**Annales
Geophysicae**

Determination of daily solar ultraviolet radiation using statistical models and artificial neural networks

F. J. Barbero[1], G. López[2], and F. J. Batlles[1]

[1]Departamento de Física Aplicada, Universidad de Almería, 04120, Almería, Spain
[2]Departamento de Ingeniería Eléctrica y Térmica, E.P.S., Universidad de Huelva, 21819, Huelva, Spain

Abstract. In this study, two different methodologies are used to develop two models for estimating daily solar UV radiation. The first is based on traditional statistical techniques whereas the second is based on artificial neural network methods. Both models use daily solar global broadband radiation as the only measured input. The statistical model is derived from a relationship between the daily UV and the global clearness indices but modulated by the relative optical air mass. The inputs to the neural network model were determined from a large number of radiometric and atmospheric parameters using the automatic relevance determination method, although only the daily solar global irradiation, daily global clearness index and relative optical air mass were shown to be the optimal input variables. Both statistical and neural network models were developed using data measured at Almería (Spain), a semiarid and coastal climate, and tested against data from Table Mountain (Golden, CO, USA), a mountainous and dry environment. Results show that the statistical model performs adequately in both sites for all weather conditions, especially when only snow-free days at Golden were considered (*RMSE*=4.6%, *MBE*= –0.1%). The neural network based model provides the best overall estimates in the site where it has been trained, but presents an inadequate performance for the Golden site when snow-covered days are included (*RMSE*=6.5%, *MBE*= –3.0%). This result confirms that the neural network model does not adequately respond on those ranges of the input parameters which were not used for its development.

Keywords. Meteorology and atmospheric dynamics (Radiative processes; General or miscellaneous) – Atmospheric composition and structure (Transmission and scattering of radiation)

1 Introduction

Ultraviolet (UV) radiation extends for wavelengths from 100 nm to 400 nm. The biological effects of UV radiation vary enormously with wavelength; by convention, the ultraviolet spectrum has been further subdivided into three regions: UV-C (100–280 nm), UV-B (280–315 nm) and UV-A (315–400 nm). The extraterrestrial ultraviolet solar spectrum contains only a small fraction (<8%) of the whole solar radiation reaching our planet. Of this fraction, the UV-A band comprises nearly 80% of the solar ultraviolet irradiance (Gueymard, 2004).

Stratospheric ozone content influences solar UV irradiance received at the Earth's surface in very different amounts through the UV range; e.g. the effective absorption cross section is 20 times less at 340 nm than at 300 nm (Molina and Molina, 1986). As a consequence, ground UV-B irradiance is only a small fraction, as low as 5%, of the whole ultraviolet irradiance. Total UV irradiance shows daily and yearly cycles that strongly depend on the geographic latitude and on other local conditions such, as altitude above the mean sea level, cloudiness, ground albedo, etc. The UV climatology at a specific site depends primarily on the time of day and day of the year (all equivalent to the air mass traversed by the solar radiation), secondly on cloudiness and thereafter on the type and amount of aerosols (Foyo-Moreno et al., 2003).

While most of the recent scientific studies have been devoted to the effects of UV-B radiation on humans, ecosystems and materials, numerous investigations also emphasize the important role of UV-A radiation:

– The location of the DNA damage in human skin suggests that long-wave UV radiation is an important carcinogen for the stem cells in the skin (Bachelor and Bowden, 2004; Agar et al., 2004). Also, photosensitivity reactions in human skin are promoted by UV-A radiation.

Correspondence to: G. López
(gabriel.lopez@die.uhu.es)

– Not only UV-B but UV-A radiation cause severe damage in some animal species in sub-artic climates (Zellmer, 1998), and both UV-B and UV-A radiation are potent modulators of the immune defence in fish (Salo et al., 2000).

– Solar UV is the main energy source driving the solar photocatalytic detoxification processes (Malato et al., 2003, 2004). On the other hand, actinic flux in the UV-A range is important in dissociating some chemical compounds in the troposphere (van Weele et al., 1995).

Due to the fact that the technologies to measure solar UV global radiation have been traditionally expensive, stations registering this spectral band are rather sparse; even when measured data were available, the series tend to cover medium or short time periods. Consequently, it is desirable to have tools to estimate it from other more commonly and routinely measured radiometric or climatological variables, as is the case of horizontal broadband global solar radiation.

Estimates of UV (295–385 nm) values from broadband global ones have been achieved by several authors in extensive literature, using hourly, daily, monthly mean hourly or monthly mean daily values. The literature reflects the wide range of hourly values for the terrestrial ratio UV/global found for different locations, varying from 2% to 8% (Baker-Blocker et al., 1984, Al-Aruri, 1988; Riordan et al., 1990). A part of this variability is explained by the strong spectral dependence of the atmospheric transmittance of the solar spectrum with the traversed optical air mass, which causes a differential behaviour of the incoming solar UV radiation and broadband global radiation. But also the ratio UV/global is affected by the local climatic and meteorological conditions. Absorption and scattering of the solar radiation by clouds and aerosols strongly influences the received solar spectrum at the surface (Ambach et al., 1991; Kerr, 2005; WCP 112, 1986). Surface UV intensity is reduced by clouds, although not to the same extent as the visible or infrared (IR) intensity. Clouds, as pure water, are a very weak absorber of UV radiation, and attenuate UV primarily by scattering. Also, cloud transmission depends on wavelength, increasing slightly towards shorter wavelengths, due to the increased multiple reflections between cloud and the air molecules (Seckmeyer at al., 1996). In short, both air mass and cloudiness affect the UV/global ratio in opposite ways; it diminishes with increasing optical air mass but reaches high values for high cloudiness. Since it is very difficult to separate both effects, the empirical models relating these variables become site specific and the value of the UV/global ratio for hourly or daily values cannot be adopted as a reference for the UV climatology. The use of this ratio for estimating UV radiation from measured global radiation should be restricted to the particular site where it was obtained.

An attempt to reduce the effect of the site and its climatology on the UV-global relationship involves the utilisation of hemispherical transparency indices (or clearness indices) K_{tUV} and K_t (Martínez-Lozano et al., 1994). These clearness indices are defined as $K_t = H_g/H_{g0}$ and $K_{tUV} = H_{UV}/H_{UV0}$, where H_g and H_{UV} are, respectively, the daily global solar broadband radiation (typically in the spectral range 0.3–3 μm) and the daily global solar UV radiation on horizontal surfaces at ground level, and H_{g0} and H_{UV0} are the corresponding horizontal daily extraterrestrial radiations. The extraterrestrial daily values in both spectral bands, H_{g0} and H_{UV0}, are defined by the conventional algorithms (Iqbal, 1983) for the geographical latitude of the site ϕ and for each day of the year (equivalent to the sun declination, δ):

$$H_{X0} = \frac{24}{\pi} I_{SC}(X) E_0 \left[\cos\delta \cos\phi \cos\omega_s + \frac{\pi}{180}\omega_s \sin\delta \sin\phi \right] (1)$$

where $I_{SC}(X)$ is the integrated value of the solar irradiation in the selected band (X = broadband global or UV spectral band), E_0 the relative Sun-Earth distance and ω_s the hourly angle of sunrise at the selected day.

Fairly linear or polynomial expressions, and also fuzzy inference techniques, relating the UV and the global clearness indices, have been derived for different sites for hourly and/or daily data. Nevertheless, all these studies show that absorption and scattering of solar radiation by clouds and aerosols affect the relationship between the clearness indices, K_{tUV} and K_t, showing a local dependence mainly on the dominant type and amount of clouds. Foyo-Moreno et al. (1999) derived empirical non-linear expressions relating both hourly clearness indices by means of fitting parameters which themselves depend on the relative optical air mass.

In this work we propose and test two different models in order to estimate daily UV values from only measured horizontal global solar broadband irradiation. In a first step and using traditional statistical techniques, we suggest an alternative relationship between both daily clearness indices, which includes the dependence on the relative optical air mass. Next, a further investigation into the relationships between UV radiation and different radiometric and atmospheric variables has been undertaken by means of artificial neural networks (ANNs), which have themselves shown to be very promising supervised learning tools for modelling complex nonlinear relationships. In particular, ANNs are beginning to be successfully applied to the estimation of solar radiation components and integrated spectral bands (López et al. 2000, 2001, 2005a; Schwander et al., 2002; Bosch et al., 2005). The aims in using ANNs are various:

– to determine whether the spread existing between UV and global irradiation around the linear relationship can be explained by means of a more complex nonlinear relationship,

– to determine whether additional radiometric and/or atmospheric variables are able to improve the UV estimation, and

Table 1. Description of the radiometric stations and their databases.

Site	Period	Valid days	Coordinates and elevations		
Almería (Spain)	1993–1997	1038	36°51' N	2°23' W	15 m amsl
Golden, CO (USA)	2001–2003	759	39°34' N	105°4' W	1829 m amsl

– to determine whether neural network models are able to perform better than those developed using standard statistical techniques.

To answer these questions the Bayesian method, known as Automatic Relevance Determination (ARD), and applied to neural networks (MacKay, 1994a, 1994b; Neal, 1996), has been utilized. This method allows ANN to determine the relative importance of the various input variables considered in calculating UV values, leading to the removal of those inputs which do not significantly explain any variance in the UV measurements.

2 Experimental set-up and databases

Data from two different radiometric stations were used. The main characteristics of the stations are summarized in Table 1, including geographic coordinates and altitudes. Latitude is similar in both cases but the sites are located at different altitudes and with different climatic conditions. The Almería station is located on a seashore site on the Mediterranean coast in southeastern Spain. The Golden (Colorado, U.S.) radiometric station is located at the top of the Southern Table Mountain; the climate is continental, with generally low relative humidity.

2.1 The Almería database

The Almería database includes measurements on a horizontal surface of global solar UV irradiance (in the spectral range 295–385 nm) by means of an UV Eppley radiometer model TUVR, broadband global and diffuse irradiances (using Kipp & Zonen pyranometers, model CM11, one of them equipped with a Eppley shadow band model SBS, 300–3000 nm), photosynthetically active radiation –PAR– Q (using a LI-COR silicon quantum sensor, 400–700 nm) and atmospheric downward longwave irradiance L_{DW} (using a Eppley pyrgeometer, model PIR, 3.5–50 μm). Air temperature T and relative humidity Φ are also available. Data were recorded at 20-s intervals and then averaged every 5 min. The TUVR values were temperature corrected following the calibration provided by the manufacturer, and diffuse radiation measurements were corrected following Batlles et al. (1995), to take into account the sensor screening by the shadow band. Daily values for all the radiometric variables were obtained by integration during the day.

Next, the database was extended with the daily clearness indices. The extraterrestrial daily integrated values in both bands, H_{g0} and H_{UV0} (in MJ m^{-2} day^{-1}), were obtained by using Eq. (1), where the $I_{SC}(X)$ values were derived from the AM0 synthetic spectrum from Gueymard (2004). The calculated solar constant for this AM0 spectrum is 1366.1 W m^{-2}. In the UV range, the AM0 spectrum is primarily based on the ATLAS-3/SUSIM spectrum, resulting finally in a spectral resolution of 0.5 nm. The equivalent ultraviolet solar constant in the 295 nm to 385 nm band is I_{SC}(UV)=80.83 W m^{-2}.

The relative optical air mass m_r was also added to the database. This parameter is defined as the ratio between the optical mass at the actual zenith distance and the value at the zenith of the site. In first approximation, for a plane-parallel atmosphere, the relative optical air mass may be written as:

$$m_r = \frac{1}{\cos \theta_z}, \qquad (2)$$

where θ_z is the solar zenith angle in degrees. As we are concerned with the calculation of daily integrated values we used the relative optical air mass at noon (corresponding to the minimal zenith angle in the day), where the irradiance is highest and therefore the contribution to the daily dose is largest. For very low Sun angles an expression for the air mass including the Earth curvature effect should be used (Kasten and Young, 1989). Equation (2) is applicable to a standard pressure of 1013.25 mbar at sea level; for other pressures, it should be modified (Iqbal, 1983). In the case of Almería this correction is not needed. For Golden, which is located at 1829 m above the mean sea level, the atmospheric pressure is reduced around 20%. The relative optical air mass should then be corrected. However, this is not needed since we have seen that the effect of the site altitude on the models is taken into account by means of an increase in global irradiation.

Finally, precipitable water content w_p, daily sunshine fraction S, daily diffuse fraction K and daily direct transmittance K_b were also calculated and added to this database. These were considered to analyse the dependence of the UV-global relationship on other radiometric and atmospheric variables.

For the purpose of developing and testing the models, the whole Almería database was divided into two subsets following a uniformly random distribution. The training data set contained 725 days and the test data set comprised 363 days.
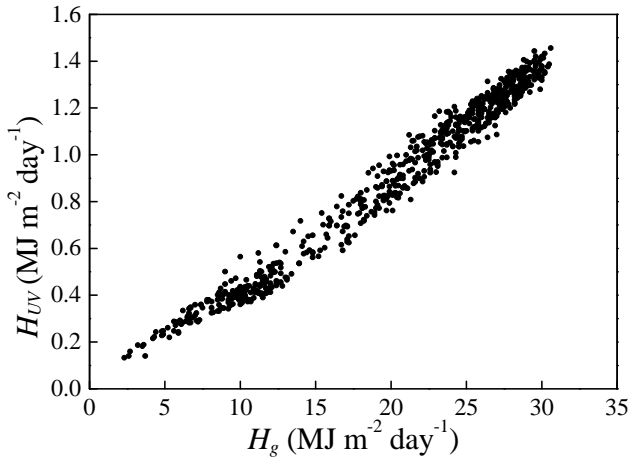
Fig. 1. Daily solar UV radiation against global irradiation for the Almería train data set.
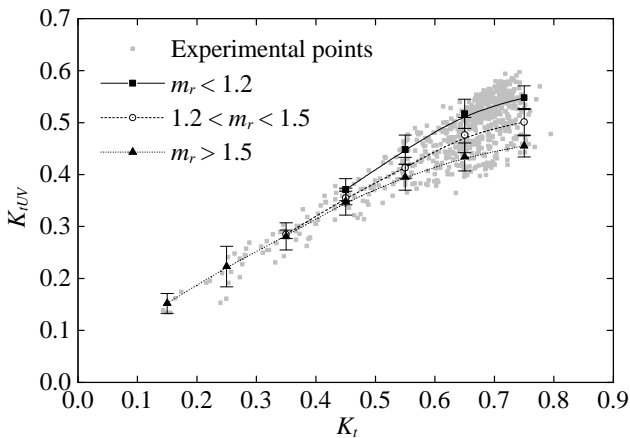

Fig. 2. Daily UV clearness index against daily global clearness index for the Almería train data set. Overlayered lines are for mean values of extreme low and high air masses (see text).

2.2   The Golden database

Database from the Golden radiometric station contains records from broadband horizontal upwelling and downwelling global solar irradiances, measured by two pyranometers Eppley (model PSP). Horizontal global solar UV irradiance was obtained by means of a radiometer Kipp & Zonen, model UV-S-A-T. The broad spectral response of this radiometer corresponds to (315–400) nm. However, the calibration factor was modified to obtain UV values in the spectral band (295–385) nm. This was achieved by comparing the radiometer signal output (mV) to the integrated spectral irradiance between 295 nm and 385 nm, measured by an Optronic Laboratories OL-754 (double monochromator UV spectroradiometer).

Assuming that the relative spectral irradiance of the Sun is invariant with respect to solar elevation and atmospheric

conditions, the calibration factor can also be assumed to be constant. Comparisons of solar spectra recorded at different solar elevations and sky conditions have demonstrated that this assumption approximately holds for the wavelength region 330 nm to 400 nm (UV-A) but does not hold for the UV-B region (280–315 nm) (Johnsen et al., 1997). Since, as was noted in the Introduction, the UV-B radiation of the solar spectrum accounts for approximately 5% of the total UV radiation (280–400 nm), the calibration factor will therefore be relatively insensitive to spectral changes in the UV-B.

Daily values were calculated from 3-min averages. The clearness indices, the relative optical air mass during the day and the local ground albedo were also computed. Ground broadband albedo is obtained from the upwelling and downwelling global solar irradiances. This parameter is useful to detect days affected by surface snow cover. After screening out those suspicious readings, we were left with 759 valid days. The number of days with ground albedo lower than 0.3 was 669. We will refer to these days as snow-free days.

In order to evaluate the predictive capability of the derived models, we used the performance parameters mean bias error, *MBE,* and root-mean square error, *RMSE,* defined as:

$$MBE = \frac{1}{N} \sum_{i=1}^{N} (y_i - x_i) \qquad RMSE = \frac{1}{N} \sum_{i=1}^{N} (y_i - x_i)^2, \quad (3)$$

where y_i are the predicted values, x_i the measured ones and N the number of data. The respective percent values are obtained dividing these expressions by the mean value of the measured variable.

3 K_{tUV}-K_t model

The scatter plot between daily solar UV and global irradiations for the Almería training data set is shown in Fig. 1. A clear linear dependence is observed between both variables, although a remarkable dispersion is noted in almost all the range of H_g values. A linear fit between H_{UV} and H_g values for the training data at Almería, forced to pass through the origin, results in the relationship H_{UV}=0.0450 H_g, with performance parameters *MBE*=1.0% and *RMSE*=6.8%. The corresponding slope value for Golden is 0.0446. We may see that, in both stations, the mean daily irradiation H_{UV} is nearly 4.5% of the mean daily H_g value. Hereafter, we refer to this simple relationship between daily irradiation values as model A.

Figure 2 shows the scatter plot between the clearness indices K_{tUV} and K_t for the Almería training data set. Although as shown, there is a close relationship between the K_{tUV} and K_t indices for low K_t values, a high spread of the experimental data points is evident for clear or near-clear skies. A linear fitting of the whole data through the origin provides a slope (mean K_{tUV}/K_t) of 0.74 at Almería and 0.71 at Golden. In both cases, the coefficient of determination r^2 was around 0.9.

Figure 2 also includes the averaged K_{tUV} values on intervals of 0.1 steps in K_t for three ranges of the relative optical air mass, along with their standard deviations. It is observed that K_{tUV} values, corresponding to low m_r values (associated with summer months), are clustered in the upper region of high values of K_t. As m_r increases, K_{tUV} values shift to lower regions and expand to lower values of K_t. It may be thus considered that this $K_{tUV} - K_t$ diagram simultaneously reflects the influence of both the cloudiness and the Sun position (through the relative optical air mass) on the clearness indices. As a consequence, we suggest including the additional dependence on the relative optical air mass into the $K_{tUV} - K_t$ relationship in order to explain its dispersion.

The proposed model is based on the parameterization of the daily clearness indices in the form: $\log K_{tUV} = a + b \log K_t$, along with an explicit dependence on the fitting parameters on the relative optical air mass. This allows the ultraviolet clearness index to be a function of the global clearness index but modulated by the relative optical air mass.

The fitting procedure has been carried out on the Almería training database in ranges of m_r. Results show that the a parameter of the fit clearly exhibits an strong inverse linear dependence on the logarithm of the optical air mass ($r^2 = 0.99$), while the values of the b parameter are not systematic in m_r, being represented as $b = 0.77 \pm 0.01$. The fitting process results in the following expression:

$$K_{tUV} = 0.712 \, K_t^{0.771} \, m_r^{-0.254} \qquad (4)$$

with $RMSE = 4.8\%$ and $MBE = -0.3\%$. The statistical parameters indicate that this model (hereafter model B) clearly improves the estimates by the previous simple model, using the daily global irradiation as the only measured input. Figure 3 illustrates the fit of Eq. (4) to the K_{tUV} values. It is observed that estimated values are distributed around the line 1:1 of perfect fit for all the range of measured K_{tUV} values. The existing dispersion underlines the need for additional predictive variables or a more complex relationship between variables involved.

Daily H_{UV} values may now be calculated from the K_{tUV} clearness index obtained from Eq. (4) and the corresponding extraterrestrial irradiation values for that day. The ability of this method to estimate daily ultraviolet values for the Almería test data set and the other stations is evaluated in Sect. 5.

4 Neural network model

Artificial neural networks were implemented using a combination of custom-designed MATLAB functions (The MathWorks, Natick, Mass.) and the Netlab (Nabney, 2001) toolkit for MATLAB. A standard multilayer perceptron (MLP) architecture with three fully interconnected layers (input, hidden, and output) was employed. The hyperbolic tangent
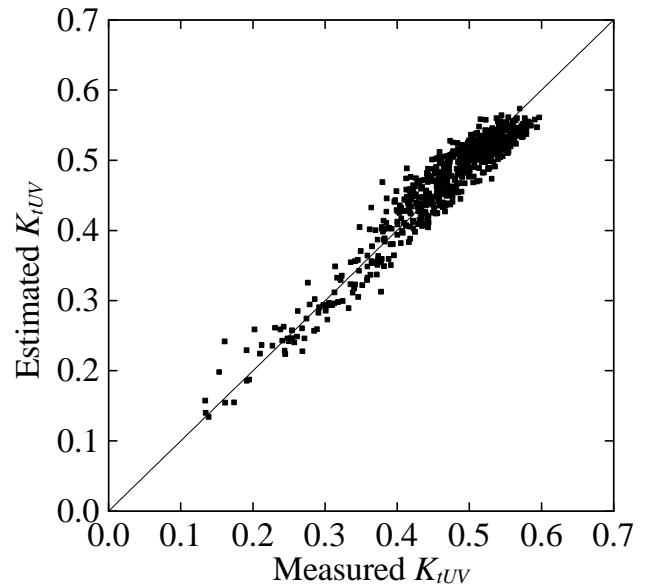

Fig. 3. Estimated K_{tUV} values by model B against measured ones.

transform was the nonlinear activation function in the hidden layer, and the identity function was selected as the activation function in the output layer. Such a network determines a nonlinear mapping from an input vector (radiometric and atmospheric variables) to the output (UV irradiation) parameterized by a set of network weights. All of the ANNs were trained with the scaled conjugate gradient algorithm (Møller, 1993; Marwala, 2001).

The ability of ANNs to detect complex, nonlinear mappings between dependent and independent variables without any a priori information about the existing relationship between them, has led us to analyse firstly the influence of several radiometric and atmospheric variables on UV radiation, and to try to improve the results given by Eq. (4). These are global irradiance H_g, daily clearness index K_t, daily diffuse fraction K, daily direct transmittance K_b, atmospheric downward longwave irradiation L_{DW}, global PAR Q, sunshine fraction S, relative optical air mass m_r, temperature T, relative humidity Φ and precipitable water content w_p. The use of global irradiance, daily clearness index and relative optical air mass has been well founded in the previous sections. The use of daily diffuse fraction and direct transmittance tries to take into account the attenuating effects of clouds and atmospheric constituents. The sunshine fraction is introduced as an alternative variable related to the presence of clouds, as well as atmospheric longwave irradiation (Goforth et al., 2002; Iziomon, 2003; Galli, 2004). PAR is considered in the analysis if this spectral region improves estimates, since visible wavelengths are attenuated by water vapour absorption to a much lower degree than the broadband global radiation. Temperature, relative humidity and precipitable water accounts for the effect of water vapour on the transmission
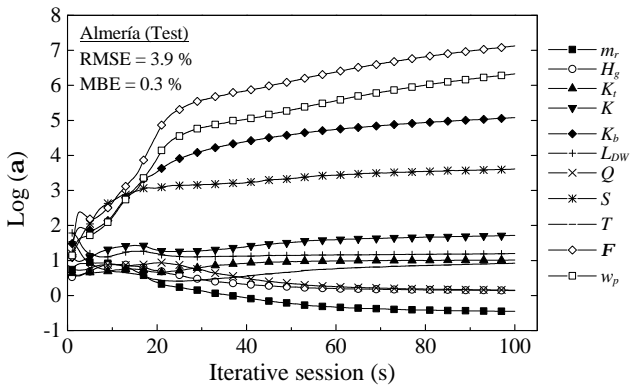
Fig. 4. ARD results.

of global solar radiation (McArthur et al., 1999; Fioletov et al., 2003). Notice that several variables provide equivalent information or are not independent of each other (e.g. daily clearness index, diffuse fraction and direct transmittance, or temperature, relative humidity and precipitable water). The reason for their inclusion as inputs is to analyse which of them are more suitable for the estimation of UV by means of ANN.

To analyse the relevance of the inputs considered, we used the Bayesian technique of Automatic Relevant Determination (ARD) for MLP networks (MacKay, 1994a, 1994b; Neal, 1996). This is based on a probabilistic interpretation of network training, providing a useful method for assessing the relevance of input parameters, as well as for avoiding overfitting (Thodberg, 1996; Vivarelli and Williams, 1997; Penny and Roberts, 1999; Lampinen and Vehtari, 2001). ARD method uses multiple regularization constants, called hyperparameters α, one associated with each input. By applying Bayes' theorem, the regularization constants for non-informative inputs are automatically inferred to be large, preventing those inputs from causing overfitting. Determination of hyperparameters is undertaken during the network training by means of the evidence procedure (MacKay, 1994a). The overall network training then consists of two steps: the first updates the ANN weights according to the Scaled Conjugate Gradient (SCG) algorithm during a fixed number of training cycles, after which the hyperparameters are updated in a second re-estimation step. These two steps are carried out in "s" number of iterative sessions.

The Almería training data set was employed for analysing the input relevance. No validation data subset is used since the ARD method prevents itself from overfitting. Due to the different intervals exhibited for each input and output variables and since weights are bounded by a regularizer term, all variables were first normalized to have zero mean and unity variance, and after were linearly scaled in the interval [0.1, 0.9]. The scaling was performed using the minimum and maximum values for each input and output parameter. This pre-processing of the data is only applied to the train-

ing data. The ANN weights are rescaled after training, so that the network can work on unscaled data. The training using the ARD method was carried out considering an ANN with three hidden neurons. One hundred iterative sessions were undertaken, each one consisting of one hundred SCG training cycles for the first step, and four hyperparameter re-estimations for the second step. These values were selected based on previous studies (López et al., 2005a).

Figure 4 shows the log-values of the hyperparameters α versus the number of the iterative session, s. We have also included the values of the statistical tests *RMSE* and *MBE* from the regression analysis between estimated and measured UV values corresponding to the Almería test data set. They are expressed as a percentage of the corresponding mean value of the measured UV irradiation. The overall performance of such a complex model as that given by this neural network indicates that UV irradiation may be calculated with *RMSE* $\sim$ 4% and a *MBE* $\sim$0%. However, results indicate that relative humidity Φ, precipitable water content w_p, daily direct transmittance K_b and daily sunshine fraction S explain nothing of the measured UV irradiation variance, as they present the higher values of the hyperparameters. Variables accounting for the water vapour absorption on the solar broadband radiation, relative humidity and precipitable water are thus superfluous in the estimate of daily UV radiation. This finding could be due to the use of data corresponding to all weather conditions, where clouds dominate the solar radiation attenuation against the effect of water vapour absorption, along with the inclusion of other more predictive variables.

In contrast to these inputs, relative optical air mass m_r, global irradiation H_g and photosynthetically active radiation Q are seen to be the more important parameters. This result is consistent with the existing correlation between UV and global total irradiation. The inclusion of photosynthetically active radiation Q as a relevant input variable may be related to its strong correlation with global H_g, since the selective effect of water vapour absorption under all weather conditions was not sensitive to the ANN and PAR relevance is not higher than that exhibited by global irradiation. In this case, this variable would be redundant.

Finally, daily diffuse fraction K, daily clearness index K_t, air temperature T and longwave radiation L_{DW} show similar relevance, with values of the hyperparameters around ten times higher than that from relative optical air mass m_r. A second analysis is needed to assure their inclusion in the ANN model. For this purpose, several ANN training trials, taking different input configurations and using two hidden neurons, were performed. Results indicated that the inclusion of the daily diffuse fraction, atmospheric downward longwave radiation and temperature allowed no improvement in the UV estimates. The small relevance for air temperature and atmospheric longwave radiation derived from the ARD method may be due to their relationship to the relative optical air mass (which is the most important input parameter) and therefore act as a proxy of m_r. No additional

Table 2. Statistical results of the model performances in percents.

Model	Almería Test		Golden (all days)		Golden (snow-free days)	
	RMSE	MBE	RMSE	MBE	RMSE	MBE
$A : H_{UV} = f(H_g)$	6.0	1.0	6.9	0.5	6.4	1.0
$B : H_{UV} = f(K_t, m_r)$	4.7	0.0	5.5	–1.0	4.6	–0.1
$ANN: \{m_r, H_g, K_t\}$	4.4	0.0	6.5	–3.0	5.1	–2.0

information with regards to water vapour absorption or cloud effects could be explained by these inputs. It should be noted that the removed variables could be relevant if a data set containing records from sites with different climatic conditions was used.

Once the input variables are determined $\{m_r, H_g, K_t\}$, the next step is to obtain a simple ANN model (in the sense of minimum number of weights) which is able to generalise the relationship which exists between inputs and output. To search the minimum number of hidden neurons needed for an optimal NN model performance, several training trials were carried out, increasing the number of hidden neurons from 2 to 10. In addition, to overcome potential different model performances related to the random initialization of the weights, for each ANN architecture, ten training runs were accomplished and the performance values (*RMSE* and *MBE*) were the mean. Results show that an ANN model with two hidden neurons is enough to perform satisfactorily. This small number of hidden neurons indicates that the relationship between UV and the input variables selected is very simple and therefore a very simple equation like Eq. (4) would be sufficient to estimate UV.

5 Analysis of model performances

Table 2 displays the global performance of the models for the two sites in terms of the statistical parameters *RMSE* and *MBE* in percent values. In this table, the first row corresponds to the simpler model, derived from the linear relationship between the irradiation values in the UV and global bands (model A). We decided to keep the information from this model to show how, even when applied to places with very similar mean UV/global ratios, the linear $H_{UV} - H_g$ model is not enough to accurately estimate H_{UV} values. The second and third rows correspond to the $K_{tUV} - K_t$ model and the ANN based model, respectively.

We can see that model B provides the best overall performance. If only snow-free days in Golden are considered, the *RMSE* values are around to 4.6% and almost null *MBE* values at both locations. We see how the introduction of a more suitable parameter as m_r in model B, leads to improved perfor-
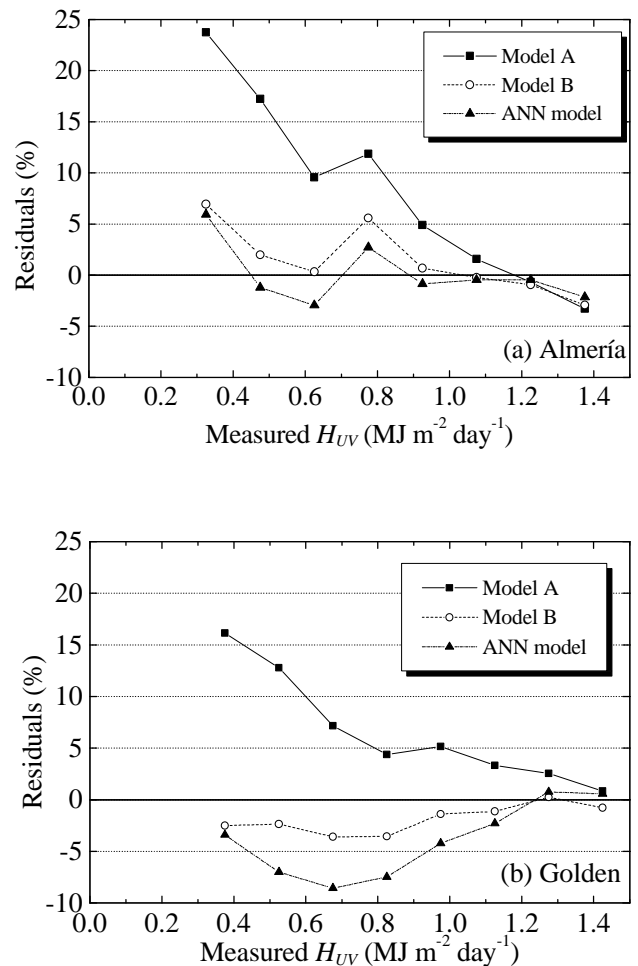

Fig. 5. Average relative differences (expressed as a percentage of the measured mean value) for each model and site under clear-sky conditions.

mance of the linear model A, which presents *RMSE* values of about 6% and a slight overestimation of 1% at both sites. The ANN model presents statistical results similar to the model B at Almería, but performance deterioration is observed at Golden, where an underestimation of 2% is present. If days
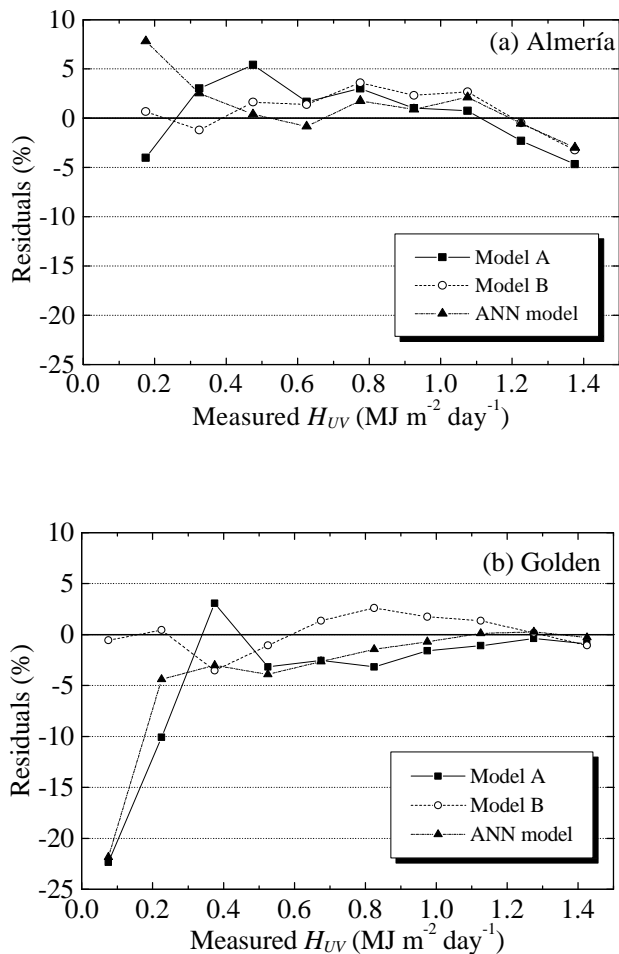
Fig. 6. Average relative differences (expressed as a percentage of the measured mean value) for each model and site under partially cloud and overcast conditions.

with snow are considered at Golden, a systematic additional underestimation of around 1% with respect to the model performances using snow-free days is found in both model B and the ANN model. *RMSE* values are also increased as a consequence of these higher underestimations. This result indicates the higher increasing effect of snow cover on UV than on global solar radiation (McArthur et al., 1999). Statistical models based on the UV-global relationship should therefore include snow cover information to avoid being site dependent. Although both statistical and ANN models are derived from Almería data, these results shows that the ANN model may be considered as more site dependent than model B, accurately estimating the UV values in sites whose data have been used for its development or with similar statistical characteristics.

To complete the study of the performance of the models, we have analysed the residual differences (calculated as estimated minus measured UV values) against measured H_{UV} and for different sky conditions, using snow-free days at

Golden. Figures 5 and 6 show the residuals expressed as a percentage of the mean value of the measured UV, respectively, for clear-sky[1] days and for partially and totally cloudy days. Under clear-sky conditions (Figs. 5), it can be observed that the simple linear model A presents the higher relative deviations in both sites, overestimating clear-sky daily UV irradiation for low to intermediate values.

Models B and ANN behave similarly to each other for Almería clear skies, with a good general fitting, but overestimating for winter clear days. At Golden, both models perform satisfactory for higher clear-sky H_{UV} values, but underestimate for intermediate clear-sky H_{UV} values. This underestimation is more noted for the ANN model. This is mainly due to the existing differences of the relationship between clear-sky H_{UV} and m_r at Almería and Golden, respectively. In this regard, measured intermediate clear-sky H_{UV} values are typically 10% higher at Golden than at Almería for the same m_r interval. However, the measured clear-sky H_g values are also increased at Golden with respect to Almería. In model B, the underestimation caused by the differences in the $H_{UV} - m_r$ relationship is partially neutralized by the increase in the global irradiation values. However, the ANN model is less sensitive to variations in H_g and K_t values than to variations in m_r, leading to the observed higher underestimation.

In Fig. 6 we can see the behaviour of the models for partially cloudy and overcast days. Under these conditions, overall the three models exhibit similar residual trends at each site, the differences found between model A and the other ones being reduced under clear skies. This is because the major differences in the data points between both sky conditions in the $H_{UV} - H_g$ diagram are seen for intermediate and low values of global irradiation, where the percentage of data corresponding to clear skies becomes lower. This will balance the performance of model A so as to perform better under partially cloudy and overcast days and for days with high daily UV values, as is observed in Figs. 5 and 6. In more detail, model A seems to be an adequate estimator for intermediate H_{UV} values, but with higher deviations in both lower and higher values, suggesting that the relationship between H_{UV} and H_g is not strictly linear but rather dependent on the local cloudiness.

Model B tends to overestimate for intermediate H_{UV} values (nearly 4%). The ANN model exhibits adequate performance for the two stations and in all the observed H_{UV} ranges, with the unique exception of the lowest H_{UV} values, corresponding to overcast conditions, where the model presents the highest relative residuals. These different ANN model trends in the two locations for low daily ultraviolet irradiation values (overcast conditions) are related to a poor fitting by this model in this H_{UV} region as a consequence of

---

[1] Clear-sky days were selected following the clear-sky parameterization of the daily global radiation given by López and Batlles (2005b).

the very low number of data available at Almería (less than 5% of all the data) for training the neural network successfully.

6 Conclusions

Two different methodologies, based on traditional statistical techniques and artificial neural networks, have been tested in order to estimate daily horizontal solar ultraviolet values from measured horizontal broadband global ones. The common and simple linear relationship between UV and global values has shown to be very deficient for estimating daily UV values under clear-sky conditions, with larger overestimations for non summer months. However, it appears as an adequate alternative if only cloudy or partially cloudy conditions are involved. Inclusion of the relative air mass, along with a new reformulation of the relationship between UV and global irradiations, but based on the corresponding clearness indices, has allowed one to correct of the drawbacks presented in the previous algorithm, and to improve the estimates of UV values under all-sky conditions. In addition, this new simple formulation has shown to be valid for sites with disparate climatology. Nevertheless, there is evidence of a variable or correction factor in the models to account for the presence of surface snow covers.

Finally, in order to improve the UV estimates and explain the existing point dispersion in the $H_{UV} - H_g$ diagram, the neural network method based on the automatic relevance determination was used. A large number of potential predictive variables were analysed. Relative optical air mass, global irradiation and clearness index were shown to be the more relevant predictors. Variables related to the atmospheric attenuation of incoming solar radiation by water vapour and aerosols were shown to be irrelevant. The ANN based model developed using the three relevant input variables consisted of a multilayer perceptron network with two hidden neurons. This extremely small number indicates that the relationship between UV and the input variables is very simple and that a very simple equation, like model B, is sufficient to predict UV. Moreover, the neural network model presents an inadequate performance on those ranges of the input parameters which were not used for its development. The use of the neural network methodology should thus be limited to those cases where large data sets covering all possible real atmospheric conditions were available.

References

Agar, N. S., Halliday, G. M., Barnetson, R. C., Ananthaswamy, H. N., Wheeler, M., and Jones A. M.: The basal layer in human squamous tumors harbours more UVA than UVB fingerprint mutations: A role for UVA in human skin carcinogenesis, Proceedings of the National Academy of Sciences, USA, 101, 4954–4959, 2004.

Al-Aruri, S., Rasas, M., Al-Jamal, K., and Shaban, N.: An assessment of global UV solar radiation in the range (0.290–0.385 μm) in Kuwait, Solar Energy, 41, 159–162, 1988.

Ambach, W., Blumthaler, M., and Wendler, G.: A comparison of ultraviolet radiation measured at an arctic and an alpine site, Solar Energy, 47, 121–126, 1991.

Bachelor, M. A. and Bowden, G. T.: UVA-mediated activation of signalling pathways involved in skin tumour promotion and progression, Seminars in Cancer Biology, 14, 131–138, 2004.

Baker-Blocker, A., DeLuisi, J. J., and Dutton, E.: Received ultraviolet radiation in the South Pole, Solar Energy, 32, 659–662, 1984.

Batlles, F. J., Olmo, F. J., and Alados-Arboledas, L.: On the shadow band correction methods for diffuse irradiance measurements, Solar Energy, 54, 105–114, 1995.

Bosch, J. L., López, G., and Batlles, F. J.: Correlating daily solar global irradiation values over complex terrain using artificial neural networks, Proceedings of the "International Forum of Experts in Solar Radiation, SOLARIS 2005, 2nd Joint Conference", Athens, Greek, May 26–27, 2005.

Galli, C., Nardino, M., Levizzani, V., Rizzi, R., and Georgiadis, T.: Radiative energy partition and cloud radiative forcing at a Po valley site, Atmos. Res., 72, 329–351, 2004.

Goforth, M. A., George W. Gilchrist, G. W., and Sirianni, J. D.: Cloud effects on thermal downwelling sky radiance, Proceedings of SPIE, Volume 4710, Thermosense XXIV, Xavier P. Maldague, Andres E. Rozlosnik, Editors, 203–213, 2002.

Gueymard, C.: The sun's total and spectral irradiance for solar energy applications and solar radiation models, Solar Energy, 76, 423–453, 2004.

Fioletov, V. E., Kerr, J. B., McArthur, L. J. B., Wardle, D. I. and Mathews, T. W.: Estimating UV index climatology over Canada, J. App. Meteorol., 42, 417–433, 2003.

Foyo-Moreno, I., Vida, J., and Alados-Arboledas, L.: A simple all-weather model to estimate ultraviolet solar radiation (290–385 nm), J. Appl. Meteorol., 38, 1020–1026, 1999.

Foyo-Moreno, I., Alados, I., Olmo, F. J., and Alados-Arboledas, L.: The influence of cloudiness on UV global irradiance (295–385 nm), Agr. Forest Meteorol., 120, 101–111, 2003.

Iqbal, M.: An introduction to solar radiation, Academic Press, 1983.

Iziomon, M. G., Mayerb, H., and Matzarakisb, A.: Downward atmospheric longwave irradiance under clear and cloudy skies: Measurement and parameterization, J. Atmos. Sol-Terr. Phy., 65, 1107–1116, 2003.

Johnsen, B., Futsæther, C., Dahlback, A., Grimnes, A., Hannevik, M., Thorseth, T. M., Kjeldstad, B., Lange, T., Skartveit, A., Opedal, L., Mikkelborg, O., and Storsveen, G.: The 1995 intercomparison of UV- and PAR instruments at the University of Oslo, StrålevernRapport 1997:7, Østerås: Norwegian Radiation Protection Authority, 1997.

Kasten, F. and Young, A. T.: Revised optical air mass tables and approximation formula, Appl. Optics, 28, 4735–4738, 1989.

Kerr, J. B.: Understanding the factors that affect surface UV radiation, Opt. Eng., 44, 1–9, 2005.

Lampinen, J. and Vehtari, A.: Bayesian approach for neural networks. Review and case studies, Neural Networks, 14, 257–274, 2001.

López, G., Martinez, M., Rubio, M. A., Tovar, J., Barbero, J., and Batlles F. J.: Estimation of the hourly diffuse fraction using a neural network based model, 2a Asamblea Hispano-Portuguesa de Geodesia y Geofísica, Lagos, Portugal, 8–12 February 2000, Ed. Instituto Geofisico do Infante D. Luis (Portugal), ISBN: 0870-2748 (in Spanish), 425–426, 2000.

López, G., Rubio, M. A., Martínez, M., and Batlles, F. J.: Estimation of hourly global photosynthetically active radiation using artificial neural network models, Agr. Forest Meteorol., 107, 279–291, 2001.

López, G., Batlles, F. J., and Tovar, J.: Selection of input parameters to model direct solar irradiance by using artificial neural networks, Energy, 30, 1675–1684, 2005a.

López, G. and Batlles, F. J.: A new simple parameterization of global solar radiation under cloudless sky conditions, Proceedings of the International Forum of Experts in Solar Radiation, SOLARIS 2005, 2nd Joint Conference, Athens, Greek, 26–27 May, 122–127, 2005b.

McArthur L. J. B., Fioletov, V. E., Kerr, J. B., McElroy, C. T. and Wardle, D. I.: Derivation of UV-A irradiance from pyranometer measurements, J. Geophys. Res., 104, 30 139-30 151, 1999.

MacKay, D. J. C.: Bayesian methods for backpropagation networks, in: Models of Neural Networks III, edited by: Domany, E., van Hemmen, J. L., and Shulten, 211–254, Springer-Verlag, New York, 1994a.

MacKay, D. J. C.: Bayesian non-linear modelling for the energy prediction competition, ASHRAE Transaction, 100, 1053–1062, 1994b.

Malato, S., Blanco, J., Vidal, A., Alarcón, D., Maldonado, M. I., Cáceres, J., and Gernjak, W.: Applied studies in solar photocatalytic detoxification: an overview, Solar Energy, 75, 329–336, 2003.

Malato, S., Blanco, J., Vidal, A., Maldonado, M. I., Fernández, P., Alarcón, D., Collares-Pereira, M., Farinha, J., and Correia de Oliveira, J.: Engineering of solar photocatalytic collectors, Solar Energy, 77, 513–524, 2004.

Martínez-Lozano, J. A., Casanovas, A. J., and Utrillas, M. P.: Comparison of global UV (290 - 285 nm) and global irradiation measured during the warm season in Valencia, Spain, Int. J. Climatol, 14, 93–102, 1994.

Marwala, T.: Scaled conjugated gradient and Bayesian training of neural networks for fault identification in cylinders, Comput. Struct., 79, 2793–2803, 2001.

Molina, L. T. and Molina, M. J.: Absolute absorption cross section of ozone in the 185 to 350 nm wavelength range, J. Geophys. Res., 95, 14 501–14 508, 1986.

Møller, M.: A scaled conjugate gradient algorithm for fast supervised learning, Neural Networks, 6, 525–533, 1993.

Nabney, I. T.: NETLAB: algorithms for pattern recognition, Springer, London, United Kingdom, 325–361, 2001.

Neal, R. M.: Bayesian Learning for Neural Networks, Lecture Notes in Statistic, Vol. 118, Springer, New York, 204, 1996.

Penny, W. D. and Roberts, S. J.: Bayesian neural networks for classification: how useful is the evidence framework?, Neural Networks, 12, 877–892, 1999.

Riordan, C. J., Hulstrom, R. L., and Myers, D. R.: Influence of atmospheric conditions and air mass on the ratio of UV to total solar radiation, SERI/TP-215-3895, U.S. Department of Energy, 28 pp, 1990.

Salo, H. M., Jokinen, E. I., Markkula, S. Eveliina, Aaltonen, T. M., and Pettila, H. T.: Comparative effects of UVA and UVB irradiation on the immune system of fish, J. Photoch. Photobio. B, 56, 154–162, 2000.

Schwander, H., Koepke, P., Kaifel, A., and Seckmeyer, G.: Modification of spectral UV irradiance by clouds, J. Geophys. Res., 107, D16, doi:10.1029/2001JD001297, 2002.

Seckmeyer, G., Erb, R., and Albold, A.: Transmittance of a cloud is wavelength-dependent in the UV-range, Geophys. Res. Lett., 23, 2753–2755, 1996.

Thodberg, H. H.: A review of Bayesian neural networks with application to near infrared spectroscopy, IEEE T. Neural Networks, 7, 56–72, 1996.

Van Weele, M., Vilá-Guerau de Arellano, J., and Kuik, F.: Combined measurements of UVA actinic flux, UVA irradiance and global radiation in relation to photodissociation rates, Tellus, 47B, 353–364, 1995.

Vivarelli, F. and Williams, K.I.: Using Bayesian neural networks to classify segmented images, Fifth IEEE International Conference on Artificial Neural Networks, 263–268, 1997.

WCP-112: World Climate Program. A preliminary cloudless standard atmosphere for radiation computation, WMO/TD-24, World Meteorological Organization, Geneva, Switzerland, 1986.

Zellmer, I. D.: The effect of solar UVA and UVB on sub arctic Daphnia pulicaria in its natural habitat, Hydrobiologia, 379, 55–62, 1998.