VISIBILITY ESTIMATES FROM ATMOSPHERIC AND RADIOMETRIC VARIABLES USING ARTIFICIAL NEURAL NETWORKS

GABRIEL LÓPEZ[1], JUAN LUIS BOSCH[1], INMACULADA PULIDO-CALVO[2] & CHRISTIAN A. GUEYMARD[3]
[1]Dpto. Ingeniería Eléctrica y Térmica, de Diseno y Proyectos, University of Huelva, Spain
[2]Dpto. Ciencias Agroforestales, University of Huelva, Spain
[3]Solar Consulting Services, Colebrook, NH, USA

ABSTRACT

Visibility is traditionally needed for air quality monitoring or air traffic control, and has become a key input to determine the transmission losses of solar radiation propagating between heliostats and the receiver of solar tower power (STP) plants. Recent studies suggest that haze can reduce visibility and increase these losses up to 25% compared to clear conditions. Monitoring visibility would thus be needed for proper design and operation of STPs, but this is usually not done at all potential sites. Here, visibility's magnitude and variability are analyzed in terms of more common atmospheric variables: temperature, humidity, pressure, wind speed and precipitable water. To that effect, 1-min observations from a visibility meter located in Huelva (southwestern Spain) are analyzed over a 2-month period. Relative humidity is linearly correlated with visibility and explains 40% of its variability. This correlation is strongest under cloudless and daytime conditions. Using standard statistical techniques, no significant correlation is found between visibility and other atmospheric variables. Artificial neural networks (ANN) are thus investigated here for mapping the complex and non-linear relationships between visibility and multiple atmospheric inputs. This improves results significantly, increasing the explained visibility variability up to 72% and reducing RMSE from 40% to 30%. Moderate improvements in visibility estimation are further obtained when radiometric information (direct and diffuse solar irradiances) are added as ANN predictors. These findings show that visibility can be estimated from local atmospheric and radiometric observations using ANN, despite the complex and non-linear relationships between them.
Keywords: visibility, atmospheric and radiometric variables, modeling, artificial neural networks.

1 INTRODUCTION

Visibility may be defined in colloquial terms as the maximum distance an observer can separate a structure from its background. Visibility degradation is caused by the extinction of light by particles and gases in the atmosphere, and is thus directly related to the total atmospheric extinction coefficient. This relationship allows visibility to be evaluated by indirect extinction measurements. According to the WMO recommendation, the meteorological optical range (MOR) is the more suitable parameter to express that relationship. MOR represents the greatest distance at which a black object of suitable dimensions, situated near the ground, can be seen and recognized when observed against a bright background.

Although visibility was first defined for meteorological purposes, visibility is nowadays of interest in several fields such as air quality monitoring, civil aviation or road traffic. After fog, haze (caused by aerosols such as pollutants, dust, smoke, etc.) constitutes one major cause of visibility reduction. Water vapor and aerosols close to the surface are the main atmospheric constituents contributing to the slant-path solar transmission losses occurring between the heliostats and the receiver in a solar tower power (STP) plant [1]. This has prompted the use of horizontal visibility as one key input parameter in computer codes that have been developed

for the design and operation codes of STP plants [2]. However, whereas information on the total atmospheric columns and atmospheric attenuation effects of water vapor and aerosols is usually available at STP locations, visibility is not measured at most actual or potential STP sites. Moreover, measurements of the vertical profiles of aerosols or water vapor are not fully representative of the actual surface distribution.

Due to this lack of information, visibility estimates from other variables would help improve the design and operation of STP plants. In addition, other applications that require visibility data could employ such models to fill gaps in measured time series, for instance.

In this paper, the dependence of visibility on readily available atmospheric and radiometric variables is examined. These variables consist of 1-min values of relative humidity RH, temperature T, atmospheric pressure p, wind speed WS and precipitable water PW, direct normal irradiance G_b, and diffuse horizontal irradiance G_d.

The relationship between poor visibility and high humidity has already been noticed several decades ago (e.g. [3]). Humidity has a direct effect on the aerosol extinction coefficient, and consequently on visibility, due to the water uptake behavior of soluble particles. Different functional relationships between visibility and relative humidity have been described in several studies. Chen et al. [4] proposed a parameterization of the extinction coefficient derived from visibility on the basis of the relative humidity and the aerosol volume concentration by means of a power function type. Using numerical models, Han et al. [5] proposed an exponential dependence of visibility on relative humidity, at least for visibilities above 10 km. Usman et al. [6] analyzed a 5-year record of monthly-mean visibility and found a linear relationship between it and RH, and between visibility and T. The influence of temperature, air pressure or wind speed on visibility has also been reported in various studies [7]. However, due to the intricate relationship between visibility and all these variables, the development of a multivariate analytical solution is quite difficult, and remains elusive, at least form a quantitative standpoint.

In order to remove the limitations associated with standard statistical techniques, a more powerful methodology based on artificial neural networks (ANN) is used here to investigate the functional dependence of visibility on the variables listed above. It is well known that ANNs are able to detect complex nonlinear mappings between dependent and independent variables without any a priori information about possible existing relationships between them. In addition, two radiometric variables, the direct beam and the diffuse horizontal irradiances, are included here as additional inputs to the ANN. Including these irradiances has been reported to improve the prediction of visibility in previous studies [8, 9].

2 EXPERIMENTAL SETUP AND DATABASES

The experimental setup is located at the *El Arenosillo* Atmospheric Sounding Station ($37.1°$ N, $6.7°$ W, 40 m a.s.l.) in Huelva (southwestern Spain). It is a seashore site close to the Atlantic Ocean. The main aerosol types present at that station are coastal maritime, continental and desert dust [10]. The experimental dataset consists of 1-min averaged values of visibility VIS, T, RH, p, WS, G_b, and G_d. The measured period is from Dec 1st 2016 until Jan 28th 2017.

Visibility was observed with a BIRAL SWS-250 sensor. This visibilimeter is a forward-scatter meter, which measures the amount of light scattered at an angle of $45°$ by small particulates suspended in, or large particles passing through, its sample volume.

Visibility is reported by this instrument in terms of Meteorological Optical Range (MOR). The observable MOR range of the BIRAL SW-250 is 10 m to 75 km, with an accuracy that is a strong function of MOR: $±5.1\%$ at 2 km; $±12.5\%$ at 15 km; and $\leq 20\%$ at 30 km.

Additional atmospheric variables were measured using a WS-500 sensor manufactured by Lufft GmbH. The direct normal irradiance was measured with an EKO MS-56 pyrheliometer, and diffuse irradiance with an EKO MS-802 pyranometer and a tracking shadow ball attachment. Both radiometers are mounted on an EKO automatic solar tracker, model STR-22G.

The atmospheric precipitable water content P W was estimated from the air temperature and relative humidity [11], and added to the dataset. For the purpose of developing and testing the models, the whole database was divided into two subsets, based on an uniformly random distribution. The training and testing data sets comprised around 33000 and 8500 values, respectively. In addition, a second database of about 1000 values corresponding to cloudless conditions was also generated. Cloudless conditions were selected from the cloud-screened Level-1.5 dataset of one co-located Cimel sunphotometer belonging to NASA's AERONET (AErosol RObotic NETwork).
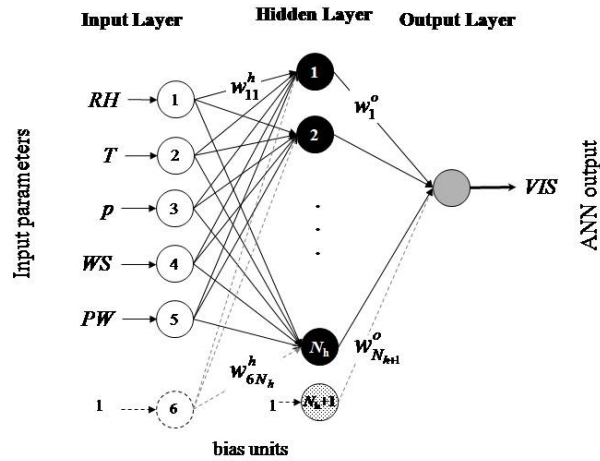

Figure 1: Description of the ANN architecture used here.

3 ANN DESCRIPTION

Artificial neural networks are implemented here using a combination of custom-designed MATLAB functions [12] in conjunction with several routines developed elsewhere [13]. A standard multilayer perceptron (MLP) architecture with three fully interconnected layers (input, hidden and output) is employed, as shown in Fig. 1. The hyperbolic tangent transform is chosen as the nonlinear activation function in the hidden layer, and the identity function is selected as the activation function for the output layer. Such a network determines a non-linear mapping from an input vector $\{x_1, x_2, \ldots, x_{Ni}\}$ with N_i input variables (constituted of the atmospheric and radiometric variables) to the output, i.e., VIS:

$$VIS = \sum_{j=1}^{N_h} w_j^o \tanh\left(\sum_{i=1}^{N_i+1} w_{ij}^h x_i\right) + w_{N_h+1}^o \qquad (1)$$

where N_h is the number of hidden neurons, w_{ij}^h are the weights connecting the input and hidden layers and w_j^o are the weights connecting the hidden and output layers.

Whereas the number of input and output neurons is determined by the respective number of independent and dependent variables involved in the current problem, the choice of the

number of hidden neurons N_h is usually not trivial. In current applications, the number of intermediate neurons is often decided in a heuristic way. In this work, the optimal number of intermediate neurons is determined empirically as the minimum number of neurons for which the performance (evaluated on the validation set) is found satisfying.

All weights are randomly initialized within the range $(-0.5, 0.5)$. Among several existing training algorithms, a Gauss–Newton based Levenberg–Marquartd method is selected due to its rapid convergence properties and robustness [14].

Considering the different intervals exhibited by the input and output variables, these are first normalized to have zero mean and unity variance, and then linearly scaled to remain in the interval $[0.1 - 0.9]$. The scaling is performed using the minimum and maximum values for each input and output parameter. This pre-processing of the data is only necessary in the ANN training step since the weights are back-computed afterwards to allow for raw inputs/outputs.

4 ANALYSIS AND RESULTS

4.1 Standard statistical techniques

The functional dependence of visibility on various inputs is initially examined using standard statistical techniques. To this end, visibility is regressed against every atmospheric and radiometric variable to search for any evident relationship between each other. Fig. 2 shows VIS versus RH for two different situations: (a) daytime and nighttime data for all weather conditions, and (b) daytime and cloudless conditions only. It is observed that visibility is linearly anti-correlated with relative humidity. This relationship is strongest under daytime and cloudless conditions, which are typically those encountered in STP applications. The visibility variability explained by the relative humidity is around 40% in the first case and 57% in the second one, based on the coefficient of determination r^2 obtained in the linear regression fit. This functional dependence has also been described in the literature, albeit based on monthly values only [6, 7].
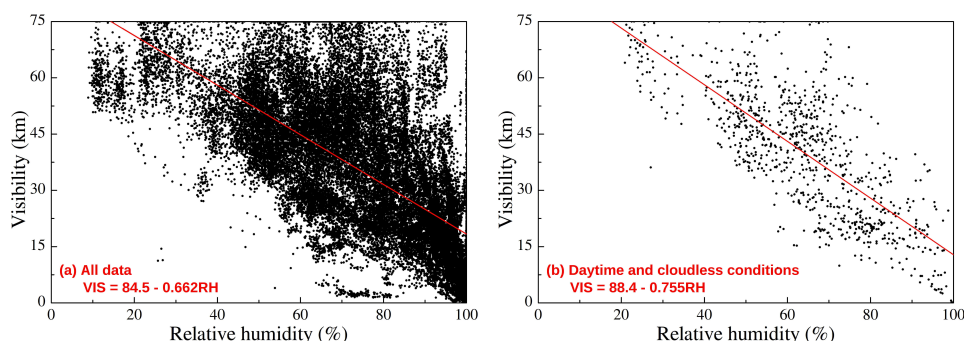

Figure 2: Visibility versus relative humidity for (a) all day and all weather conditions; and for (b) daytime and cloudless conditions. The lines and equations obtained by linear regression are also displayed in red.

On the other hand, a visual inspection of additional plots suggests that visibility is independent from the other atmospheric or radiometric variables. For instance, Fig. 3 shows the lack of correlation between VIS and T or p.
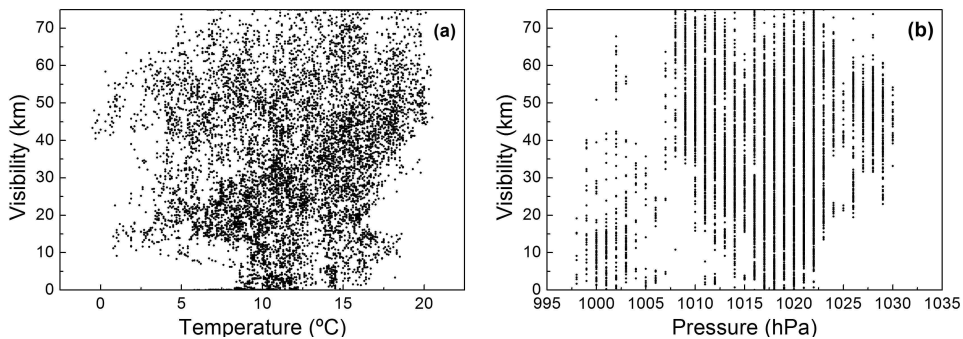
Figure 3: Visibility versus (a) temperature, and (b) pressure.

4.2 Artificial neural networks

The accuracy of the ANN predictions is further evaluated in terms of the coefficient of determination (r^2), the mean bias deviation (MBD) and the root mean square deviation (RMBD), resulting from the statistical analysis between estimated and measured visibilities. MBD and RMSD are calculated in percent of the mean measured visibility. For each ANN architecture (with different input number or different hidden neuron number), 10 trainings are performed. Moreover, the statistical indicators are the mean of all 10 statistics (r^2, RMSD or MBD), along with the corresponding standard deviation of the mean. Using this approach, local minima problems due to the random weight initializations can be avoided.

The ANN architecture is complete when hidden units are fixed to a certain optimal value and the inputs variables are selected to be the best configuration for calculating visibility (i.e., after removal of any redundant or non-relevant variable). The first task is achieved by analyzing the ANN performance when the number of hidden neurons is increased from 2 to 35, and when using all input parameters. For $N_h > 15$, the ANN ensembles perform similar to each other with $r^2 = 0.72$, RMSD= 29%, and an almost null MBD. On the other hand, decreasing the number of hidden units yields an increase of RMSD, up to $\sim 33\%$. Based on these results, N_h is set to 15.

The last step is to select the best input configuration and to analyse the relevance of the atmospheric and radiometric variables. To that end, a backward pruning is undertaken, first using all seven inputs. The ANN is trained and validated ten times using the above training and test datasets. The ANN performance is analysed based on the averaged RMSD and MBD values, along with the standard deviation corresponding to the validation data. This task is repeated using different input configurations. All cases are listed in Table 1. MBD values for the test data are in the range $(-0.2 \pm 0.2)\%$ for all cases. This statistical test is thus not suitable for selecting the relevance of the best input configuration. Nevertheless, it shows the ability of the ANN to adjust the weights so as to almost completely avoid bias in the predicted output.

Compared to the base case (RH being the single input), it is clear that an improvement in visibility estimates is obtained when the amount of input information (i.e., the number of inputs) is increased. The only non-relevant input variable is found to be PW. This is certainly related the fact that PW is a strong function of temperature and, even more so, relative humidity, and thus does not explain any additional variability in visibility. Adding temperature, pressure and wind speed to RH improves the accuracy of the estimates. RMSD

Table 1: Statistical results of the ANN performances using different input configurations. The standard deviation of the averaged RMSD and MBD is also added.

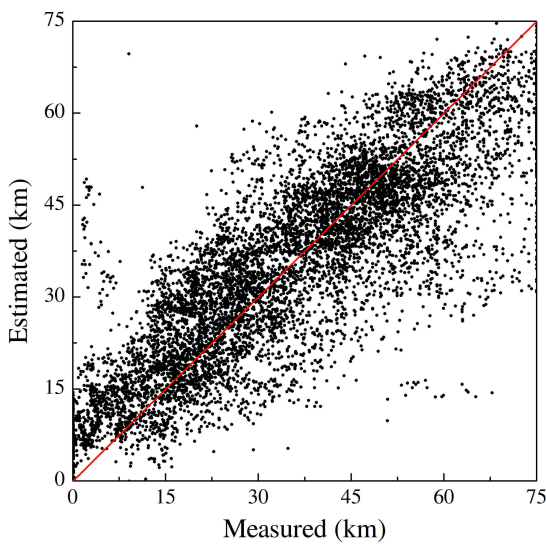| Input variables	r^2	RMSD(%)	MBD(%)
RH	0.51	38.1±0.1	-0.2±0.1
RH, T	0.53	37.2±0.1	-0.1±0.0
RH, T, p	0.63	33.5±0.7	-0.5±0.3
RH, T, p, WS	0.65	32.7±0.9	-0.2±0.2
RH, T, p, PW	0.63	33.1±1.2	-0.3±0.3
RH, T, p, WS, PW	0.65	32.7±0.9	-0.4±0.2
RH, T, p, WS, G_b	0.70	30.4±0.4	-0.2±0.3
RH, T, p, WS, G_b, G_d	0.72	29.7±1.2	0.0±0.2

Figure 4: Visibility estimated using the ANN model with atmospheric and radiometric inputs versus the measured one for the test dataset.

is indeed reduced from $\sim 38\%$ to $\sim 33\%$. The coefficient of determination is also improved from 0.51 to 0.65.

The inclusion of direct and diffuse components of solar radiation lead to a moderate improve of the estimates, decreasing RMSD from $\sim 33\%$ to $\sim 30\%$. Direct irradiance appears the more relevant of the two radiometric variables. The overall coefficient of determination reaches 0.72. Fig. 4 shows the scatter plot of the ANN-estimated visibility against the measured one. The data points are centered around the line 1:1 of perfect fit. Fig. 5 shows the daily evolution of both estimated and measured visibilities during January 2017. The ANN model is able to reproduce the daily excursions on visibility quite correctly. Visibility exhibits a higher variability when it is high (under very clear conditions). This is a consequence of the increasing uncertainty of the visibilimeter under those conditions. The ANN model is not able to reproduce such a noisy behavior.
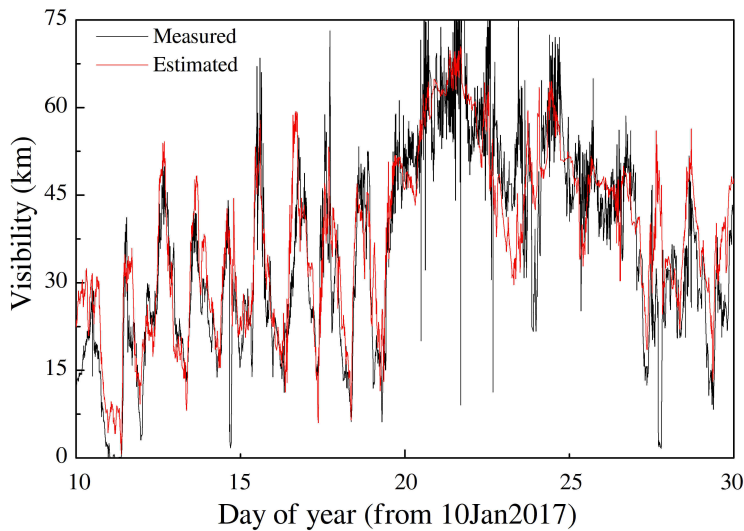
Figure 5: Daily evolution/Time series of the measured (black line) and estimated (red line) visibility using the ANN with all atmospheric and radiometric input parameters for January 2017.

Finally, the estimation of visibility under daytime and cloudless conditions leads to a significant improvement compared to the general case. The RMSD decreases to 22%. These conditions represent the norm for STP plants, since they cannot operate if the sun is obscured by thick clouds.

5 CONCLUSIONS

Two different methodologies based on traditional statistical techniques and artificial neural networks have been used to estimate visibility from high-frequency (1-min) observations of atmospheric and radiometric variables. The simple linear relationship between visibility and relative humidity was shown more relevant under daytime and cloudless skies than under all-weather conditions. No other obvious dependence of visibility on other atmospheric and radiometric variables could be detected by simple visual inspection of the corresponding scatter plots. The use of ANN has revealed that visibility could be estimated with satisfactory accuracy from a number of atmospheric and radiometric variables. This is possible even at a 1-min temporal resolution, despite the large variability and noise this implies. The qualitative knowledge of the influence of temperature, air pressure or wind speed on visibility that was shown in previous studies is quantitatively analyzed here in more detail. Adding two commonly measured radiometric variables(direct and diffuse solar irradiances) leads to a modest improvement in visibility estimates. The main improvement occurs under daytime and cloudless conditions. These conditions precisely correspond to those required for the operation of STP plants. Hence, the ANN model proposed here can provide the accurate visibility estimates that, in turn, are required to analyze the irradiance transmission losses throughout the heliostat-tower field. The proposed ANN model can also be useful in more general meteorological applications, for instance to fill gaps in visibility time series if the relevant atmospheric or radiometric variables are measured.

REFERENCES

[1] Gueymard, C.A., López, G. & Rapp-Arrarás, I., Atmospheric transmission loss in mirror-to-tower slant ranges due to water vapor. Presented at *SolarPACES 2016 Conference*, Abu Dhabi, United Arabian Emirates, 2016.

[2] Pitman, C.L. & Vant-Hull, L.L.,Atmospheric transmittance model for a solar beam propagating between a heliostat and a receiver. Report SAND83-8177, prepared by Sandia National Laboratories and Livermore (USA), 1984.

[3] Lundgren, D.A. & Cooper, D.W., Effect of humidify on light-scattering methods of measuring particle concentration. *Journal of the Air Pollution Control Association*, **19**(4), pp. 243–247, 1969. DOI: 10.1080/00022470.1969.10466482

[4] Chen, J., Zhao, C.S., Ma, N., Liu, P.F., Göbel, T., Hallbauer, E., Deng, Z.Z., Ran, L., Xu, W.Y., Liang, Z., Liu, H.J., Yan, P., Zhou, X.J. & Wiedensohler, A., A parameterization of low visibilities for hazy days in the North China Plain. *Atmospheric Chemistry and Physics*, **12**, pp. 4935–4950, 2012. DOI:10.5194/acp-12-4935-2012

[5] Han, X., Zhang, M., Tao, J., Wang, L., Gao, J. & Wang, S., Modeling aerosol impacts on atmospheric visibility in Beijing with RAMS-CMAQ. *Atmospheric Environment*, **72**, pp. 177–191, 2013.

[6] Usman, A., Olaore, K.O. & Ismaila, G.S., Estimating visibility using some meteorological data at Sokoto, Nigeria. *International Journal of Basic and Applied Science*, **1**(4), pp. 810–815, 2013.

[7] Buba, D., Anjorin, F.O. & Jacob, A., The analysis of influence of weather conditions on atmospheric extinction coefficient over Bauchi, north eastern Nigeria. *Journal of Atmospheric Pollution*, **3**(1), pp. 31–38, 2015. DOI:10.12691/jap-3-1-6

[8] Liepert, B.G. & Kukla, G.J., Decline in global solar radiation with increased horizontal visibility in Germany between 1964 and 1990. *Journal of Climate*, **10**, pp. 2391–2401, 1997.

[9] Bason, F., Diffuse solar irradiance and atmospheric turbidity. Presented at *EuroSun 2004 Conference*, Freiburg, Germany, 2004.

[10] Sorribas, M., de la Morena, B.A., Wehner, B, López, J.F., Prats, N., Mogo, M., Wiedensohler, A. & Cachorro, V.E., On the sub-micron aerosol size distribution in a coastal-rural site at El Arenosillo Station (SW Spain), *Atmospheric Chemistry and Physics*, **11**, pp. 11185–11206, 2011.

[11] Gueymard, C.A., Analysis of monthly average atmospheric precipitable water and turbidity in Canada and Northern United States. *Solar Energy*, **53**, pp. 57–71, 1994.

[12] MatLab, The MathWorks, Inc. Natick, MA, USA, 1999.

[13] Norgaard, M., *Neural network based system identification toolbox*, Technical Report 97-E-851, Department of Automation, Technical University of Denmark, 1997.

[14] Fletcher, R., *Practical methods of optimization*, John Wiley & Sons, Chichester, 1987.