Renewable Energy 50 (2013) 168–176

Contents lists available at SciVerse ScienceDirect

Renewable Energy

journal homepage: www.elsevier.com/locate/renene

On the use of niching genetic algorithms for variable selection in solar radiation estimation

A. Will [a,d,*], J. Bustos [a,d], M. Bocco [b], J. Gotay [a,d], C. Lamelas [c]

[a] Facultad de Ciencias Exactas y Tecnologías, Universidad Nacional de Tucumán, Argentina
[b] Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba, Argentina
[c] Estación Experimental Agroindustrial Obispo Colombres, Tucumán, Argentina
[d] Centro de Investigaciones en Tecnologías Avanzadas de Tucumán, U.T.N., F.R.T., Argentina

ARTICLE INFO

Article history:
Received 22 December 2010
Accepted 18 June 2012
Available online 20 July 2012

Keywords:
Niching genetic algorithms
Solar radiation
Variables selection
Meteorology

ABSTRACT

Prediction of climatic variables, in particular those related to wind and solar radiation, has developed a huge interest in recent years, mainly due to its applications to renewable energy. In many cases there is a large number of factors that influence the climatic variable of interest, and the researcher chooses the most relevant ones (based on previous knowledge of the region, availability, etc.) and runs a series of experiments combining the available data in order to find the combination that provides the best prediction.

In this work we present two applications of Niching Genetic Algorithms to solve the problem of selection of variables for the estimation of Solar Radiation. On one hand, this methodology is able to estimate a given climatic variable using databases with missing data, since the algorithm can compensate it by the use of others. On the other hand, we present a methodology that allows us to select the relevant input variables for a given climatic variable estimation or prediction problem, in a systematic way, using the same Genetic Algorithm with different parameters.

Both methods were tested in the estimation of daily Global Solar Radiation in El Colmenar (Tucumán, Argentina), using linear regression on data from 14 weather stations spread along the north of Argentina. The results obtained show that the methodology is appropriate, providing an RMSE = 2.36 [MJ/m^2] and R = 0.926 using an average of 64 out of 329 initial variables, on a 70 individuals/85 generations combination. For a 200 individuals/150 generations combination it obtained an RMSE = 2.34 [MJ/m^2] and R = 0.928 using an average of 54 variables.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Prediction of climatic variables, in particular those related to wind and solar radiation, has developed a huge interest in recent years, mainly due to its applications to renewable energy. Solar radiation is one of the climatic variables whose lack of data is quite generalized. Worldwide, about one in 500 weather stations make measurements of incident solar radiation Raichijk et al. [1]. Solar radiation affects the growth of crops and it is used in numerical models to estimate soil moisture, photosynthesis and potential evapotranspiration (Ball et al. [2]). There are many agricultural regions of Argentina lacking data on radiation and, they need to be estimated (Al-Alawi and Al-Hinai [3]).

Empirical models (Noia et al. [4] and Tovar and Baldasano [5]), statistical approaches coming from time-series analysis (Ji et al. [6]), linear regression (Lin and Gao [7], Ji et al. [6], Bocco et al. [8]), and neural networks (Al-Alawi and Al-Hinai [3], Mohandes et al. [9], Kalogirou et al. [10], Bocco et al. [11] and Rehman and Mohandes [12]) are some of the main tools that have been applied to estimate and predict the solar radiation that reaches the earth's surface. Studies report that linear regression has better performance than statistical models (Ji et al. [6]), and neural network outperforms linear regression (Bocco et al. [11], Bocco et al. [8]), but it takes a longer processing time. These methods have also been used to predict other weather phenomena (Brahm and Varas [13], Bilgili et al. [14], Kusiak and Li [15]).

In all the mentioned works, the selection of the models and the variables that are used for the prediction, is an analysis of combinations among climatic variables, based on a list of possibilities

* Corresponding author. Facultad de Ciencias Exactas y Tecnologías, Universidad Nacional de Tucumán, 4000 San Miguel de Tucumán, Argentina. Tel.: +54 381 154575737; fax: +54 381 4364093.
E-mail addresses: adrian.will.01@gmail.com, awill@herrera.unt.edu.ar (A. Will).

0960-1481/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.renene.2012.06.039

proposed by the researcher. Linear correlation is sometimes used (Al-Alawi and Al-Hinai [3], Kusiak and Li [15]), even though the relation among the variables might not be linear, as in Neural Networks. It is also reported that the quality of the estimates largely depends on the degree of correlation between the weather stations used. Moreover, the number of variables used in the model can be increased in an attempt to improve the accuracy of the prediction. The drawback is that it significantly increases the number of combinations, making it harder to determine which variables have the greatest influence on the climate variable to be estimated, and also increasing the number of tests needed.

The objective of this work is to use a niching genetic algorithm to determine which climatic variables and weather stations have the greatest influence in the estimation of a climatic variable. This allows an objective analysis of large amounts of data, as a previous step to other analysis that, like neural networks, would be prohibitively expensive. In our approach, all non-relevant variables can be eliminated, and the critical variables identified, greatly reducing the number of variables involved in a prediction or estimation.

It is important to distinguish between the proposed methodology and sensitivity analysis. Sensitivity analysis studies how error or uncertainty in the output variables can be related to error or uncertainties in the input variables (Saltelli et al. [16], (Saltelli [17])). Although more sophisticated methods exists, most methods found in the literature are local or one-at-a-time (Saltelli et al. [16]), and the technique is mainly used to determine which variable has the greatest influence in the output (see Varmuza and Filzmoser [18] and Vázquez Piqué et al. [19]). The proposed methodology is oriented to provide different combinations of the input variables that will obtain the prediction with the least possible error. The objective in this case is to identify the variables with the greatest influence in the model, and to allow the use of the prediction methodology when some of the variables are missing.

We apply the methodology to the case of using Linear Regression to estimate Global Solar Radiation for El Colmenar, Tucumán, Argentina, on the basis of data from 13 other weather stations spread along the North of Argentina.

2. Material and methods

2.1. Data collection

Data from 14 different weather stations were used in this work, corresponding to locations in the Northwest of Argentina, and from El Colmenar (province of Tucumán, Argentina). The data from the 13 stations were provided by the Centro de Informaciones Meteorológicas (S.M.N.), containing the following variables: day of year, daily average temperature [°C], air humidity [%], atmospheric pressure [mbar], cloudiness [x/8], and sunshine hours [%]. The data from El Colmenar were provided by Estación Experimental Agroindustrial Obispo Colombres, and the corresponding database contains: daily minimum and maximum temperature [°C], sunshine hours [%], and global solar radiation [MJ/m^2]). Both databases correspond to the period 01/01/2000–12/31/2005. See Fig. 1 and Table 1.

Despite the fact that these are well maintained stations, some of them showed lack of isolated data in some variables. This is due to the accidental interruption of the normal data acquisition process. This situation is to be expected in weather stations, and they tend to be fixed within a few days. In our case, less than 1% of the data was missing, in different variables and bases. Those data were filled up

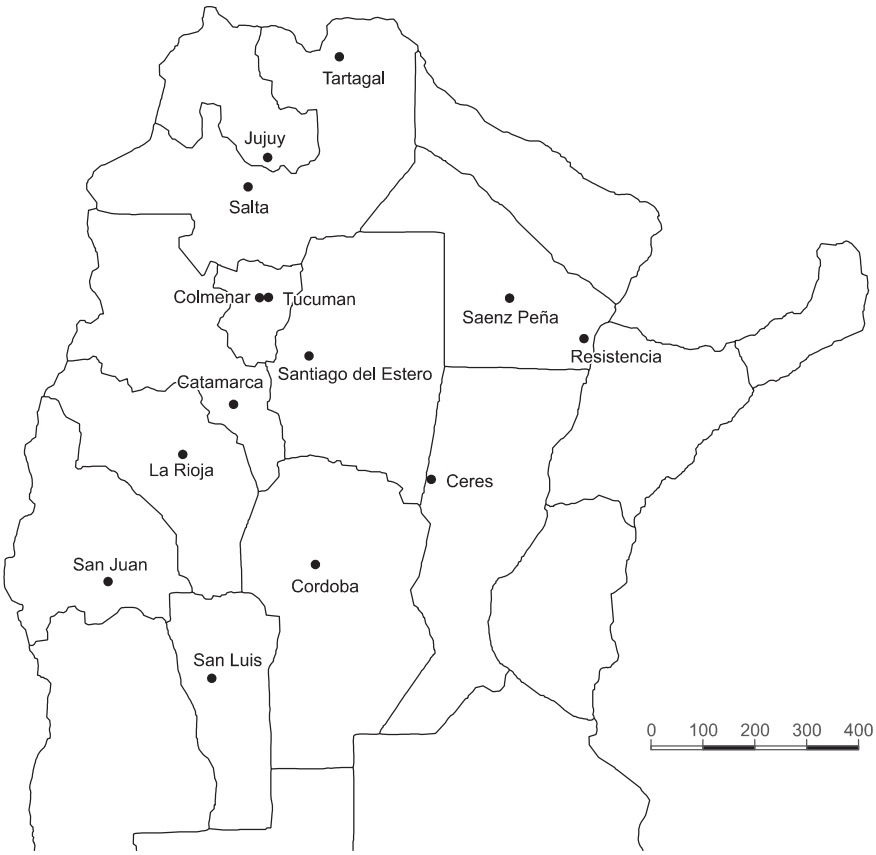

Fig. 1. Meteoroligical stations in NW Argentina.

دریافت فوری ← متن کامل مقاله

ISIArticles
مرجع مقالات تخصصی ایران

✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
✓ امکان دانلود نسخه ترجمه شده مقالات
✓ پذیرش سفارش ترجمه تخصصی
✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
✓ دانلود فوری مقاله پس از پرداخت آنلاین
✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات