Forecasting of global and direct solar irradiance using stochastic learning methods, ground experiments and the NWS database

Ricardo Marquez, Carlos F.M. Coimbra *

Mechanical Engineering and Applied Mechanics Program, School of Engineering, University of California, 5200 N. Lake Road, Merced, CA 95340-3222, USA

Abstract

We develop and validate a medium-term solar irradiance forecasting model by adopting predicted meteorological variables from the US National Weather Service's (NWS) forecasting database as inputs to an Artificial Neural Network (ANN) model. Since the inputs involved are the same as the ones available from a recently validated forecasting model, we include mean bias error (MBE), root mean square error (RMSE), and correlation coefficient (R^2) comparisons between the more established forecasting model and the proposed ones. An important component of our study is the development of a set of criteria for selecting relevant inputs. The input variables are selected using a version of the Gamma test combined with a genetic algorithm. The solar geotemporal variables are found to be critically important, while the most relevant meteorological variables include sky cover, probability of precipitation, and maximum and minimum temperatures. Using the relevant input sets identified by the Gamma test, the developed forecasting models improve RMSEs for GHI by 10–15% over the reference model. Prediction intervals based on regression of the squared residuals on the input variables are also derived.
© 2011 Elsevier Ltd. All rights reserved.

Keywords: Solar irradiance forecasting; Gamma test; Model input selection; Solar meteorology

1. Introduction

Although the radiation that reaches the top of the layers of the atmosphere is well defined and can be easily calculated, the solar irradiance that reaches the ground level where solar collectors (thermal and photovoltaic) operate depends strongly on localized and complex atmospheric conditions. From the operational standpoint, the balancing of supply and demand peaks in the electrical grid requires detailed consideration of the availability of solar power. Forecasting the available insolation is therefore an enabling technology for the success of any policy to include solar power as an important contribution to both centralized and decentralized systems. Cloud cover, aerosol content, and the presence of atmospheric gases in the troposphere (water vapor, carbon dioxide) and in the stratosphere (ozone) can reduce the availability of direct insolation at the ground level to a small fraction of the solar irradiance that reaches the upper atmosphere. These effects are particularly strong on the Direct Normal Irradiance (DNI). Because cloud cover corresponds to the strongest effect on ground insolation, no statistical method that ignores micro-scale ($\leqslant 2\,km$) or meso-gamma scale ($2–20\,km$) weather systems can succeed in estimating real-time and/or forecasting solar power availability.

The problem is substantially more complicated for the characterization of DNI, which is the component of the total solar irradiation that is most useful for solar concentrators. Solar concentrators that can concentrate in excess

---

* Corresponding author. Tel.: +1 209 228 4058; fax: +1 209 228 4047.
    E-mail address: ccoimbra@ucmerced.edu (C.F.M. Coimbra).

of 300 times the solar flux are considered viable technology alternatives for solar utilization, both for direct photovoltaic and solar-thermal conversion systems. In the case of concentrated PV, the rationale for concentration is primarily due to economics: the high-efficiency multi-junction photovoltaic cells that can achieve in excess of 40% direct radiation-to-electricity conversion in laboratory environments cost slightly more than 100 times the common PV cell. Recent studies also point out that the Energy Pay Back Time (EPBT) for multi-junction cells may be substantially shorter (see e.g., Fthenakis and Alsema, 2006; Pacca et al., 2007) than previously estimated. For solar-thermal conversion, high concentration designs are needed to achieve higher thermodynamic cycle efficiencies. Utility-scale solar thermal power system also use primarily DNI for both solar-tower and solar-trough power plants. Because most commercially viable, utility-scale solar technologies use high concentration technologies, DNI characterization is of critical importance in designing, implementing and deploying such power plants.

The purpose of this work is to produce hourly medium-term forecasting models for both the Direct Normal Irradiance and the Global Horizontal Irradiance (GHI) based on meteorological predictions from the National Weather Service's National Digital Forecasting Database (NDFD). The NDFD produces same-day up and to 7-days ahead forecasts of meteorological variables, not including solar irradiance. We refer to these forecasting time horizons as medium-term to be consistent with Perez et al. (2007), whom have recently developed and validated a forecasting model for GHI based on the sky cover data supplied from the NDFD. When applied to the Sacramento and New York areas, their forecasting model obtained accuracies consistent with preliminary analysis using Multiple Output Statistics and meso-scales models elsewhere (e.g. Heinemann, 2004). More recently, their model has been evaluated in a number of locations with widely different climates (Perez et al., 2010) and thus further validating the generalization of the simple model. The present study extends theirs by considering the use of additional meteorological variables in order to enhance the forecasting capabilities of the models. Because this study considers the use of a larger number of variables, we present and employ an objective strategy which helps to select inputs that contain the most predictive information.

The Gamma test (GT) is an appropriate tool for identifying relevant inputs, as this test has been demonstrated to be able to provide information regarding the relationship between input and output data sets, even prior to model development (Durrant, 2001; Jones, 2004; Wilson et al., 2004; Moghaddamnia et al., 2009; Remesan et al., 2008). The GT is an algorithm designed to provide an estimation of the lowest possible mean squared error (MSE) attainable by a continuous and differentiable model for some output, y, based on the inputs, $\mathbf{x}$. The criteria for choosing the content in $\mathbf{x}$ can be based on keeping the dimension of $\mathbf{x}$ small (to remove insignificant inputs), meanwhile retaining a low estimate for the MSE. The GT is a relatively simple algorithm to apply and is used here as a basis for the selection of relevant inputs. The selection procedure also involves a genetic algorithm (GA) search in order to efficiently explore possibly significant input combinations. In Section 3 the GT and GA procedures are described, and in Section 4 the input selection algorithms are evaluated.

After applying the GT and GA to identify which subsets of inputs are potentially useful, Artificial Neural Networks (ANNs) are used to construct the forecasting models. ANNs are described and applied in Sections 5 and 6. In Section 7, prediction intervals are derived which take into account the forecasting uncertainty based on the predicted NDFD-based meteorological conditions.

2. Data

The University of California Merced solar observatory is equipped with several total and spectral solar instruments acquired from Eppley Labs. Three integrating instruments were used extensively in the present study: 1 PSP (Precision Spectral Pyranometer) for benchmarking the global irradiance, 1 NIP (Normal Irradiance Pyrheliometer) for benchmarking the direct irradiance, and 1 SDK (Shaded Disk Kit) that shades another PSP mounted on a solar tracker to measure the diffuse irradiance. From measurements of two components irradiance the third can be calculated, (e.g., global and diffuse data can be used to calculate the direct). We used all three measurements to constantly monitor the quality of the data. Expected accuracies of the irradiance measurements are estimated to be on the order of 5% for PSPs, and 3% for the NIPs. Data acquisition, logging and processing is automated with the aid of a Campbell Scientific's CR1000 data-logger. The data sampling rate is two samples per second. The data logger calculates and stores averaged values for each quantity every 30 s – hourly averages are calculated from the logged values.

The National Weather Service (NWS) manages the National Digital Forecast Database (NDFD) which provides gridded digital forecasts of weather parameters for the entire country at high resolutions of up to 2.5-km spatial and 1-h temporal (Glahn and Ruth, 2003; Perez et al., 2007). Local forecasts from NWS Weather Forecast Offices (WFOs) are generated by national model outputs, meso-scale model runs and human input. These local forecasts are then merged and assembled on a national grid (Glahn and Ruth, 2003; Perez et al., 2007). Weather forecasted values are publicly available and can be accessed electronically in Extensible Markup Language (XML) format through the NWS website: http://www.weather.gov/ndfd (Schattel and Bunge, 2008). The forecasted meteorological elements provided by the NWS include ambient temperature, dew point temperature, relative humidity, sky cover, wind speed and direction, probability of precipitation, significant wave height, weather type, and snow amount. Daily maximum and minimum ambient temperatures are also provided by the NDFD.

NDFD meteorological data was collected daily at approximately 12:00 p.m. (Pacific local time) for the period of this study which spans the dates November 1, 2008–November 30, 2009. The NDFD data sets collected on each day contain same-day forecasted values and forecasted values for the next 6 days. Since the forecasting methodology is essentially a regression on the NDFD data sets, the solar irradiance forecasts generated here will also be available as the NDFD forecasts are produced (same-day forecasts as well as a few days ahead).

3. Model inputs and input selection

To develop the irradiance forecasting models, 11 input variables are considered. Nine of the variables are predicted meteorological variables (described in Section 2) and listed in Table 1. The other two variables are geometric/temporal variables: the cosine of the solar zenith angle ($\cos Z$) and the normalized hour angle ($\bar{\omega}$, defined later). The geometric variables are important because they describe the deterministic diurnal variations of clear-sky solar irradiation. One of the objectives of this work is to make a determination of which auxiliary meteorological variables are useful for predicting solar irradiation and which ones can be discarded to prevent over complication of the model. In order to do so, the GT is used as a criteria for determining the relevant inputs. With a total of 11 inputs there are $2^{11} - 1 = 2047$ possible subsets of inputs to perform the GT with. Rather than evaluating all possible combinations, a GA search is employed to reduce the input space. The description of the GT and GA procedures are provided after a describing the data preprocessing and introducing the definition of $\bar{\omega}$.

3.1. Variable representation and preprocessing

Table 1 lists the input variables that are used for model input selection. In order to keep the notation concise, the input variables are represented as x_j, where j is the associated variable number indicated in the first column of Table

1. The set of input variables consists of the meteorological variables from the NDFD database the solar geometry variables: $\cos Z$ and $\bar{\omega}$. To avoid scaling issues that can complicate the Gamma test or neural network training, all meteorological input variables are normalized by subtracting the variables by their corresponding minimum values, then dividing by their range (maximum–minimum) so that the values vary from 0 to 1.

3.2. Dimensionless time scale as an input

We introduce an additional input for modeling DNI which we refer to as the normalized hour angle defined as

$$\bar{\omega} = \omega / (\omega_{max}), \tag{1}$$

where ω is known as the hour angle and ω_{max} is equal to absolute value of the maximum hour angle for a given day while the sun is up; $\omega_{max} = \omega_{sun \, set} = -\omega_{sun \, rise}$. These variables are commonly used for conversions between standard time and solar time and are readily calculated using algebraic relations given in Duffie and Beckman (2006). By definition, $\omega = 0$ at solar noon when the sun is due south and is at its highest elevation for a given day. Observations of DNI daily profiles for clear-sky conditions versus $\bar{\omega}$ are shown in Fig. 1. This figure shows that $\bar{\omega}$ is a potentially useful input variable because clear-sky DNI profiles for any part of the year have the same basic shape with respect to $\bar{\omega}$.

3.3. Residual variance estimation: Gamma test

The GT was developed by Koncar (1997) and Stefansson et al. (1997), and further advanced by Evans and Jones (2002) and Jones (2004). The idea behind the GT is to suppose there exists a continuous and differentiable function, f, relating the output, y, to some m-dimensional input vector, $\mathbf{x}$, and has the general form

Table 1
List of inputs for modeling. Note: the range of hourly ambient temperature is equal to the maximum of the daily Max Temperature and minimum of the daily Min Temperatures. Wind direction data is given in increments of 10°.

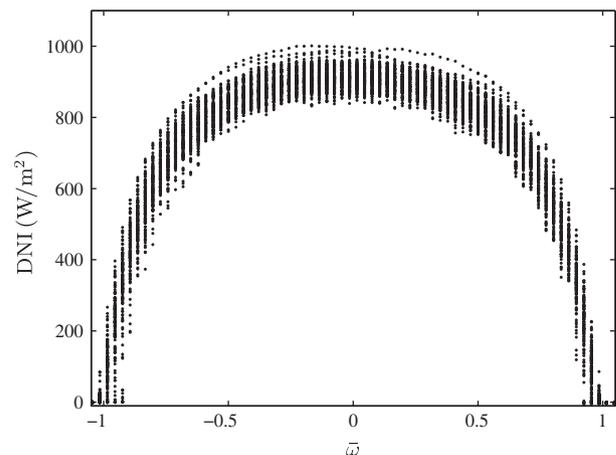| Number	Name	Minimum	Maximum
1	Max temperature (°C)	7.22	40.55
2	Temperature (°C)	−2.78	40.55
3	Dew point temperature (°C)	0	37.78
4	Relative humidity (%)	0	100
5	Sky cover (%)	0	100
6	Wind speed (m/s)	0	22.5
7	Wind direction (deg)	0	350
8	Probability of precipitation (%)	0	100
9	Min temperature (°C)	−2.78	22.22
10	$\cos(Z)$	0	1
11	Normalized hour angle, $\bar{\omega}$	0	1

Fig. 1. DNI for clear-sky conditions plotted versus normalized time index $\bar{\omega}$. DNI profiles have same basic shape when viewed on this time scale for all parts of the year.

$$y = f(\mathbf{x}) + r, \tag{2}$$

where the residual term, r, represents the uncertainty or noise in the model, and has a mean of 0 and variance σ_r^2.

The procedure of the GT is to first construct the $k^{th}(1 \leqslant k \leqslant k_{max})$ nearest neighbor list $\mathbf{x}_{N[i,k]}(1 \leqslant i \leqslant M)$ of input vectors $\mathbf{x}_i(1 \leqslant i \leqslant M)$, where k_{max} is the maximum number of nearest neighbors (typically $k_{max} = 10$) and M is the number of data points. The following steps are then applied:

- for $(1 \leqslant k \leqslant k_{max})$ compute:

$$\delta_M(k) = 1/M \sum_{i=1}^{M} |\mathbf{x}_{N[i,k]} - \mathbf{x}_i|^2, \tag{3}$$

and

$$\gamma_M(k) = 1/(2M) \sum_{i=1}^{M} |y_{N[i,k]} - y_i|^2. \tag{4}$$

- construct the linear regression for relationship for the data $(\delta_M(k), \gamma_M(k))$:

$$\gamma_M(k) = \Gamma + A\delta_M(k). \tag{5}$$

Provided that the regression line is a good fit, the intercept of Eq. (5), Γ, provides a close estimate of the residual variance σ_r^2 and therefore the lowest attainable MSE. This condition is satisfied when $M \to \infty$ as proven in Evans and Jones (2002). The slope A gives a crude estimate of the complexity of the unknown surface of the regression function, $f(\mathbf{x})$. The GT is non-parametric technique and the results apply regardless of the particular methods used to subsequently build a model. This test is particularly suited for the analysis of non-linear regression problems.

An extension of the Gamma statistic Γ is the V_{ratio} parameter defined as the ratio between the variance of the residuals and the variance of the output, $V_{ratio} = \Gamma/\text{Var}(y)$. This parameter can be considered a normalization of the expected MSE and is analogous to the commonly used model quality measure, R^2:

$$R^2 \approx 1 - V_{ratio}. \tag{6}$$

In this paper, we use the V_{ratio} rather than Γ to evaluate the GT outputs.

A sequence of GTs performed for an increasing number of data points is referred to as an M-test (Jones, 2004). The M-test is useful because allows one to evaluate the reliability of Γ as an estimate of σ_r^2, and to determine whether enough data is available for model construction. For instance, during initial GT evaluations when the number of data points is small, the M-test plots fluctuate, indicating that the MSE estimates do not agree and are therefore not reliable. As more data is included, the M-test plots begin to stabilize to consistent MSE predictions. If the M-Test plots stabilize, then a degree of confidence that the Γ estimate is reasonably accurate is achieved (Jones, 2004). In these cases the input/output data will be relatively simple to

model. In Section 4, the M-test will be applied to verify that enough training data is available and to compare different model inputs.

For a fuller discussion on the GT and the implications for modeling we refer the reader to the reference (Jones, 2004).

3.4. Genetic algorithm search for model input selection

The selection of the optimal subset of inputs can be accomplished with genetic algorithms (GAs). GAs are optimization search techniques inspired by the process of biological evolution. The algorithms involve optimization searching patterns where alleles (features) of individuals in a population of potential solutions are altered by crossover and mutation operators over generations/cycles so that the fittest individuals continue to evolve (Wilson et al., 2004). The proportion of individuals that either mutate or crossover are fixed by the mutation and crossover parameters. In the current problem of input selection, the individual fitness is based on minimizing the objective function adopted from Durrant (2001) and Wilson et al. (2004), which consist of three penalty terms:

$$G(\mathbf{x}_i) = w_\Gamma g_\Gamma(\mathbf{x}_i) + w_A g_A(\mathbf{x}_i) + w_L g_L(\mathbf{x}_i). \tag{7}$$

The three penalty terms g_Γ, g_A, and g_L, are monotonically increasing functions of Γ, A, and the ratio of number of inputs in $\mathbf{x}_i$ to total number of inputs, respectively. When selecting the most relevant subset of inputs, $\mathbf{x}_i \subset \mathbf{x}$, the objective function ($G(\mathbf{x}_i)$) encourages competition among three optimality measures. That is, the optimal subset of inputs should (1) produce low prediction errors (the Γ value the subset should be low); (2) keep the complexity of the input and output relationship low (value of A should be small); and (3) include only the relevant features. Correspondingly, the fitness scaling parameters in this objective function which must be modified include: (1) intercept fitness weight, w_Γ; (2) gradient fitness weight, w_A; and (3) length fitness weight, w_L.

The scale, w_Γ, corresponds to emphasis on the Gamma statistic, Γ. In choosing a combination of inputs, it is preferable to keep inputs which contribute to producing low values of Γ (equivalently, low values of $g_\Gamma(\mathbf{x}_i)$). Therefore, the scale w_Γ is set at a high value of $w_\Gamma = 1$.

As mentioned in Section 3.3, the gradient returned by the GT is a crude estimation of the complexity of the model, $f(\mathbf{x})$. Given that A is only a crude estimate, we put little emphasis on this measure, so the scale for this term ($g_A(\mathbf{x}_i)$) in the objective function is assigned a low value of $w_A = 0.1$.

The length fitness parameter, w_L corresponds to the number of inputs to be considered in the model. The value to assign for w_L is not as straight forward because too little emphasis on the length fitness leads to the retention of insignificant or counterproductive inputs. On the other hand, too much emphasis on $g_L(\mathbf{x_i})$ may lead to an overall increase of Γ for the population of inputs in the GA search.

Table 2
Genetic algorithm search parameters for model input selection.

Parameter	Value
Population size	100
Crossover rate	0.5
Mutation rate	0.05
Intercept fitness, w_Γ	1
Gradient fitness, w_A	0.1
Length fitness, w_L	variable $(0.1, 0.2, \ldots, 1)$
Run time(s)	600

Therefore, w_L is left variable and subsets of inputs are considered optimal based on which inputs remain relevant as w_L is varied from 0.1 to 1 in increments of 0.1.

Values of all the GA optimization parameters are listed in Table 2.

4. Evaluation of input selection algorithms

The GA search is applied using the WINGAMMA computer software. The length fitness scale, w_L, is varied from 0.1 to 1 in increments of 0.1 and the results obtained by the GA search are used to produce the bar plots in Fig. 2a and b. The height of the bars represent the collective frequency (in percent) that each input is included in the best 10% of input subsets which minimized the objective function ($G(\mathbf{x}_i)$). These results indicate that the variable x_{10} ($\cos Z$) is the most important input for modeling GHI. This is expected because $\cos Z$ dictates the deterministic part of the global horizontal solar irradiation. The next most relevant inputs for GHI are x_5 (sky cover), x_8 (probability of precipitation), and x_9 (minimum temperature). The variables which seem to be the least important are x_7 (wind direction), and x_4 (relative humidity) which have low frequency in the best 10% of input combinations.

Fig. 2a can be used for choosing which input subsets to consider. Based on Fig. 2a, the combination of inputs to be selected for GHI are subset 1: $[x_8, x_{10}]$; subset 2:

$[x_5, x_9, x_8, x_{10}]$; subset 3: $[x_1, x_2, x_3, x_5, x_6, x_8, x_9, x_{10}]$ and subset 4: [all variables]. The inputs in subset 1 corresponds to the simplest case where the two variables that were predicted by the GA to have the most significance are used. The size of the input set is incremented in subset 2 by including the next significant variables: x_5 (minimum temperature) and x_9 (probability of precipitation). Input subset 3 increments subset 2 by including the next most significant variables, and subset 4 contains all the available variables. These input subsets are listed in the second column of Table 3.

The relevant inputs for DNI are slightly different than for GHI. The most relevant inputs appear to be x_5 (sky cover), x_9 (minimum temperature), and x_{11} (normalized hour angle). The four subsets selected for modeling DNI are subset 5: $[x_5, x_9, x_{11}]$; subset 6: $[x_1, x_3, x_5, x_8, x_9, x_{11}]$; subset 7: $[x_1, x_2, x_3, x_5, x_6, x_8, x_9, x_{11}]$; and again subset 4: [all variables]. These input subsets are listed in the second column of Table 3.

An M-test is performed on the selected inputs and the results are plotted in Fig. 3a and b. As described in Section 3.3 the reliability of the GT can be inferred from the stability of the M-test plots. All GHI M-test plots appear to have stabilized indicating that enough data is obtained and that the GT results are reliable. It is clear from Fig. 3a that the variables x_8 (probability of precipitation) and x_{10}($\cos Z$) alone will not be enough to produce the most accurate models. The M-test results for DNI (Fig. 3b) are similar in that the variables x_5 (sky cover), x_9 (probability of precipitation), and x_{11}($\bar{\omega}$) will not give the best possible model accuracy.

5. Artificial Neural Networks

Artificial Neural Networks (ANNs) are useful tools for approximating complicated mapping functions for problems in classification and regression (Bishop, 1994), and have also been used extensively in many areas including
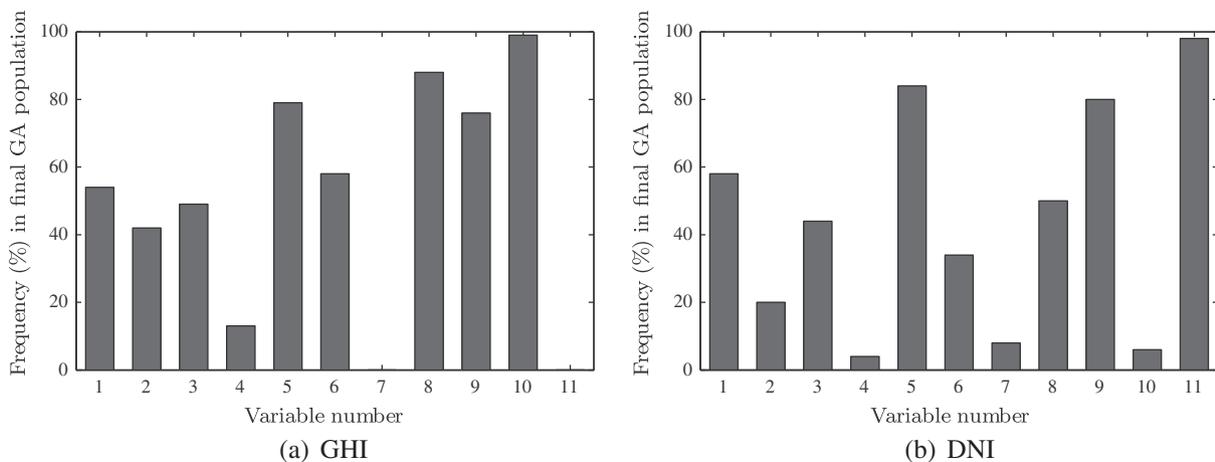

Fig. 2. Genetic algorithm results for model input selection. High frequency of inclusion indicates that the variable is useful for minimizing the objective function, Eq. (7).
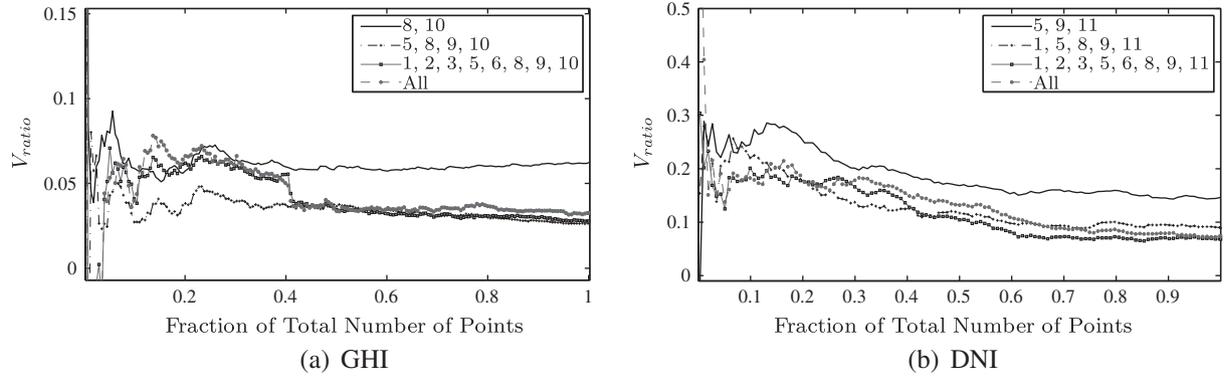
Fig. 3. M-tests for various input subsets (as indicated in the legend) which were selected based on results obtained from GA results. The V_{ratio} on the vertical axis is an estimate of the normalized MSE, i.e. $\approx 1 - R^2$. The total number of data points for GHI is 4472 and for DNI is 3738.

solar radiation modeling. Mellit (2008) gives a review of over 40 applications of ANNs and other artificial intelligence techniques applied to solar radiation modeling. The advantage of ANNs is that no assumptions are required about the underlying process relating input and output variables. However, because ANNs are universal approximating functions, their mapping capabilities can potentially lead to problems such as over-fitting training data (Bishop, 1994), and thus leading to poor generalization on new data sets. To prevent overtraining (Lauret et al., 2006), for example, used an approach to construct network architectures based on Bayesian and sensitivity analysis where coincidentally the methodology identifies relevant inputs as well as an appropriate neural network architecture in the case of modeling the Direct Normal Irradiance. Another approach which addresses the issue of overtraining and model input selection simultaneously is based on the GT (see e.g., Moghaddamnia et al. (2009) and Remesan et al. (2008) for application on modeling daily values of GHI). The principal advantage of the GT approach over the Bayesian approach is the simplicity of the theoretical underpinnings behind the methodologies. The GT results obtained in Section 4 are used here to prevent overtraining with ANNs.

An ANN is a specific representation of y in terms of some input variables $\mathbf{x}$. The ANN representation is based on signals being sent through elements called neurons in such a way that the processing of the inputs signals produces an output, y, that is sufficiently close to the desired target value of t. Neurons are arranged in layers, where the first layer contains the set of inputs, the last layer contains the output, and the layers in between, referred to as hidden layers, contain hidden neurons. A feedforward neural network with N inputs and N_h neurons in one hidden layer with a linear output activation function can be expressed as

$$y = f(\mathbf{x}, \mathbf{w}) = \sum_{i=1}^{N_h} w_k f_k^{hidden} \left(\sum_{j=1}^{N} w_{j,k} x_j + w_0, k \right) + w_0 \qquad (8)$$

where f_k^{hidden} are sigmoidal functions, such as the hyperbolic tangent function. A characteristic of feedforward neural

networks is that the neurons are successively interconnected from layer to layer where neurons in one layer affect all neurons in the next but do not affect other neurons in the same layer or any preceding layers. Numerical optimization algorithms such as back-propagation, conjugate gradients, quasi-Newton, and Levenberg–Marquardt have been developed to efficiently adjust the weights, $w_{k,j}$ and w_k in the feedforward neural networks so that the minimization of some performance function is achieved on some training data. Typically the performance function used for adapting the weights is the MSE.

$$MSE = \frac{1}{M} \sum_{i=1}^{M} (y_i - t_i)^2. \qquad (9)$$

6. Model development and evaluation

The MATLAB NEURAL NETWORK TOOLBOX was used to construct regression models for GHI and DNI where the inputs consists of same-day forecasted meteorological variables. The same-day forecasted NDFD data was randomized and split into two categories consisting of a training set and a test set. The proportions of each set are based on the results of the M-test (Fig. 3a and b). Accordingly, the training set consists for 60% of the randomly selected points of the entire data set to train the GHI neural networks and 80% to train the DNI neural networks. The neural network architectures are based on the feed-forward structure described in Section 5 where the input layer consist of the inputs from the four subsets suggested in Section 4.

6.1. Forecasting model training

Eight ANN models were trained and tested to forecast solar irradiation. Models 1–4 forecast GHI and Models 5–8 forecast DNI. In each training cycle (epoch), all the training data is presented to the ANN in random batches and the MSEs are computed. The weights are adapted using Levenberg–Marquardt learning algorithm and the stopping criteria for neural network training is based on

continued training until the R^2 reaches the Gamma test output, V_{ratio}, or for a maximum of 100 epochs. The number of neurons in the hidden layer was kept at 10–20 neurons. When more neurons were used, larger discrepancies of the model quality statistics between the training and testing sets were observed and as a result the ANN models generalized more poorly.

6.2. Benchmark reference models: 24-h persistent model and Perez model

For the purpose of comparisons, we include a persistent model defined as the model that estimates the solar irradiance as having the same hourly values as the previous day, i.e., the persistent model used here is a 24-h persistent model.

The Perez GHI forecasting model (Perez et al., 2010) which has been validated at a number of locations is also applied here. The Perez model relates GHI to sky cover (SC) as:

$$\text{GHI} = \text{GHI}_{\text{clear}}(1 - 0.87 SC^{1.9}), \tag{10}$$

where SC is the sky cover values obtained from the NDFD forecasts. Estimates for $\text{GHI}_{\text{clear}}$ are obtained using a Linke Turbidity factor of 3 as per Perez et al. (2002).

6.3. Statistical evaluation metrics

The eight models differ by their inputs as is indicated in Table 3. The model quality metrics used for comparisons of the different models include the R^2, RMSE, and relative RMSE (rRMSE), which are calculated using the following equations:

$$R^2 = 1 - \frac{\sum_{i=1}^{M}(y_i - t_i)^2}{\sum_{i=1}^{M}(t_i - \bar{t})^2}, \tag{11}$$

$$\text{RMSE} = \sqrt{\sum_{i=1}^{M}\frac{1}{M}(y_i - t_i)^2}, \tag{12}$$

$$\text{rRMSE} = \text{RMSE}/\bar{t}, \tag{13}$$

$$\text{MBE} = \sum_{i=1}^{M}\frac{1}{M}(y_i - t_i), \tag{14}$$

and

$$\text{rMBE} = \text{MBE}/\bar{t}. \tag{15}$$

In the above definitions, y_i are the predictions from the ANNs, t_i, are measured values of GHI or DNI, and $\bar{t}$ are the averages of GHI or DNI for the training or test set. Night values (values which occur when $\cos Z < 0$) are removed when computing each of the statistical quantities.

6.4. Forecasting model evaluations

All models proposed here show major improvements over the 24-h persistent model. The improvement is particularly accentuated for the DNI forecasting models (see Table 3). By comparing the statistical measures for training and test sets of each model, we notice there are significant improvements over the simpler models with the least inputs. As can be the case in modeling with ANNs, better training performance is achieved when more inputs are used. However, the use of more inputs does not necessarily translate to better generalization or ability to forecast. The results in Table 3a and b show that models with intermediate complexity perform best on both training and test sets.

As indicated by the RMSE and R^2 values, Perez's 2010 model (Perez et al., 2010) compares favorably with Model 1, which includes the same inputs (note that clear-sky GHI is approximated well by $\cos Z$). In this case, Perez's model should be favored over Model 1 because the later employs fewer parameters than Model 1, and no additional errors due to excessive computations are incurred. The Perez model has only two parameters whereas the ANN method requires over 20 parameters and is therefore overly complex in this case. However, a comparison between the MBE values for the ANN and Perez's models show that there is room to improve over Perez's simple model. If we take Perez's model as the baseline for RMSE comparisons, Model 1 underperforms by 6%, whereas Models 2–3

Table 3
Statistical summary of trained neural network models and comparison with persistent and Perez Model. Best values are denoted in bold fonts.

Model	Inputs	Training			Test		
		MBE [W/m²] (%)	RMSE [W/m²] (%)	R^2	MBE [W/m²] (%)	RMSE [W/m²] (%)	R^2
(a) [GHI]							
1	$[x_8, x_{10}]$	**+0.38 (+.1)**	77 (19.3)	0.939	**+0.60 (+.1)**	89 (22.8)	0.915
2	$[x_5, x_8, x_9, x_{10}]$	−1.35 (−0.3)	67 (17.2)	0.952	−3.23 (−0.6)	**72 (17.7)**	**0.947**
3	$[x_1, x_2, x_3, x_5, x_6, x_8, x_9, x_{10}]$	−2.25 (−0.5)	63 (15.7)	0.959	−3.10 (−.6)	73 (18.5)	0.945
4	[All]	−1.94 (−0.5)	**58 (14.8)**	**0.965**	−2.10 (−0.5)	74 (18.6)	0.942
Persistent GHI		–	–	–	−1.5 (−0.4)	123.1 (31.1)	0.854
Perez model		–	–	–	−10.73 (−2.50)	84 (21.7)	0.921
(b) [DNI]							
5	$[x_5, x_9, x_{11}]$	**−1.5 (−0.3)**	171 (35.0)	0.763	**−2.4 (−0.5)**	162 (32.3)	0.781
6	$[x_1, x_3, x_5, x_8, x_9, x_{11}]$	−2.7 (−0.5)	142 (29.1)	0.835	−3.4 (−1.3)	**156 (31.2)**	**0.801**
7	$[x_1, x_2, x_3, x_5, x_6, x_8, x_9, x_{11}]$	−6.3 (−9.5)	148 (30.1)	0.821	−7.1 (−10.3)	161 (33.2)	0.788
8	[All]	−1.3 (−1.9)	**138 (28.1)**	**0.845**	−4.2 (−8.3)	158 (32.0)	0.797
Persistent DNI		–	–	–	−8.40 (−1.7)	270 (54.5)	0.404

improve over Perez's model by over twice the same amount, indicating that the stochastic approach with more inputs represents a true gain that compensates for the additional complexity in computation. Again, some of these results were expected from the M-tests (see Fig. 3a).

It is also interesting to compare results from Table 3 with the M-test results in Fig. 3a and b. In comparing the actual training R^2s of each model with those predicted from the M-tests, we see that they are quite close, in particular, for the GHI models. For example, according to the M-test for Model 1, we could have expected an $R^2 \approx 0.945$, which is close to the actual $R^2 = 0.939$. The training R^2s for GHI are generally much closer to what was predicted in the M-test than for DNI. This is likely because the M-test plots have stabilized with a relatively smaller proportion of data (at $\approx 45\%$) as compared to DNI (M-test plots do not stabilize until after $\approx 80\%$ of total number of data points used). This also indicates GHI is much easier to predict than DNI and that additional data will be required to obtain better models for DNI.

One concern for using a larger input set is that the prediction errors of each input directly propagates to errors in forecasting the output. In order to observe the propagation of the NDFD uncertainty, the forecasting errors for all models for same-day and several days ahead are shown in Fig. 4a and b. These figures illustrate that forecasting errors increase with forecasting lead times as a result of uncertainty in the next few days of the inputs from the NDFD. This point is explained more explicitly in the next section. A general observation from these figures is that models with fewer inputs have lesser nominal increases in forecasting errors with respect to forecasting horizon. In addition, as the forecasting horizon increases to 4 or more days ahead there is less of a difference between each of the models, and so the preferred models would be Model 1 and Model 5 because fewer inputs are involved.

The rRMSEs given so far are a measure of model quality over the entire data set. In some cases it may be important to know if there are certain periods when the solar radiation is more predictable. The rRMSEs for GHI Model 3 and DNI Model 7 are calculated on a monthly basis to produces Fig. 5a and b. For both models the rRMSEs are lower than the aggregated rRMSE during the months between and including March through September. It is also apparent from these figures that during the summer months solar irradiation (GHI and DNI) is much more forecastable.

7. Prediction intervals

As a post-analysis step, prediction intervals are provided in order to examine how model and forecasting uncertainty depends the type of sky situation. For instance, as shown in Fig. 5a and b, model and forecasting accuracy is generally much better during summer months when there are few cloudy days. Similar as in Bacher et al. (2009) and Lorenz et al. (2009), prediction intervals are derived here by analyzing the residuals.

To construct the prediction intervals we make a few assumptions regarding the distribution of the residuals. One of the assumptions is that the expected value, or mean value, of the residuals (r) given some value of the inputs $\mathbf{x}$, denoted as $\mathsf{E}(r|\mathbf{x})$, is equal to zero, $\mathsf{E}(r|\mathbf{x}) = 0$. The next assumption is that the residuals with the expectation value $\mathsf{E}(r|\mathbf{x}) = 0$ and variance $\mathrm{Var}(r|\mathbf{x}) = \mathsf{E}(r^2|\mathbf{x})$ are normally distributed. Then letting $s(\mathbf{x}) = \sqrt{\mathsf{E}(r^2 \mid \mathbf{x})}$, the prediction limits of the models can be computed as

$$y_{prediction\ limits} = f(\mathbf{x}) \pm z_{\alpha/2} s(\mathbf{x}), \tag{16}$$

where $z_{\alpha/2}$ is the z-score for the confidence level $1 - \alpha$. For 95% confidence levels, $z_{\alpha/2} = 1.96$. In Lorenz et al. (2009), $s(\mathbf{x})$ was modeled using a fourth degree polynomial using the predicted clearness index and cloud cover. Similarly, we applied regression using ANNs to produce model estimates of $s(\mathbf{x})$.

Fig. 6a shows time-series plots of measured and modeled GHI with prediction bands for same-day forecasts. The upper limits of the bands were limited to a maximum value of 1000 W/m². Although approximately 95% of all the observed points fell within the confidence intervals, most
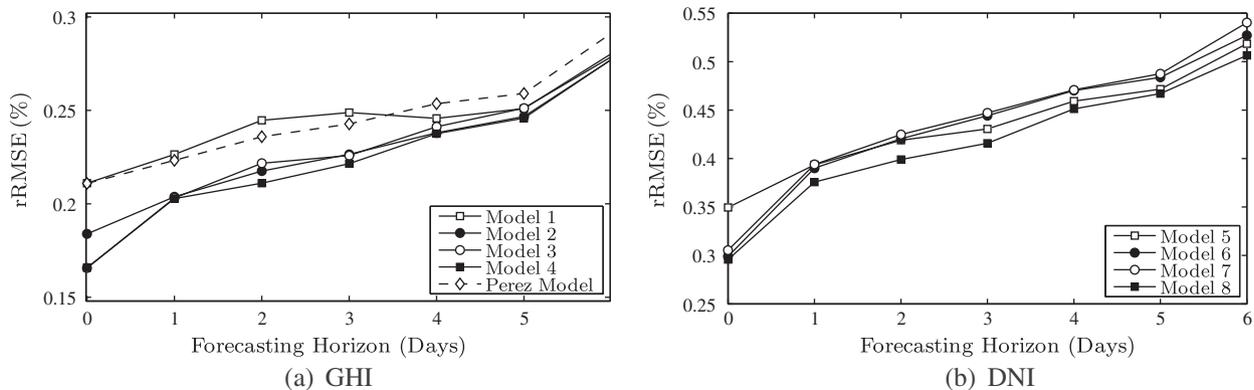

Fig. 4. The relative RMSE (%) for forecasts using models with different inputs as indicated in the legend on graph. The horizontal axis represents the number of days ahead where 0 denotes same-day.
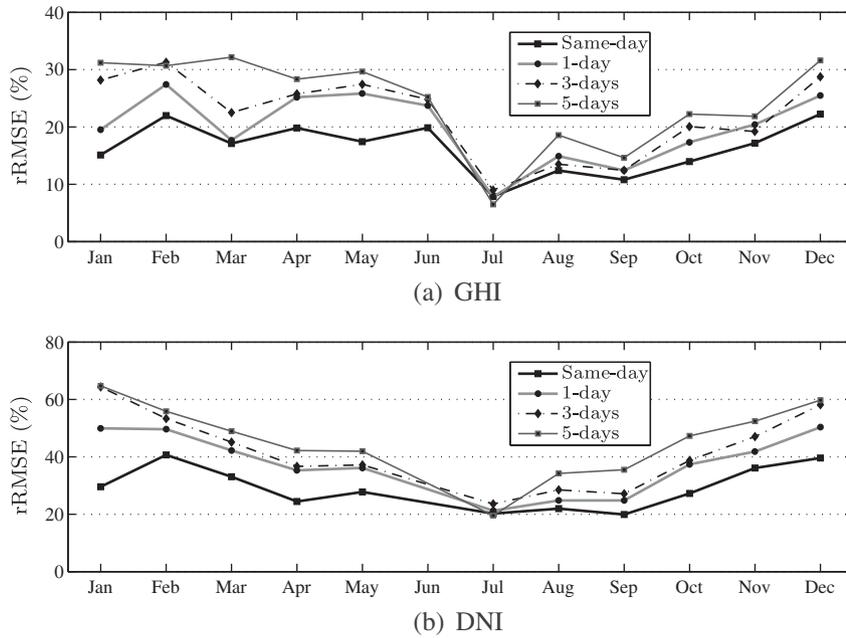
(a) GHI

(b) DNI

Fig. 5. The relative RMSE by month from beginning of November 2008 to end of November 2009. The rRMSEs are normalized with respect to $GHI_{mean} = 474$ W/m² and $DNI_{mean} = 541$ W/m². The rRMSEs are for modeling DNI for same-day and forecasts 1–6-days ahead as indicated by legend on graph. The DNI data sets for June 2009 were deemed unreliable due to mechanical problems with the measurement tracking system, and therefore these values are not reported here.

of the departures were clustered on a few days, while most days had no departures. These systematic departures are probably due to errors in the inputs for which the estimates of prediction bands are based on.

For clear days, the prediction bands are relatively narrow indicating that GHI is very predictable for clear-sky situations. For days with more sky cover, however, irradiance is much less predictable as indicated by the wider prediction bands. Fig. 6b shows 1-day ahead predictions of GHI for the same days as in Fig. 6a. These results coincide with previous results from Bacher et al. (2009) and Lorenz

et al. (2009), where they characterized larger forecasting uncertainty during partly cloudy to overcast conditions for solar power and irradiance model predictions, respectively.

Fig. 8 shows sky cover values for the same days used in the prediction interval plots. For the day 4/24/2009, the 1-day ahead forecasted hourly values are underestimated, and thus the prediction bands are smaller for the 1-day ahead irradiance predictions than the same-day predictions. The opposite is true for the sky cover and predictions bands for 4/25/2009. These figures show that forecasting

(a) Same-day forecasted hourly values and measured values.

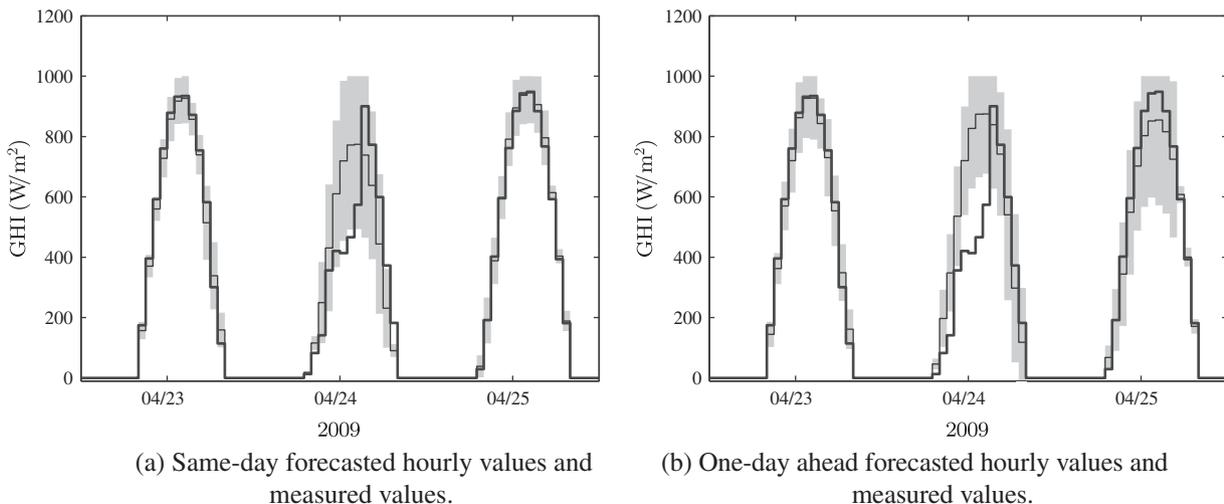(b) One-day ahead forecasted hourly values and measured values.

Fig. 6. Forecasts from Model 3 with prediction intervals for hourly GHI. The dark-bold lines are the measured values, the thin lines are the predictions, and the shaded regions are the 95% confidence intervals.
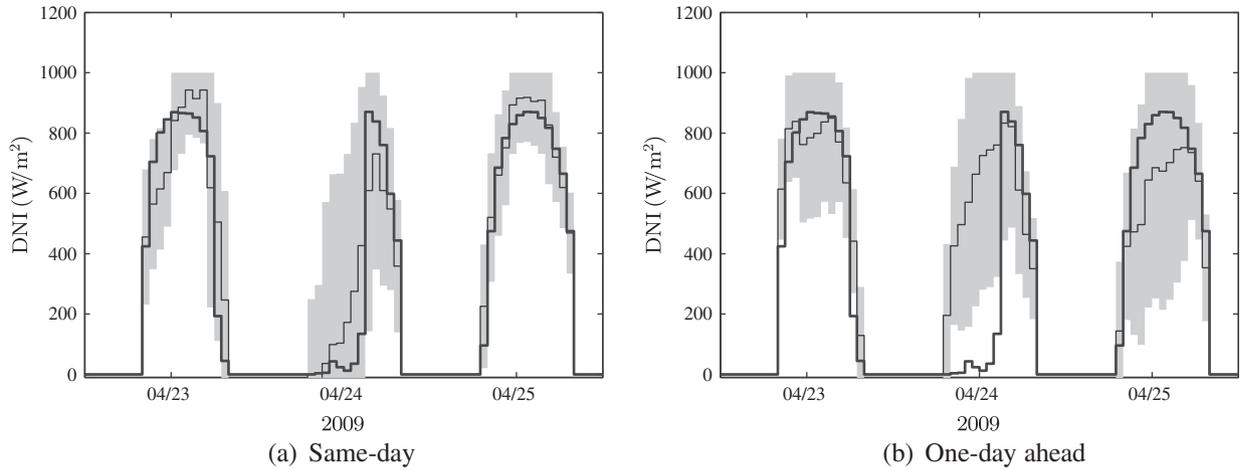
Fig. 7. Forecasts with prediction intervals for hourly DNI. The dark-bold lines are the measured values, the thin lines are the predictions, and the shaded regions are the 95% confidence intervals.
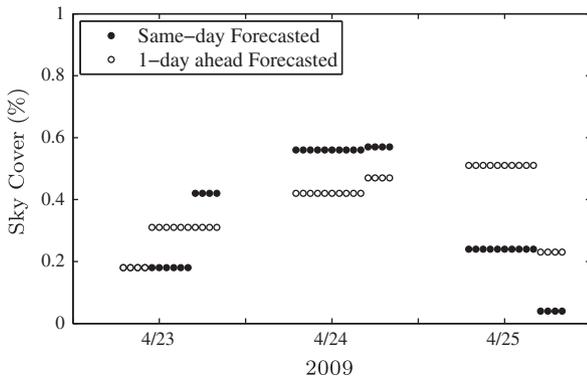

Fig. 8. Hourly values of sky cover for same days used in the prediction interval plots. Night values were removed.

uncertainty is proportional to sky cover and also that the irradiance predictions and prediction intervals are directly related to sky cover forecasting uncertainty.

One can also compare the prediction intervals for same-day forecasts with 1-day ahead forecasts for DNI, Fig. 7a and b. The prediction intervals for DNI are much wider than for GHI, further illustrating that DNI is relatively more difficult to reliably predict.

8. Conclusions

We developed forecasting models for hourly solar irradiation using Artificial Neural Networks for lead times of up to 6 days. Model inputs included current and forecasted meteorological data obtained from the US National Weather Services forecasting database, and solar geotemporal variables. The normalized hour angle was introduced as an effective input for modeling and forecasting solar irradiation. An input selection scheme was applied using the Gamma test combined with a master genetic algorithm-based method to reveal the most relevant set of inputs which included the solar geometry variables, sky

cover, probability of precipitation, and minimum and maximum temperatures.

For same-day forecasts of GHI, rRMSEs range from 15% to 22% for different models constructed on 13 month data set. The same-day forecasts show that the models constructed here compare favorably to those of satellite-based models (Perez et al., 2002; Schillings et al., 2004; Vignola et al., 2007), and other medium-range (lead times of 1–6-days) forecasting models (Lorenz et al., 2009; Bacher et al., 2009; Breitkreuz et al., 2009; Perez et al., 2007, 2010). Implementation of Perez's latest GHI forecasting model to the data set used here yields an RMSE of 84.5 W/m^2 (rRMSE = 21.7%) for same-day forecasts which compares well with one of our simple models that uses the same inputs. Models with slightly larger sets of inputs generally perform better for same-day and 1-day ahead forecasts. The DNI forecasting models were not compared to any established models because their is none available to compare. However, this study finds DNI to be generally much more difficult to predict reliably: rRMSEs obtained on same-day forecasts are in the range of 28–35%. This trend is in agreement with previous satellite-based studies (see e.g., Vignola et al., 2007).

In general, and as expected, the models lose accuracy with increasing forecasting horizon but less sharply during the summer than during the winter months. This is due to long periods of consecutive clear days in the summer and a large number of cloudy to overcast days during the winter for California's Central Valley. These weather characteristics are important for the siting and operation of solar farms, since the high irradiation summer months coincide with peak power demand in the region. Solar power is not only abundant in the region studied, but it is also very predictable during the peak demand season.

References

Bacher, P., Madsen, H., Nielsen, H.A., 2009. Online short-term solar power forecasting. Solar Energy 83, 1772–1783.

Bishop, C., 1994. Neural networks and their application. Review of Scientific Instruments 65, 1803–1832.

Breitkreuz, H., Schroedter-Homscheidt, M., Holzer-Popp, T., Dech, S., 2009. Short-range direct and diffuse irradiance forecasts for solar energy applications based on aerosol chemical transport and numerical weather modeling. Journal of Applied Meteorology and Climatology 48, 1766–1779.

Duffie, J.A., Beckman, W.A., 2006. Solar Engineering of Thermal Processes, third ed. John Wiley & Sons, Inc., Hoboken, New Jersey.

Durrant, P.J., 2001. *winGamma^T M*: A Non-linear Data Analysis and Modelling Tool for the Investigation of Non-linear and Chaotic Systems with Applied Techniques for a Flood Prediction System. Ph.D. Thesis, Department of Computing Science, Cardiff University.

Evans, D., Jones, A., 2002. A proof of the Gamma test. Proceedings of the Royal Society of London Series A – Mathematical Physical and Engineering Sciences 458, 2759–2799.

Fthenakis, V., Alsema, E., 2006. Photovoltaics energy payback times, greenhouse gas emissions and external costs: 2004-early 2005 status. Progress in Photovoltaics: Research and Applications 14, 275–280.

Glahn, H., Ruth, D., 2003. The new digital forecast database of the national weather service. Bulletin of the American Meteorological Society 84, 195–201.

Heinemann, D., 2004. Forecasting of solar irradiation. In: Proceedings of the International Workshop of Solar Resource from the Local Level to Global Scale in Support of the Resource Management of Renewable Electricity Generation, Institute for Environment and Sustainability, Joint Research Center, Ispra, Italy.

Jones, A., 2004. New tools in non-linear modelling and prediction. Computational Management Science 1, 109–149.

Koncar, N., 1997. Optimisation Methodologies for Direct Inverse Neurocontrol. Ph.D. Thesis, Department of Computing Imperial College of Science, Technology and Medicine, University of London.

Lauret, P., David, M., Fock, E., Bastide, A., Riviere, C., 2006. Bayesian and sensitivity analysis approaches to modeling the direct solar irradiance. Journal of Solar Energy Engineering – Transactions of the ASME 128, 394–405.

Lorenz, E., Hurka, J., Heinemann, D., Beyer, H.G., 2009. Irradiance forecasting for the power prediction of grid-connected photovoltaic systems. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 2, 2–10.

Mellit, A., 2008. Artificial Intelligence technique for modelling and forecasting of solar radiation data: a review. International Journal of Artificial Intelligence and Soft Computing 1, 52–76.

Moghaddamnia, A., Remesan, R., Kashani, M.H., Mohammadi, M., Han, D., Piri, J., 2009. Comparison of LLR, MLP, Elman, NNARX and ANFIS models-with a case study in solar radiation estimation. Journal of Atmospheric and Solar-Terrestrial Physics 71, 975–982.

Pacca, S., Deepak, S., Keoleian, G.A., 2007. Parameters affecting the life cycle performance of PV technologies and systems. Energy Policy 35, 3316–3326.

Perez, R., Ineichen, P., Moore, K., Kmiecik, M., Chain, C., George, R., Vignola, F., 2002. A new operational model for satellite-derived irradiances: description and validation. Solar Energy 73, 307–317.

Perez, R., Moore, K., Wilcox, S., Renne, D., Zelenka, A., 2007. Forecasting solar radiation – preliminary evaluation of an approach based upon the national forecast database. Solar Energy 81, 809–812.

Perez, R., Kivalov, S.J.S., Hemker, K.J., Renne, D., Hoff, T., 2010. Validation of short and medium term operational solar radiation forecasts. Solar Energy 84, 2161–2172.

Remesan, R., Shamim, M.A., Han, D., 2008. Model data selection using gamma test for daily solar radiation estimation. Hydrological Processes 22, 4301–4309.

Schattel Jr., J.L., Bunge, R., 2008. The national weather service shares digital forecasts using web services. Bulletin of the American Meteorological Society 89, 449–450.

Schillings, C., Meyer, R., Mannstein, H., 2004. Validation of a method for deriving high resolution direct normal irradiance from satellite data and application for the Arabian Peninsula. Solar Energy 76, 485–497.

Stefansson, A., Koncar, N., Jones, A., 1997. A note on the Gamma test. Neural Computing & Applications 5, 131–133.

Vignola, F., Harlan, P., Perez, R., Kiniecik, M., 2007. Analysis of satellite derived beam and global solar radiation data. Solar Energy 81, 768–772.

Wilson, I., Jones, A., Jenkins, D., Ware, J., 2004. Predicting housing value: genetic algorithm attribute selection and dependence modelling utilising the Gamma test. In: Binner, J.M., Kendall, G., Chen, S.H. (Eds.), Applications of Artificial Intelligence in Finance and Economics, vol. 19 of Advances in Econometrics: A Research Annual. JAI-Elsevier Sci BV, Sara Burgerhartstraat, PO Box 211, 1000 AE Amsterdam, Netherlands, pp. 243–275.