dataset will be considered from now on.

Several statistical analyses and simulation tests were performed applying different subsets of the 36 predictors in ANNs, in order to find a reduced set of variables that can led to a performance similar that obtained with the use of all the predictors. It was found a set of 8 predictors: solar radiation flux at TOA (rtoa), relative humidity (rh2m), surface temperature (tsfc), precipitable water amount (agpl), zonal wind speed at 10-m height (u10m), and predictors for cloud fractions (cbnt, hinv and mdnv). The ANNs using 36 and 8 predictors will be called *ANN-36* and *ANN-8p*, respectively.

Table 4 presents the performance indices obtained for ANN-36 and ANN-8p estimates (P00UT-Rad18UT) taking the investigation dataset for both ground sites. It can be noted the very similar performance in terms of correlation (R) and RMSE deviations. However, the ANN-8p has generated estimates with almost half ME than the estimates provided by the ANN-36p in both sites. Comparing Tables 3 and 4, it can be noted the best correlation coefficients and the lowest ME% and RMSE% produced by ANNs over model Eta/CPTEC. The ANN-8p have presented the lower systematic deviation while the estimates provided by model Eta/CPTEC showed the lower correlation and larger deviations (ME and RMSE) for both ground sites.

Figures 4 and 5 present scatter-plots, where the forecast values were compared with observations. Besides the scatter-plots for Eta model, ANN-36p and ANN-8p, it is also showed a plot for a method called *persistence forecast*. Persistence forecast is the simplest method to predict meteorological data and it consists in to take the value observed in a previous day, as the forecast for the current day. Any forecast method is useful if it can lead to results better than the persistence forecast.

According to Figures 4 and 5, the radiation solar fluxes estimates provided by model Eta/CPTEC are better than persistence estimates, in general. However, it can be observed the positive bias mentioned before. The Eta/CPTEC estimates are overestimated, especially for cloudy days when solar radiation flux at the surface is lower. The scatter-plots for ANNs showed better agreement between estimates and observations, and most of the points are located near the perfect-forecast line (blue line). No clear differences are observed between ANN-36p and ANN-8p, indicating that most of the 36 predictors are not necessary to forecast the solar radiation fluxes at the surface.

Figure 6 shows a small temporal series taken from the investigation dataset of FLN and SMS sites. Estimates from model Eta/CPTEC and ANNs are put together with observations for the days in Winter/2005 and Summer/2004-2005. It can be observed that the ANNs forecasts are closer to the ground data than the overestimated forecasts from Eta model. The deviations for each day were calculated and it is presented in Figure 7. It is clear from both figures that an important improvement in solar radiation flux estimates was achieved by using ANNs supplied with the future atmospheric-state data generated by model Eta/CPTEC. However, no significant differences were observed between ANN-36p and ANN-8p. Again, it is clear that just eight predictors are enough for a good performance of ANN. To quantify the improvement acquired by the use of ANNs, the *skill* values were calculated using RMSE% score, and the results are presented in Table 5. In general, the ANNs lead to skill-scores higher than 30% in RMSE% when compared to model Eta/CPTEC.

4. Conclusions

Currently, the renewable sources of energy are getting more importance into electricity generation systems. In reason of that, there is an increasing demand from the energy sector for accurate forecasts of solar energy resources in order to support the electricity generation and distribution systems. The forecasts provided by numerical weather models could supply this information demand but, in general, these forecasts present large deviations reducing its confidence and reliability. In Brazil, the model Eta/CPTEC generated solar radiation flux estimates with bias around 25%. It was observed lower deviations when estimates are provided by ANNs that uses a dataset of meteorological variables predicted by model Eta/CPTEC for the near future as predictors. The comparison among estimates and ground data acquired in the SONDA sites located in Florianópolis (FLN) and São Martinho da Serra (SMS), showed a similar performance between the ANNs using 36 and 8 predictors, and both these models provide estimates for solar radiation flux at the surface better than the model Eta/CPTEC, in general. The biases observed in the Eta/CPTEC estimates are larger than the one presented by ANNs results. The skill-score indices show that ANNs have improved the confidence (taking in account the RMSE%-reduction) in estimates for solar radiation in more than 30% when compared to model Eta/CPTEC. The improvements in predictability were observed in correlation coefficients as well: from 0.72 to 0.80 in FLN, and from 0.78 to 0.85 in SMS.

Acknowledgments

The SONDA project was possible thanks to the FINEP (project 22.01.0569.00) and Petrobras/CENPES (project 0050.0019.104.06.2). Thanks are also due to the following colleagues: Silvia V. Pereira, Sheila A. B. Silva, Rafael Chagas and all the colleagues from *Laboratório de*

Instrumentação Meteorológica (LIM/CPTEC). The following institutional acknowledgment is due to Center for Weather Forecast and Climatic Studies (CPTEC/INPE) and CNPq (grants 151700/2005-2, 141844/2006-0, 132148/2004-8).

References

Bentley, R.W., 2002: Global Oil and Gas Depletion: An Overview. *Energy Policy*., **30**, 189-205.

Black, T. L., 1994: The new NMC mesoscale Eta model: Description and forecast examples. *Weather and Forecasting*, **9**(2), 265-278.

Chou, S. C., Tanajura C. A. S., Xue, Y., and Nobre, C. A., 2002: Validation of the coupled Eta/SsiB model over South América. *J. Geophys. Res.*, **107**(D20), LBA 56, 1-20.

Colle, S.; Pereira, and E. B., 1998: *Atlas de irradiação solar do Brasil (primeira versão para irradiação global derivada de satélite e validada na superfície).* INMET, 50pp.

Fels, S. B., and Schwarzkopf, M. D., 1975: The simplified exchange approximation: a new method for radiative transfer calculations. *J. Atmosph. Sci.*, **32**(7), 1475-1488.

Geller, H. S., 2003: *Revolução Energética: políticas para um futuro sustentável*. Relume Dumará, 299pp.

Goldemberg, J.; Villanueva, L. D., 2003: *Energia, Meio Ambiente & Desenvolvimento*. 2ª ed. Edusp, 232pp.

Guarnieri, R. A., 2006: *Emprego de redes Neurais Artificiais e Regressão Linear Múltipla no refinamento das previsões de radiação solar do modelo Eta.* Insituto Nacional de Pesquisas Espaciais, 171pp.

Haykin, S., 1994: *Neural networks: a comprehensive foundation*. Macmillan College Publishing Company.

Hinkelman, L. M., Ackerman, T. P., and Marchand, R. T., 1999: An evaluation of NCEP Eta model predictions of surface energy budget and cloud properties by comparison with measured ARM data. *J. Geophys. Res.*, **104**(D16), 19535-19549.

Iqbal, M., 1983: *An Introduction to Solar Radiation*. Academic Press Canada, 390pp.

Janjić, Z. L., 1994: The step-mountain Eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Monthly Weather Review*, **122**(5), 927-945.

Kidder, S. Q., and Vonder Haar, T. H., 1995: *Satellite Meteorology: an introduction*. Academic Press, 466pp.

Kipp & Zonen, 2006: *Kipp & Zonen Co. Webpage: manufacturer of meteorological and industrial sensors.* Available at <http://www.kippzonen.com/>.

Lacis, A. A., and Hansen, J. E., 1974: A parameterization of the absorption of solar radiation in earth's atmosphere. *J. Atmosph. Sci.*, **31**, 118-133.

Liou, K. N., 1980: *An Introduction to Atmospheric Radiation*. Academic Press Inc., 392pp.

Mesinger, F., 1984: Ablocking for representation of mountains in atmospheric models. *Rivista di Meteorologia Aeronautica*, **44**, 195-202.

Mesinger, F.; Janjić, Z. L.; Ničković, S., and Deaven, D. G., 1988: The step-mountain coordinate: Model description and performance for cases of Alpine lee ciclogenesis and for a case of Appalachian redevelopment. *Monthly Weather Review,* **116**, 1493-1518.

Ničković, S., Mihailović, D., and Rajković, B., 1988: *Scientifc documentation of the Eta model.* Word Meteorological Organization.

Pereira, E. B., Martins, F. R., Abreu, S. L. and Ruther R., 2007. *Atlas Brasileiro de Energia Solar*. Instituto Nacional de Pesquisas Espaciais, 70pp.

INPE ePrint: sid.inpe.br/mtc-m18@80/2008/07.22.19.39 v1 2008-07-23

Riedmiller, M., and Braun H., 1993: A direct adaptive method for faster backpropagation learning: The RPROP algorithm. Proceedings, *IEEE International Conference on Neural Networks*.

Robinson, N. (Ed.), 1966: *Solar Radiation*. Elsevier Publishing Company, 347pp.

Rumelhart, D. E., Hinton G. E., and Williams, R. J., 1986: Learning representations by back-propagation errors. *Nature*, **323**, 533-536.

List of Table captions.

Table 1. The 33 atmospheric variables, provided by model Eta/CPTEC, used as predictors in ANNs.

Table 2. Number of artificial neurons in each layer of ANN model implemented for this work. The number of layers and neurons were determined in preliminary tests in order to obtain the best performance of ANN.

Table 3. Performance indices for estimates provided by model Eta/CPTEC (P00UT-Rad18UT) using all data and only the investigation dataset.

Table 4. Performance indices for *ANN-36p* and *ANN-8p* estimates (P00UT-Rad18UT) using the investigation dataset.

Table 5. Skill-score calculated with RMSE% score for ANNs taking model Eta/CPTEC and persistence as reference forecast methods.

Table 1. The 33 atmospheric variables, provided by model Eta/CPTEC, used as predictors in ANNs.

VARIABLE	**DESCRIPTION (UNITS)**
rh2m	Relative humidity at 2m-height (0 to 1 – adimensional)
pslc	Pressure at surface (hPa)
tp2m	Temperature at 2m-height above the surface (K)
dp2m	Dew Point Temperature at 2m above the surface (K)
u10m	Zonal wind at 10m-height above the surface (m s^{-1})
v10m	Meridional wind at 10m-height above the surface (m s^{-1})
wnds	Wind velocity at 10m-height above the surface (m s^{-1})
prec	Total rainfall (kg m^{-2} dia^{-1})
prcv	Convective rainfall (kg m^{-2} dia^{-1})
prge	Large scale rainfall (kg m^{-2} dia^{-1})
clsf	Latent Heat Flux at the surface (W m^{-2})
cssf	Sensible Heat Flux at the surface (W m^{-2})
ghfl	Heat Flux in the soil (W m^{-2})
tsfc	Surface Temperature (K)
qsfc	Specific humidity at the surface (kg(H_2O) kg(ar) $^{-1}$)
lwnv	Cloud cover Index for low clouds (0 a 1 - adimensional)
mdnv	Cloud cover Index for average clouds (0 a 1 - adimensional)
hinv	Cloud cover Index for high clouds (0 a 1 - adimensional)
cbnt	Mean Cloud cover Index (0 a 1 - adimensional)
ocis	Downward shortwave radiation flux at the surface (W m^{-2})
olis	Downward longwave radiation flux at the surface (W m^{-2})
oces	Upward shortwave radiation flux at the surface (W m^{-2})
oles	Upward longwave radiation flux at the surface (W m^{-2})
roce	Upward shortwave radiation flux at the TOA (W m^{-2})
role	Upward longwave radiation flux at the TOA (W m^{-2})
albe	Albedo (0 a 1 - adimensional)
cape	Available potential convective energy (m^2 s^{-2})
cine	Energy to avoid convection (m^2 s^{-2})
agpl	Instantaneous precipitable water amount (kg m^{-2})
pcbs	Pressure at the bottom of the clouds (hPa)
pctp	Pressure at the top of the clouds (hPa)
tgsc	Soil temperature at the surface layer (K)
ussl	Soil humidity at the surface (0 a 1 - adimensional)

Table 2. Number of artificial neurons in each layer of ANN model implemented for this work. The number of layers and neurons were determined in preliminary tests in order to obtain the best performance of ANN.

	ANN-36p.	**ANN-8p.**
Input layer	36	8
First hidden layer	36	16
Second hidden layer	18	8
Output layer	1	1

ANN-36p – ANN using 36 variables as predictors

ANN-8p – ANN using 8 variables as predictors

Table 3. Performance indices for estimates provided by model Eta/CPTEC (P00UT-Rad18UT) using all data and only the investigation dataset.

Scores	**Florianópolis**		**São Martinho da Serra**	
	N =1150	N =287*	N =472	N =118*
R	0.747	0.720	0.790	0.775
R^2	0.558	0.519	0.624	0.600
ME%	24.7%	24.6%	27.8%	28.0%
RMSE%	39.7%	40.0%	41.9%	43.2%

* - results obtained using investigation dataset.

Table 4. Performance indices for *ANN-36p* and *ANN-8p* estimates (P00UT-Rad18UT) using the investigation dataset.

Scores	**Florianópolis**		**São Martinho da Serra**	
	ANN-36p	ANN-8p	ANN-36p	ANN-8p
R	0.804	0.790	0.839	0.848
R^2	0.646	0.625	0.704	0.720
ME%	-2.1%	-0.8%	-1.7%	-0.7%
RMSE%	26.2%	26.9%	28.8%	27.6%

All results obtained using investigation dataset.

Table 5. Skill-score calculated with RMSE% score for ANNs taking model Eta/CPTEC and persistence as reference forecast methods.

Indices	**Florianópolis**		**São Martinho da Serra**	
	ANN-36p	ANN-8p	ANN-36p	ANN-8p
Skill(RMSE%,persistence)	0.429	0.414	0.464	0.487
Skill(RMSE%,Eta)	0.344	0.328	0.333	0.361

* - results obtained using investigation dataset.

List of Figure captions:

Figure 1. Localization map of the ground data acquisition sites of SONDA network operating in Brazilian Territorry. The ground data acquired at Florianópolis (FLN) and São Matinho da Serra (SMS) were used here to compared with forecasts provided by model Eta/CPTEC and ANNs.

Figure 2. Symbolic representation of an artificial neuron and its parameters.

Figure 3. Schematic diagram of a Feedforward ANN used in this simulation. These ANN model has presented the best performance in preliminary tests using predictors dataset with 33 or 8 variables.

Figure 4. Scatter-plots of solar estimates for radiation flux at surface in FLN site against ground data: (a) estimates provided by persistence method, (b) estimates provided by model Eta/CPTEC, (c) estimates provided by ANN-36p, and (d) estimates provided by ANN-8p.

Figure 5. Scatter-plots of solar estimates for radiation flux at surface in SMS site against ground data: (a) estimates provided by persistence method, (b) estimates provided by model Eta/CPTEC, (c) estimates provided by ANN-36p, and (d) estimates provided by ANN-8p.

Figure 6. Temporal series comparing estimates and ground data for solar radiation flux at surface in both FLN and SMS sites. The series corresponds to days in the investigation dataset.

Figure 7. Deviations among estimates and ground data obtained for the same days presented in Figure 6.

INPE ePrint: sid.inpe.br/mtc-m18@80/2008/07.22.19.39 v1 2008-07-23

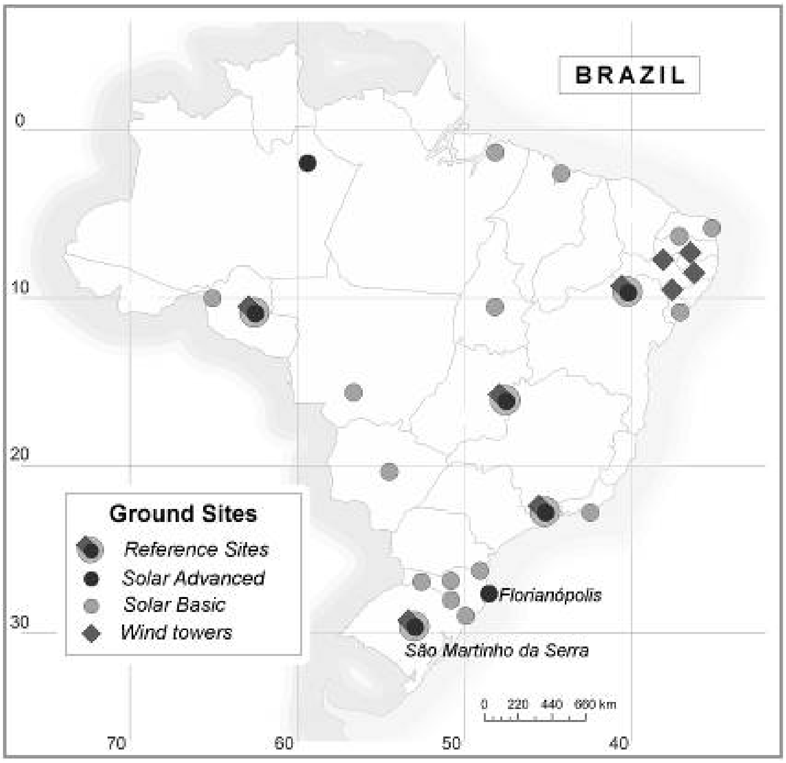

Figure 1. Localization map of the ground data acquisition sites of SONDA network operating in Brazilian Territorry. The ground data acquired at Florianópolis (FLN) and São Matinho da Serra (SMS) were used here to compared with forecasts provided by model Eta/CPTEC and ANNs.

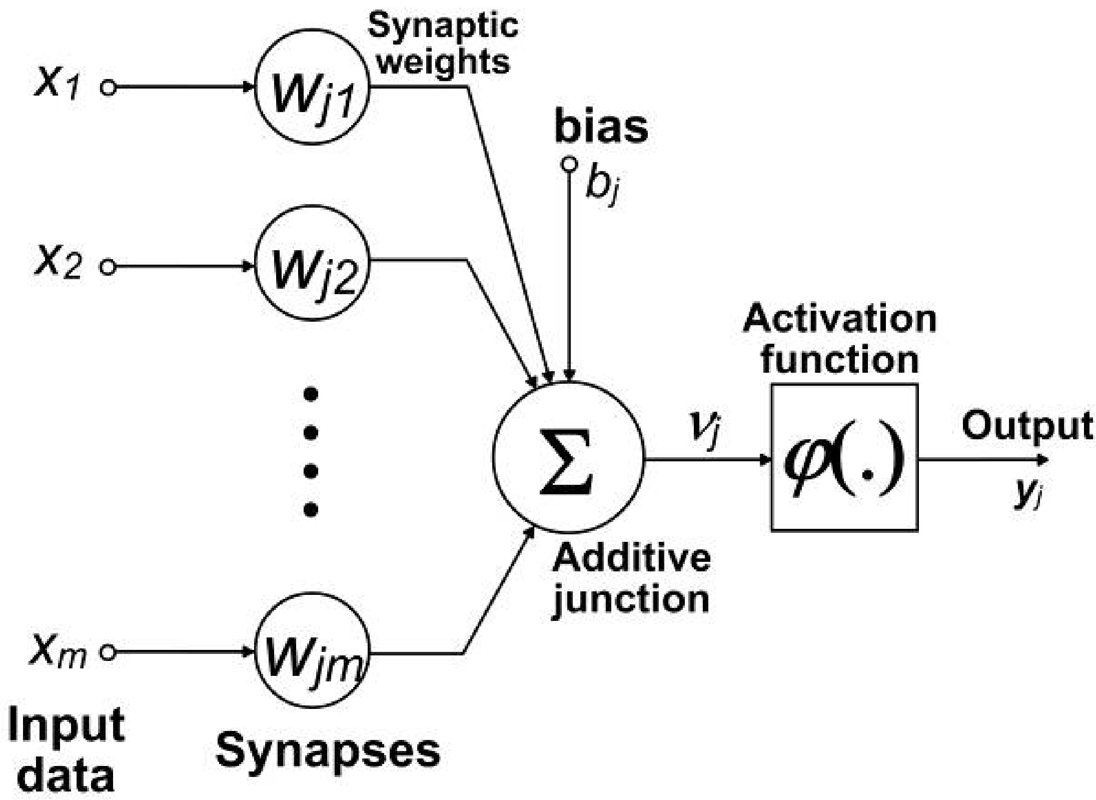

Figure 2. Symbolic representation of an artificial neuron and its parameters.

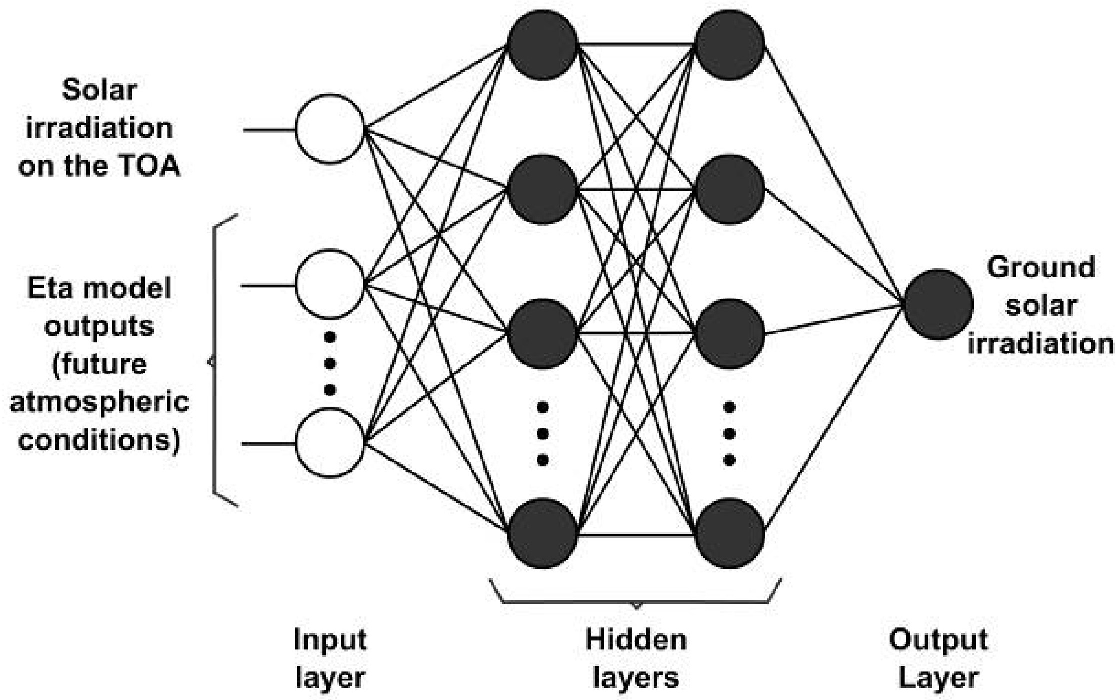

Figure 3. Schematic diagram of a Feedforward ANN used in this simulation. These ANN model has presented the best performance in preliminary tests using predictors dataset with 33 or 8 variables.

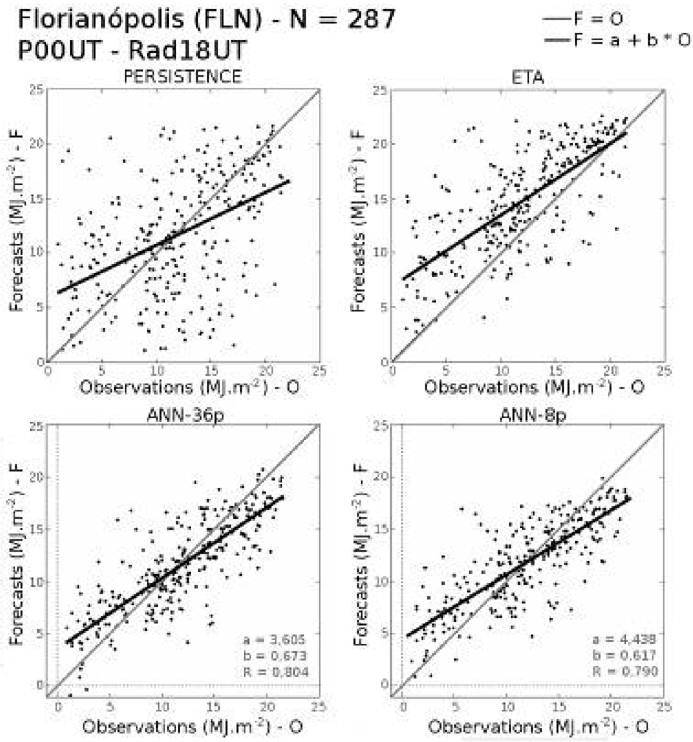

Figure 4. Scatter-plots of solar estimates for radiation flux at surface in FLN site against ground data: (a) estimates provided by persistence method, (b) estimates provided by model Eta/CPTEC, (c) estimates provided by ANN-36p, and (d) estimates provided by ANN-8p.

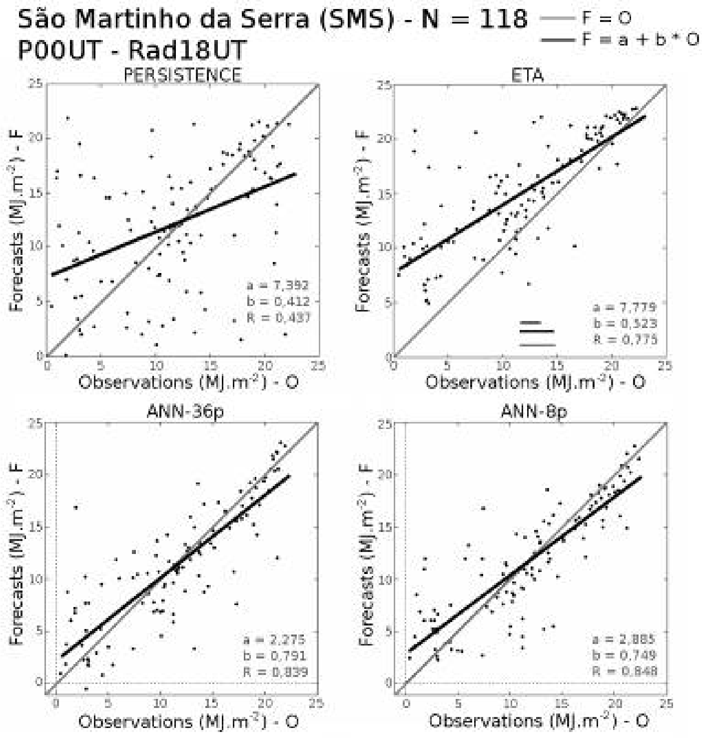

Figure 5. Scatter-plots of solar estimates for radiation flux at surface in SMS site against ground data: (a) estimates provided by persistence method, (b) estimates provided by model Eta/CPTEC, (c) estimates provided by ANN-36p, and (d) estimates provided by ANN-8p.

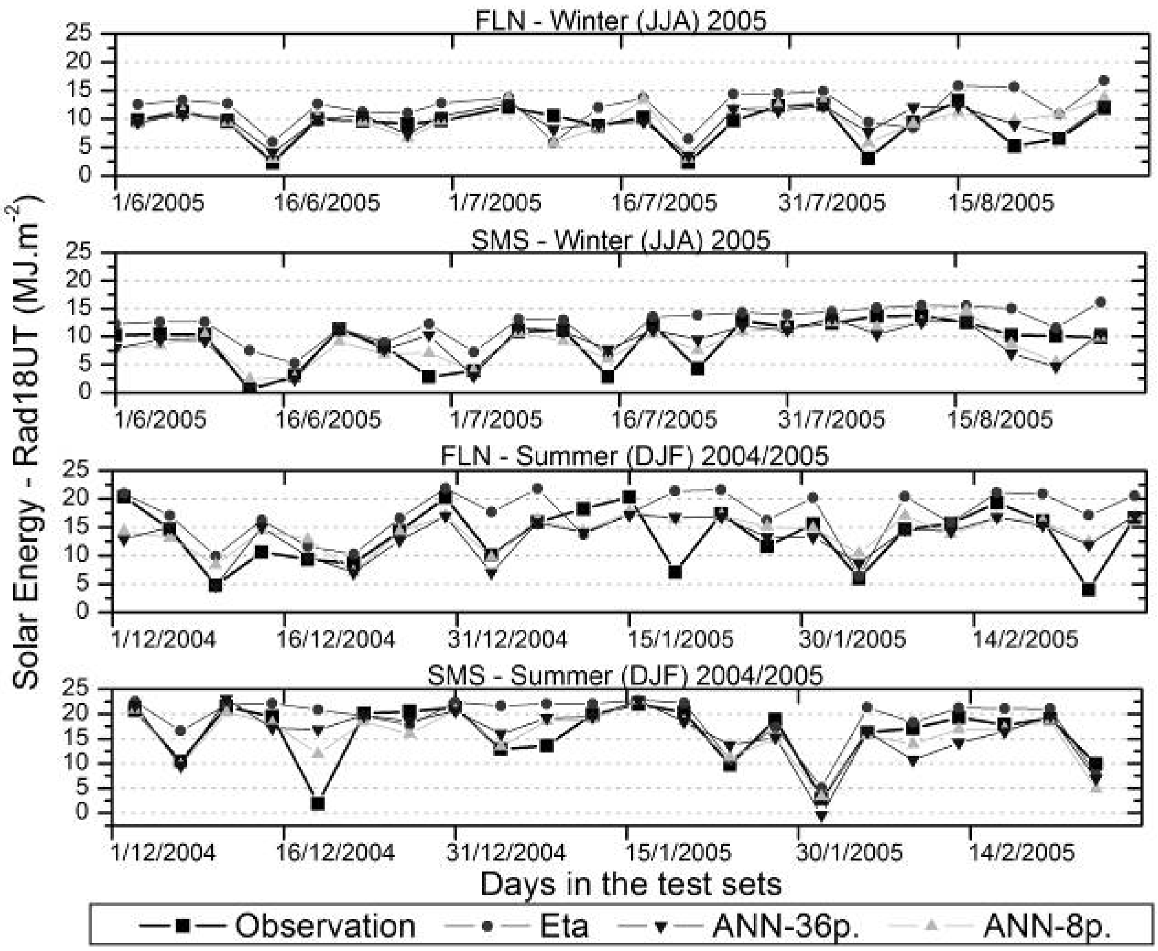

Figure 6. Temporal series comparing estimates and ground data for solar radiation flux at surface in both FLN and SMS sites. The series corresponds to days in the investigation dataset.

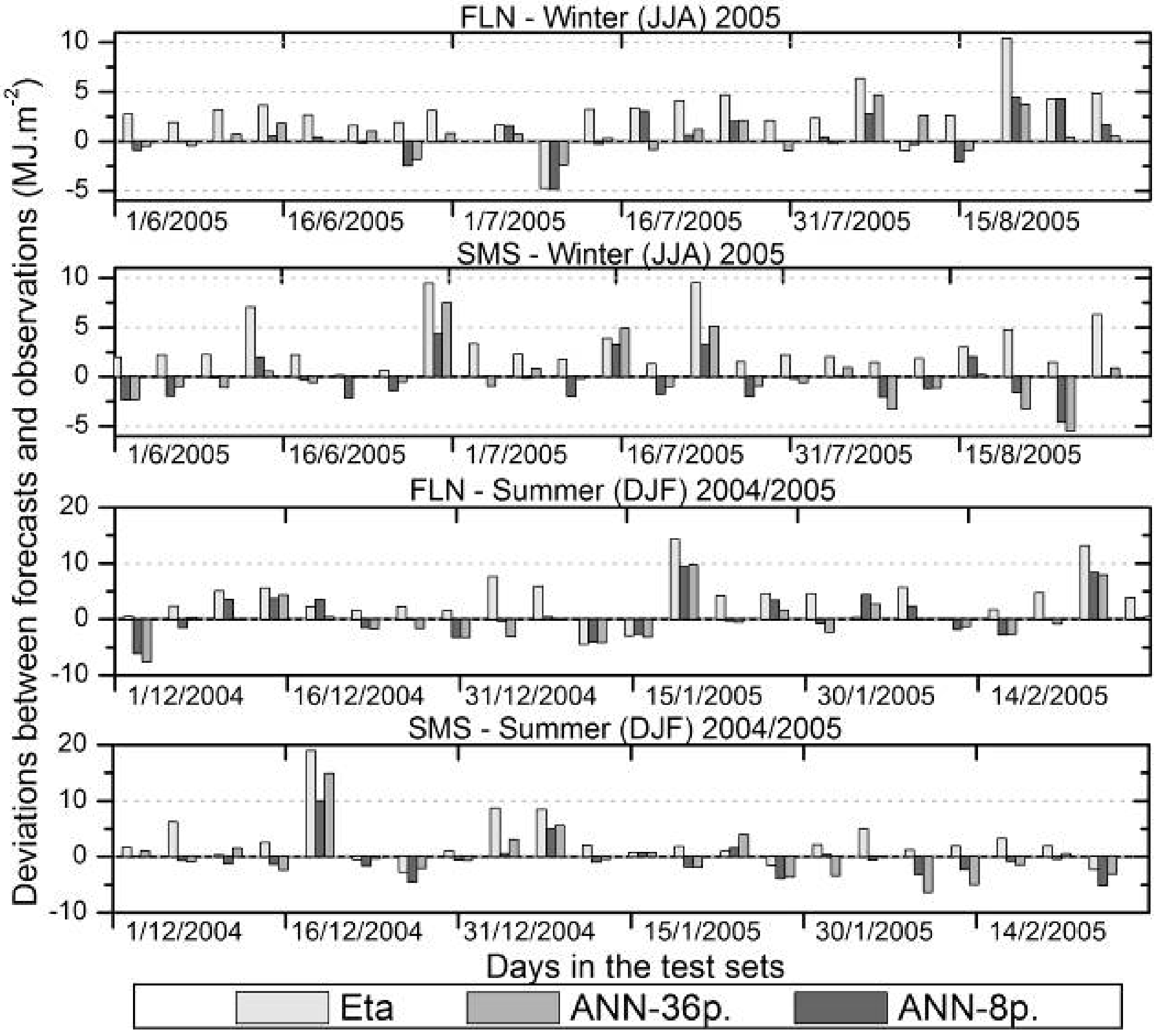

Figure 7. Deviations among estimates and ground data obtained for the same days presented in Figure 6.