Progress in Energy and Combustion Science 39 (2013) 535–576

Contents lists available at SciVerse ScienceDirect

Progress in Energy and Combustion Science

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

Review

Solar forecasting methods for renewable energy integration

Rich H. Inman, Hugo T.C. Pedro, Carlos F.M. Coimbra*

Department of Mechanical and Aerospace Engineering, Jacobs School of Engineering, Center of Excellence in Renewable Energy Integration and Center for Energy Research, University of California, 9500 Gilman Drive, La Jolla, CA 92093, USA

ARTICLE INFO

Article history:
Received 14 January 2013
Accepted 16 June 2013
Available online 26 July 2013

Keywords:
Weather-dependent renewable energy
Solar forecasting
Solar meteorology
Solar variability
Solar energy integration
Evolutionary forecasting methods

ABSTRACT

The higher penetration of renewable resources in the energy portfolios of several communities accentuates the need for accurate forecasting of variable resources (solar, wind, tidal) at several different temporal scales in order to achieve power grid balance. Solar generation technologies have experienced strong energy market growth in the past few years, with corresponding increase in local grid penetration rates. As is the case with wind, the solar resource at the ground level is highly variable mostly due to cloud cover variability, atmospheric aerosol levels, and indirectly and to a lesser extent, participating gases in the atmosphere. The inherent variability of solar generation at higher grid penetration levels poses problems associated with the cost of reserves, dispatchable and ancillary generation, and grid reliability in general. As a result, high accuracy forecast systems are required for multiple time horizons that are associated with regulation, dispatching, scheduling and unit commitment. Here we review the theory behind these forecasting methodologies, and a number of successful applications of solar forecasting methods for both the solar resource and the power output of solar plants at the utility scale level.
© 2013 Elsevier Ltd. All rights reserved.

Contents

* Corresponding author. Tel.: +1 858 534 4285.
E-mail address: ccoimbra@ucsd.edu (C.F.M. Coimbra).

0360-1285/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.pecs.2013.06.002

1. Introduction

The global demand for renewable energy integration to the power grids highlights the importance of economic and technological issues associated with growing levels of flat-panel Photo-Voltaic (PV), Concentrated Solar Power (CSP), and Concentrated PV (CPV) penetrations into the power grid. These concerns arise from the variable nature of the solar resource, seasonal deviations in production and load profiles, the high cost of energy storage, and the balance between grid flexibility and reliability [1,2]. As a result, solar plants are often backed by ancillary generators for periods of high variability, which increases the capital and operational costs of solar generation. Accurate solar forecasts over several time horizons are required so that Independent System Operators (ISOs) or equivalent grid balancing authorities are able to successfully integrate increased levels of solar power production while maintaining reliability [3–5]. Solar forecasts on multiple time horizons become increasingly important as solar penetration grows for the purposes of grid regulation, load-following production, power scheduling and unit commitment. Short-term, intra-hour solar forecasts are particularly useful for power plant operations, grid balancing, real-time unit dispatching, automatic generation control (AGC) and trading. Forecasts for longer time horizons are of interest to utilities and ISOs for unit commitment, scheduling and for improving balance area control performance. Ultimately, a spectrum of solar forecasts is required to address the planning,

operational and balancing needs of both the distribution and the transmission grids. Solar forecasting is therefore an enabling technology for the integration of ever increasing level of solar penetration into the grid because it improves the quality of the energy delivered to the grid and reduces the ancillary costs associated with weather dependency. The combination of these two factors (better energy quality through information that is capable of lowering integration and operational costs) has been the driving motivation for the development of a complex field of research that aims at producing better solar forecasting capabilities for the solar resource at the ground level and for the power output from different solar technologies that depend on the variable irradiance at the ground level. Solar, wind and load forecasting have become integral parts of the so-called 'smart grid concept'.

To date, high-fidelity, robust solar forecast systems that work for widely different microclimates remain evasive. The problem is of great complexity due to the non-linear and chaotic effect of cloud motion on solar irradiance at the ground level. However, a number of promising approaches have been developed in the past few years, and the incipient research field of solar meteorology for renewable generation has grown considerably by aggregating diverse areas of knowledge such as atmospheric physics, solar instrumentation, machine learning, forecasting theory and remote sensing in its quest to better predictive skills. This work presents an overview of the forecasting methods for solar resourcing and solar power generation, as well as the theoretical basis for the most promising methods, and a discussion on their effectiveness for operational use.

2. Fundamental considerations

Here we review some fundamental concepts and methods including solar irradiance components, airmass, Linke turbidity, clear sky models, clear sky and clearness indices, persistence forecasts, and a metric for the evaluation of solar forecasting skill. These topics lay the foundation for the development of solar forecasting methodologies that are covered in the following sections.

2.1. Solar irradiance components

The total extraterrestrial beam irradiance incident on the earth's atmosphere I_0 fluctuates about an average value of approximately 1360 Wm^{-2} [6]. This incident radiation is attenuated as it negotiates its way to the ground level through a complex series of multiple reflections, absorptions and re-emissions due to interactions with atmospheric constituents [7]. This results in the division of the incident extraterrestrial beam radiation into two distinct components: Direct Normal Irradiance (DNI) and Diffuse Horizontal Irradiance (DHI). The geometric sum of which is the Global Horizontal Irradiance (GHI) which can be written as

$$\mathrm{GHI} = \mathrm{DHI} + \mathrm{DNI}\cdot\cos\theta \tag{1}$$

where θ is the solar zenith angle.

2.2. Airmass and Linke Turbidity

Two coefficients which are employed quite often in solar engineering are the atmospheric airmass (AM) and the Linke Turbidity (T_L). In the context of this review, airmass refers to the optical path length of the atmospheric boundary layer through which radiation must propagate in order to reach ground level. When the sun is directly overhead, $\theta = 0$, the airmass has a value of unity, AM = 1. As the solar zenith angle θ increases, so does the airmass. Many models of the airmass as a function of solar zenith angle, AM(θ), exist (see Ref. [8] for a more detailed review). It is convenient to note that the irradiance at AM = 0 is the extraterrestrial irradiance $I_0 = 1360\ \mathrm{Wm}^{-2}$ [6].

The Linke turbidity coefficient represents the number of clean and dry atmospheres that would be required to produce the observed irradiance at ground level. In 1922 Linke proposed that the total optical thickness of a cloudless atmosphere could be expressed as a function of the optical thickness of a clear and dry atmosphere, δ_{cda}, and the Linke turbidity coefficient, T_L, as

$$\mathrm{DNI} = I_0\cdot\exp(-\delta_{\mathrm{cda}}\cdot T_L\cdot\mathrm{AM}). \tag{2}$$

It should be noted that several improvements have been made to this model including Kasten and Young's 1989 correction [9] and Ineichen's airmass independent formulation [10]. Again, the reader is referred to [8] for more information.

2.3. Clear sky models

Many solar irradiance forecasting models require the knowledge of clear sky conditions. Clear sky conditions are characterized by the absence of clouds. Clear sky models are used for persistence forecasts, discussed in Section 2.5, and for normalization of the forecasting skill metric described in Section 2.6. Satellite based forecasts, covered in Section 5, also use clear sky models to derive ground solar irradiance components through normalization of satellite data and are frequently used for locations where no other measurements exist.

Clear sky models are typically developed using one of several Radiative Transfer Models (RTMs) and require local meteorological inputs such as ozone content, water vapor content and/or Linke turbidity in combination with solar geometry. A comprehensive comparison of eight clear sky models against 16 independent data banks was published by Ineichen in 2006 [11]. Ineichen concluded that the input parameters, most notably turbidity, had the highest influence on model accuracy. In addition, Ineichen found that accuracy is not highly dependent on the model itself. Therefore, the model selection criteria can be tailored to the application and availability of data. If access to comprehensive climactic input parameters are not available, the ESRA or Ineichen model would be appropriate due to the requirement of only a single input parameter. If access to input parameters is not an issue or if spectral resolution is required, the Solis model would be an appropriate choice as it gives the overall best results according to [11]. The next section briefly describes ten different clear sky models, eight of which were included in Ineichen's comprehensive comparison.

2.3.1. Solis model

The Solis model is a spectrally resolved clear sky transmittance model that was developed by Mueller et al. [12] in 2004 for the European Heliosat-3 project. The Solis model makes use of an RTM, Beer–Lambert functions and integration over the solar spectrum in order to obtain irradiance components. Necessary inputs include ozone content, water vapor content and aerosol optical depth at 550 nm.

2.3.2. European Solar Radiation Atlas (ESRA) model

The European Solar Radiation Atlas (ESRA) model was developed for the Heliosat-2 project through the work of Rigollier et al. [13] and refined by Geiger et al. [14]. ESRA was later adapted by Zarzalejo et al. [15] and Badescu [16]. The model is derived from the Linke turbidity factor and Kasten's Rayleigh optical depth parameterization [17]. The only required input parameter is the Linke turbidity factor at an airmass of 2.

2.3.3. Bird and Hulstrom model

As the name implies, Bird and Hulstrom [18] developed this model using RTM schemes based on McClatchey and Selby's LOWTRAN [19]. The authors referred to this new scheme as SOLTRAN. Three parameters are required as inputs: water vapor column, aerosol optical depth (at 380 and 500 nm) and the ozone column. For more information see Refs. [11,18].

2.3.4. Molineaux model

The Molineaux model was developed in 1998 by Molineaux [20] and is based on a Moderate Resolution model for LOWTRAN (MODTRAN) developed by Berk et al. [21] and Smarts2 which is a product of Gueymard's work [22]. Because Molineaux's model is based on the equivalence of pyrheliometric and mono-chromatic aerosol optical depths, it cannot be used to calculate GHI. Two inputs are required: the aerosol optical depth at 700 nm and the water vapor column.

2.3.5. Ineichen model

Previous to the development of this model, the Linke turbidity coefficient had the disadvantage of being dependent on airmass. This widely used clear sky model was developed by Ineichen and Perez in 2002 [10] to establish the Linke turbidity coefficient independently of the airmass. The model possesses the attractive property that only the Linke turbidity is required as an input.

2.3.6. CPCR2 model

The CPCR2 model was developed by Gueymard in 1989 [23] and is a dual band technique for modeling solar radiation which parameterizes the transmittance of each extinction layer. The components of solar radiation (DNI and GHI) are then calculated as functions of the extinction layer transmittances. Inputs to the model include the Ångstrom size coefficient, the Ångstrom turbidity coefficient and the aerosol optical depth.

2.3.7. REST2 model

REST2 is a dual-band model developed by Gueymard in 2004 [24] which is essentially a modified version of the CPCR2 model. The REST2 incorporates the latest extraterrestrial spectral distribution, solar constant value and updated versions of the transmittances functions in the CPCR2 model. This model requires water vapor content, Ångstrom turbidity coefficient and reduced NO_2 and O_3 vertical path lengths as inputs.

2.3.8. Kasten model

This model is based on Kasten's pyrheliometric formula developed in Ref. [25]. The model calculates ground irradiance components by considering the absorption and scattering that takes place at altitudes of 8000 and 2500 m. Like the Ineichen model, it conveniently only requires the Linke turbidity at airmass 2 as an input. A full description can be found in Kasten's 1984 work [26].

2.3.9. Polynomial fit

One of the simplest clear sky models is a location specific polynomial of the cosine of the solar zenith angle $\cos(\theta)$ given by

$$I^{\mathrm{clr}_p} = \sum_{n=0}^{N} c_n(\cos(\theta))^n. \tag{3}$$

Coimbra and Marquez [27] showed that typically a third order polynomial fit ($N = 3$) is sufficient. The third order polynomial can be written

$$I_t^{\mathrm{clr}_p} = c_3(\cos\theta_t)^3 + c_2(\cos\theta_t)^2 + c_1(\cos\theta_t) + c_0, \tag{4}$$

where the coefficients are determined by fitting the polynomial in (4) to the clear sky days in the data set. The limitation of this model is that a historical database of clear days is required for the specific location. Typically such a historical database is not available. If this is the case, one of the other clear sky models should be employed.

2.3.10. ASCE evapotranspiration model

R. L. Snyder [28] described the steps required to estimate reference evapotranspiration using hourly weather data in a method that was adopted by the American Society of Civil Engineers (ASCE-EWRI, 2004). The method, which employs the Penman-Monteith (hourly) reference evapotranspiration equations, calculates, as an intermediate step, clear sky total global solar radiation at the Earth's surface. This model combines Duffie and Beckman's equations [29] for extraterrestrial radiation and a clear sky GHI model from Allen et al. [30]. Only location (latitude, longitude and altitude) is required as an input.

2.4. Clear sky & clearness indices

Two parameters which are frequently used in solar forecasting are the clear sky index and the clearness index. They are similar and differ only in the normalization factor used in their definition. The clear sky index k_t is defined as the ratio of the measured irradiance to the modeled clear sky irradiance at ground level

$$k_t = \frac{I_t}{I_t^{\mathrm{clr}}}. \tag{5}$$

The clearness index K_t is similarly defined with normalization of the ground level irradiance with respect to the extraterrestrial irradiance

$$K_t = \frac{I_t}{I_t^{\mathrm{EX}}}. \tag{6}$$

Extraterrestrial irradiance is much easier to model than the clear sky irradiance due to the absence of physical atmospheric fluctuations. A simple extraterrestrial irradiance model can be written as

$$I_t^{\mathrm{EX}} = I_0\cos(\theta_t), \tag{7}$$

where $I_0 = 1360\ \mathrm{W/m^2}$ [6] is the solar constant and θ_t is the solar zenith angle at time t. In contrast to the clear sky models, this extraterrestrial irradiance model does not involve any location specific fitting parameters except for θ_t which depends on latitude.

Because k_t and K_t are dependent on I_t^{clr} and I_t^{EX} respectively, there can be several clear sky or clearness indices depending on which clear sky model or extraterrestrial model is used. Fig. 1 illustrates the relationship between the clear sky and clearness indices. It is important to note that the clear sky index k_t is always greater than the clearness index K_t due to the difference in normalization factors used.

2.5. Persistence forecasts

Persistence forecasts are some of the most simple forecasts to implement and are often used as a baseline for the performance evaluation of other forecast engines. Persistence forecasts, as the name implies, are defined as having the clear sky or clearness index persist for the next time step. For the clear sky persistence forecast we define the following persistence condition

$$k_{t+\Delta t} = k_t = \frac{I_t}{I_t^{\mathrm{clr}}} \tag{8}$$

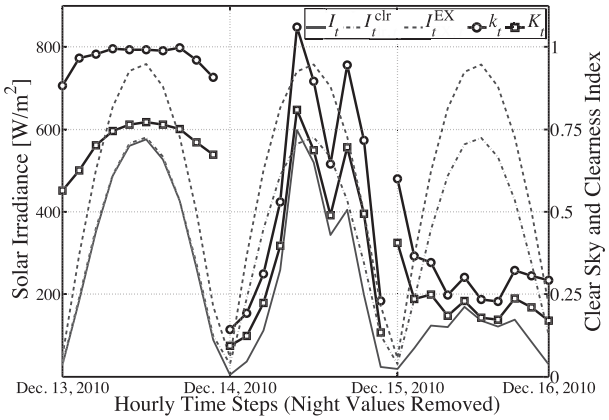


Fig. 1. Time-series of measured global horizontal irradiance values I_t, estimated clear sky irradiance I_t^{clr}, estimated extraterrestrial irradiance I_t^{EX} with corresponding clear sky index k_t and clearness index K_t. Note that the clearness index is always less then the clear sky index by definition. The Ineichen model was used for I_t^{clr} and Equation (7) was used for I_t^{EX}. Measured hourly data is for December 13–15, 2010 at UCSD EBUII in La Jolla, CA (courtesy of Jan Kleissl). Night values have been removed for clarity.

where Δt is the forecast horizon. Similarly, for the clearness persistence forecast we define a corresponding persistence condition

$$K_{t+\Delta t} = K_t = \frac{I_t}{I_t^{\mathrm{EX}}}. \tag{9}$$

From these indices, the forecast of the solar irradiance at the next time step is computed as

$$I_{t+\Delta t}^{\mathrm{pers}_k} = k_t I_{t+\Delta t}^{\mathrm{clr}} \tag{10}$$

and

$$I_{t+\Delta t}^{\mathrm{pers}_K} = K_t I_{t+\Delta t}^{\mathrm{clr}}. \tag{11}$$

Again, since these forecasts rely on the definition of a clear sky or clearness index, there can be several persistence forecasts depending on how I_t^{clr} and I_t^{EX} are estimated. Two noteworthy properties of the persistence forecasts are shown in Fig. 2. First, during cloudless periods the persistence forecast performs quite well, however, large forecasting errors occur during rapid changes in irradiance I which result from the passing of opaque clouds. Secondly, the persistence forecast displays an obvious time delay due to the implied persistence of the clear sky indices.

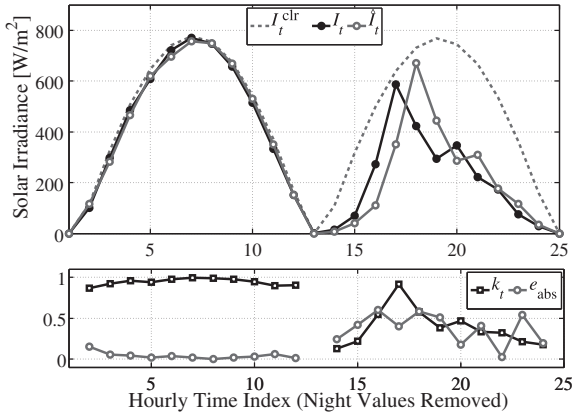


Fig. 2. Example of persistence forecast performance for a clear and cloudy day (Oct. 11, 2010 and Oct. 13, 2010) at UCSD EBUII in La Jolla, CA (data courtesy of Jan Kleissl). The persistence forecast performs relatively well on the clear day. However, large errors which occur during abrupt changes in measured irradiance and a 'time delay' are visible in the persistence forecast during the cloudy day. Clear sky index k_t and absolute error $e_{\mathrm{abs}} = \left|I_t - \widehat{I}_t\right|/I_t$ of the persistence forecast are shown on the lower part of the graph for the same time steps.

2.6. Evaluation of solar forecasting skill

Historically, traditional power generation technologies such as fossil and nuclear power which were designed to run in stable output modes, have resulted in the majority of power grid variability originating from demand fluctuations [3,27]. On the other hand, energy sources that rely on the solar resource exhibit a high degree of variability (see Fig. 3) which introduces fluctuations also on the energy supply side. Recent studies suggest that in order to facilitate higher market and grid penetration of solar power, ISOs need accurate forecasts of solar irradiance on multiple time horizons [3–5]. Despite the large number of forecasting methods described in this work, the comparison of results and evaluation of relative advantages between models has been evasive. These difficulties arise from the fact that solar irradiance is inherently dependent on geographic location, time of year and climate as well as the different evaluation techniques used by various authors for the quality assessment of their models and forecasts, denoted by $\widehat{I}$. Traditional statistical metrics to characterize model quality include the coefficient of determination which compares the variance of the errors to the variance of the modeled data

$$R^2 = 1 - \frac{\mathrm{Var}\left(\widehat{I} - I\right)}{\mathrm{Var}(I)}, \tag{12}$$

the Mean Absolute Error (MAE) which considers only the absolute value of the errors

$$\mathrm{MAE} = \frac{1}{N}\sum_{t=1}^{N}\left|\widehat{I}_t - I_t\right|, \tag{13}$$

the Mean Absolute Percentage Error (MAPE)

$$\mathrm{MAPE} = \frac{100\%}{N}\sum_{t=1}^{N}\left|\frac{\widehat{I}_t - I_t}{I_t}\right|, \tag{14}$$

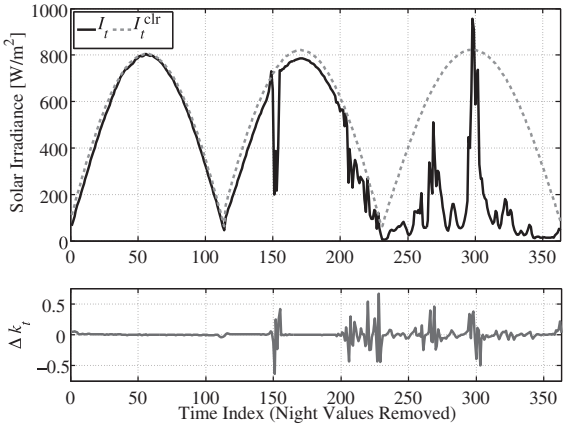


Fig. 3. Measured GHI values at 5 min resolution, Ineichen model clear sky irradiance values and the calculated values of step-changes of the clear sky index Δk_t (Data for Sep. 28–30, 2010 at EBUII in La Jolla, CA; courtesy of Jan Kleissl). During clear periods the solar variability is near zero. Ramps in the solar irradiance coincide with ramps in the Δk_t signal.

and the correlation coefficient

$$\rho = \frac{\left(\mathrm{Cov}\left(\widehat{I}, I\right)\right)^2}{\mathrm{Var}\left(\widehat{I}\right)\mathrm{Var}(I)}. \tag{15}$$

In recent years, the European and International Energy Agency (IEA) developed recommendations for reporting irradiance model accuracy based on three validation metrics: Root Mean Square Error (RMSE), Mean Bias Error (MBE), and Kolmogorov–Smirnov Integral (KSI) [31–34]. The RMSE is a measure of the average spread of the errors

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(\widehat{I}_t - I_t\right)^2}, \tag{16}$$

the MBE is a measure of the average bias of the model

$$\mathrm{MBE} = \frac{1}{N}\sum_{t=1}^{N}\left(\widehat{I}_t - I_t\right), \tag{17}$$

and the KSI aims to quantify the ability of a model to reproduce observed statistical distributions

$$\mathrm{KSI} = \int_{x_{\min}}^{x_{\max}} D_n \mathrm{d}x \tag{18}$$

where D_n is the difference between two cumulative distribution functions. Trapezoidal integration is typically employed as a result of D_n being a discrete variable. Respectively, these metrics provide a measure of the model's dispersion (RMSE), average bias (MBE), and ability to reproduce frequency distributions (KSI) [35].

However, in many contexts relative error is more commonly desired than absolute error, especially for users in the utility industry. Simplified reporting approaches for relative (percent) versions RMSE and MBE were proposed in a recent paper by Hoff et al. [35]. Additionally a method for reporting %KSI was proposed by Perez et al. [36].

According to Hoff et al. [35], the MAE normalized by the average value of the irradiance provides the best practical measure of relative dispersion error based on the evaluation criteria used in their study. This measure can be written as

$$\mathrm{MAE/Avg.} = \frac{\frac{1}{N}\sum_{t=1}^{N}\left|\widehat{I}_t - I_t\right|}{\frac{1}{N}\sum_{t=1}^{N} I_t} = \left(\frac{1}{\sum_{t=1}^{N} I_t}\right)\sum_{t=1}^{N}\left|\widehat{I}_t - I_t\right| \tag{19}$$

and is independent of the number of observations as well as being simple to understand. Hoff et al. also found that normalization of the RMSE by the maximum nominal irradiance, or capacity $C = 1000\ \mathrm{Wm}^{-2}$, which can be written as

$$\begin{aligned}\mathrm{RMSE/Capacity} &= \frac{\sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(\widehat{I}_t - I_t\right)^2}}{C} \\ &= \left(\frac{1}{C\sqrt{N}}\right)\sqrt{\sum_{t=1}^{N}\left(\widehat{I}_t - I_t\right)^2}\end{aligned} \tag{20}$$

is also desirable and currently widely accepted in the wind power industry.

Normalization of the KSI is typically defined as

$$\mathrm{KSI} = \frac{\int_{x_{\min}}^{x_{\max}} D_n \mathrm{d}x}{a_c} \tag{21}$$

where the critical area a_c is calculated as

$$a_c = V_c(\Delta x_{\max} - x_{\min}) \tag{22}$$

and the critical value V_c is calculated for a 99% level of confidence as

$$V_c = \frac{1.63}{\sqrt{N}}, \quad N \geq 35 \tag{23}$$

according to [37]. In order to circumvent the strong dependence of V_c on the number of samples, Perez et al. [36] suggest using the %KSI or absolute integrated difference between the modeled and measured normalized cumulative distribution functions.

More recently, Marquez and Coimbra presented a novel approach for the evaluation of the quality of forecast models based on the comparison of solar resource variability and forecast uncertainty [27]. This relationship between variability and uncertainty provides a consistent metric that is independent of the time horizon. However, the metric requires a clear sky model for normalization purposes. As discussed in Section 2.3, the accuracy of clear sky models is not dependent upon the model itself, but rather the selection of input parameters the model is dependent upon [10]. Assuming that access to accurate input parameters for clear sky modeling is available, we now address the two quantities used to determine the solar forecasting metric.

2.6.1. Solar variability

Variability of solar irradiance at ground level is affected by a number of factors, but most importantly solar position and cloud cover [27]. Fluctuations due to solar position are completely deterministic and are typically calculated through the use of clear sky or clearness models, see Section 2.3. Cloud induced fluctuations, on the other hand, are stochastic processes difficult to model and predict. As a result, the accuracy of the solar irradiance forecasting models depends almost exclusively on the ability to forecast the stochastic component induced by cloud cover [5,38,39]. For this reason, in the definition of solar irradiance variability, the authors in Ref. [27] addressed only the stochastic component and removed fluctuations due to annual and diurnal changes of solar position. This was accomplished by defining solar irradiance variability V as the standard deviation of the step-changes of the clear sky or clearness index as

$$V = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(\frac{I_t}{I_t^{\mathrm{clr}}} - \frac{I_{t-\Delta t}}{I_{t-\Delta t}^{\mathrm{clr}}}\right)^2} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(\Delta k_t)^2}. \tag{24}$$

The definition of solar variability above is essentially synonymous with those used by Kleissl and Lave [40] and Hoff [41] except for the refinement to include the deterministic changes as done in Refs. [38,39]. Fig. 3 shows a sequence of clear and cloudy days with the associated values of Δk_t. For a clear day the removal of the deterministic variability yields very small fluctuations of Δk_t whereas for the cloudy days the large ramps manifest themselves in the Δk_t curve.

2.6.2. Forecast uncertainty

Uncertainty of the forecast, as defined in Ref. [27], is a normalized version of the RMSE used in Ref. [42], which is similar to the

relative RMSE used in Ref. [43,44]. The difference here is that normalization is performed with respect to I_{clr}, rather than the average irradiance, as

$$U = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(\frac{\widehat{I}_t - I_t}{I_{clr,\,t}}\right)^2}. \quad (25)$$

This definition of forecast uncertainty is the second required quantity for the determination of the performance metric described in the next section.

2.6.3. Performance metric

Now that solar variability V and forecast uncertainty U have been defined, we can write the performance metric suggested in Ref. [27] as the difference between V and U normalized with respect to V

$$s = \frac{V-U}{V} = 1 - \frac{U}{V}, \quad (26)$$

where it is implied that U and V are calculated using the same data set. The authors in Ref. [27] point out that when $s = 1$ the forecast performs perfectly due to the vanishing uncertainty, however, when $s = 0$ the forecast is dominated by solar variability. Close inspection of Equations (10), (24) and (25) reveals that the persistence forecast, by definition, should have a performance metric value of $s = 0$ (see Table 1). As a result, the ratio U/V can be thought of as a measure of the quality performance of a forecast with the persistence model as the baseline. If a developed forecast model produces a performance metric for which $s < 0$, the model performs worse than a persistence forecast. This means that a typical forecast model should be characterized with values ranging from 0 to 1, with values closer to 1 indicating a higher quality forecast [27].

As a result of U and V being random variables, s is also a random variable. Thus, the authors in Ref. [27] suggest using the average value of the metric $\langle s\rangle$ for a representation of forecasting skill. In order to do this, the data is partitioned into several time windows of equal length N_w. Values of U_j and V_j are calculated for each of the j windows and an average value of the metric $\langle s\rangle$ is obtained. An important feature of the performance metric is that periods of low variability do not positively bias the value of $\langle s\rangle$. That is to say, when a time window has many clear days (low variability) the forecasting error will also be low, therefore preserving the relative amount of U and V, as seen in Fig. 4. The opposite is also true. Events of high variability which are completely predictable, such as an eclipse, will not negatively bias the value of the performance metric because uncertainty is near zero.

Marquez and Coimbra applied the preceding analysis to two forecasting models based on Artificial Neural Networks (NAR and NARX) for hourly data collected from Jan. 1–Oct. 31 of 2010, and found that the metric is a statistical invariant which is preserved over a wide variety of time horizons [27]. This makes the metric an attractive and robust candidate for the comparison of solar forecast model performance. Fig. 4 shows scatter plots of U_j versus V_j computed for each jth time window for which $N_w = 50$, 100, 150 and 200, and Table 1 gives corresponding numerical values of $\langle s\rangle$ obtained using $N_w = 200$. The ESRA clear sky model was employed for normalization purposes. General trends in the scatter plots demonstrate that the performance of the NAR forecast model is similar to the persistence forecast's. On the other hand, the NARX model shows significant forecasting improvement over the persistence as many of the points fall below the 1:1 reference line.

Table 1
Forecasting quality metrics for the persistent, NAR, and NARX models on validation and training data sets from Ref. [27].

Model	R^2	RMSE (W/m^2)	$\langle s\rangle$
Training set			
$\widehat{I}_{ESRA}$	0.969	55.5	0
NAR	0.972	53.2	2.27%
NARX	0.977	48.8	12.02%
Validation set			
$\widehat{I}_{ESRA}$	0.926	59.5	0
NAR	0.924	60.2	−2.53%
NARX	0.949	49.4	16.25%

3. Regressive methods

Prior to stochastic techniques solar radiation models focused on the extrapolation of data from long term averages and steady state values resulting in essentially static models which described only seasonal and diurnal changes. Landsberg [45] incorporated GHI data from over 300 locations into a world map of average solar energy available at the Earth's surface. Whillier [46] derived graphical models from correlations of measured GHI in several parts of the world at latitudes ranging from the equator to 50° north and south. Swartman and Ogunlade [47,48] derived relationships correlating solar radiation intensity with sunshine and relative humidity for tropical conditions in the Toronto area. Lund [49] correlated nine years of daily radiation data taken at Blue Hill, Massachusetts, month by month with observations of snow cover, wind, sunshine, sky cover, pressure and precipitation in order to find a linear combination of parameters to best estimate GHI. These models, whether graphical or mathematical, ignored the short-term time dependent patterns of solar radiation data which result from changing local weather conditions and cause fluctuations on timescales ranging from seconds to days. Forecasting of such fluctuations is essential to the operation of, for example, concentrating photovoltaic systems due to their relatively small apertures and strong dependence on DNI [50]. Any forecasting model which ignores the short term stochastic characteristics of solar radiation data is clearly deficient.

A second shortcoming of the pre-stochastic models was the use of common regression techniques to develop mathematical relationships. The statistical theory at the foundation of these regression techniques assumes that the individual observations of solar irradiance change independently [51]. However the reverse is true and changes due to local weather conditions generate data time-series which are strongly dependent. The correlated nature of solar irradiance data is another essential characteristic which must be addressed by any successful forecasting model. It is evident from the discussion that models which are based on long term averages and employ independent regression techniques could be improved through the use of stochastic models which account for short term fluctuations and the correlated nature of the data.

In the 1920s G. Udny Yule began to establish new approaches for the analysis of stationary time-series. Yule was prompted by the quintessential criticism of strict periods which is at the heart of the hypothesis of the classical methods of Fourier and Schuster [52]. While these methods are pervasive in almost all areas of the physical sciences, the rigid nature of strict periods limits the applications for which these models are adequate. In response to the call for a modified method, in 1922 Yule pioneered the first of two new approaches for stationary time-series analysis which he called Moving Averages (MA), see Fig. 5(c). The theory originates from two papers regarding the regular fluctuations of purely random series which are given, for instance, by throwing dice [53,54]. The second of Yule's approaches, linear Auto-Regression (AR), Fig. 5(b), came in 1927 in a well known paper which discussed the periodicities of disturbed series with special reference to Wolfer's sunspot

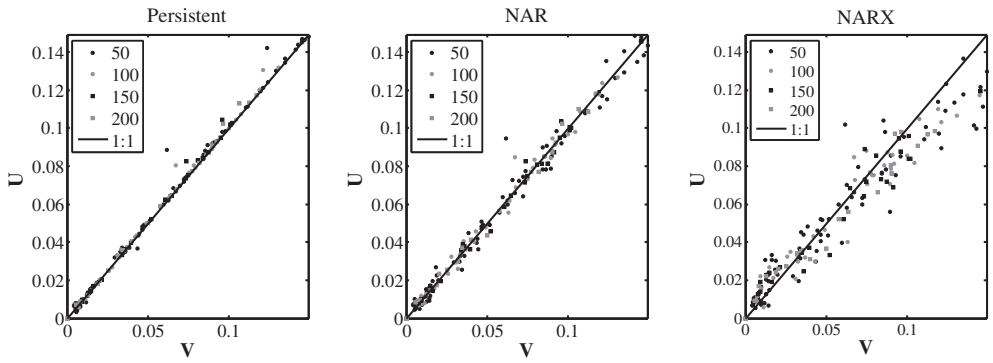


Fig. 4. Scatter plot of U and V using ESRA clear sky model. The NAR forecast shows little improvement with respect to the performance of the persistence forecast. However, the NARX forecast shows a clear improvement over the persistence as many of the points fall below the 1:1 reference line. Adapted from Ref. [27].

numbers [55]. Shortly after, in the 1930s, the work of Yule was furthered by H. Wold who applied the theory to business cycles and econometric statistics [52]. It wasn't until the 1960s that G. E. P. Box and G. M. Jenkins, motivated by the principle of parsimony, would popularize what would become two of the most widely used models in time-series analysis [56]. The first, known as Auto-Regressive Moving Averages (ARMA), advanced from the combination of the schemes established by Yule and Wold for stationary time-series analysis, Fig. 7(a). The second, utilized a evolutive component and gave way to a class of models known as Auto-Regressive Integrated Moving Averages (ARIMA) which are useful for the analysis of non-stationary processes, Fig. 7(b) [52]. The use of ARMA and ARIMA provide a basis for many problems outside the realm of solar irradiance forecasting including economic and business planning, production planning, inventory and production control and optimization of industrial processes [56].

Before beginning a discussion of the various stochastic models it will be useful to review some simple operators and terminology. One operator which will be used frequently is the forward shift or advance operator q which, when applied to a time-series $\{z\}$ at time t, is defined as

$$qz_t = z_{t+1}. \tag{27}$$

The forward shift operator can be applied successively to yield

$$q^k z_t = z_{t+k}. \tag{28}$$

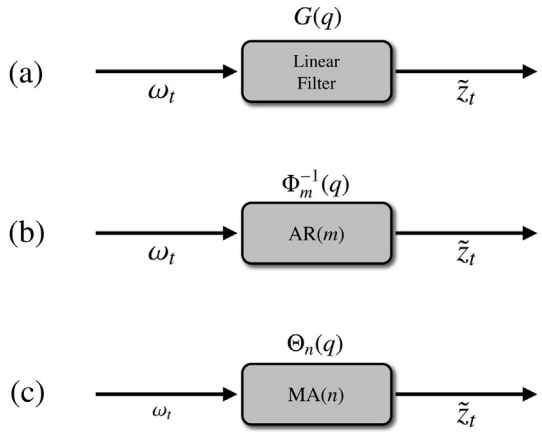


Fig. 5. (a) Linear filter model of a time-series generated from a white noise process with transfer function $G(q)$. (b) AR(m) process modeled as a linear filter with transfer function $\Phi_m^{-1}(q)$. (c) MA(n) process modeled as a linear filter with transfer function $\Theta_n(q)$.

The inverse of the forward shift operator is the backward shift or delay operator q^{-1} which, when applied to a time-series, is given by

$$q^{-1}z_t = z_{t-1}, \tag{29}$$

which in turn yields,

$$q^{-k}z_t = z_{t-k}. \tag{30}$$

The backward shift operator can be used to construct the backward difference operator D as follows,

$$Dz_t = z_t - z_{t-1} = \left(1 - q^{-1}\right)z_t. \tag{31}$$

Its inverse, the summation operator S, can be written as

$$Sz_t = D^{-1}z_t = z_t + z_{t-1} = \left(1 + q^{-1}\right)z_t. \tag{32}$$

The stochastic models that follow are based on the result from Yule [55] that if a strong dependency exists between successive terms in a time-series which is, for example, a characteristic of solar irradiance data, it can be effectively generated by a series of independent shocks which are assumed to have non-permanent effects and are drawn at random from a stationary distribution having zero mean and variance σ_ω^2. Any series of such random shocks $\{\omega\}$ is

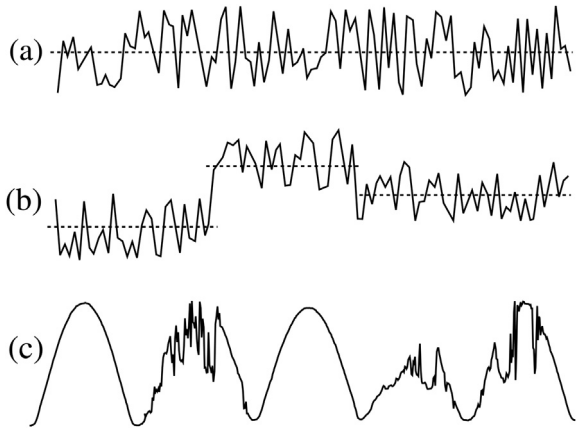


Fig. 6. (a) An example stationary time-series fluctuating about a static level μ which can be represented by the model $\Phi_m(q)\tilde{z}_t = \Theta_n(q)\omega_t$. (b) An example time-series showing non-stationary behavior in local level can be represented by the model $\Phi_m(q)\nabla z_t = \Theta_n(q)\omega_t$. (c) GHI data at 5 min resolution with night values removed showing non-stationary behavior in local level and slope which can be represented by the model $\Phi_m(q)\tilde{\nabla}^2 z_t = \Theta_n(q)\omega_t$.

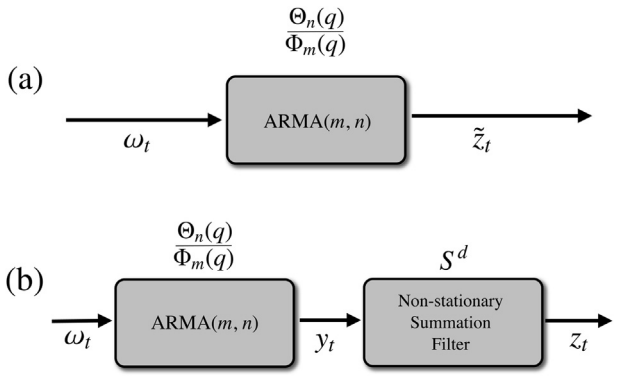


Fig. 7. (a) ARMA(n,m) modeled as a linear filter with transfer function $\Phi_m^{-1}(q)\Theta_n(q)$. (b) ARIMA($n$,$d$,$m$) modeled as a stationary ARMA(m,n) linear filter in series with a non-stationary summation filter S^d.

referred to as a white noise process. A linear filter can be used to transform the white noise process into the time-series $\{z\}$, which is depicted in Fig. 5(a). The linear filter is defined by the following operation

$$z_t = \mu + \omega_t + g_1\omega_{t-1} + g_2\omega_{t-2} + \ldots = \mu + G(q)\omega_t \tag{33}$$

where μ is a parameter which determines the level of the process. The level of a stationary process can be thought of as the average value about which the series fluctuates. However, the level of a non-stationary process is time dependent and varies from section to section, as exemplified in Fig. 6. The operator that transforms ω_t into z_t is called the transfer function $G(q)$ of the filter and is given by

$$G(q) = \sum_{k=0}^{\infty} g_k q^{-k} = 1 + g_1 q^{-1} + g_2 q^{-2} + \ldots \tag{34}$$

where $g_0 = 1$. It should be noted that we choose the less-obvious q as the argument of G rather than q^{-1} for clarity. The sequence of weights g_1,g_2,… may be finite or infinite which gives us an interpretation for the level μ and determines the nature of the process: stationary or non-stationary.

3.1. Linear stationary models

Observational series that describe a changing physical phenomenon with time can be classified into two main categories; stationary and non-stationary. If the sequence of weights in Equation (34) is finite, or infinite and convergent, the linear filter is said to be stable and the time-series $\{z\}$ to be stationary. Stationary time-series are static with respect to their general shape. The fluctuations may appear, to the user, as clearly ordered or totally random, nonetheless, the character of the series is, on average, quite similar in different segments. In this case, the parameter μ may be interpreted as the average value about which the series fluctuates as seen in Fig. 6(a). Stationary time-series find applications in many areas of the physical sciences. For instance, observational time-series and series involving deviations from a trend are often stationary [52]. In fact, the stochastic portion a solar radiation data set is often framed as a stationary process [57].

3.1.1. Auto-Regressive (AR) models

The AR models get their name from the fact that the current value of the process can be expressed as a finite, linear combination of the previous values of the process and a single shock ω_t. Thus, the process is said to be regressed on the previous values. If we define the stochastic portion of the time-series $\tilde{z}_t, \tilde{z}_{t-1}, \tilde{z}_{t-2}, \ldots$ as deviations from the mean value μ as

$$\tilde{z}_t = z_t - \mu \tag{35}$$

then the Auto-Regressive process of order m can be written as

$$\tilde{z}_t + \phi_1\tilde{z}_{t-1} + \phi_2\tilde{z}_{t-2} + \ldots + \phi_m\tilde{z}_{t-m} = \omega_t \tag{36}$$

We can simplify the previous expression by defining the Auto-Regressive transfer function of order m, $\Phi_m(q)$, as

$$\Phi_m(q) = \sum_{k=0}^{m} \phi_k q^{-k} = 1 + \phi_1 q^{-1} + \phi_2 q^{-2} + \ldots + \phi_m q^{-m}, \tag{37}$$

where $\phi_0 = 1$. It follows that the AR(m) model may be written conveniently as

$$\Phi_m(q)\tilde{z}_t = \omega_t. \tag{38}$$

Here it is clear that the process is regressed on the previous values of $\{\tilde{z}\}$. In order to implement this model one must determine the $m + 2$ unknown parameters ϕ_1,ϕ_2,…,ϕ_m,μ, and the variance of the white noise process σ_ω^2. Typically these are calculated from the data using the techniques covered in Section 3.1.3. It is illustrative to note that Equation (38) implies

$$\tilde{z}_t = \Phi_m^{-1}(q)\omega_t. \tag{39}$$

Therefore, it is helpful to think of the AR(m) process as the output of a linear filter with transfer function $\Phi_m^{-1}(q)$ and white noise ω_t as the input (see Fig. 5(b)).

In order for the AR(m) process to be stationary a set of conditions must be satisfied. In Ref. [56] the authors point out that the general AR(m) process has the inverse transfer function

$$\Phi_m(q) = \left(1 - \Gamma_1 q^{-1}\right)\left(1 - \Gamma_2 q^{-1}\right)\ldots\left(1 - \Gamma_m q^{-1}\right) \tag{40}$$

which allows expansion of the process in partial fractions,

$$\tilde{z}_t = \Phi_m^{-1}(q)\omega_t = \sum_{i=1}^{m} \frac{\kappa_i}{\left(1 - \Gamma_i q^{-1}\right)}\omega_t \tag{41}$$

where it is clear that if $\Phi_m^{-1}(q)$ is to be a convergent series for $|q^{-1}| \leq 1$, then we must have $|\Gamma_i| < 1$, where $i = 1,2,\ldots,m$. This is equivalent to saying that the roots of the equation $\Phi_m(q) = 0$ must lie outside the unit circle. For a discussion of stationary conditions of AR(m) processes see Refs. [52,56,58].

3.1.2. Moving Average (MA) models

While the AR techniques model the stochastic portion of the time-series $\{\tilde{z}\}$ as a weighted sum of previous values $\tilde{z}_{t-1}, \tilde{z}_{t-2}, \ldots, \tilde{z}_{t-m}$, MA methods model $\{\tilde{z}\}$ as a finite sum of n previous shocks ω_t,ω_{t-1},ω_{t-2},…,ω_{t-n}. The Moving Average process of order n, MA(n), is defined as

$$\tilde{z}_t = \omega_t + \theta_1\omega_{t-1} + \theta_2\omega_{t-2} + \ldots + \theta_n\omega_{t-n}. \tag{42}$$

Let us pause here and note that the term moving average can be a bit misleading due to the fact that the weights in Equation (42) do not, in general, need to be positive nor does their sum necessarily equal unity [56]. Nonetheless, the name is used for historic convention. The MA(n) operator is defined as

$$\Theta_n(q) = \sum_{k=0}^{n} \theta_k q^{-k} = 1 + \theta_1 q^{-1} + \theta_2 q^{-2} + \ldots + \theta_n q^{-n}, \quad (43)$$

where $\theta_0 = 1$. As a result we can write the MA model in an economic fashion

$$\tilde{z}_t = \Theta_n(q)\omega_t. \quad (44)$$

The MA process can be thought of as the output $\tilde{z}_t$ of a linear filter whose transfer function is $\Theta_n(q)$, with white noise ω_t as the input (see Fig. 5(c)). Like its counterpart, the MA model contains $n + 2$ undetermined parameters $\theta_1,\ldots,\theta_n,\mu$, and the variance of the white noise process σ_ω^2 which must be determined from the data using the techniques described in the next section. Unlike AR models, MA models are unconditionally stable [52].

3.1.3. Mixed Auto-Regressive Moving Average (ARMA) models

Linear processes represented by an infinite or an extraneous number of parameters are clearly not practical. However, it is possible to introduce parsimony and still obtain useful models. A well known result in time-series analysis is the relationship between the Θ weights and Φ weights [56]. Operating on both sides of Equation (38) by $\Theta(q)$ and making use of Equation (44), yields

$$\Theta(q)\Phi(q)\tilde{z}_t = \Theta(q)\omega_t = \tilde{z}_t \quad (45)$$

which implies

$$\Theta(q)\Phi(q) = 1 \quad (46)$$

that is

$$\Phi^{-1}(q) = \Theta(q). \quad (47)$$

Equation (47) indicates that the Φ weights may be arrived at from knowledge of the Θ weights, and vice-versa. Thus the *finite* MA(n) process $\tilde{z}_t = \Theta_n(q)\omega_t$ can be written as an *infinite* AR process

$$\tilde{z}_t = \omega_t - \theta_1\tilde{z}_{t-1} - \theta_1^2\tilde{z}_{t-2} - \ldots \quad (48)$$

However, if the process were really MA(n), we would arrive at a non-parsimonious representation in terms of the AR expression. By the same reasoning, a *finite* AR(m) model could not be parsimoniously represented using an *infinite* MA expression. Therefore, in practice, in order to realize a parametrization which is parsimonious, both AR and MA terms are often used in the model development. Hence,

$$\tilde{z}_t + \phi_1\tilde{z}_{t-1} + \ldots + \phi_m\tilde{z}_{t-m} = \omega_t + \theta_1\omega_{t-1} + \ldots + \theta_n\omega_{t-n} \quad (49)$$

or

$$\Phi_m(q)\tilde{z}_t = \Theta_n(q)\omega_t. \quad (50)$$

Equation (50) is referred to as the mixed Auto-Regressive Moving Average (ARMA) process of order (m, n). It is illustrative to note that the ARMA(m, n) model can be written as

$$\tilde{z}_t = \frac{\Theta_n(q)}{\Phi_m(q)}\omega_t = \frac{1 + \theta_1 q^{-1} + \ldots + \theta_n q^{-n}}{1 + \phi_1 q^{-1} + \ldots + \phi_m q^{-m}}\omega_t \quad (51)$$

where it is clear that $\tilde{z}_t$ is the output from a linear filter, whose transfer function is the ratio of two polynomials $\Theta_n(q)$ and $\Phi_m(q)$, with white noise ω_t as the input (see Fig. 7(a)).

In practice, it is frequently true that adequate representation of actually occurring stationary time-series can be obtained from models in which n and m are not greater than two and often less than two [56,57]. The order of the model, that is the values of m and n, is determined using the sample auto-correlation function and partial auto-correlation function of the time-series [59]. The model parameters are estimated by least squares methods and the resulting model is said to adequately describe the statistical information contained in the series in a parsimonious manner.

These techniques have been used since the 1970s to analyze time-series resulting from solar irradiance data. Boileau [60] and Guerrier [61] evaluated several stochastic models in the field of solar meteorology for 24-h time steps. Later, Guerrier, Boileau and Bernard [62] would employ seasonal and non-seasonal stochastic techniques for solar irradiance prediction which were originally developed by Box and Jenkins. Guerier [63] showed the usefulness of small correlations of consecutive day solar irradiance data through an optimization of stochastic techniques. Brinkworth [64] used the simple ARMA processes to model irradiance in cloudy areas of the U.K. in order to predict the output of solar thermal systems. Benard et al. [65] estimated empirical values for stochastic models from irradiance measurements performed at three sites: a temperate site (Trappes, France, 49N), a Mediterranean site (Carpentras, France, 44) and an equatorial altitude site (Huallao, Peru, 13S, 3 km). Subsequently, these values were used to develop ARMA models for the prediction of daily solar irradiance [66]. Mustacchi et al. [67] compared ARMA models with factor analysis, Markov transition-matrices, Gaussian mapping and transmittance transition tensor techniques. Boch et al. [68] attempted to predict solar irradiance at hourly time steps using stochastic techniques for locations in France. ARMA processes have also been used to model hourly irradiance for the calculation of the optimal control of buildings in France and Japan [69–71]. Aguiar and Collares-Pereira modeled daily sequences of hourly irradiance and clearness index K_t using ARMA techniques, where K_t was obtained by multiplying a clear sky value by a non-stationary fluctuation, with probability depending on the hour of the day [72]. Later, Aguiar and Collares-Pereira used an ARMA(1, 0) model to generate hourly series of solar irradiance values which were transformed backwards to generate synthetic sequences of clearness index K_t values [73]. Al-Awahdi et al. proposed an ARMA model which used a bilinear time-series to generate daily irradiance models for Kuwait [74]. Mora-Lopez et al. [75] employed a multiplicative ARMA model for global radiation time-series with diurnal and seasonal components. The multiplicative nature of these models enabled the acquisition of two types of relationships observed in recorded hourly series of global irradiation: the relationship between the value at 1 h and the value at the previous hour, and the relationship between the value at 1 h in one day and the value at the same hour in the previous day. Moreno-Muñoz et al. [76] more recently used multiplicative ARMA models to generate instantaneous series of global irradiance in southern Spain during a four year period. Lately, however, ARMA methods typically find applications as components of a robust hybrid system architecture [77–81] as explained in Section 8.

3.1.4. Mixed Auto-Regressive Moving Average models with exogenous variables (ARMAX)

All of the linear stationary stochastic techniques discussed so far have been univariate, meaning the technique uses previous values of only the time-series it is attempting to model. However, the accuracy of ARMA(m, n) models may be improved by including information external to the time-series under analysis. For example, in the case of solar irradiance forecasting, the error of a forecasting model may be reduced by including information about the evolution of the local temperature, relative humidity, cloud cover, wind speed, wind direction, etc. Variables such as these, which are independent of the model but affect its value, are referred to as exogenous variables. We can include into the

ARMA(m, n) models p exogenous input terms which allow us to write the ARMAX(m, n, p) process as

$$\tilde{z}_t + \phi_1\tilde{z}_{t-1} + \ldots + \phi_m\tilde{z}_{t-m} = \omega_t + \theta_1\omega_{t-1} + \ldots + \theta_n\omega_{t-n} + \lambda_1 e_{t-1} + \ldots + \lambda_p e_{t-p}. \quad (52)$$

The above model contains AR(m) and MA(n) models as well as the last p values of an exogenous time-series e_t. Defining the exogenous input operator of order p as

$$\Lambda_p(q) = \sum_{k=1}^{p} \lambda_k q^{-k} = \lambda_1 q^{-1} + \lambda_2 q^{-2} + \ldots + \lambda_p q^{-p} \quad (53)$$

allows us to write the ARMAX(m, n, p) model conveniently as

$$\Phi_m(q)\tilde{z}_t = \Theta_n(q)\omega_t + \Lambda_p(q)e_t. \quad (54)$$

The careful reader might already be aware of the fact that all of the linear stationary models discussed so far have a similar structure. In fact, many models in linear system analysis can be considered a special case of the general discrete time model structure

$$\Phi(q)\tilde{z}_t = \frac{\Theta(q)}{\Psi(B)}\omega_t + \frac{\Lambda(q)}{\Xi(q)}e_t \quad (55)$$

where $\Phi(q)$, $\Theta(q)$, $\Lambda(q)$, $\Psi(q)$ and $\Xi(q)$ are polynomials of the shift operator q [58,82]. Table 2 summarizes some of the commonly used discrete time models which can be considered special cases of (55), however, the current study is limited to the models we have discussed so far.

3.2. Non-linear stationary models

So far we have only considered general classes of linear stationary models. However, non-linear methods would enable powerful structures with the ability to accurately describe complex non-linear behavior such as chaos, hysteresis and saturation effects or a combination of several non-linear problems [82]. A step towards non-linear modeling is made by introducing the Non-linear AR-eXogenous (NARX) model as

$$\tilde{z}_t = f(\tilde{z}_{t-1}, \tilde{z}_{t-2}, \ldots, \tilde{z}_{t-m}, e_{t-1}, e_{t-2}, \ldots, e_{t-n}) + \omega_t. \quad (56)$$

In much the same way one can also convert the ARMAX model into a Non-linear ARMAX model (NARMAX) as follows

$$\tilde{z}_t = f(\tilde{z}_{t-1}, \tilde{z}_{t-2}, \ldots, \tilde{z}_{t-m}, e_{t-1}, e_{t-2}, \ldots, e_{t-n}, \omega_{t-1}, \omega_{t-2}, \ldots, \omega_{t-p}) - \omega_t. \quad (57)$$

These non-linear input–output models find many applications in the field of engineering, especially in the parametrization of Artificial Networks which are discussed in Section 4.1.

Table 2
Special cases of the generalized linear model (55). Polynomials of the shift operator which are not listed for a model are assumed to equal unity.

Polynomials	Model
Φ,($\Lambda/\Xi = 0$)	AR (Auto-Regressive)
Θ,($\Lambda/\Xi = 0$)	MA (Moving Average)
Θ, Φ,($\Lambda/\Xi = 0$)	ARMA (Auto-Regressive Moving Average)
Φ, Λ	ARX (AR-eXogenous)
Θ, Φ, Λ	ARMAX (ARMA-eXogenous)
Φ, Θ, Ψ	ARARX (Auto-Regressive-ARX)
Φ, Θ, Λ, Ξ	ARARMAX (Auto-Regressive ARMAX)
Θ, Ψ, Λ, Ξ	BJ (Box–Jenkins)

3.3. Linear non-stationary models

If the sequence of weights in Equation (34) is infinite but not convergent, the linear filter's transfer function $G(q)$ is said to be unstable and the process z_t to be non-stationary. In this case, μ has no physical meaning except as a reference to the level of the process (Fig. 6(b)). Non-stationary processes are different in one or more respects throughout the time-series due to the time dependent nature of the level. As a result, in the analysis of non-stationary time-series, time must play a fundamental role, for example, as the independent variable in a progression function, or as a normalization factor in the analysis of the evolution of a phenomenon from an initial state [52]. Several observed time-series behave as if they have no specified mean about which they fluctuate, for example, daily stock prices or hourly readings from a chemical process [56].

3.3.1. Auto-Regressive Integrated Moving Average (ARIMA) models

While non-stationary processes do not fluctuate about a static mean, they still display some level of homogeneity to the extent that, besides a difference in local level or trend, different sections of the time-series behave in a quite similar way. These non-stationary processes may be modeled by particularizing an appropriate difference, for example, the value of the level or slope, as stationary, (Fig. 6(b)(c)). What follows is a description of an important class of models for which it is assumed that the dth difference of the time-series is a stationary ARMA(m, n) process.

We have seen that the stationarity condition of an ARMA(m, n) process is that all roots of $\Phi_m(q) = 0$ lie outside the unit circle, and when the roots lie inside the unit circle, the model exhibits non-stationary behavior. However, we have not discussed the situation for which the roots of $\Phi_m(q) = 0$ lie on the unit circle. Let us examine the following ARMA(m, n) model

$$\Phi_m(q)\tilde{z}_t = \Theta_n(q)\omega_t \quad (58)$$

and specify that d of the roots of $\Phi_m(q) = 0$ lie on the unit circle and the residuum lie outside. We can then express the model as

$$\Phi_m(q)\tilde{z}_t = \Theta_n(q)\left(1 - q^{-1}\right)^d \tilde{z}_t = \Theta_n(q)\omega_t \quad (59)$$

where $\Phi_m(q)$ is a stationary and invertible AR(m) operator. Seeing that $S^d\tilde{z}_t = S^d z_t$ when $d \geq 1$, we can write

$$\Phi_m(q)S^d z_t = \Theta_n(q)\omega_t. \quad (60)$$

Defining $y_t = S^d z_t$ allows one to express the model in a more illustrative way

$$\Phi_m(q)y_t = \Theta_n(q)\omega_t \quad (61)$$

where it is clear that the model is in agreement with the assumption that the dth difference of the time-series can be regarded as a stationary ARMA(p, q) process. We refer to (60) as the Auto-Regressive Integrated Moving Average (ARIMA) process. Because the AR transfer function $\Phi_m(q)$ is of order m, d differencing steps are taken, and the MA transfer function $\Theta_n(q)$ is of order n in (60), we refer to the process as ARIMA(m, d, n). In practice, d is typically 0, 1 or at most 2 [56]. As mentioned above, the ARIMA(m, d, n) model is equivalent to representing the process z_t as the output of a linear filter with transfer function $\Phi_n^{-1}S^d\Theta_n$ and takes white noise ω_t as an input (see Fig. 7).

In the past ARIMA models have found applications in many areas of research, however, only recently have they been employed to model problems involving solar radiation. Craggs et al. [83] used

ARIMA models to compare 10 min, 20 min, 30 min and 1 h averages of solar irradiance levels for a site in the UK. Santos et al. [84] used an ARIMA(1, 1, 1) model to calculate synthetic daily solar radiation values using the monthly average radiation as the input for locations in Spain. Kärner [85] carried out an ARIMA analysis for satellite-based global tropospheric and stratospheric temperature anomaly and solar irradiance data sets. Kärner's results emphasized a dominating role of the solar irradiance variability in variations of the tropospheric temperature. Yürekli et al. [86] imposed ARIMA based techniques to generate solar irradiance, temperature and relative humidity forecasts for use in the control strategy of agricultural facilities. Stanhill and Cohen [87] used a first order ARIMA model to describe the time course of annual sunshine duration and global irradiance to study trends and changes in solar forcing at the Earth's surface and reported a 0.5% increase in solar forcing per decade during the 20th century. Kärner [88] later employed ARIMA models to compare the variability of the total solar irradiance at the top of the atmosphere to surface air temperature series, which he reports are strongly correlated. Reikard [89] compared ARIMA models with a number of other methods, including transfer functions, neural networks and hybrid models, for six data sets at resolutions of 5, 15, 30 and 60 min using the global horizontal component of solar irradiance. Reikard found that, in nearly all the tests, the best results are obtained using the ARIMA in logs, with time-varying coefficients. In Ref. [90] the authors used daily solar radiation measurements carried out in Bogotá, Columbia, from 2003 to 2009 as inputs for an ARIMA(1, 0, 0) model for forecasting the mean daily global solar radiation. Like its stationary counterpart, the ARIMA model has, as of late, also found many applications in the construction of hybrid systems [79–81,89].

3.3.2. Auto-Regressive Integrated Moving Average models with exogenous variables (ARIMAX)

In a similar way to the ARMAX(m, n, p) model, the previous p values of an exogenous time-series e_t may also be included into the ARIMA(m, d, n) model to yield the ARIMAX process of order (m, d, n, p)

$$\tilde{z}_t = \phi_1 S^d z_{t-1} + \ldots + \phi_m S^d z_{t-m} + \omega_t + \theta_1 \omega_{t-1} + \ldots + \theta_n \omega_{t-n} + \lambda_1 e_{t-1} + \ldots + \lambda_p e_{t-p} \tag{62}$$

As we did before, defining $y_t = S^d z_t$ in terms of the backwards shift operator allows us to express the model in a more compact form

$$\Phi_m(q) y_t = \Theta_n(q)\omega_t + \Lambda_p(q) e_t \tag{63}$$

which again looks very similar to Equation (55).

4. Artificial Intelligence (AI) techniques

The development of Artificial Intelligence (AI) techniques began in the early 1950s with a number of experiments conducted by Herbert Simon, Allen Newell and Cliff Shaw [91]. Simon was consulting at the RAND Corporation when he saw a printer using customary text and punctuation symbols to produce images [92]. Motivated by this, Simon recognized that machines could be used to simulate the decision making process and possibly, if given enough resources, the human thought process. Simon enlisted the help of Allen Newell, the RAND corporation logistics scientist who generated the code that printed the images, and Cliff Shaw to develop a program with the ability to provide proofs to mathematical theorems. The resulting program, called Logic Theorist, used heuristics to draw from a knowledge base of previously proved axioms to discover new proofs and would ultimately solve thirty-eight of the fifty-two problems presented by Bertrand Russell and Alfred Whitehead in their Principia Mathematica [93]. Meanwhile, Claude Shannon demonstrated how a machine could be used to play a reasonable game of chess [94]. Shannon's chess program relied on the optimization of a weighted function dependent on the position of the chess pieces using a minmax procedure. Admitting that the work of Simon et al. and Shannon was the earliest in the field artificial intelligence, the topic's origin is frequently associated with another event.

The first coordinated undertaking in the field of machine intelligence was the Artificial Intelligence Conference organized by John McCarthy, Marvin Minsky, Nathaniel Rochester and Claude Shannon at Darthmouth College in New Hampshire in 1956 [91]. It was at this conference that John McCarthy, who would develop the popular LISP family of computer programming languages two years later, suggested the term "Artificial Intelligence", which is still frequently used today to describe a wide variety of machine learning methods. The AI Conference of 1956 is commonly viewed as the event responsible for the motivation to examine computers to solve problems rather than focusing on hardware that simulated intelligence [95].

Several AI methods were developed in the years following the conference in New Hampshire. In 1959 Simon et al. developed a General Problem Solver (GPS) program capable of proving mathematical theorems like Logic Theorist, playing chess reasonably like Shannon's algorithm and, in addition, solving elaborate geometric puzzles. GPS accomplished this by generating heuristics through means-ends analysis [96]. However, the predominant criticism of the GPS code was that it could not learn from problems solved in advance [91]. This paved the way for McCarthy's LISP programming language to become the standard in AI development [97].

In 1960 additional codes were developed that mimicked key aspects of human thought, such as language and vision processing. Kenneth Colby, an American psychiatrist working at Stanford University, developed a family of computer programs known as "chatterbots" which were able to simulate intelligent conversation with people through a process known as Natural Language Processing (NLP). Soon after, Joseph Weizenbaum wrote a program known as ELIZA which would become one of the most popular chatterbots [91]. ELIZA used rules correlated with words like 'I', 'you', 'like' etc., which were prompted if one of these words was encountered [98]. Bobrow [99] and Winograd [100] also made contributions to the emerging field of NLP, most notably through the development of programs with the ability to make assumptions and to learn from previously solved problems [91]. Parallel to this work, Marvin Minsky and his group at MIT, which supported the idea that AI is fundamentally a symbol processing exercise [101], developed a computer program that could perform visual analogies by identifying visual relationships from two figures and locate another set of figures with similar relationships from a larger set [102].

All of the methods mentioned above use a "symbolic" approach to intelligent systems. This symbol processing architecture focuses on knowledge representation through the translation of the user's expert knowledge into a set of formal symbols. Once a formal encoding is established the system functions in the traditional Von Neumann computer pattern: INPUT → PROCESS → OUTPUT. The advantage lies in the enormous representational power of these systems which stems from their generality. However, the central criticism is that only the user's own knowledge comes into effect [103]. In addition, these systems are limited by the fact that all the responses must be explicitly programmed beforehand. These symbolic approaches are referred to as brittle due to the "crashing" that occurs when an unusual set of circumstances not anticipated in the response coding occurs. Even so, symbol processing remained the popular motif in AI systems development well into the 1980s.

However, there was another school of thought that defended that intelligent systems should be modeled after living organisms with brains and the ability to adapt.

A seminar paper in the Bulletin of Mathematical Biology in 1943 titled "A Logical Calculus of Ideas Imminent in Nervous Activity" [104] was the result of a joint effort between Warren McCulloch and Walter Pitts. In this work, which would later be included in Hebb's famous Organization of Behavior [105], the authors proposed the first artificial neuron called the Threshold Logic Unit. In 1958 Frank Rosenblatt, a computer scientist working at Cornell Aeronautical Laboratories, published a book titled "Principles of Neurodynamics" [106]. In this book, Rosenblatt discussed machines with the ability to learn how to classify information by adapting weights, which he called perceptrons. The Pandemonium architecture, which was proposed by Oliver Selfridge in 1959 [107] and saw success in modeling human pattern recognition, suggested that the brain works like a collection of daemons, or background processes, which work in parallel and are each responsible for a single task. Soon after, at Stanford University, Professor Bernard Widrow and his graduate student Ted Hoff developed the ADALINE (ADAptive LInear NEuron) and the Least Means Square (LMS) rule for training [108]. This work resulted in the emergence of a new discipline of AI techniques known as Artificial Neural Networks (ANNs). ANNs are essentially attempts to create an electrical analog to the biological neural networks present in the human brain.

Another set of AI techniques that were motivated by a living organism's ability to adapt and evolve was developed in the 1970s by John Holland, a professor of Psychology and Electrical Engineering at the University of Michigan, Ann Arbor. In his book titled "Adaptation in Natural and Artificial Systems" [109], Holland, motivated by Darwinian views of evolution, conceptualized an algorithm which would identify an optimal solution through the combination and reproduction of the strongest individuals in a solution space. Holland's work was the first in a branch of AI techniques known as Genetic Algorithms (GAs). Neural and genetic computing remained relatively unpopular during the 1970s, however, with an increase in available computing power and technological advancements such as multi-layer perceptrons, the 1980s saw a renewed interest in ANNs and GAs [95,110].

In contrast to the symbolic approach, these "evolutionary" approaches emphasize learning and adaptation over the representational power of symbolic processing. The "intelligence" of these systems is attributed to the ability to distinguish between two general classes of patterns without being directly programmed [103]. For example, a set of solar irradiance measurements in one location forms a class of patterns which is different from a set at another location, as long as these sites are far enough away from each other as to not be correlated through local weather patterns [40]. Evolutionary techniques are much more systematic than symbolic approaches and it should be noted that the internal structure of the system itself is critical inasmuch as it determines the type of function the ANN is able to approximate [103].

So far, we have discussed several AI techniques. It is beneficial to pause here and take note of the several branches into which these techniques can be classified.

- *Problem Solving and Planning*: systematic refinement of goal hierarchy and plan revision mechanisms, e.g. Shannon's chess code.
- *Expert Systems*: uses knowledge-base of previously established rules of thumb for well-defined problems for complex decision-making, e.g. Simon's Logic Theorist and GPS
- *Natural Language Processing*: text processing and generation, speech synthesis and analysis, and machine translation; e.g. Weizenbaum's ELIZA
- *Computer Vision*: Image understanding, facial recognition, motion derivation and intelligent visualization; e.g. Minsky's visual analogy program.
- *Genetic Algorithms*: Evolutionary algorithms with an inherent learning capacity.
- *Artificial Neural Networks*: Combination of pattern recognition, deductive reasoning and numerical computations to simulate learning in the human brain.
- *Hybrid Systems*: Any number of combinations of the above branches.

The present study, however, focuses on ANNs, GAs and their applications in various Hybrid Systems (HS). For a comprehensive review of AI techniques see Refs. [91,95]. Before moving on, it should be noted that while there have been several attempts to define AI [111–114], no single definition has been universally accepted. The authors in Ref. [95] suggest that this is likely due to the various definitions of "intelligence", which is an abstract and immeasurable quantity.

4.1. Artificial Neural Networks (ANNs)

ANNs were motivated from the observations of a special type of cell known as a neuron. Neurons are responsible for a number of signal processing tasks in our bodies such as responding to touch, sound, light or other external stimuli and receiving signals from the brain and spinal chord which control muscles, glands and arteries. Neurons are located in the brain, spinal chord and the autonomic ganglion found just outside of the spinal chord. Neurons are furnished with a number of antenna-like structures that stretch out from the cell body, or soma, which allow the cell to send and receive signals from other cells and the environment. The structures which allow the neuron to accept input signals are called dendrites, while the structures which carry signals away from the neuron are called axons. A neuron may posses numerous dendrites but it never has more than one axon. Nonetheless, the dendrites and axon may branch hundreds of times before they terminate forming complex tree-like structures as seen in Fig. 8(a). By actively regulating calcium, chloride, potassium and sodium ion concentrations inside the cell, the neuron is able to maintain electrical potential gradients across its membrane. When a neuron's electric potential exceeds a specific threshold, an all-or-none electro-chemical impulse called an

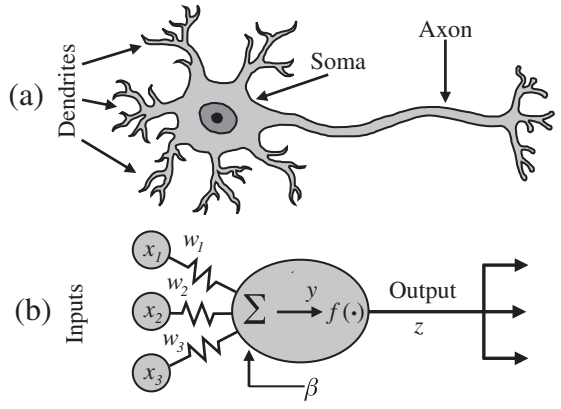


Fig. 8. (a) Simplified diagram of a biological neuron. Antenna-like structures which extend from the cell body, or soma, allow the neuron to communicate with other cells. The structures which allow the neuron to accept input signals are called dendrites. The structure which carries signals away from the neuron is called an axon. A neuron may posses numerous dendrites but it never has more than one axon. (b) Artificial neuron with inputs x_1, x_2 and x_3 weighted by w_1, w_2 and w_3. The neuron has an embedded net function $\beta + \sum_{i=1}^{3} w_i x_i$ and transfer function $f(\cdot)$ which are used to calculate output z.

action potential is generated. The action potential speeds down the neuron's axon and triggers synaptic communications at the tips of the axonic tree. This process is known as the "firing" of a neuron. The rate at which the neuron fires, which can reach up to 300 Hz, is regulated by adrenaline which acts as a bias for the neuron, making it much more likely to fire in the presence of a stimulus [115].

Once the architecture of a biological neuron is understood, one can begin to construct an artificial neuron. Similar to their biological counterparts, artificial neurons can be connected to each other to form a network. A simple model of an artificial neuron which takes inputs x_1,x_2 and x_3 from three other neurons and processes a corresponding output z is detailed in Fig. 8(b). The quantity β is known as the bias and is used to model adrenaline's ability to lower the threshold. Because not every input is equally relevant to whether the neuron fires or not, we assign weights w_1,w_2 and w_3 to each of the inputs. Here we use a linear net function to sum the inputs which is given by

$$y = \beta + \sum_{i=1}^{3} w_i x_i. \tag{64}$$

The firing rate of the action potential is modeled through the use of a transfer function $f(\cdot)$. Several different transfer functions have been suggested over the years (see Table 3). The most commonly used transfer function is the sigmoid or logistic function due to its attractive mathematical properties such as monotonicity, continuity and differentiability [95,115]. Employing the sigmoid transfer function with the linear net function, the output z of the artificial neuron in Fig. 8(b) can be written in a very simple manner

$$z = \frac{1}{1 + e^{-\left(\beta + \sum_i w_i x_i\right)}} \tag{65}$$

The attractiveness of ANNs stem directly from their likeness to their biological counterpart. Specifically, ANNs allow for very low level programming (net and transfer functions) to solve a wide variety of complex, non-linear, non-analytic, non-stationary, stochastic or general mathematically ill-defined problems in a self-organizing manner that requires little or no interference with the program itself. Conventional computing algorithms apply elaborate sets of equations to solve specific and well defined problems through the programmer's imposed organization. Clearly, this is not the most parsimonious expression of a solution. When a pesky fly avoids a dedicated flyswatter, it has no time to carefully solve complex differential equations concerning Newton's laws of motion. Nonetheless, its non-specific decision making process allows it to consistently stay one step ahead of the flyswatter, in addition to carrying out a wide number of other tasks. In order to be as efficient and versatile as possible biological systems must converge to the simplest algorithmic architecture [116].

Table 4 summarizes the early progress in ANN structure development which is discussed in the next section. Careful examination of Table 4 reveals a twenty-two year dormant period in ANN development between 1960 and 1982. This hibernation resulted from the fact that networks with hidden layers were not considered. These single layer ANNs were unable to solve the popular two state eXclusive-Or (XOR) problem and researchers lost interest as a result [82]. Despite that, the 1980s saw an explosion in renewed interest in ANNs with the development of multi-layer perceptrons and back propagation algorithms. Thus, as one can see, the inner structure of ANNs is particularly relevant. The inner-structures of ANNs are typically described with what is commonly known as a directed graph. Directed graphs consist of neurons and directed arcs which describe the synaptic links. The following subsections cover the most widely used ANN structures in more detail.

4.2. Early networks

4.2.1. Threshold Logic Unit (TLU)

The earliest formulation of an artificial neuron was the Threshold Logic Unit (TLU) developed by McCulloch and Pitts in 1943. The TLU employed, as a transfer function, the threshold or Heaviside step function (see Table 3). In Ref. [104] the authors stated five assumptions which governed their neuron:

1. The activity of a neuron is an "all-or-none" process.
2. A certain fixed number of synapses must be excited within the period of latent addition in order to excite a neuron at any time, and this number is independent of previous activity and position on the neuron.
3. The only significant delay within the neural system is synaptic delay.
4. The activity of *any* inhibitory synapse absolutely prevents the excitation of the neuron at that time.
5. The structure of the network does not change with time.

Although these assumptions are some of the earliest in ANN development, they do not necessarily apply to current ANNs, especially the fifth assumption which is commonly replaced with the Hebbian Learning Law [105], allowing for changes in the structure of the network.

4.2.2. Simple perceptron

Arguably one of the first ANNs, simple perceptrons were developed by Frank Rosenblatt to act as electronic analogs to the human retina. Rosenblatt connected the input layer of the perceptron to a rectangular array of light sensors. A directed graph of a simple perceptron network is shown in Fig. 9(a). The weight connecting input node i to output node j is denoted by the matrix component w_{ij}. The goal of Rosenblatt's system was to trigger the relevant response unit given a class of input patterns. The input layer used a unipolar linear transfer function and the output was passed to the response layer by way of trainable weights.

The simple perceptron is said to use a form a supervised learning due to the fact that the weights were adjusted when an undesired response was triggered. The perceptron learning rule begins with a definition of the change made to the weight matrix component w_{ij} during the nth training set,

$$\Delta w_{ij}^{(n)} = \alpha\left(d_j^{(n)} - z_j^{(n)}\right)x_i^{(n)} = \alpha\varepsilon^{(n)}x_i^{(n)}, \tag{66}$$

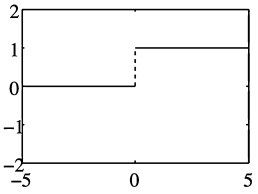
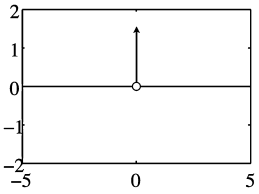
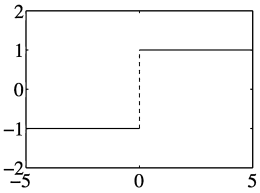
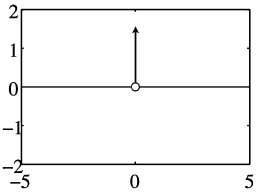
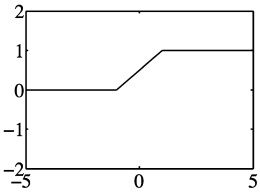
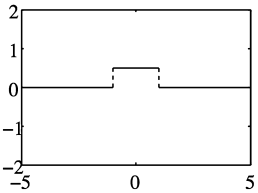
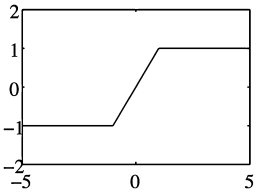
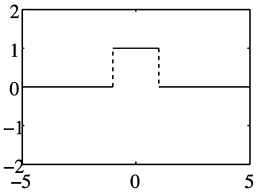
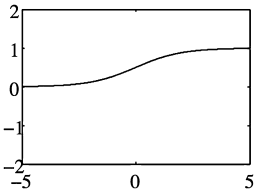
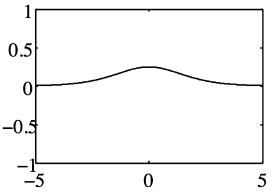
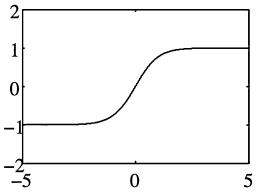
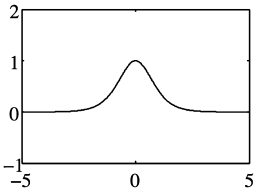
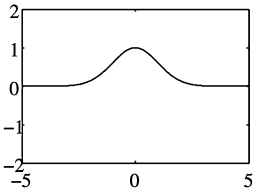
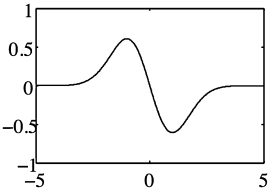
where $d_j^{(n)}$ is the desired output value for input $x_i^{(n)}$, $z_j^{(n)}$ is the actual output from the jth neuron with input $x_i^{(n)}$, $\varepsilon^{(n)}$ is the output error of the nth set and α is referred to as the learning rate coefficient. Updating the weight matrix $\mathbf{W}$ proceeds as

$$\mathbf{W}^{(n+1)} = \mathbf{W}^{(n)} + \Delta\mathbf{W}. \tag{67}$$

The result after n training sets is that each of the weights w_{ij} will have been updated according to the rule outlined in Table 5. The perceptron learning rule can be optimized through the tuning of the learning rate coefficient α. Small α corresponds to a stable and slow learning scheme. Therefore, in practice, one would like to make α as large as possible to ensure quick learning without introducing unstable oscillations about the desired value which result from over-relaxation.

In addition to the perceptron learning rule, Rosenblatt also proved a perceptron convergence theorem which states: given a finite set of inputs $\mathbf{x}_N$ and desired output training sets $\mathbf{d}_N$, each presented with a positive probability, the perceptron learning rule

Table 3
Artificial neuron transfer functions.

Classification	Expression	Derivative	$f(z)$	$f'(z)$
Unipolar step or heaviside (threshold)	$f(y) = H(y) = \begin{cases} 1, & \text{if } y > 0 \\ 0, & \text{if } y < 0 \end{cases}$	$\delta(y) = \begin{cases} 0, & \text{if } y \neq 0 \\ \infty, & \text{if } y = 0 \end{cases}$		
Bipolar step (threshold)	$f(y) = \text{sign}(y) = 2H(y)-1$	$\delta(y) = \begin{cases} 0, & \text{if } y \neq 0 \\ \infty, & \text{if } y = 0 \end{cases}$		
Unipolar linear	$f(y) = \begin{cases} 0, & \text{if } y < -1 \\ \frac{1}{2}(y+1), & \text{if } \lvert y \rvert < 1 \\ 1, & \text{if } y > 1 \end{cases}$	$½[H(y+1)-H(y-1)]$		
Bipolar linear	$f(y) = \begin{cases} -1, & \text{if } y < -1 \\ y, & \text{if } \lvert y \rvert < 1 \\ 1, & \text{if } y > 1 \end{cases}$	$H(y+1)-H(y-1)$		
Unipolar sigmoid (logistic)	$f(y) = (1/1+e^{-y})$	$f(y)(1-f(y))$		
Bipolar sigmoid (hyperbolic tangent)	$f(y) = \tanh(y)$	$(1-\lvert f(y)^2 \rvert)$		
Gaussian radial basis	$f(y) = \exp(-\lVert y-m \rVert^2/\sigma^2)$	$-2(y-m)f(y)/\sigma^2$		

provides guaranteed convergence of the weight matrix to values which give the correct outputs if and only if the said set of weights exists. For a detailed outline of the proof see Ref. [117].

However, there are many tasks for which a set of weights do not exists and, as mentioned before, in 1969 Minsky and Pappert published a book concerning the strengths and limitations of single layer perceptron networks [101]. Included in this book, and the less popular earlier text by E. B. Carne [118], was a criticism of the single layer perceptron's inability to solve the simple two-state XOR problem which is described in Table 6. Solution of the XOR problem

Table 4
Milestones in ANN development.

Number	Authors	References	Year	Structure
1	McCulloch & Pitts	[104]	1942	Threshold Logic Unit, the first Artificial Neuron (not a network).
2	Hebb & Pitts	[105]	1942	First learning rule; an object can be memorized by adapting the weights (not a network).
3	Rosenblatt	[106]	1958	Perceptron, which is the earliest ANN.
4	Lee	[281]	1959	Artron, which is a statistical switch closely related to Adaline.
5	Widrow and Hoff	[108]	1960	Adaline (ADAptive LInear NEuron), an early artificial neuron (not a network) and LMS rule for training.
6	Hopfield	[282]	1982	Hopfield Network. Different in many important ways, especially its recurrent feedback between nodes.
7	Widrow and Winter	[119]	1988	Madaline (Many Adaline) which is a network formulation of the Adaline neuron.
8	Rumelhart et al.	[283]	1986	Back-Propagation network, a multi-layer perceptron based ANN.
9	Hecht-Nielsen	[284]	1987	Counter-Propagation network uses Self-Organizing Mapping to accelerate unsupervised learning.
10	Chua & Yang	[285]	1988	Cellular networks in which neurons are connected to their nearest neighbors only.

would require a non-linear partitioning of even parity points from odd parity points. This is impossible for single layer perceptrons because they can only partition regions into two linearly separable spaces. As a result, researchers lost interest in neural computing and work on perceptrons was effectively wiped out for almost two decades.

4.2.3. Adaptive Linear Neuron (ADALINE)

Shortly after the conception of Frank Rosenblatt's perceptron, Bernard Widrow and Ted Hoff developed the ADAptive LInear NEuron (ADALINE) while at Stanford University. Like the TLU, the ADALINE is not a network, but rather a single logic neuron. Nonetheless, these artificial neurons can be connected in a single layer in much the same way as perceptrons (Fig. 9(a)). Unlike the unipolar step transfer function of the perceptron, ADALINEs employed a bipolar step function (Fig. 3), which gave an output of ±1. Inputs of ADALINEs are also typically bipolar. The main contribution of

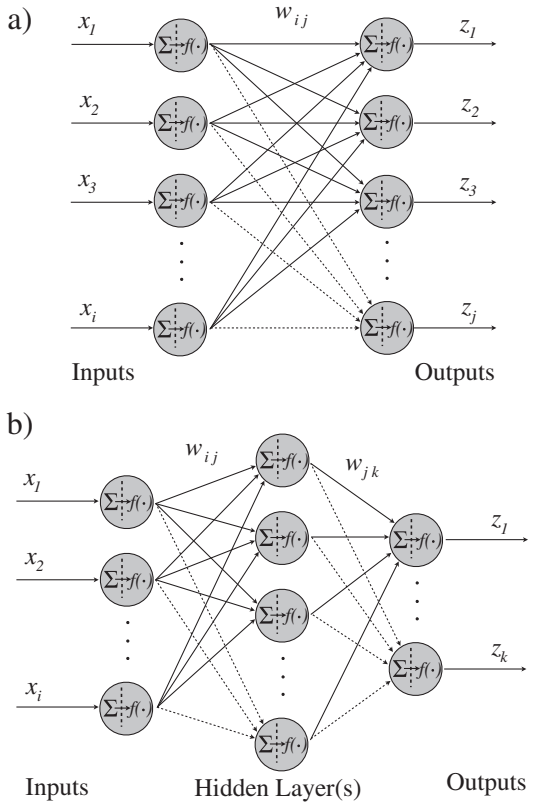


Fig. 9. (a) Simple perceptron logic only capable of mapping which require linear separability. (b) Multi-Layer ANN capable of mapping which require non-linear separability.

Widrow and Hoff was the LMS training rule in which the weights linked with each input node are adjusted with an adaptive learning algorithm as follows: given a finite number N of training sets $\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_N$ and desired outputs $d_1,d_2,\ldots,d_N$, Widrow and Hoff defined a training cost given by

$$J(\mathbf{w}) = E\left[\varepsilon_n^2\right] \cong \frac{1}{N}\sum_{n=1}^{N}\varepsilon_n^2, \tag{68}$$

where $E[\bullet]$ indicates the expectation and ε_n denotes the training error at the nth set, which is simply $d_n - z_n$. Combining the definitions $E[\mathbf{x}d] = \boldsymbol{\rho}$ and $E[\langle \mathbf{x},\mathbf{x}\rangle] = P$, where $\langle \bullet,\bullet\rangle$ denotes the inner product, with the notation from above yields

$$E\left[\varepsilon_n^2\right] = E\left[d_n^2\right] + \mathbf{w}^T P\mathbf{w} - 2\mathbf{w}^T\boldsymbol{\rho}. \tag{69}$$

Combining this result with (68), gives a gradient ∇J which is used to minimize the cost function as

$$\nabla J = \frac{\partial J(\mathbf{w})}{\partial \mathbf{w}} = 2P\mathbf{w} - 2\boldsymbol{\rho} = 0. \tag{70}$$

Thus, the optimal Least Mean Square (LMS) setting of $\mathbf{w}$ is given by

$$\mathbf{w}^{\mathrm{LMS}} = \frac{1}{P}\boldsymbol{\rho}. \tag{71}$$

Because the LMS rule for training makes use of expectations, the sample averages tend to be erroneous for a small number of N training sets. True convergence requires $N \to \infty$. The simple ADALINE processing element has found applications in many engineering problems such as adaptive filtering, echo suppression, pattern recognition and prediction [117].

4.3. Multi-layer networks

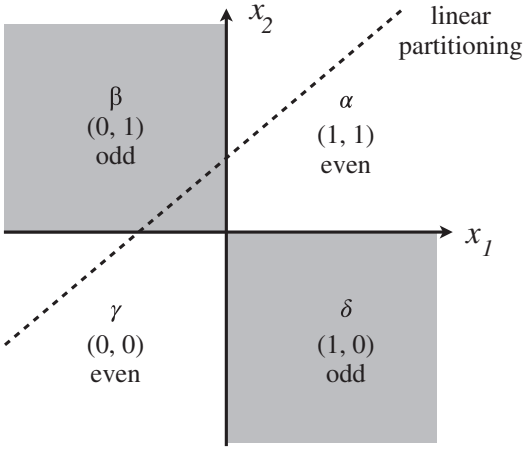
4.3.1. MADALINE

In the 1980s, almost twenty years after Minsky and Pappert's criticism of the simple perceptron's inability to solve the XOR problem, Bernard Widrow and Rodney Winter considered a new multi-layer network which consisted of Many ADALINEs (MADALINE) [119]. In addition to the input and output layers of simple ANN structures in Fig. 9(a), the MADALINE architecture included

Table 5
Perceptron learning rule.

$w_{ij}^{(n+1)} = w_{ij}^{(n)} + \alpha x_i^{(n)}$	If the output is ZERO and should be ONE
$w_{ij}^{(n+1)} = w_{ij}^{(n)} - \alpha x_i^{(n)}$	If the output is ONE and should be ZERO
$w_{ij}^{(n+1)} = w_{ij}^{(n)}$	If the desired output was achieved

Table 6
XOR state table.



State	Quadrant	Inputs		Output z	Parity
		x_1	x_2		
α	1	1	1	0	Even
β	2	0	1	1	Odd
γ	3	0	0	0	Even
δ	4	1	0	1	Odd

$(x_1 \text{ or } x_2) \cap (\bar{x}_1 \text{ or } \bar{x}_2)$, where $\bar{x} \equiv \text{not}(x)$.

additional layers as in Fig. 9(b). These additional neural layers are termed hidden due to the fact that no partial outputs are available for the training of these nodes. The use of multiple layers is one way of circumventing the problem of computing mappings which require non-linear separability. It is the introduction of hidden layers that enabled the MADALINE architecture to solve the XOR problem and renew interest in neural computing.

However, the history of the application of multi-layer ANNs to mappings which require non-linear separability actually began in 1900 with Hilbert's formulation of his famous 23 problems for the coming century [120]. In particular, Hilbert's 13th problem came in the form of the following supposition: there exist analytic functions of three variables that *cannot* be represented by a finite superposition of continuous functions of only two variables. This supposition was discredited by Kolmogorov in 1957 [121] whose reasoning was later refined by Sprecher in 1965 [122] who provided the following theorem [82].

For each integer $n \geq 2$, there exists a real, monotonically increasing function $\psi(x)$ which is dependent on n and possesses the property: for each pre-allocated number $\delta > 0$ there exists a rational number ε, $0 < \varepsilon < \delta$, such that every real continuous function of n variables $f(\mathbf{x})$, can be represented by

$$f(\mathbf{x}) = \sum_{j=1}^{2n+1} \chi \left[\sum_{i=1}^{n} \lambda^i \psi(x_i + \varepsilon(j-1)) + j - 1 \right], \tag{72}$$

where the function χ is real and continuous and λ is a constant which is independent of f.

The coupling of the previous theorem and multi-layer ANNs was pointed out by Hecht-Nielsen in 1987 [122] when he demonstrated that, as a result of the Sprecher theorem, any continuous mapping f can be regarded as a form of multi-layer ANN with two hidden layers. The output function of the first and second layers are given by Ψ and χ respectively [122,123]. In addition to the work of Sprecher, Hornik et al. [124], Funahashi [123] and Cybenko [125] independently demonstrated that a general multi-layer ANN with sufficient hidden layers is adequate for the approximation of a continuous non-linear function on a finite closed interval.

Widrow and Winter also developed a training algorithm known as the MADALINE adaptation Rule (MR). Widrow reasoned that the weights should be modified at each training set to reduce the error with as little disturbance to the representations learned by the previous training sets which he called the *principle of minimal disturbance* [117,119]. The MR training algorithm has been refined several times yielding procedures labeled MR, MRII, MRIII etc. The procedure outlined below is MRIII.

The decision of whether or not the weights for a given layer should be adjusted requires an input vector **x** and the appropriate target output **d** in order to calculate the sum of the squared output errors. Widrow changed the input to the kth neuron by some small amount Δs and investigated the change in the sum squared output error given by

$$\Delta\varepsilon^2 = \Delta\left(\sum_k \varepsilon_k^2\right). \tag{73}$$

The gradient of the sum squared output error term with respect to its weight vector is estimated through the use of finite differences as

$$\nabla_k = \frac{\partial\left(\varepsilon_k^2\right)}{\partial s_k} x_k \cong \frac{\Delta\left(\varepsilon_k^2\right)}{\Delta s} x_k, \tag{74}$$

which is subsequently used to reduce the error in a direction directly opposite of the gradient;

$$w_{k+1} = w_k - \alpha \frac{\Delta\left(\varepsilon_k^2\right)}{\Delta s} x_k. \tag{75}$$

Finally, defining $f' = f(1-f)$ as the derivative of the sigmoid function (see Table 3), the MRIII training rule given is by

$$w_{k+1} = w_k + 2\alpha\varepsilon_k f'(s_k) x_k, \tag{76}$$

where α is again the learning rate coefficient. It should be noted that this training algorithm is similar to the back-propagation training algorithm discussed in the next section. MADALINEs have found applications in many areas outside of solar irradiance modeling and forecasting, such as invariant pattern recognition, missile guidance and detonation and general computations of well-behaved functions [117].

4.3.2. Multi-layer perceptron

Multi-Layer Perceptrons (MLPs) with Back Propagation-Learning (BPL) are some of the most established ANN architectures due to their ability to perform arbitrary non-linear mappings. MLPs are also sometimes referred to as Multi-Layer Feed-Forward (MLFF) networks to emphasize the forward flow of information with respect to the backward direction of weight adjustment used in BPL algorithm. The structure of MLPs are similar to MADALINEs, with layered neurons possessing only forward connections to successive layers (see Fig. 9(b)). However, MLPs differ from MADALINEs in the derivation of the popular BPL algorithm outlined next. As we have seen before, the LMS and Perceptron Learning rules perform in very similar ways. Nonetheless, neither of these learning rules can be applied to MLPs because these methods do not identify how to make adjustments to weights associated with hidden-layers. As a result, a new method using upstream variables was developed.

Several authors independently derived the BPL method in the 1970's and 1980s. Paul J. Werbos proposed the BPL method first in his 1974 Harvard University doctoral dissertation [126]. Almost ten years later, in 1985, D. B. Parker re-derived the BPL method in his MIT technical report [127]. Even so, credit for developing the BPL algorithm into a realizable procedure is typically assigned to David Rummelhart and the other members of his distributed processing group at the University of California, San Diego, in 1985 [128].

In order to begin a discussion of BPL, we consider an MPL with a single hidden layer which has been modified for the BPL algorithm as shown in Fig. 10. The results can then be generalized to an MLP with an arbitrary number of hidden layers. As before, we adopt the following convention for weight indexing: the synaptic weight connecting input neuron i to hidden layer neuron j is denoted by w_{ij} and the weight connecting hidden layer neuron j to output neuron k is denoted by w_{jk}. We will also define the net input to the hidden and output layers' neurons as,

$$H_j = \sum_i w_{ij} x_i \tag{77}$$

and

$$O_k = \sum_j w_{jk} y_j, \tag{78}$$

respectively. This is similar to an unbiased version of Equation (64). Using the above notation we can write the outputs of node j of the hidden layer and unit k of the output layer as

$$z_j = f(H_j) \tag{79}$$

$$z_k = f(O_k) \tag{80}$$

In order to begin training one must have differentiable transfer functions, such as the sigmoid function, and access to a training data set consisting of N pairs of input patterns. As usual, a training method is developed in order to reduce the total system error for all training patterns through an adjustment of the weights. The total system error E_{sys}, defined as the average of the output errors over all training patterns, is

$$E_{\text{sys}} = \frac{1}{N} \sum_{n=1}^{N} E_n. \tag{81}$$

As has been done before, weights are adjusted on each successive pattern presentation proportional to the negative of the error gradient. Therefore, at the $k+1$ step we can write

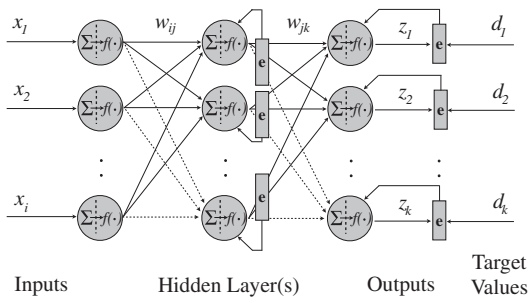


Fig. 10. Modification of an MLP network with a single hidden layer for the application of the BPL algorithm. The use of partial derivatives of errors with respect to the weights upstream of the flow of information allows the algorithm to overcome the problem of inaccessible of hidden layer data and is at the core of the BPL. The resulting algorithm may then be generalized to an MLP with an arbitrary number of hidden layers.

$$\Delta W^{(k+1)} = -\alpha \frac{\partial E_n}{\partial \mathbf{W}^{(k)}}, \tag{82}$$

where α is the familiar learning rate coefficient. The total system error can be written as

$$\frac{\partial E_{\text{sys}}}{\partial \mathbf{W}} = \frac{1}{N} \sum_{n=1}^{N} \frac{\partial E_n}{\partial \mathbf{W}}. \tag{83}$$

Here we define E_n as the mean square error,

$$E_n = \frac{1}{2} \sum_{n=1}^{N} (d_n - z_n)^2. \tag{84}$$

In order to express the update rule in terms of the system parameters, we must successively apply the chain rule. Since we know for output unit k

$$\frac{\partial E}{\partial z_k} = -(d_k - z_k) \tag{85}$$

and

$$\frac{\partial z_k}{\partial I_k} = f'(I_k), \tag{86}$$

we can write

$$\frac{\partial E}{\partial I_k} = \frac{\partial E}{\partial z_k} \frac{\partial z_k}{\partial I_k} = -(d_k - z_k) f'(I_k). \tag{87}$$

By defining the previous operation as a generalized delta term,

$$\delta_k = (d_k - z_k) f'(I_k), \tag{88}$$

we can now write an update rule for the output units as

$$\Delta w_{jk} = -\alpha \frac{\partial E}{\partial w_{jk}} = \alpha \delta_k y_j. \tag{89}$$

The same reasoning is applied to the hidden layer weights. We want an expression for the update rule given by

$$\Delta w_{ij} = -\alpha \frac{\partial E}{\partial w_{ij}} = -\alpha \frac{\partial E}{\partial H_j} \frac{\partial H_j}{\partial w_{ij}}. \tag{90}$$

Examination of the first term reveals that another application of the chain rule is required,

$$\frac{\partial E}{\partial H_j} = \frac{\partial E}{\partial z_j} \frac{\partial z_j}{\partial H_j} = \frac{\partial E}{\partial z_j} f'(H_j), \tag{91}$$

while the second term can be written explicitly from Equation (77) as,

$$\frac{\partial H_j}{\partial w_{ij}} = \sum_i \frac{\partial}{\partial w_{ij}} (w_{ij} x_i) = x_i. \tag{92}$$

Using the fact that we can express the output from node z_k for a given input pattern as

$$\begin{aligned} z_k = f(I_k) &= f\left(\sum_j w_{jk} z_j \right) = f\left(\sum_j w_{jk} f(H_j) \right) \\ &= f\left(\sum_j w_{jk} f\left(\sum_i w_{ij} x_i \right) \right) \end{aligned} \tag{93}$$

along with the error defined in Equation (84), we are able to directly differentiate the $\partial E/\partial z_j$ term in Equation (91), obtaining

$$\frac{\partial E}{\partial z_j} = \frac{1}{2}\sum_k \frac{\partial\left(d_k - f\left(\sum_j w_{jk} z_j\right)\right)^2}{\partial z_j} \tag{94}$$

$$= -\sum_k (d_k - z_k) f'(I_k) w_{jk}. \tag{95}$$

As done before, the previous expression may be conveniently written as a general delta operator

$$\delta_j = f'(H_j)\sum_k \delta_k w_{jk}, \tag{96}$$

and the update rule for the hidden layer can be written as

$$\Delta w_{ij} = \alpha \delta_j x_i. \tag{97}$$

The results for the BPL learning rule are summarized in Table 7.

It is useful here to pause and more closely examine the previous result. For a given input pattern x_n, the information flows forward through the network until it reaches the output layer where node k calculates its output z_k. Subsequently, errors are calculated from knowledge of the desired output pattern d_n and weights w_{jk} are adjusted according to Equation (89) as

$$w_{jk}^{(n+1)} = w_{jk}^{(n)} + \Delta w_{jk}^{(n)}, \tag{98}$$

which is very similar to the perceptron learning rule summarized in Table 5. Turning our interest to the hidden layer nodes, we do not have target values to use in the computation of an error like we did with the output layers. As a result, in this case, we must distribute the output errors in some way in order to adjust the weights connecting input node i to hidden node j. This is accomplished through the use of δ_j

$$w_{ij}^{(n+1)} = w_{ij}^{(n)} + \Delta w_{ij}, \tag{99}$$

which was discussed earlier. Thus, one can see that the BPL algorithm takes partial derivatives of errors with respect to the weights upstream, which allows it to overcome the inaccessibility of hidden layer data. The resulting errors are then propagated "backwards" through the network every time a training pattern is presented. For a detailed derivation of the BPL algorithm see Refs. [82,116,117,126–128].

4.4. Applications of ANNs

Since the late 1990s ANNs have seen increased application in the field of solar forecasting. Al-Alawi and Al-Hinai used climatological variables as inputs to an ANN to predict monthly values of GHI over a year [129]. Sfetsos and Coonick developed a forecast model for mean hourly GHI based on several AI-based techniques including linear, feed-forward, recurrent and radial basis ANNs alongside an adaptive neuro-fuzzy inference scheme which showed improvements over traditional linear methods [78]. Cao and Cao combined a recurrent ANN with wavelet analysis for the forecast of daily solar irradiance [130,131]. Hontoria et al. used supervised Recurrent Neural Networks (RNNs) for the generation of synthetic hourly GHI time-series [132]. Sözen et al. used meteorological data from 4 years (2000–2003) from twelve cities spread over Turkey as training (nine stations) and testing (three stations) data for ANNs used to predict mean monthly solar irradiance [133]. Cao and Lin combined an RNN with a Wavelet Neural Network (WNN) to develop a Diagonal Recurrent Wavelet Neural Network (DRWNN) for hourly GHI forecasting [134,135]. Yona et al. used MLPs, RNNs, and RBNNs for 24-h-ahead power output forecasting for PV systems [136]. Paoli et al. recently used MLPs for the prediction of daily GHI [137,138]. Azadeh et al. also used MLPs for forecasts, however their analysis was based on monthly average meteorological data from six cities in Iran [139]. Mellit and Pavin proposed an MLP-model to forecast solar irradiance on a 24 h basis using the current values of the mean daily solar irradiance and air temperature for the operation of a grid connected PV plants in Italy [140]. Mellit et al. employed a Field Programmable Gate Array (FPGA) and an MLP to predict daily GHI values over a year [141]. Chen et al. made use of historic power measurements and meteorological forecasts of relative humidity and temperature with an ANN to provide online 24 h ahead forecasts of GHI appropriate for operational planning of transmission system operation [142]. ANNs have also been used in many other areas of energy science including modeling of solar steam-generators, solar water heating systems, Heating Ventilating and Air Conditioning (HVAC) systems, wind speed predictions, control in power generation systems, load forecasting and refrigeration [143].

Table 7
Back propagation learning rule.

Output units	$\Delta w_{jk} = \alpha \delta_k z_j$	where $\delta_k = (d_k - z_k) f'(I_k)$
Hidden units	$\Delta w_{ij} = \alpha \delta_j x_i$	where $\delta_j = f'(H_j)\sum_k \delta_k w_{jk}$

4.5. k-nearest neighbors (k-NN)

Despite the similarity in the acronym, the k-nearest neighbors (k-NN) method is unrelated to ANNs. Although both methods can be considered Artificial Intelligence techniques, the origin of the k-NN stems from classification theory and its origins can be tracked back to an unpublished technical report by Fix & Hodges regarding the density estimation for classification theory. Even though their work has not been published in peer-review literature Fix & Hodges have received ample recognition in classification theory textbooks and articles [144–146].

As an algorithm for pattern classification in classification theory the k-NN method is one of several nonparametric techniques for estimating the density functions from sample patterns. Within the confines of classification theory the k-NN has a very specific place and its mathematical derivation and properties can be derived precisely from the underlying theory. Outside of the confines of classification theory the k-NN is often defined more heuristically by the following statement (or some version thereof): an initial set of objects is created and each object is classified by a property or value; new objects are classified based on the decision rule that assigns to an unclassified object the class/value of the nearest k objects (hence the designation k-nearest neighbors), where k is a positive integer, typically small. If $k = 1$, then the object is simply assigned to the class of its nearest neighbor. Fig. 11 depicts this heuristic definition graphically.

The k-NN is one of the simplest methods among the machine learning algorithms. In contrast to the statistical methods presented before that attempt to find models from the available data, the k-NN uses the training data as the model, and unlike the ANN models there are no unknown model coefficients that need to be determined through training. Despite the fact that the k-NN model was originally developed for pattern classification its simplicity allows it to be applied to regression problems for time-series [147], such as the forecast of solar radiation. Although there are very few articles in literature that apply k-NN to the forecasting of solar irradiation (see for instance [138,148]), k-NN has been extensively applied as a forecasting technique to problems such as: the forecast of electricity load and electricity market price [149,150], the

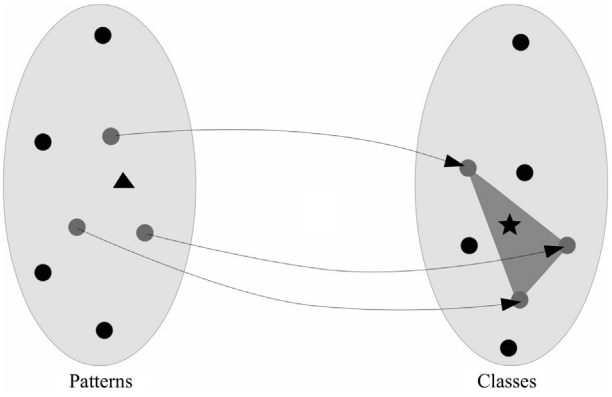


Fig. 11. The schematic of the k-NN classification. The triangle represents the unclassified pattern and grey circles represent the 3 nearest neighbors. The class for the new pattern can be computed from the weighted average from the k-NN classes. Depending on the weights, the new class – represented by the star symbol, can assume values anywhere in the shaded area.

forecast of daily river water temperature [151], the forecast of water inflow [152] and weather forecast [153], just to cite some of the most recent publications. In the field of meteorology and weather forecasting the k-NN method is also known as the analog method [154–156]. The analog method searches a database of past weather patterns that resemble most closely the current weather conditions. The patterns are described either in terms of circulation patterns or in terms of various surface weather parameters. The selected past weather conditions, also known as analogs, are then used for predicting weather and its various components such as maximum, minimum and ambient temperatures, average wind speed, atmospheric pressure, etc. Likewise, for the propose of forecasting time-series, the k-NN model consists of looking into the time-series history and identifying the timestamps in the past that resemble the "current" conditions most closely – the nearest neighbors. Once they are determined the forecasting is computed from the time-series values subsequent to the matches. In essence, the k-NN model resembles a lookup table for which previous patterns are used as indicators of sequential behavior.

The first step in developing a k-NN model is to construct the database of features that will be used in the comparison with "current" conditions. For a univariate k-NN (no exogenous variables), examples of features used are: values of the time-series, averaged values of the time-series, entropy of the time-series, the cumulative sum of the time-series, etc. If significant exogenous variables are available they can easily be added to the set of patterns in the k-NN database. Fig. 12 exemplifies some of the patterns derived from an univariate time-series – in this case the GHI clear sky index, k_t.

With respect to the implementation of k-NN the algorithm can be easily summarized thus: assuming that the n features or patterns for time t are assembled in the vector $\mathbf{p}$ with components p_i $i = 1,\cdots,n$ and that the features for historical data are assembled in a matrix A_{ij} whose rows correspond to the vector of features for each time in the historical data set, the k nearest neighbors to p_i can be determined in two steps:

1. compute the distance between p_i and all the rows from A_{ij}. Using the Euclidean distance that yields:

$$d_j = \sum_i \sqrt{(p_i - A_{ij})^2}; \tag{100}$$

2. extract the k indices that return smallest elements of d_j and their associated k timestamps $\{\tau_1,\cdots,\tau_k\}$.

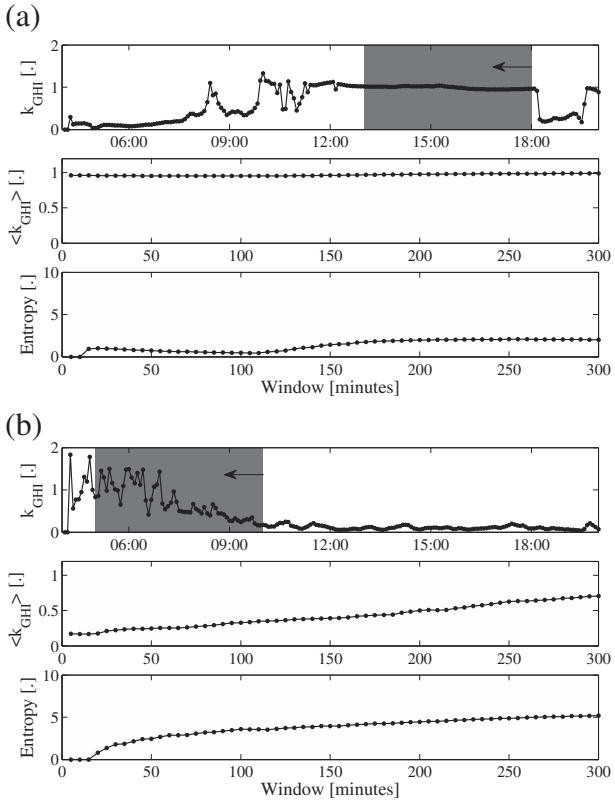


Fig. 12. Examples of preprocessing operations used to create patterns from the time-series. Only the last 300 min are used in the computation of the patterns. (a) Examples of patterns of the k_{GHI} time-series for a section of the time-series with low variability. *Top*: the time-series; *middle*: the Backward cumulative sum for windows from 5 to 300 min; *bottom*: the Entropy for windows from 5 to 300 min. (b) Same as above but for a section of the time-series with high variability.

Once the set of nearest neighbors is determined the forecast can be obtained as:

$$z_{t+\Delta t} = \frac{\sum_{i=1}^{k} \alpha_i z_{\tau_i+\Delta t}}{\sum_{i=1}^{k} \alpha_i} \tag{101}$$

In other words, the prediction $z_{t+\Delta t}$ is computed from a linear combination of the time-series values that follow the nearest neighbors τ_i. The weights α_i often are a function of the distance d_j – Fig. 11.

5. Remote sensing models

Solar forecasts are also used in engineering applications such as the design and operation of solar power plants and the modeling of agricultural variables including vegetation and crop growth, crop yield, soil moisture, surface energy and moisture balance [16,50,157–160]. However, all of the solar forecasting methods discussed so far require accurate measurement of the solar resource. Dissatisfyingly, precise pyranometer measurements have historically been rangebound to scattered weather stations and universities due to the expense and labor required to procure, maintain and quality check instruments and data. For this reason, when no data was available for a given location it was common to extrapolate data from the nearest ground measurement station [161]. This approach is appropriate for sites in proximity to measurement stations to constitute strong micro-climatic and environmental correlation. Having said that, it is a well known result

that the error of this method increases with extrapolation distance and temporal refinement [161] thereby limiting its application. An alternative to a national or global scale ground sourced monitoring network is provided through satellite based irradiance measurements of the Earth and its atmosphere.

Origins of the meteorological satellite can be traced back to a paper published in 1919 by Robert H. Goddard titled "A Method of Reaching Extreme Altitudes" [162]. A decade later, in 1929, Goddard launched a liquid fueled rocket carrying instrumentation which included a thermometer, a barometer and a camera [163]. Advancements in rocket technology over the two decades following Goddard's work led to the first images of clouds from high altitudes [164,165] and eventually the launch of Sputnik-1; the first artificial satellite to be put into Earth's orbit by the Soviet Space Program. Shortly after, in 1958, the U.S. formed the National Aeronautics and Space Administration (NASA) which would launch a number of meteorological satellite missions; many of which would carry instrumentation useful for atmospheric modeling.

5.1. Orbits

Before discussing the assortment of satellite based methods it is useful to address some foundational principles. There are two basic types of orbits which a meteorological satellite may occupy: Low Earth Orbit (LEO) and GeoStationary Orbit (GSO). Each type of orbit presents its own set of advantages and disadvantages. LEOs get their name from their relatively low altitude of approximately 750–850 km or about 10% of the radius of the Earth. This low orbiting altitude implies a short period, typically on the order of 90–100 min, which leads to about 15 orbits per day. Satellites in a LEO can be used to measure ozone concentrations or atmospheric temperatures among many other quantities.

LEOs may be further classified as sun-synchronous, polar or both. A satellite in a Sun-Synchronous Low Earth Orbit (SSLEO) integrates altitude and inclination in such a way that the satellite passes over a given location on the earth at the same local solar time. For example, a satellite in an SSLEO might cross over the equator fifteen times a day each time at 11:00 a.m. local time. Satellites in an SSLEO also need to process approximately one degree per day to keep up with the revolution of the Earth around the sun. SSLEOs are used for satellites which require constant illumination for imagery. Satellites that make use of visible band radiation are placed in SSLEOs with bright sunlight while satellites which measure long-wave radiation are placed in darkness.

A satellite in a Polar Low Earth Orbit (PLEO) has an inclination of almost 90° and passes over (or near) the planet's poles on each revolution. PLEOs remain fixed with respect to the rotation of the Earth which will cause the satellite to pass over the equator at a different longitude on successive orbits. This allows the satellite to image many different parts of the Earth's surface as it rotates underneath the satellite. The shortcoming of these orbits is that they image different parts of the Earth at different times making observation of one location difficult. The exception to this is if the satellite is in what is known as a Polar Sun-Synchronous Low Earth Orbit (PSSLEO). Satellites in a PSSLEO may be used to conduct high resolution imagery and radiometry of a single location a finite number of times per day.

GSOs are orbits in which the satellite circles the Earth at the same rate as the Earth spins which places them at an altitude of approximately 36,000 km or about one tenth of the distance to the moon. Satellites in GSO are located near the equator since there is a constant force of gravity due to the bulging of the Earth at this latitude. Consequently, the satellite appears stationary over a certain point on the equatorial plane which allows frequent full disk observations. However, because the satellites are almost fifty times farther than their LEO counterparts the image resolution is typically much lower, on the order of 3–5 km. Fig. 13 illustrates some of the fundamental differences between LEOs and GSOs. Once the various types of meteorological satellite orbits are understood we can move on to a discussion of the specific satellite missions.

5.2. Satellites

NASA launched its first LEO meteorological satellite, the Television InfraRed Observation Satellite (TIROS-1), on the first of April, 1960. TIROS-1 was equipped with a five-channel radiometer designed by Hanel [166]. The radiometer's third channel measured solar radiation reflected by Earth and its atmosphere in a spectral range of 0.2–6 μm. TIROS-1 would be succeeded by nine other similar satellites (TIROS 2-10) launched between 1960 and 1966.

Shortly after the initial TIROS program, NASA developed the Environmental Science Services Administration (ESSA) satellite program which lasted four years and included the launch of nine satellites. ESSA was essentially an extension of the mission objectives from TIROS; namely to provide cloud-cover photography to national meteorological centers. Resolution was improved over the TIROS program and ESSA satellites included a "cartwheel" feature initially tested on TIROS-9. This cartwheel configuration combined with two cameras mounted 180° opposite each other and an SSPLEO allowed the ESSA series to image a given point on the surface every time the satellite rotated about its axis. However, radiometers were deployed only on ESSA-7 and ESSA-9.

As a result of the success of the early ESSA missions the NIMBUS program was initiated in 1964 and included seven satellites. The NIMBUS satellites were launched into SSLEO and used a new automatic picture transmission to more than 60 low-cost data acquisition stations on the ground. The Advanced Vidicon Camera System (AVCS), which was used for daylight imaging, and the infrared Flat Plate Radiometer (FPR), which was employed to monitor night time cloud cover conditions, were contributed to the ESSA-9 mission. In addition, the NIMBUS program allowed for the first global radiation budget which permitted the refinement and validation of climatological models. Other LEO satellites include NASA's Aqua and Terra series used to monitor the Earth's water cycle and changing climate respectively [167].

Current operational LEO satellites include the U.S. National Oceanic and Atmospheric Administration (NOAA) series, also known

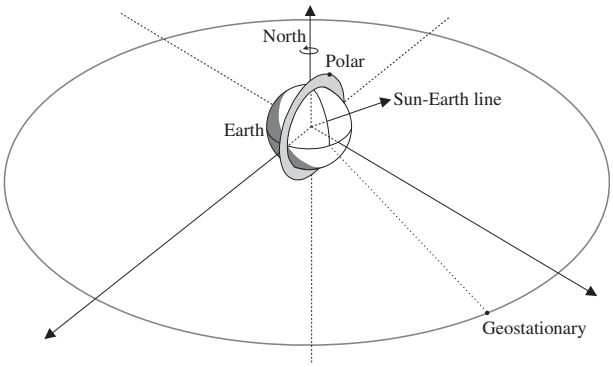


Fig. 13. Illustration of a Polar Sun-Synchronous Low Earth Orbit (PSSLEO) satellite and a GeoStationary Orbit (GSO) satellite. The PSSLEO satellite has an altitude which is nearly 50 times lower than the GSO and an inclination of nearly 90° with respect to the GSO and equatorial plane. As a result, the PSSLEO satellite completes approximately fifteen orbits per day tracing out the disk shown in the image while the GSO satellite remains fixed above a single point on the equator. The PSSLEO satellite is able to conduct high resolution imagery and radiometry of a location of interest only a finite number of times per day. On the other hand, the GSO satellite can provide continuous full disk observations, but at a substantially reduced resolution.

as the Advanced TIROS-N (ATN), and the European Organization for the Exploration of Meteorological Satellites (EUMETSAT) MetOp series [167]. The NOAA series was initially developed as the Improved TIROS Operational System (ITOS) and then the Next generation TIROS (TIROS-N) programs which were launched between 1970 and 1994. Both the NOAA and MetOp series of LEO satellites are equipped with the third generation of the visible/infrared Advanced Very High Resolution Radiometer (AVHRR) and the Advanced Microwave Sounding Unit (AMSU) [167]. Other LEO satellite programs include Russia's Meteor satellite series and China's Feng Yun satellites. In October 2011 NASA's Suomi National Polar-observing Partnership (NPP) launched a new satellite which carried five instruments: the Advanced Technology Microwave Sounder (ATMS), the Cross-track Infrared Sounder (CrIS), the Ozone Mapping and Profiler Suite (OMPS), the Visible Infrared Imaging Radiometer Suite (VIIRS) and Clouds and the Earth's Radiant Energy System (CERES).

While LEO satellites are well suited for the high resolution monitoring of specific locations at finite intervals, continuous full disk observations were required for applications of weather analysis. Accordingly, interest shifted from PLEO to GSO satellites. In 1966 and 1967, two experimental Applications Technology Satellites (ATS-1 and ATS-3) were launched by NASA in order to examine the practicality of placing a satellite into a GSO. The ATS series was primarily intended for communications, but also carried meteorological related equipment including a spin-scan cloud camera. Shortly after, in 1974 and 1975, two more GSO satellites were launched by NASA and operated by NOAA, Synchronous Meteorological Satellites (SMS-1 and SMS-2). The SMS series carried a visible infrared spin-scan radiometer (VISSR) for day/night measurements of cloud cover and temperatures of the earth and atmosphere.

The high-altitude GSO techniques developed during the ATS and SMS programs paved the way for current operational GSO satellites which include the NOAA Geostationary Operational Environmental Satellite (GOES), the European METEOSAT series and the Japanese Multifunctional Transport Satellites (MTS) [167]. The instrumentation on these operational GSO satellites tend to vary by operator, however the capabilities of the instrumentation share a number of characteristics including the ability to measure visible and infrared radiation. Typically, images are acquired every 30 min although the METEOSAT's Spinning Enhanced Visible and Infrared Imager (SEVIRI) possesses a nominal sampling frequency of 15 min [167,168] and the GOES imager can scan small areas of interest in as little as 30 s [167,169].

5.3. Radiation budget

Prior to the launch of TIROS-1 it was impossible to make direct measurements of the radiation that originated from the sun and was subsequently scattered back into space by the earth and its atmosphere [170]. This scattered fraction is referred to as the planetary albedo (not to be confused with the atmospheric or terrestrial albedo discussed later). Previous attempts to quantify planetary albedo made use of indirect measurements of the dark side of the moon, which was consequently illuminated by a fraction of the sunlit earth [171–173]. Other approaches involved the estimation of cloud distribution over the planet and the assignment of specific reflectivities to various cloud types [174–176].

Eventually, satellites equipped with radiometers allowed for direct measurement of the planetary albedo. Fig. 14 illustrates the fundamental relationship between satellite observation of planetary albedo and ground level irradiance. It should be noted that a slightly different notation is used in this section in order to stay consistent with notations commonly used in the field. By the conservation principle, the difference between the net irradiance at the top of the atmosphere $I_{\text{in}}^{\text{TOA}}$ and the net irradiance at the ground level I_{in}^{g} is equal to the flux lost (either reflected, scattered or absorbed) in the volume joining the two,

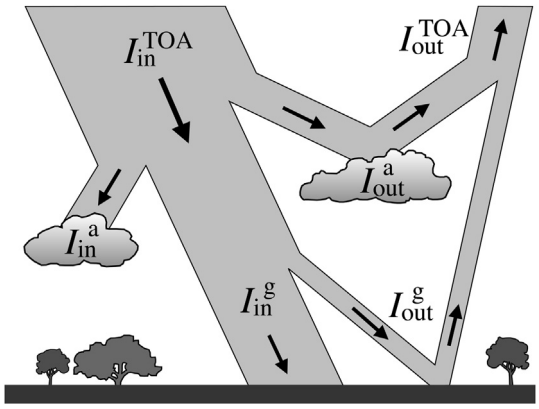


Fig. 14. Illustration of the fundamental relationship between satellite observation of planetary albedo and ground level irradiance. Satellites equipped with radiometers allow for the direct measurement of the planetary albedo and by the conservation principle, the difference between the net irradiance at the top of the atmosphere $I_{\text{in}}^{\text{TOA}}$ and the net irradiance at the ground level I_{in}^{g} must be equal to the flux lost (either reflected, scattered or absorbed) in the volume joining the two.

$$I_{\text{in}}^{\text{TOA}} - I_{\text{in}}^{g} = I_{\text{in}}^{a} + I_{\text{out}}^{\text{TOA}}. \tag{102}$$

In the previous equation I_{in}^{a} is the flux lost due to absorption in the atmosphere and $I_{\text{out}}^{\text{TOA}}$ is the total irradiance reflected (or scattered) back out of the atmosphere into space, a combination of the irradiance reflected from the ground I_{out}^{g} and the irradiance reflected from the atmosphere I_{out}^{a}:

$$I_{\text{out}}^{\text{TOA}} = I_{\text{out}}^{g} + I_{\text{out}}^{a}. \tag{103}$$

Substitution of (103) into (102) combined with normalization with respect to $I_{\text{in}}^{\text{TOA}}$ and a separation of atmospheric and terrestrial terms yields,

$$\frac{I_{\text{in}}^{g}}{I_{\text{in}}^{\text{TOA}}} + \frac{I_{\text{out}}^{g}}{I_{\text{in}}^{\text{TOA}}} = 1 - \frac{I_{\text{in}}^{a}}{I_{\text{in}}^{\text{TOA}}} - \frac{I_{\text{out}}^{a}}{I_{\text{in}}^{\text{TOA}}}. \tag{104}$$

The right hand side of the previous equation can be written as,

$$\frac{I_{\text{in}}^{g}}{I_{\text{in}}^{\text{TOA}}} + \frac{I_{\text{out}}^{g}}{I_{\text{in}}^{\text{TOA}}} = \left(1 + \frac{I_{\text{out}}^{g}}{I_{\text{in}}^{g}}\right) \frac{I_{\text{in}}^{g}}{I_{\text{in}}^{\text{TOA}}}, \tag{105}$$

which gives the following relationship

$$\frac{I_{\text{in}}^{g}}{I_{\text{in}}^{\text{TOA}}} = \frac{1 - \frac{I_{\text{in}}^{a}}{I_{\text{in}}^{\text{TOA}}} - \frac{I_{\text{out}}^{a}}{I_{\text{in}}^{\text{TOA}}}}{1 + \frac{I_{\text{out}}^{g}}{I_{\text{in}}^{g}}}. \tag{106}$$

The observant reader may recognize the clearness index K_t on the left hand side of the equation along with the atmospheric absorptivity,

$$\alpha_a = \frac{I_{\text{in}}^{a}}{I_{\text{in}}^{\text{TOA}}}; \tag{107}$$

the atmospheric albedo,

$$\rho_a = \frac{I_{\text{out}}^{a}}{I_{\text{in}}^{\text{TOA}}}; \tag{108}$$

and the terrestrial albedo,

$$\rho_g = \frac{I^g_{\text{out}}}{I^g_{\text{in}}}. \quad (109)$$

Employing the previous definitions reveals the relationship between clearness index and quantities which can be measured by the satellite's radiometers in a compact form,

$$K_t = \frac{1-\alpha_a-\rho_a}{1+\rho_g}. \quad (110)$$

As one would expect, the atmospheric albedo intensifies with increasing atmospheric turbidity and cloudiness. In addition, heightened atmospheric albedo also implies the attenuation of the radiation reaching the ground. If we are able to estimate the atmospheric absorption coefficient and know the terrestrial albedo, a priori, we would be able to approximate the clearness index K_t from the measured satellite data using Equation (110).

The method used to approximate the aforementioned parameters typically varies by author. Nonetheless, the relationship between the clearness index and the atmospheric and terrestrial parameters in Equation (110) is at the foundation of all satellite based irradiance forecasting models. Measurement of $I^{\text{TOA}}_{\text{out}}$ combined with an approximation of I^a_{in} (through the knowledge of the Linke turbidity coefficient, for example) and a priori knowledge of I^g_{out} (from historical measurements of clear day planetary albedo) allows for the estimation of ground level irradiance from information regarding the clearness index K_t.

Depending on the treatment on the interaction of solar irradiance and the atmosphere satellite models may be classified into two categories: physical models and statistical models [177,178]. The benefits of physical models include their generality and the elimination of the requirement of ground data through the use of RTMs. Correspondingly, RTMs introduce the requirement of precise and comprehensive measurement of the atmospheric morphology as well as careful calibration of the satellite data. Statistical models, on the other hand, rely on straightforward statistical regressions between satellite and ground measurements. As a result, statistical models are less computationally expensive than RTM based models, however, they suffer from their loss of generality and requirement of ground data.

5.4. Physical satellite models

Physical satellite models for solar irradiance forecasting are based on the interactions between solar radiation and participating atmospheric components such as gasses (CO_2, H_2O, N_2, O_2, O_3, etc.) and aerosols. Similar to a number of the clear sky models covered in Section 2.3, the physical interactions are typically modeled by way of RTMs. Thus, physical satellite models can be thought of as an improvement upon RTM based clear sky models through the addition of information regarding current atmospheric conditions. Atmospheric conditions are accounted for through the measurement of local meteorological data. This eliminates the need for solar irradiance data at the surface, but, because these models need to convert digital counts from satellite based radiometers into a corresponding flux densities, accurate and frequent calibration of the instrument is required [178].

5.4.1. Gautier-Diak-Masse model

One of the earliest physical models was developed by Gautier, Diak and Masse (GDM) in 1980 [179]. The GDM model considered clear and cloudy conditions separately. The determination of whether a given pixel was clear or cloudy was accomplished by selecting a brightness threshold obtained through the selection of the minimum value at every pixel for every hour from the past several days. One shortcoming of the original GDM model was the absence of variations in terrestrial albedo with changing solar zenith angle. Raphael and Hay [180] included the T-minimum brightness determination [177] in order to correct for the previous consideration.

The GDM clear sky RTM included several input parameters: reflection coefficient for diffuse radiation, atmospheric transmissivity as a function of the reflection coefficient for DNI, the absorption coefficient for slant water vapor path and the solar zenith angle, the atmospheric albedo as a function of the irradiance received by the satellite, and a second absorption coefficient for slant water vapor path for the satellite zenith angle. Reflection coefficients were calculated using the results from Coulson [181,182] while the absorption coefficients made use of the expressions from Paltridge [183]. Improvements to GDM model include the absorption of ozone and aerosols by Gautier and Frouin [184] in addition to multiple reflections by Tanrè et al. [185].

The cloudy sky RTM allowed for the simple treatment of stratiform low and middle altitude clouds [179]. Absorption above and below the clouds were considered separately and further classified as either upwelling or downwelling. Similar inputs for the cloudy sky RTM were considered: cloud albedo as a function of the absorption of short wave radiation above and below the cloud, and cloud absorption coefficient estimated on the basis of the satellite's measurement of the visible brightness of the cloud. The authors assumed a simple linear relationship between measured visible brightness and absorption ranging from zero for no cloud to a maximum of 0.2 for very bright clouds [179].

5.4.2. Marullo-Dalu-Viola model

Marullo, Dalu and Viola (MDV) re-evaluated the GDM model in 1987 using data for the METEOSAT data for the Italian peninsula [186]. Like its predecessor, the MDV model also considered clear sky and cloudy conditions separately. However, the terminology "standard atmosphere" and "real atmosphere" were used in place of clear sky and cloudy sky respectively.

The MDV "standard atmosphere" model was similar to the GDM clear sky model except for the addition of a reflecting non-absorbing layer which accounted for the presence of aerosols in the atmospheric column. The MDV RTM was adapted from Schmetz [187] and required information regarding the temperature profile of the atmosphere, water vapor content and a three-layer aerosol column which describes regional clear sky conditions. However, like the GDM model the MDV model did not account for the variation of surface albedo with solar zenith angle. Planetary albedo for a standard atmosphere was approximated though the use of regional clear sky data and assumed to be uniform for the region, varying only with solar zenith angle.

Any significant deviation from the standard atmosphere model was assumed to be a consequence of atmospheric particle loading. The atmospheric loading in the real atmosphere was resolved by a thin reflecting non-absorbing layer assumed to be higher than the particles responsible for scattering in the standard atmosphere.

5.4.3. Möser-Raschke model

Möser and Raschke (MR) also used METEOSAT images to estimate ground level irradiance [188], but the authors used the RTM developed by Kerschgens [189] which was more complex than the previous models. Improvements to the MR method include the addition of more parameters to accurately describe the state of the atmosphere and infrared data which is used to estimate the cloud top height.

Input parameters include the solar zenith angle, cloud top height, optical depth of the clouds, terrestrial albedo, boundary

layer structure, climatological profiles of temperature, pressure, humidity, ozone concentration and cloud droplet size distribution. One significant result of this model was the demonstration that clouds, rather than aerosols, have the greatest impact on irradiance reaching ground level.

5.4.4. Dedieu-Deschamps-Kerr model

Unlike the previous modes, the model developed by Dedieu, Deschamps and Kerr (DDK) in 1987 [190] used a single equation valid for both clear and cloudy conditions. This was accomplished by considering a clear sky model in conjunction with a model which considered only the effect of clouds on solar irradiance. The assumption was made that the other effects, such as aerosol loading was constant.

Input parameters include a sky transmissivity factor, which accounted for Mie and Rayleigh scattering as well as gaseous absorption using the formulae of Lacis and Hansen [191] together with the RTM of Tanrè [185], and planetary and terrestrial albedo determined from with the METEOSAT radiometer data. Multiple reflections between the cloud base and the ground were assumed to behave isotropically. It should be noted that as a consequence of uniformity of the aerosol content in both the clear sky and cloudy conditions the model treats an unusually strong concentration of aerosols as a cloud [177].

5.5. Statistical satellite models

Statistical satellite models are based on regressions between solar irradiance measured by pyranometers at the ground level and simultaneous digital counts provided by satellite based instrumentation. Rather than use RTMs, a number of independent parameters are included in various regression equations: solar zenith angle, cloud cover index, atmospheric transmissivity, current brightness, minimum brightness and maximum brightness of each pixel. As pointed out in Ref. [177] there are two main difficulties which arise when comparing satellite and ground data. The first is given by the errors associated with the localization of the ground based pyranometer sites on the satellite images. The second arises from the fundamental difference in the measurement technique: satellite data are instantaneous “snapshots” over a small solid viewing angle while ground data have been historically integrated over time (typically an hour) over a large solid viewing angle (2π). Several authors have suggested solutions to these problems, including the incorporation of more pixels in the definition of target areas as well as enhancing satellite resolution.

5.5.1. Hay-Hanson model

One of the simplest statistical satellite models was developed by Hay and Hanson (HH) in 1978 [192]. The model was developed for the Global Atmospheric Research Program’s Atlantic Tropical Experiment to generate maps of the shortwave radiation (0.55–0.75 μm) reaching the surface of the ocean. The HH model is based of a statistical linear regression of the clearness index and atmospheric absorptivity:

$$K_t = a - b\alpha_a. \tag{111}$$

Substituting Equations (7) and (111) into (6) and solving for I_t yields,

$$I_t = K_t I_0 \cos(\theta_t) = (a - b\alpha_a) I_0 \cos(\theta_t). \tag{112}$$

Hay and Hanson [192] originally determined regression coefficients a and b as

$$a \simeq 0.79 \quad b \simeq 0.71$$

These values were later re-evaluated by Raphael and Hay [180] to be

$$a \simeq 0.788 \quad b \simeq 1.078$$

which gives a better agreement with their dataset.

It has been pointed out [177] that this relationship fails under unusually high surface albedo which results from a snow- or ice-covered surface. In addition, despite what has been mentioned about statistical methods, this approach requires the calibration of reported digital satellite counts i order to determine visible radiance.

5.5.2. Tarpley & Justus-Paris-Tarpley models

Tarpley used a set of coincident satellite and ground pyranometer data sets taken by the National Environmental Satellite Data and Information Services (NESDIS) and the Great Planes Agricultural Council over the central U.S. in late 1970s [193]. This study made use of statistical regressions against measurements from GOES VISSR. Three separate cases were considered based on the value of the cloud index defined by Tarpley as

$$n = \frac{0.5N_2 + N_3}{N}, \tag{113}$$

where N is the total number of pixels included in the target area, and N_2 and N_3 are the number of pixels in partly cloudy and cloudy conditions respectively. The Tarpley regression model was defined as

$$I = \begin{cases} a + b\cos(\theta) + cK_t + dn + e\left(\frac{B_m}{B_0}\right)^2 & \text{if } n \le 0.4 \\ a + b\cos(\theta) + cn\left(\frac{B_{cld}}{B_n}\right)^2 & \text{if } 0.4 < n < 1, \\ a + b\cos(\theta) + c\left(\frac{B_{cld}}{B_n}\right) & \text{if } n = 1 \end{cases} \tag{114}$$

where B_m is the mean target brightens, defined as the mean brightness of a 7×6 pixel array, B_{cld} is the mean cloud brightness, estimated through an average of the brightness values of all the pixels in the target area brighter than a specified threshold, and $B_n = B_0(\theta = 45^\circ, \phi_s = 105^\circ)$ is the normalized clear brightness which is a special case of the clear brightness B_0 obtained from the following regression,

$$B_0 = a + b\cos(\theta) + c\sin(\theta)\cos(\phi_s) + d\sin(\theta)\cos^2(\phi_s).$$

Raphael and Hay [180] also estimated their own regression coefficients for this model which are different from Tarpley’s treatment (see Table 8).

This model was later refined by Jutus, Paris and Tarpley (JPT) [194] for part of the Agriculture and Resources Inventory Surveys through Aerospace Remote Sensing (AgRISTARS) program. This new model replaced the three equations of Tarpley’s model with the following single equation,

$$I = I_0\left(\frac{r_0}{r}\right)^2 \cos(\theta)\left[a + b\cos(\theta) + c\cos^2(\theta)\right] + d\left(B_m^2 - B_0^2\right), \tag{115}$$

where B_m is again the mean observed target brightness and B_0 is defined by the following relationship,

Table 8
Top: Regression coefficients for the Tarpley model as determined by Tarpley [193] and Raphael and Hay [180], in parenthesis. Bottom: Regression coefficients for the Justus-Paris-Tarpley model as determined by Justus et al. [194].

Model	*a*	*b*	*c*	*d*	*e*
Tarpley ($n < 0.4$)	−809.54 (−195.67)	3646.91 (3722.93)	1155.10 (85.98)	−438.90 (151.10)	−266.78 (−90.86)
Tarpley ($0.4 \leq n < 1$)	−400.79 (−199.30)	3959.34 (4047.97)	−319.13 (−329.30)	− (−)	− (−)
Tarpley ($n = 1$)	−274.73 (−49.80)	3672.04 (2187.16)	−314.10 (−168.80)	− (−)	− (−)
Justus-Paris-Tarpley	0.4147	0.7165	−0.3909	−1.630	−

$$B_0 = \begin{cases} B'_0 & \text{if } B_m \geq B_{\max} \\ w_1 B + 0' + (1 - w_1) B_m & \text{if } B'_0 < B_m < B_{\max} \\ B_m & \text{if } B'_0 - 2 < B_m \leq B'_0. \\ w_2 B + 0' + (1 - w_1) B_m & \text{if } B_{\min} \leq B_m < B'_0 - 2 \\ B'_0 & \text{if } B_m < B_{\min} \end{cases}$$

As before, the authors in Ref. [194] assumed that the brightness for clear sky conditions B'_0 and the measured target mean brightness B_m were known. The weights w_1 and w_2 are values between 0 and 1 which were empirically determined. Each of the cases above approximates various conditions of the atmosphere. The first and fifth cases correspond to the likely presence of clouds and the insufficient scene illumination for radiation forecasts respectively. Each of these cases leaves the clear brightness unaltered. The second case allows for seasonal variation in the clear brightness due to snow- or ice-cover. The third case is to account for clearer than normal days while the fourth case allows for the removal of erroneous effects from the satellite image on B_0 [177]. Regression coefficients for both the Tarpley model and the JPT model can be found in Table 8.

5.5.3. Cano-HELIOSAT model

Cano developed a model for the French HELIOSAT project in 1982 which used visible band METEOSAT data [195]. The Cano-HELIOSAT model proposes a simple linear relationship between the clearness index K_t and the cloud index n_t at the same point in time and space. This is accomplished by considering local values of K_t and n_t at each pixel as

$$K_t(i,j) = A(i,j) n_t(i,j) + B(i,j), \tag{116}$$

where A and B are matrices of regression coefficients [14]. The cloud cover index was defined as

$$n_t(i,j) = \frac{\rho_t(i,j) - \rho_0(i,j)}{\rho_c - \rho(i,j)}, \tag{117}$$

where ρ_t is the measured ground albedo, ρ_0 is the reference ground albedo and ρ_c is the average albedo of the top of the clouds. The reference ground albedo was calculated using Bourges model [196] and a recursive procedure which minimized the variance of the errors of the clear sky model.

Refinements to the Cano-HELIOSAT model include use of the ESRA clear sky model to correct the estimation of the terrestrial and atmospheric albedos by Rigollier et al. [197]. These corrections were subsequently used to derive the following relationship between the cloud index n_t and a clear sky index k_t,

$$k_t = \begin{cases} 1.2 & \text{if } n_t < -0.2 \\ 1 - n_t & \text{if } -0.2 \leq n_t < 0.8 \\ 2.0667 - 3.6667 n_t + 1.6667 n_t^2 & \text{if } 0.8 \leq n_t < 1.1 \\ 0.05 & \text{if } n_t \geq 1.1. \end{cases}$$

More recent developments of the Cano-HELIOSAT model include consideration of the three dimensional structure of cloud in the determination of the cloud index [198], modification of the previous k_t–n_t relationship to include moments of the cloud index distribution [199], corrections for non-Lambertian reflectivity and the backscattering of clouds [200], and integration of the SOLIS-RTM platform [12].

5.5.4. Perez Operational model

One of the most widely used statistical satellite models is the operational model of Perez [201]. The Perez model uses a modified version of Kasten's clear sky model (see Section 2.3.8) which defines a Linke turbidity coefficient independent of airmass [10]. The model also allows for the modification of the algorithm based on real time measurements of snow- or ice-cover as well as the correction of sun-satellite angle effects for each pixel [16].

The model relates hourly global irradiance I_t and cloud index n_t through a simple regression:

$$I_t = I_{\text{clr},\,t} f(n_t) \left[a I_{\text{clr},\,t} + b \right], \tag{118}$$

where $f(n_t)$ is a fifth order polynomial of the cloud index given by

$$f(n_t) = c_5 n_t^5 + c_4 n_t^4 + c_3 n_t^3 + c_2 n_t^2 + c_1 n_t + c_0. \tag{119}$$

Values of the coefficients, as calculated by Perez, are given in Table 9. This model was also modified by Perez and Ineichen to forecast DNI from GHI forecasts provided by the operational model as well as corrections for locations presenting complex arid terrain [202].

6. Numerical Weather Prediction (NWP)

As discussed in the introduction of Section 5, the sparse nature of ground based pyranometer measurement stations limits the coverage and applicability of traditional time-series based forecasting models. Numerical Weather Predictions (NWPs) provide yet another alternative to a national or global scale ground based monitoring network. The development of NWPs began in the early 20th century with a paper written by Norwegian scientist Vilhelm Bjerknes titled, "The Problem of Weather Forecasting, Considered from the Viewpoints of Mechanics and Physics" [203]. Bjerknes knew that successful forecasting relied on sufficiently accurate knowledge regarding the state of the atmosphere at some initial time and the physical laws which governed the transition of the atmosphere from one state to another. This argument is still valid today and is at the heart of the NWP methodology. First a physical domain is chosen, then discretized at a desired resolution, and

Table 9
Regression coefficients for the Perez Operational Satellite Model [16].

a			*b*		
0.0001			0.9		
c_0	c_1	c_2	c_3	c_4	c_5
1	−0.58	−2.36	6.22	−6.2	2.36

physical laws of motion and thermodynamics are then numerically solved on the discrete spatial grid (see Fig. 15). It is the time marching of solutions to the physical and thermodynamic equations which provide the basis of NWP forecasts. However, Bjerknes recognized that the difficulty stems from the need to numerically solve a system of non-linear partial differential equations.

Four years after the work of Bjerknes, Austrian meteorologist Felix M. Exner produced reasonably accurate forecasts using an extremely simplified model based on the assumption of geostrophically balanced flow. Exner summarized his results in a textbook published in 1912 [204].

Soon after the work of Bjerknes and Exner, Lewis F. Richardson attempted to numerically solve weather forecast equations between 1916 and 1922 [205]. Richardson, as a proof of concept, computed a 6 h barotropic forecast by hand [205]. However, the method proved to be unsuccessful, resulting in a prediction two orders of magnitude too large. These errors were attributed to the imbalances in the initial data that were chosen [206–208]. In addition, the hand calculations themselves took longer than the forecast horizon of 6 h. Still, in 1922, Richardson published his findings and noted that $32 \times 2000 = 64{,}000$ computers would be needed to carry out computations faster than the weather advanced [205].

In 1928, Courant, Friedrichs and Lewy (CFL) studied the application of finite differences to weather forecasting equations and formulated the CFL condition which imposed a restriction on appropriate time steps given a spatial discretization [209,210]. The CFL condition, from a physical point of view, suggests that the maximum allowable time step must be on the same order of magnitude as the relevant time scales for the physical process in order to ensure stability in the integrations of the physical model; a necessary condition. When considering advection, for example, the maximum time scale is given by the ratio of the minimum distance between grid-points divided by the maximum velocity in the flow field [206]. As a result, as the spatial resolution increases, the time scale of the physical process and the maximum allowable time step decrease [211,212].

Furthermore, the systematic simplification of governing equations into a consistent set was originally facilitated through the development of filtered models by J. G. Charney and the quasi-geostrophic assumption [213]. Charney's work, in conjunction with the work of A. Eliassen, improved the understanding of atmospheric motions and allowed for some of the first successful computer assisted numerical weather forecasts in the 1950s [214]. These forecasts were based on the Barotropic Vorticity Equation (BVE), given by

$$\frac{\partial \zeta_b}{\partial t} + \nabla_h \cdot \left\{ (\zeta_b + f)\mathbf{v}_\psi \right\} = 0 \tag{120}$$

$$\zeta_b \equiv \nabla_h^2 \psi_b \tag{121}$$

$$\mathbf{v}_\psi \equiv \mathrm{k} \times \nabla_h \psi_b \tag{122}$$

where ζ_b is the barotropic vorticity, ∇_h is the two-dimensional horizontal gradient, ∇_h^2 is the two-dimensional horizontal Laplacian, $\mathbf{v}_\psi$ is the purely rotational two-dimensional wind velocity, ψ_b is the barotropic streamfunction, k is the unit vector normal to geopotential surfaces, and f is the Coriolis parameter defined by ([206])

$$f \equiv 2\Omega \sin \phi. \tag{123}$$

The use of the BVE and the Electronic Numerical Integrator and Computer (ENIAC) at Aberdeen, Maryland in 1950 is widely regarded as the first computer assisted NWP model [206,215].

Soon after the successful application of the BVE at Aberdeen, NWP development began to grow in popularity with improved understanding of baroclinic instability through the work of J. G. Charney [216], E. T. Eady [217] and N. A. Phillips [218], refinement of the primitive equations in the work of J. Smagorinsky [219] and K. Hinkelmann [220,221], establishment of general circulation modeling by N. A. Phillips [222] and J. M. Lewis [223], and a great body of work on spectral methods and the transform method [224–228], semi-implicit time differencing [229–231] and atmospheric initialization [232–237]. What followed this early NWP was a vast number of progressively sophisticated approaches. These developments, facilitated through escalating computational power, allowed for increasingly higher resolutions and larger domains to be covered. As their scope is too broad to be covered here see Refs. [205,206,238,239] for more information.

Modern NWP models may be divided into two main categories: global or regional, depending on the domain which is used (see Fig. 16). At the current stage in their development, NWPs are unable to predict the precise position and extent of cloud fields. Their relatively coarse spatial resolution, typically on the order of 1–20 km (see Table 10) renders NWP models unable to resolve the micro-scale physics that are associated with cloud formation. As a result, cloud prediction inaccuracy is among the largest sources of error in an NWP based solar forecast. Therefore, examination of fields explicitly predicted by the NWP are used for the diagnoses of atmospheric features such as clouds and precipitation. For example, saturation (or near-saturation) of the atmosphere is typically required for the formation of clouds. During post-processing of the NWPs' solution field, locations near atmospheric saturation are associated with cloud formation. This results in a cloud formation probability at best, which is inherently problematic for solar forecasts.

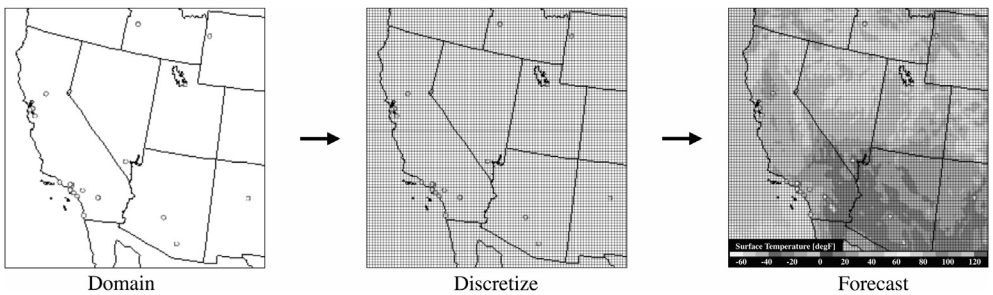


Fig. 15. Basis of all Numerical Weather Prediction (NWP models). First a domain is defined. Secondly, the domain is spatially discretized at a desired resolution. Finally, the NWP predicts desired information by solving equations of motion and thermodynamics. The forecast shown is an 18 h ahead prediction of surface temperature by the NOAA RAP model using the CONUS domain on June 1, 2012.

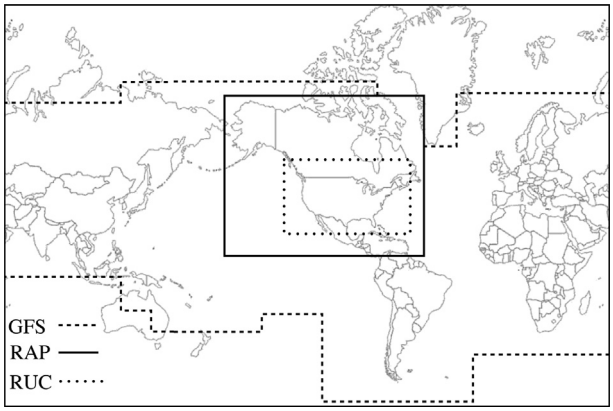


Fig. 16. Domains of the Global Forecast System (GFS), RAPid refresh (RAP) and Rapid Update Cycle (RUC) models. As a general rule, as the area of the spatial domain decreases the resolution of the NWP (both spatially and temporally) typically increases. This trend is a result of limits on available computing power and not a characteristic of the techniques themselves.

Nonetheless, NWPs provide many benefits over previously discussed forecasting methods. Among these is a relatively long time horizon of 15–240 h (see Table 10). Information regarding the propagation of large scale weather patterns is obtained through the regional and global modeling of atmospheric physics and it has been shown that NWPs provide more accurate forecasts than satellite based methods for time horizons exceeding 4 h [240,241]. Accordingly, NWPs provide the most attractive option for medium- to long-term atmospheric forecasting.

6.1. Global Forecast System (GFS)

One of the most well known global NWP models is the Global Forecast System (GFS). The GFS model is run by NOAA every 6 h and produces forecasts out to 384 h (16 days) in advance on a 28 km × 28 km grid for the global domain [242]. The GFS loop time steps are 6 h up to 180 h (7.5 days), then change to 12-h time steps out to 384 h (16 days). In addition to the 28 km × 28 km horizontal discretization, the GFS models 64 vertical layers of the atmosphere. The RTM of the GFS accepts as inputs predicted values of a fully three dimensional aerosol concentration field, predicted values of a two dimensional (horizontal) H_2O, O_2 and O_3 concentration field as well as a constant two dimensional (horizontal) CO_2 field. The GFS model also calculates wavelength specific attenuation of both upwelling and downwelling diffuse irradiances through a sophisticated scattering/absorbing scheme [243]. It should be noted that the radiant flux attenuation is dependent on H_2O phase, temperature and particle size which makes the GFS sensitive to temperature errors.

Table 10
Comparison of various NWP models.

Name	Resolution	Layers	Update period	Time horizon	Time step
GFS	28 km	64	6 h	180 h	6 h
–	–	–	–	384 h	12 h
RUC/RAP	13 km	50	1 h	18 h	1 h
NAM	12 km	60	6 h	96 h	6 h
HRRR	3 km	50	1 h	15 h	15 min
WRF	≥1 km	User Specific	User specific	User specific	User specific

6.2. Regional NWP models

Unlike global NWP models, regional NWP models simulate only a sub-domain of the global space (see Fig. 16). Regional models in the U.S. include the Rapid Update Cycle (RUC), RAPid refresh (RAP), North American Mesoscale (NAM) model, High Resolution Rapid Refresh (HRRR) and the Weather Research and Forecasting (WRF) models. Details of each are discussed in the following sections.

6.2.1. Rapid Update Cycle (RUC)/RAPid refresh (RAP) models

The RUC was a NOAA/NCEP (National Centers for Environmental Prediction) operational NWP model until May, 2012. RUC produced hourly updated 13 km × 13 km horizontally resolved forecasts with 50 atmospheric layers out to a time horizon of 18 h. The RUC loop time steps are 1 h from time of analysis out to 18 h. The RUC possessed a wavelength independent model for the absorption/scattering of radiation by water vapor only. Other atmospheric gasses and aerosols were neglected. The RUC also assumes a simplified Rayleigh scattering model which failed to capture the complex relationship between scattering intensity and radiation wavelength. In addition, only downwelling irradiance was attenuated which sometimes leads to the underestimation of diffuse irradiance due to backscattering [244].

As of May 1, 2012 the RUC was replaced with the RAPid refresh (RAP) model as the next-generation version of the NCEP hourly cycle system. The RAP model possesses the same spatial and atmospheric resolution (12 km × 12 km, 50 layers) but is based on a new rapid update configuration of the WRF model (see Section 6.2.4). As a result, the RAP benefited from the ongoing community improvements to the WRF. The domain of the RAP is also significantly larger than the previous RUC and was expanded from the CONtinental United States (CONUS) region to include Alaska as well (see Fig. 16).

6.2.2. North American Mesoscale (NAM) model

The North American Mesoscale (NAM) model is the NCEP's primary mesoscale environmental modeling tool. NAM produces 12 km × 12 km horizontally resolved forecasts with 60 atmospheric layers out to a time horizon of 96 h over North America and is updated four times daily. The NAM model loop time steps are 6 h from the time of analysis out to 84 h (3.5 days). The NAM model uses predicted water vapor concentrations, seasonally varying but zonally constant O_3 concentrations and constant CO_2 concentrations. Aerosols are not explicitly considered except for a top of the atmosphere adjustment, which is not particularly troublesome with the exception of regions with high levels of time varying aerosol concentrations. Wavelength specific attenuation of both upwelling and downwelling fluxes is accounted for.

6.2.3. High Resolution Rapid Refresh (HRRR) model

The High Resolution Rapid Refresh (HRRR) model is an NOAA operated, experimental, hourly updated, 3 km × 3 km resolution atmospheric model. The HRRR was previously only nested over the eastern 2/3 of the continental United States, but as of June 2009 coverage was expanded to the CONUS region similar to the former RUC (see Fig. 16). The HRRR model uses the 13 km resolution RUC/RAP for its initial conditions and is updated hourly. Benefits of the HRRR include the increased resolution and frequent updates which allow for shorter timescale predictions.

6.2.4. Weather Research and Forecasting (WRF) model

Some of the NWPs discussed were based on a version of the Weather Research and Forecasting (WRF) model which was created through a partnership between NOAA and the National Center for Atmospheric Research (NCAR) in 2004. The WRF has, since its

introduction, seen increased applicability in both research and operational communities. WRF software is supported by ongoing efforts including workshops and on-line documentation. One of the main goals of the WRF model is to advance mesoscale atmospheric prediction by promoting closer ties between research and operational forecasting communities. The WRF is flexible by design and intended for a wide variety of forecasting applications with a priority on spatial resolutions ranging from 1 to 10 km.

7. Local sensing

7.1. Sky imagers

Both NWP and satellite imaging techniques lack the spatial and temporal resolution to provide information regarding high frequency fluctuations of solar irradiance. An alternative is provided through ground based imaging of local meteorological conditions. An instrument used to image the hemispherical sky is called a "sky imager" or a "total sky imager". An example of a sky imager is the Total Sky Imager (TSI) manufactured by Yankee Environmental Systems shown in Fig. 17, and an example of a TSI image taken for the Solar Forecasting Initiative at the University of California Merced is shown in Fig. 18. This model was not developed for solar forecasting applications and has some limitations in the image resolution and the presence of the shadow band that blocks a significant portion of the sky. Other devices such as the one developed at the Leibniz Institute of Marine Sciences at the University of Kiel [245] or the one developed by Cazorla et al. [246] were designed to overcome these limitations.

These sky imagers have been used to model irradiance at the ground level. For instance, Kalisch and Macke [245] used the images collected on four transatlantic research cruises to optimize parameterizations of solar downward radiation at the sea surface. Martínez-Chico et al. [247] used the sky imager to characterize cloud cover and prevailing sky situations in the south of Spain in order to assess the location's suitability for sitting solar energy systems. Similarly Antón et al. [248] quantified the very short-term variability of the total solar irradiance in Granada, Spain and correlated the variability with the cloud cover obtained from a sky imager.

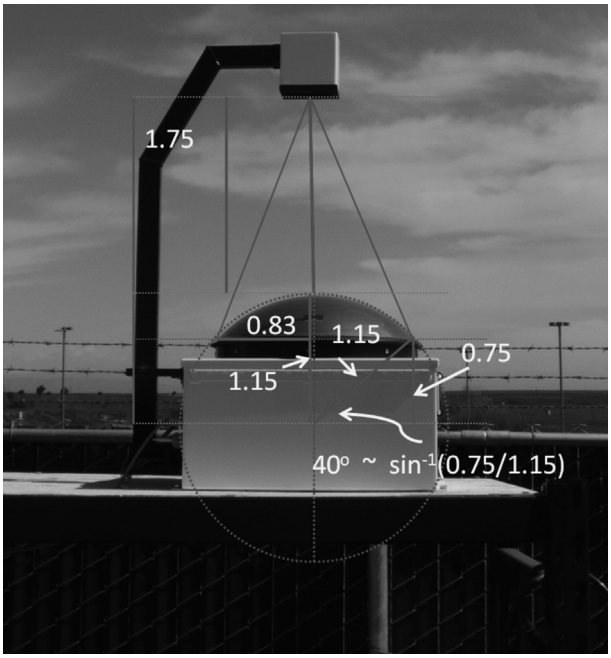


Fig. 17. TSI-880 from Yankee Environmental Systems. The instrument consists of a hemispherical mirror with a downward pointing CCD camera located above it. The mirror is outfitted with a Sun tracking shadow band that protects the optical sensor from the effects of solar reflection. Adapted from Ref. [254].

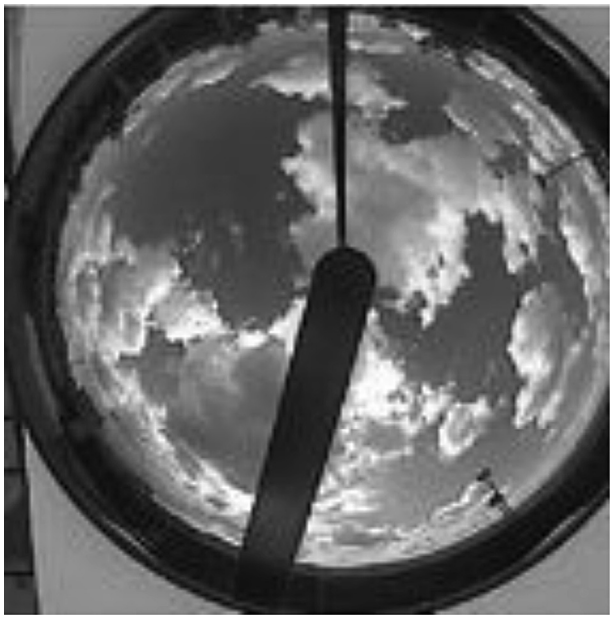


Fig. 18. Image taken with a TSI by Coimbra and collaborators at the University of California Merced on June 1, 2011. The rotating shadow band can be seen near the bottom left hand side of the image and the support for the CCD camera can be seen at the top of the image.

One of the key issues when using a sky imager is extracting cloud cover information from the images. Most of the cloud classification algorithms applied to sky images result into a binary mask (cloud/no-cloud) that cannot differentiate between thin and opaque clouds. Several researchers have tried to correct this problem. For example, Ghonima et al. [249] developed a classification algorithm that compares the pixels' red to blue ratio against the values from a clear sky library. Additionally with the aid of the aerosol optical depth measured by an AERONET photometer they added a haze correction factor to the classification model to account for variations in the level of aerosol. Similarly, Calbó and Sabburg [250] used the sky images as inputs to an automatic classification algorithm that distinguishes between different types of sky conditions: clear, low cumuliform clouds, stratiform clouds, etc. The classification is based on features such as statistical measurements of image's texture and its Fourier transform.

Besides using sky images to monitor and characterize sky conditions, researchers have used the sky images for the forecasting of solar irradiance. The methodology for processing ground based images is mostly similar to satellite based techniques. Projections of observed solar radiation conditions are based on immediate measured history while the position and impact of clouds is deduced from their motion. In the case of TSIs a CCD camera image is digitally processed in order to detect locations of the sky covered by clouds. The cloud image is then propagated forward in time resulting in a forecast. Chow et al. used images from a TSI to produce intra-hour forecasting of GHI of a distributed network of point sensors located at the University of California, San Diego [251]. Cloud fields were propagated forward in time using motion vectors generated by cross-correlating two consecutive images. The authors suggest that TSI images are useful for prediction of GHI on time horizons out to 15 min [251].

Crispim et al. [252] used AI techniques and cloudiness indices obtained from pixel classification of TSI images. These indices were used as inputs to ANN, optimized with a GA, for prediction of solar irradiance. Ferreira et al. [253] created their own portable sky imager and created ANN models optimized via a multi-objective genetic algorithm to predict GHI, cloudiness and temperature for forecasting horizons ranging from 5 min out to 4 h. Coimbra and Marquez also made use of a TSI for intra-hour DNI forecasts rather than GHI [254]. The work in Ref. [254] focused on detailed image processing procedures along with Particle Image Velocimetry (PIV) to determine grid cloud fractions in an area of interest. In agreement with the work of Chow et al. Coimbra and Marquez showed that TSIs show promise for forecasting horizons out to 15 min with the lowest error associated with time horizons of 5–6 min.

It should be noted that while these sky imaging-based methodologies make use of local meteorological information enabling intra-hour forecasts, their time horizon is restricted to approximately 30 min due to residence time of clouds over the field of view. One possible approach to extend the time horizon of ground based measurements is to distribute an array of imagers so that more information regarding local cloud fields is obtained. However, the relatively high cost associated with the TSI and the dynamic nature of local cloud fields, which may limit the correlation of successive images, pose difficulties for current ground based imaging methodologies. In addition to an upper bound on the time horizon of the TSI, a lower bound is also imposed. The lower bound is a result of circumsolar glare and limitations introduced by the shadow band, which renders time horizons shorter than 2 min inaccessible [251,254]. Nonetheless, sky imaging-based forecasts will continue to be relevant as a result of their unique temporal and spatial resolution.

7.2. Wireless sensor networks

None of the methods discussed so far possesses intra-minute time horizons. As discussed TSIs are limited to time horizons higher than 3 min, satellite and NWP models typically possess time horizons on the order of 30 min while stochastic and AI methods have not been widely applied to time horizons less than 15 min. Semiconductor point sensors are capable of very high sampling frequencies but fail to correctly characterize the distributed nature of an operational scale PV plant [255]. An alternative has been suggested by A. Cerpa and coworkers at the University of California, Merced [256]. A 1 MW PV array was outfitted with 40 TelosB nodes equipped with low cost solar irradiance sensors (see Fig. 19(a)). The authors proposed a forecasting algorithm which used multiple readings from the spatially distributed network of sensors to compute future values of the distributed power output. The forecasting approach used spatial cross-correlations between sensor nodes which provided forecasts in the range of 20–50 s. Calculated velocities agreed with TSI calculated cloud velocity field over 70% of the time. This work demonstrates the potential of wireless sensor networks as low cost and highly accurate approaches for intra-minute solar forecasting.

7.3. Pyranometer arrays

Two other methods have been developed by J. L. Bosch et al. to estimate cloud speed using radiometric measurements from sensors located at the UC San Diego Solar Energy test bed [257]. While these methods have not been successfully employed as part of a functioning forecast, the approach shows promise for irradiance prediction on intra-minute time horizons. Both methods are based on time-series analysis of GHI measurements using photodiode pyranometers (Li-200Z, Licor, Inc.) logged to a single CR1000

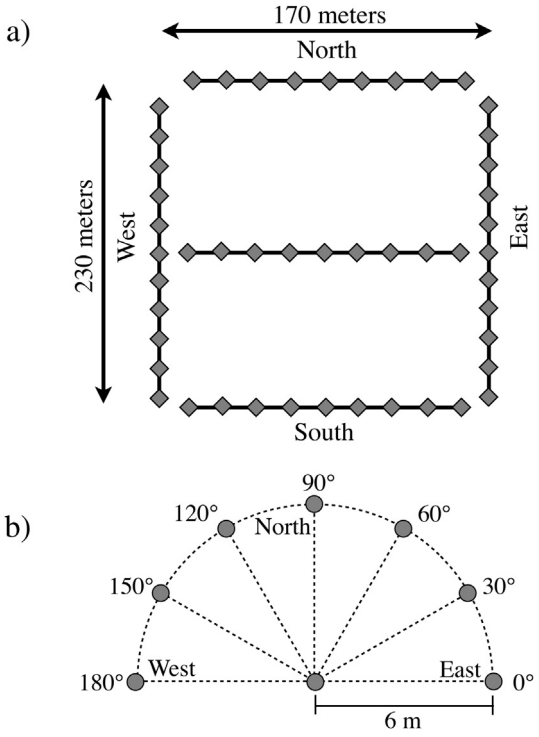


Fig. 19. (a) Birds-eye view of solar array deployment site with relative node locations. Maximum distance between successive nodes is 21 m. The solar array covers roughly 10 acres. Adapted from Ref. [256]. (b) Illustration of pyranometer configuration and principal direction angles which are measured counterclockwise about the origin sensor with a radius of 6 m. Adapted from Ref. [257].

(Campbell Scientific Inc.) datalogger at an acquisition frequency of 20 Hz. The sensors were arranged in a semicircular array with a radius of 6 m (see Fig. 19(b)).

The first method is based on the premise that GHI time-series for a pair of sensors aligned with the direction of cloud movement will be highly correlated. In addition, the largest correlation should occur when the upwind sensor time-series is lagged by the travel time of the cloud edge between the two sensors. As a result, once a time lag is determined from the most correlated pair, the direction of the cloud motion is given by the alignment of the sensors and their distance by

$$v = \frac{D}{t_{\mathrm{MCP}}} \tag{124}$$

where v is the cloud edge velocity, D is the separation between the sensors and t_{MCP} is the time lag between the most correlated pair of sensors. The time lag is calculated through a maximization of cross-correlations between time-series.

The second method used only three of the sensors in Fig. 19(b): the sensors at the origin, 0° and 90°. Consequently, the simple Equation (124) fails because the cloud edge motion directions may not necessarily be aligned with the axis of the sensor setup. Consequently, the most correlated pair method cannot be applied to the three sensors configuration. Bosh et al. developed an alternative method based on the assumptions of a linear cloud edge passing over the array of sensors and a constant cloud edge velocity (see Ref. [257]).

These methods were validated using data from October 20–21, 2011 taken at UCSD. These days were characterized by a large number of cumulus and stratocumulus clouds which resulted in relatively large variability in the GHI time-series. High variability in

the GHI signal is necessary in order to identify the most correlated pair of sensors. It should be noted that overcast and clear periods produce time-series of relatively low variability, resulting in a high correlation across all sensors, rendering the most correlated pair approach non-operational [257]. Cloud speeds determined by the two methods were compared to radiosonde data from the Integrated Global Radiosonde Archive (IGRA) [258], including profiles of pressure, wind speed and direction, and relative humidity in addition to METAR surface weather data which were obtained from the NOAA database [259]. Both methods show good agreement with the radiosonde and METAR data. The method that uses the most correlated pair has the advantages of simplicity, robustness and computational speed while the linear cloud edge method benefits from reduced installation and maintenance costs. However, it is important to note that both methods require high variability in the GHI time-series in order to generate highly correlated pairs of data.

8. Hybrid systems

A Hybrid System (HS) is characterized by a combination of any two or more of the methods described previously. As of late, several HSs have been used to produce high quality solar irradiance forecasts. One of the motivations for the development of hybrid models is that often it is possible to increase the forecasting accuracy by taking advantage of the strengths of each methodology. A typical example for this is the pairing of cloud cover information derived from satellite images with ANNs as implemented in Ref. [260]. Another reason to develop HSs concerns the situation illustrated in Fig. 20. This figure shows the approximate spatial and temporal limitations of each of the forecasting techniques described in this work. Noticeably, none of the individual methodologies is able to span all relevant areas of interest. For this reason, it is clear that a high-fidelity, robust forecasting engine would need to incorporate several of these techniques in order to appropriately forecast on several spatial and temporal resolutions.

Sfetsos and Coonick introduced a new approach for the forecasting of mean hourly GHI using traditional linear stochastic methods, several artificial-intelligence-based techniques alongside the adaptive neuro-fuzzy inference scheme [78]. Zarzalejo et al. used ANNs in conjunction with satellite-derived cloud indices for the forecasting of hourly mean GHI [199]. Mellit et al. developed a hybrid model which was used to predict the daily GHI by combining an ANN and a library of Markov transition matrices [81]. Chaabene and Ammar introduced a medium term dynamic forecasting model for irradiance and ambient temperature which consisted of a neuro-fuzzy estimator based on the meteorological parameters' behaviors during the days before as well as a short term forecast for 5 min ahead based on stochastic models and Kalman filtering [77]. Reikard performed forecasting tests using regressions in logs, ARIMA, and unobserved components models, transfer functions, neural networks and hybrid models [89]. Reikard claimed that the best results were obtained using the ARIMA in logs, with time-varying coefficients. Martín et al. tested AR, ANNs and fuzzy-logic models for application to solar thermal power plants energy production planning [44]. Mellit et al. also developed an adaptive model for predicting hourly GHI and DNI using a dataset of measured air temperature, relative humidity, and direct, diffuse and global horizontal irradiance [261]. The adaptive model's performance was compared against an MLP ANN. Ji and Chee [79] recently developed a new approach that contained two phases used to predict the hourly solar radiation series. In the detrending phase, several models are applied to remove the non-stationary trend lying in the solar radiation series. In the prediction phase, the ARMA model is used to predict the stationary residual series with a time delayed neural network. Voyant et al. [80] proposed a study of the contribution of exogenous meteorological data as time-series to their optimized MLP. To do so they compared different forecasting methods including a persistence forecast, ARIMA reference predictor, an ANN with preprocessing using only endogenous inputs and an ANN with preprocessing using both endogenous and exogenous inputs.

More recently, Marquez and Coimbra [262] developed and validated a medium-term solar irradiance forecast for both GHI and DNI based on stochastic learning methods, ground experiments and the NWS database. A genetic algorithm was used as the input selector for the ANN in order to select the most relevant input patterns. Voyant et al. used a hybrid ARMA/ANN with NWPs to predict hourly mean GHI [263]. Marquez et al. [264] also used cloud indices obtained from a TSI and cloud indices derived from infrared radiometric measurements in order to improve results for hourly forecasts of GHI. Marquez, Pedro and Coimbra have developed a

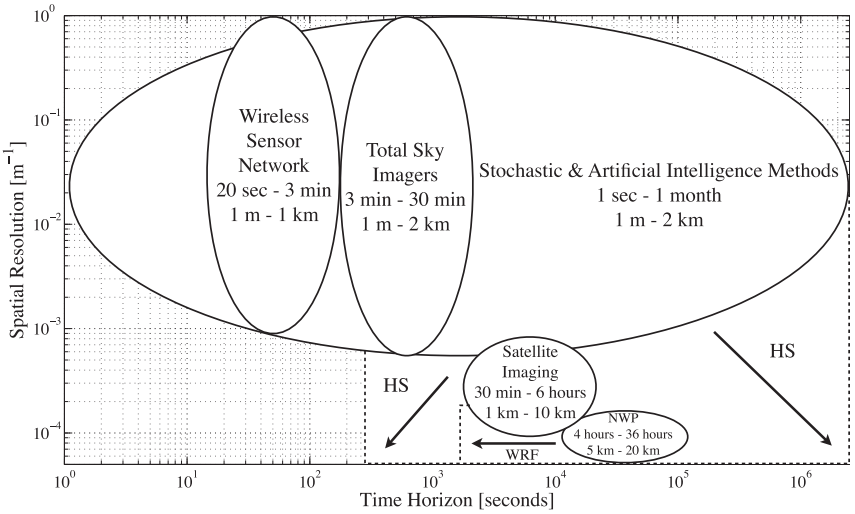


Fig. 20. Comparison of time horizon and spatial resolution. Solid lines indicate current limits of techniques while the dashed lines and arrows indicate the future progress of work. AI techniques will continue to include local, mesoscale and global meteorological data which will allow for both shorter time horizons and greater areas of interest. In addition, recent trends in NWP suggest that shorter time horizons will be available through the development of the WRF models.

forecast based on satellite images and ANNs for time-series predictions of GHI out to 2 h ahead [260].

It is noteworthy in Fig. 20 that Stochastic and ANN techniques have a much broader coverage than other techniques especially in terms of time horizon. This is a direct result from these techniques relying on observational time-series and therefore their time horizon is generally limited only by the sampling frequency used for data acquisition. However, this does not imply that an appropriate ratio between time horizon and temporal resolution should not be conserved. In other words, as the sampling frequency changes, so should the time horizon. For instance, a sampling frequency of 2 Hz would imply the same bound on temporal resolution, therefore a forecast horizon on the order of seconds would be appropriate. Although, if high-frequency data is available it is possible to preprocess the data in order to produce, for example, hourly or daily averages, lowering sampling frequencies and allowing for an effective extension of the temporal horizon. Having said that, the reciprocal argument does not hold: data initially logged at low sampling frequencies, daily averages for instance, cannot be preprocessed in a manner which would make it appropriate for forecasting short time horizons; e.g. 30 min. In conclusion, meaningful time horizons should be of the same order of magnitude as the temporal resolution of the data. The degradation of forecasting quality with increased temporal extrapolation is a common issue in the area of forecasting and is not unique to irradiance prediction. In addition to the limitations imposed on spatial and temporal resolution, it should be noted that the forecasting skill of each of the individual techniques is a function of these resolutions; most notably the time horizon. Fig. 21 shows the performance metric for several techniques from six publications as a function of time horizon. As discussed in Ref. [27], the performance metric s is very well approximated by the ratio of the model's RMSE to the RMSE of the persistence model

$$s \simeq 1 - \frac{\mathrm{RMSE}}{\mathrm{RMSE}_p}. \tag{125}$$

Here it is clear that the forecasting methods are not only rangebound but their forecasting skill is a function of time horizon. As information regarding the RMSE is available in several publications, this approximation of the performance metric was used in Fig. 21. Several methods are shown for time horizons ranging from 30 to 120 min. In this temporal range the hybrid GA/ANN from Ref. [148] performs the best.

9. Performance of forecasting techniques

In this section we review results from a selected group of recent papers that illustrate the state of the art in solar forecasting. The timeline for the works included in this comparison starts in 2008 (for articles published prior to 2008 we refer the reader to the comprehensive review by Mellit and Kalogirou [95]). Table 11 summarizes some these works and lists their main features such as the type of data used, the forecasting horizon and the methodology used. In this section we present results compiled from some of these references grouped by the type of forecasting technique. However, sometimes it is difficult to say whether a paper belongs with one category or the other based on the methodology since most of the recent research in this field seldom relies on a single family of techniques. For those cases we used our best judgment based on the emphasis given to the different techniques employed in the articles. Wherever possible we computed the forecast skill as defined above in Eq. (125) in order to enable a quantitative comparison between the different techniques.

9.1. NWP-based forecast

Even though many NWP models are run daily and available in real-time for virtually any location on earth (after all, they are the basis for most of the day-to-day weather forecasts) there are only a few studies benchmarking NWP radiation forecast against ground truth data. Understandably these studies concentrate on locations in the USA, Spain and Germany as these countries are among the world leaders in installed photovoltaic power capacity.

One of the most thorough benchmarking studies was conducted by Perez et al. (2010) [43]. In that work, cloud cover forecasts from the US National Digital Forecast Database (NDFD) were used to forecast hourly averages of GHI 1–7 days ahead of time. For short-term forecasts, 1–6 h ahead of time, GHI was predicted from extrapolated cloud cover derived from consecutive satellite images. The forecasted values were compared against a year's worth of ground measurements for seven locations in the United States. Mathiesen and Kleissl (2011) [265] used data from the same array of ground sensors to assess the forecasting performance of three NWP models – North American Model (NAM), Global Forecast System (GFS), and European Centre for Medium-Range Weather Forecasts (ECMWF). The models were used to predict intra-day (0–24 h) hourly averages of GHI. Additionally they applied a bias correction function to the NWP forecasts, obtained through Model Output

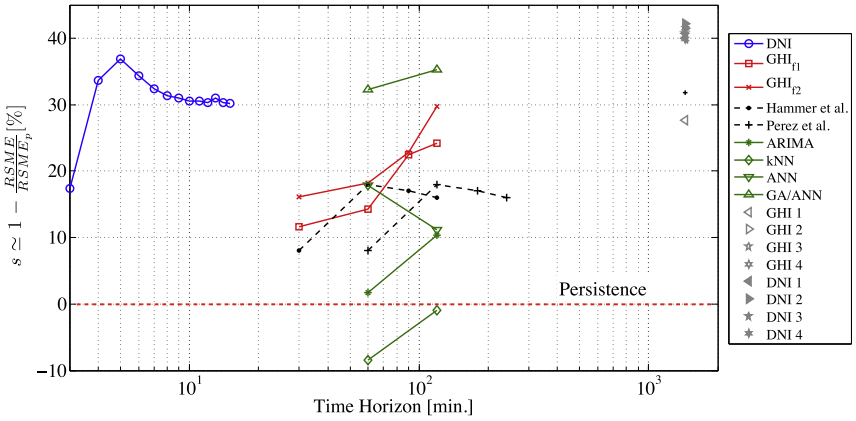


Fig. 21. Comparison of forecasting skill of various methods. Blue [254], red [260], black [43,274], green [148], grey [262]. The comparison is limited to studies which made both RMSE and RMSE$_p$ available which allowed for the approximation of forecasting skill s. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 11
Summary of recent publications in the forecasting of solar irradiance, PV power output and cloud cover.

Authors	Year	Ref.	Forecast variable	Forecast horizon	Method	Exogenous variables	Data
Hocaoglu et al.	2008	[286]	Hourly GHI	1 h	ANNs and preprocessing (hourly GHI is rendered as a 2-D image)	–	1 year of hourly GHI from Iki Eylul, Turkey
Lorenz et al.	2009	[241]	Hourly GHI	1 h to 3 days	ECMWF and post-processing	–	11 months of GHI from 200 weather stations and PV data from 11 systems, Germany
Gordon	2009	[89]	5, 15, 30, and 60 min averaged GHI	5, 15, 30 min and 1–4 h	Regressions in logs, ARIMA, ANNs, and hybrid (ARIMA and ANN)	Humidity, cloud cover, atmospheric turbulence (when available)	Hourly GHI data for 4 years in 3 locations and 1 min GHI for 1 year in 3 locations, USA
Bacher et al.	2009	[270]	Hourly PV power	Out to 36 h	AR, ARX, RX (regressive model with no endogenous variables) and clear normalization	Forecasted GHI (HIRLAM mesoscale NWP from the Danish Meteorological Institute)	1 year of 15-min PV data from 21 rooftops systems, Denmark.
Perez et al.	2010	[43]	Hourly GHI	1–6 h, 1–7 days	Empirical fit between sky cover and measured GHI	Cloud cover forecasts	1 year of hourly GHI from the SURFRAD network, USA
Martin et al.	2010	[44]	Half daily values of GHI	Out to 3 days	AR, ANNs and fuzzy logic	–	Several years of hourly GHI from the Spanish National Weather Service, Spain
Mellit et al.	2010	[261]	Hourly GHI, DNI and DHI	1 h	ANNs and AR model (with parameters that are a non-linear combinations of exogenous variables)	Sunshine duration, temperature, humidity	5 years of 5 min GHI data for Jeddah, Saudi Arabia
Paoli et al.	2010	[138]	Daily GHI	1 day	kNN, ANNs, AR, Markov chains, Bayesian inference and clear sky normalization	–	19 years of daily GHI from Ajaccio, France
Mathiesen and Kleissl	2011	[265]	Hourly GHI	1 h to 1 day	NWP (NAM, GFS, and ECMWF) and bias correction	–	Hourly GHI from the SURFRAD network, USA
Marquez and Coimbra	2011	[262]	Hourly GHI and DNI	Out to 6 days	NWP and ANN and Gamma Test for input selection	NWP weather forecast	1 year of 30 s GHI and DNI data for Merced, USA
Chow et al.	2011	[251]	Cloud cover and GHI	30 s to 5 min	Advection of the 2D cloud map; cloud motion generated from image cross-correlation; GHI obtained from empirical relation with cloud cover	–	Several days of 1 s GHI and 30 s sky-images in San Diego, USA
Ji and Chee	2011	[79]	Hourly GHI	1 h	Hybrid (ANN and ARMA) and clear sky normalization	–	1 year of 10 min data from Nanyang, Singapore
Voyant et al.	2011	[80]	Daily GHI	1 day	ANN, AR, ARIMA; clear sky normalization	Pressure, nebulosity, humidity, wind speed, etc.	9 years of 30 min data from the French Meteorological Organization in Corsica, France
Lorenz et al.	2011	[291]	Hourly PV power	Out to 2 days	ECMWF and post-processing to forecast GHI; explicit physical modeling to convert GHI into PV power	GHI	10 months of regional, hourly PV power data for Germany
Pelland et al.	2011	[267]	Hourly GHI and PV power	Out to 2 days	Canadas GEM and post-processing to forecast GHI; explicit physical modeling to convert GHI into PV power	Temperature	2 years of data: GHI obtained from 10 North-American ground stations, and PV power from 3 PV systems located in Canada
Capizzi et al.	2012	[287]	Daily GHI	1 day	Wavelet recurrent neural networks	Wind speed, humidity, and temperature	1 year of GHI and meteorological data in Catania, Italy
Chow et al.	2012	[288]	10 min PV power	10 and 20 min	ANN	Solar geometry, temperature, irradiance	2 weeks of 10 min PV data and hourly meteorological data for Hong Kong, China
Yap and Karri	2012	[289]	Monthly GHI	1 month	Linear regression, AngstromPrescottPage and ANNs	Temperature, rainfall, evaporation, sunshine hours	12 years of daily GHI and meteorological data for Darwin, Australia
Lara-Fanego et al.	2012	[269]	Hourly GHI and DNI	1, 2 and 3 days	GHI: direct from WRF model; DNI: WRF model outputs and satellite measurements of aerosol and ozone levels	–	2 years of hourly GHI from 4 ground stations, Spain
Mandal et al.	2012	[273]	Hourly PV power	1 h	ANN and wavelet decomposition	GHI, Temperature	1 year of hourly GHI, temperature and PV power from Oregon, USA
Yang et al.	2012	[271]	Hourly GHI	1 h	ARIMA and seasonal decomposition	DNI, DHI, cloud cover indices	NREL's third typical meteorological year (TMY3) for Orland and Miami, USA

Table 11 (*continued*)

Authors	Year	Ref.	Forecast variable	Forecast horizon	Method	Exogenous variables	Data
Pedro and Coimbra	2012	[148]	Hourly PV power	1 and 2 h	ARIMA, ANNs, kNN, clear sky normalization and GA optimization	–	1 year of hourly PV power from a 1 MW solar farm in Merced, USA
Voyant et al.	2012	[263]	Hourly GHI	1 h	ARMA and ANNs	NWP weather forecast	6 years of hourly data from 5 locations in the Mediterranean, France
Marquez et al.	2013	[260]	30, 60, 90, and 120 min averaged GHI	30, 60, 90, and 120 min	ANNs and satellite images	–	1 year of 1 min GHI data for Davis and Merced and hourly NOAA's GOES West satellite images, USA
Marquez et al.	2013	[264]	Hourly GHI	1 h	ANN and 3 cloud cover indices (visible radiation, sky-images and infrared radiation)	Cloud cover, IR radiation, sky-images	2 months of 1 min GHI and IR data, 20 s sky-images for Merced, USA
Marquez and Coimbra	2013	[254]	1 min DNI	3–15 min	Advection of the 2D cloud map; cloud motion generated PIV; DNI obtained from empirical relation with cloud cover	Cloud cover	Several days of 1 min DNI and 20 s sky-images for Merced, USA
Bosch et al.	2013	[257]	Cloud velocity	–	Estimates cloud speed using measurements from 8 GHI sensors	GHI	Several days of very high frequency GHI data for San Diego, USA
Voyant et al.	2013	[290]	Hourly GHI	1 h	Hybrid (ANN and ARMA); season dependent model; seasonal ANN learning; lag-error dependent model; clear sky normalization	Nebulosity, pressure, precipitation	10 years of hourly GHI in 5 locations in the Mediterranean, France
Yona et al.	2013	[292]	Hourly PV power	1 h	Fuzzy theory to predict GHI from NWP weather data; recurrent NN	Cloud cover, humidity, temperature	1 year of data for Okinawa, Japan
Jafarzadeh et al.	2013	[293]	15 min PV power	15–60 min	TSK fuzzy models	Temperature, cloud cover	1 min data from a power plant in Reno, USA

Statistics (MOS), in order to minimize the forecast's bias error. Tables 12 and 13 list the forecasting performance obtained in these two studies. Direct comparison is not possible given that the reported RMSEs in Mathiesen and Kleissl (2011) are not separated by forecasting horizon as in the case of Perez et al. (2010).

One of the main conclusions from these benchmarking studies was that most operational NWP models over-predict irradiance. In a recent study Mathiesen et al. [266] analyzed the predictions from NAM and concluded that this model predicts insufficient cloud cover on the California coast leading to the said irradiance over-prediction. In this work the authors presented and tested some techniques in order to obtain a more realistic cloud cover prediction, namely the improvement of the NAM's initial condition by assimilating realistic could cover retrieved from satellite images, and the optimization of the model's resolution and physics parameterizations. These measures largely mitigate the bias observed in NAM's GHI forecasts and improved the forecasting accuracy both for intra-day and day ahead time horizons.

Motivated by the need to predict rooftop PV power in Germany, Lorenz et al. [241] computed hourly GHI from three-hourly ECMWF forecasts out to 3 days ahead of time. The predicted GHI was then used to forecast PV power output. The forecasting performance for the time horizon of 1 day was computed using data from 200 weather stations that cover the entire German territory. With so much ground data available the authors were able to study the forecasting performance as a function of spatial averaging by assembling forecasts for distributed stations. That analysis was one of the most interesting and original aspects of that article as rarely researchers have access to such a wealth of data. Similarly, Pelland et al. [267] used the Environment Canadas Global Environmental Multiscale NWP model to forecast hourly GHI and PV power for horizons out to 48 h. They applied spatial averaging and bias removal using a Kalman filter to the NWP forecasts in order to increase the predictions' accuracy.

In these studies the researchers could simply download forecasted GHI data from the several services that provide NWP forecasts. That is not possible if one needs to predict DNI or DHI as they are rarely part of the NWP output variables. To fill this gap Breitkreuz et al. [268] proposed a model based on information provided by air-quality models to enable NWP models to compute DNI irradiance. Alternatively, Lara-Fanego et al. (2012) [269] took advantage of the fact that the WRF offers a detailed picture of the future state of the atmosphere to implement a physical post-processing technique to forecast DNI based on the WRF model outputs and satellite measurements of aerosol and ozone levels. In that work, hourly GHI and DNI forecasted with the WRF model

Table 12
Forecast skill for hourly averaged GHI vs. forecasting horizon. Table adapted from results in Ref. [43].

Horizon	Satellite						NDFD					
	1 h	2 h	3 h	4 h	5 h	6 h	2 days	3 days	4 days	5 days	6 days	7 days
RMSE [W/m^2]	87.57	102.14	116.29	127.14	135.71	159.43	157.86	164.86	168.86	174.00	180.57	191.43
Forecasting skill, s	0.08	0.18	0.17	0.16	0.17	0.11	0.06	0.18	0.21	0.23	0.21	0.14

Table 13
RMSE for NWP forecasts before and after MOS correction. Results averaged for the 7 sites analyzed [265].

NWP model	RMSE [W/m^2]	
	Original	MOS-corrected
NAM	134.2	114.1
GFS	110.5	84.6
ECMWF	123.2	106.2

were compared against ground measurements for several locations in southern Spain. The comparison was carried for forecasting horizons of 1, 2 and 3 days. Table 14 lists the average RMSE found in that work and the respective forecast skill. Similarly, Marquez and Coimbra (2011) [262] used NWP forecasts (e.g. sky cover, temperature, probability of precipitation, etc.) from the National Weather Service's (NWS) database as exogenous inputs for ANNs to predict hourly GHI and DNI out to 6 days ahead of time for Merced California. The optimal set of input variables was determined via a genetic algorithm. Table 15 shows the RMSE for the same-day forecast obtained with the best model. The authors demonstrated that hybrid models which combine data generated by NWP algorithms and machine learning algorithms can yield even better results than just post-processing the NWP forecast.

All these studies are invaluable to assess the accuracy when using NWP data in the forecasting of solar irradiation. These works also show that clever post-processing techniques, such as MOS bias correction, the improvement the NWP models' initial conditions and physical parameters or the use of machine learning techniques on top of the NWP models' outputs can improve the forecasting accuracy substantially. Furthermore, they demonstrated that even when no irradiance data is available from NWP models the other weather variables can be very useful as exogenous inputs for irradiance forecasting models. Finally, a common remark present in these articles is that the lack of temporal resolution in the NWP forecasts makes them useless in predicting the intra-hour fast ramp rates observed in solar irradiance.

9.2. Stochastic forecasts

Bacher et al. (2009) [270] tested several stochastic models to predict hourly values of PV power for horizons out to 36 h. The authors used 15 min averaged data from 21 PV rooftops systems in Denmark. The PV data was normalized using a smooth curve fit for the clear sky PV output, and forecasts of the normalized solar power were calculated using AR and ARX models. The ARX models used forecasts of GHI provided by the Danish Meteorological Institute using the HIRLAM mesoscale NWP model as an exogenous variable. The authors proposed a third model (denoted as LMnwp) which is a special case of the ARX model in the way that it includes no lagged values of the endogenous variable (normalized PV time-series). The

Table 14
WRF's GHI and DNI forecast evaluation results as a function of the forecasting horizon for two locations in southern Spain from Ref. [269].

	Horizon [days]	Cordoba			Andasol		
		RMSE [W/m^2]		s	RMSE [W/m^2]		s
		WRF	Persist.		WRF	Persist.	
GHI	1	170.0	190.0	0.11	142.0	154.0	0.08
	2	175.0	202.0	0.13	147.0	180.0	0.18
	3	189.0	199.0	0.05	152.0	191.0	0.20
DNI	1	304.0	364.0	0.16	294.0	292.0	−0.001
	2	311.0	406.0	0.23	290.0	358.0	0.19
	3	319.0	403.0	0.21	305.0	386.0	0.21

Table 15
Same-day forecast performance for the best ANN models obtained in Ref. [262].

	Persistence	Best model	
	RMSE [W/m^2]	RMSE [W/m^2]	Forecast skill, s
GHI	123.10	72.00	0.42
DNI	270.00	156.00	0.42

short- and medium-term forecast skills for the 3 models are listed in Table 16.The authors concluded that for short-term forecast (out to 2 h) the most important input is the lagged PV values, while for longer horizons the NWP GHI becomes more important.

Yang et al. (2012) [271] constructed three ARIMA forecasting models for next-hour GHI including cloud cover effects. The main difference in the three models tested concerns the inputs used: GHI in the first model, DNI and DHI in the second model and cloud cover (CC) in the third model. In the first two models the irradiance inputs were treated with a additive seasonal decomposition in order to account for deterministic irradiance variations. The authors used the third typical meteorological year (TMY3) data from the National Solar Radiation Data Base [272] to estimate the ARIMA models and to validate the forecasting accuracy.

9.3. AI forecasting models

Gordon (2009) [89] ran forecasting experiments for GHI using several data sets, at resolutions of 5, 15, 30, and 60 min. Forecasting tests were run using regressions in logs, ARIMA, and neural networks among other models to produce GHI forecasts 5, 15, 30 min and 1–4 h ahead of time. Hybrid models that combine ARIMA and ANN were also evaluated. The forecasting skill for the different models and forecasting horizons are listed in Table 17. The forecast skill is in reference to a simple forecasting model, which is a regression in logs. The authors argue that taking logs significantly reduces the forecast error at the transition points between sunrise and sunset.

Paoli et al. (2010) [138] used several stochastic and machine learning techniques to predict daily averaged global solar radiation one day ahead of time. Additionally, the authors applied two preprocessing strategies in order to normalize solar radiation. The one first was based on the clearness index and the second one used the clear sky index. Forecasting skills for some of the models developed by Paoli et al. (2010) are listed in Table 18.

Martin et al. (2010) [44] used AR, ANNs and fuzzy logic models to predict half-day values of GHI (averaged GHI from sunrise to noon and form noon to sunset) out to 3 days ahead of time. The forecasting accuracy was determined with ground solar radiation from the Spanish National Weather Service. Forecasting skill versus the time horizon for the different models is shown in Table 19.

Mandal et al. (2012) [273] used an ANN with exogenous variables to forecast the hourly PV power for forecasting horizon of 12 h. In order to improve the forecasting performance the authors applied a wavelet decomposition to the PV time-series, which is then input into the ANN, whose outputs are then passed through a wavelet reconstruction to determine the PV predictions. The substantial improvement obtained with this technique is reported in Table 20.

Table 16
Forecast skill for models proposed in Ref. [270].

	Forecasting skill, s	
	1–6 h	19–29 h
AR	0.27	0.17
LMnwp	0.25	0.36
ARX	0.35	0.36

Table 17
Forecast skill for several models as function of location and forecast horizon from Ref. [89].

Forecast horizon	Clark, Nevada			National wind database			SRRL database		
	ARIMA	NN	Hybrid	ARIMA	NN	Hybrid	ARIMA	NN	Hybrid
5 min	0.19	0.13	0.19	−0.11	−0.012	−0.01	−0.004	−0.026	0.0081
15 min	0.28	0.21	0.28	0.19	0.12	0.19	0.21	0.14	0.21
30 min	0.39	0.28	0.39	0.23	0.034	0.23	0.001	−0.013	0.016
1 h	0.29	0.23	0.29	0.38	0.30	0.38	0.38	0.32	0.38
2 h	0.16	0.12	0.16	0.24	0.23	0.24	0.25	0.22	0.25
3 h	0.13	0.12	0.12	0.15	0.12	0.15	0.16	0.13	0.16
4 h	0.085	0.053	0.085	0.09	0.061	0.088	0.11	0.083	0.11

Pedro and Coimbra (2012) [148] applied several stochastic and AI techniques (ARIMA, kNN, ANN) to predict the 1- and 2-h averaged power output of a 1 MW solar power plant in Merced, California. In that study no exogenous variables were used. The authors created a clear sky model for the power output by smoothing the PV data. Additionally a GA was employed to determine the ANNs' architecture and input selection such that the forecasting performance was maximized. A detailed discussion of the seasonal dependency of the accuracy is presented in the article. Table 21 lists the forecasting skills for the models developed as a function of the forecast horizon.

9.4. Hybrid forecasting models

Marquez et al. (2013) [260] tested hybrid forecasting models that combine information from processed satellite images with ANNs. Their work is an extension to the methods proposed by Hammer et al. (1999,2003) [274,275] to extrapolate the temporal development of the clouds based on satellite images (obtained from the NOAA GOES West). The models in Marquez et al. (2013) were developed to forecast GHI at temporal horizons of 30, 60, 90, and 120 min. The tests were carried out with data gathered at UC Davis and UC Merced in California's Central Valley. Table 22 lists the forecast skill for the persistence forecast and 2 ANN-based models. Input variables for Model 1 include only data derived form satellite images, while Model 2 also includes lagged GHI data in the set of input variables.

10. Other challenges of solar integration

As mentioned in the introduction, achieving higher levels of renewable generation on the electrical power grids requires the solution of several problems simultaneously [276]. Researchers, industry experts and operators have identified a number of major challenges including: 1) unpredictable and steep ramps, 2) making up for errors in the forecasting of renewable resources, 3) intra-hour variability, and 4) over generation in the middle of the night. The first three issues are closely connected to the ability to forecast the resource and power plant output accurately. In fact, a perfect forecast would eliminate these problems entirely. Given that the perfect forecast is impossible to achieve, two solutions to mitigate these problems are currently under investigation: storage and demand response [277,278].

Energy storage is generally expensive, but offers the possibility of a definitive answer to the variability problem [279], but even in this solution accurate forecasting is still likely to be a critical component. Indeed optimized storage systems and load-following as well as resource-following forecasts have been shown to increase the lifetime of batteries by reducing the number of deep discharging events. However, energy storage is a very broad field and covering this topic is beyond the scope of this review.

Demand response programs have been in place for several decades [280]. In general, demand response programs rely on the end-users deliberately altering the use of equipment and systems. In practice, this means changing lifestyle or reducing comfort, or changing operating procedures in exchange for some financial compensation or improved reliability of power supply. With the advent of the smart grid concept the industry is moving towards the concept of automated demand response (autoDR) where users are equipped with automated power management tools that respond to requests to decrease or increase the demand according to commands received from the systems operator. In this way, large consumers of energy are essentially energy reservoirs that can be tapped when increased load or diminished generation is expected elsewhere. Fig. 22 shows a block diagram for a type of automated demand response system. The ISO demand response automated

Table 18
Forecast skill for some of the models in Ref. [138].

Model	RMSE [MJ/m^2]	Forecasting skill, s
Persistence	4.65	–
kNN	4.48	0.04
AR	3.77	0.19
ANN	3.73	0.20

Table 19
Forecast skill for several model from Ref. [44].

Horizon [hours]	AR	ANN	ANFIS
0–12	0.12	0.12	0.11
12–24	0.16	0.17	0.17
24–36	0.18	0.20	0.19
36–48	0.19	0.21	0.20
48–60	0.20	0.23	0.21
60–72	0.21	0.23	0.22

Table 20
Forecast performance from models in Ref. [273].

Model	RMSE[a]
BPNN	2.07
RBFNN	1.35
WT + BPNN	1.35
WT + RBFNN	0.78

[a] Averaging the 8 separated days analyzed there.

Table 21
Forecast skill for the models employed in Ref. [148].

Forecast	Persistence	Forecasting skill, s			
Horizon [hours]	RMSE [kW]	ARIMA	kNN	ANN	ANN + GA
1	107.48	0.02	−0.08	0.18	0.31
2	160.79	0.10	−0.01	0.11	0.35

Table 22
Forecast skill for several models from Ref. [260].

Horizon [min]	RMSE persistence	Forecast skill, s	
		Model 1	Model 2
30 min	84.70	0.12	0.16
60 min	97.40	0.14	0.18
90 min	109.00	0.23	0.23
120 min	118.00	0.24	0.30

server (DRAS) sends commands over the public Internet to the end-user. The demand response controller on the end-user side actuates over power elements such as HVAC, lighting, pumps, etc. essentially reducing the load in a participating facility.

The successful implementation of autoDR programs will make the power grid much more flexible and efficient at the expense of added system complexity. For instance, Fig. 23 shows a situation where the total electricity generation is insufficient for supplying the load demanded from the grid. This is a plausible scenario as the penetration of renewables increases resulting in an increase in the variability on the supply side. The figure inset exemplifies how automated demand response would actuate on the demand side to compensate for the decrease in the generation. In order to determine how much load needs to be shed, the DRAS requires the forecasting of both generation and the load *before* the autoDR event. Therefore, in both storage and demand response scenarios, more accurate forecasts are paramount.

Load forecasting is a different subject than solar forecasting but many of the techniques reviewed in this paper can be and are regularly used for the foresting of power demand. All the stochastic models discussed in Section 3 and the AI techniques discussed in Section 4 are general techniques for the forecasting of time-series and can be readily applied to the forecast of electrical power loads (for more on ADR, please see Ref. [277] and references within).

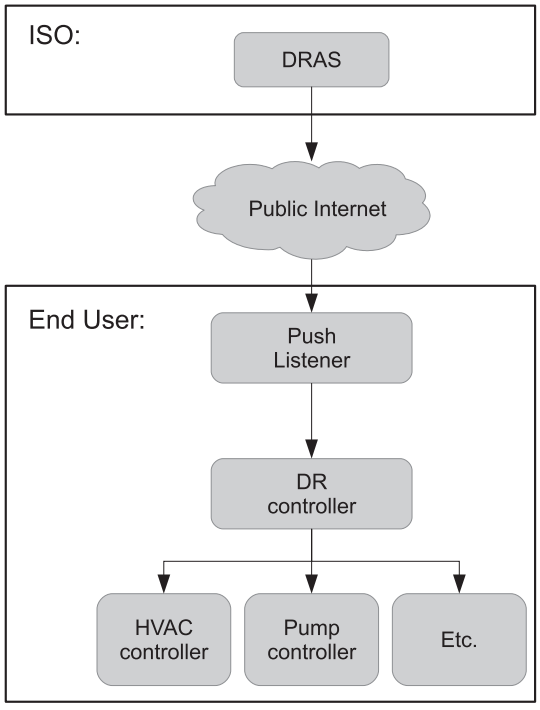


Fig. 22. Block diagram for the automated demand response scenario.

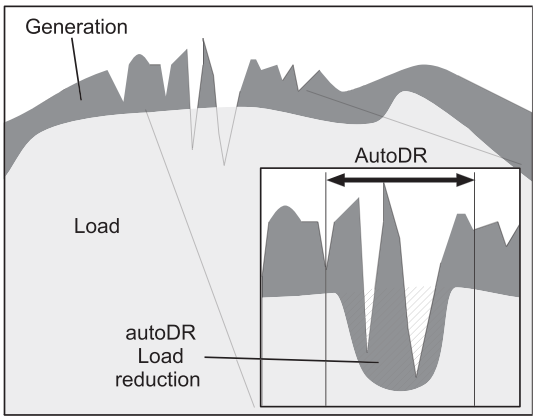


Fig. 23. Automated Demand Response compensating for fluctuations in weather-dependent resources.

11. Conclusion

Solar forecasting methods and applications are covered in this work. From the description of the numerous approaches presented, it is clear that some level of predictive success has been experienced for a number of different spatial and temporal resolutions. Although this research field is incipient, a robust number of techniques and best practices have been developed in the past few years.

Regressive methods, which take advantage of the correlated nature of the irradiance observations tend to work well in both data-poor and data-rich environments. In data-poor environments, one might only have access to historical point sensor or power output data, in which case endogenous stochastic methods can be used such as AR, MA, ARMA and/or ARIMA. On the other hand, in a data rich environment one may have access to a wide number of additional data such as the time evolution of observations of physical quantities. In these environments, exogenous inputs are included through a careful input selection process. These stochastic learning methods have been applied to a wide variety of time horizons ranging from intra-hour to yearly averages as explained in Section 3.

Artificial Neural Network (ANN) modeling offers improved non-linear approximator performance and provides an alternative approach to physical modeling for irradiance data when enough historical data is available. ANNs are generic non-linear approximators that deliver compact solutions for several non-linear, stochastic and multivariate problems. Like regressive methods, ANNs perform well in both data-rich/poor environments and are not typically temporally limited. These techniques have successfully modeled irradiance on intra-hour to yearly time horizons as demonstrated in Section 4.

Satellite imaging provides an alternative to expensive ground based pyranometer networks and allows for the forecasting of irradiance in environments where no other data is available. Elaborate physical RTMs or straightforward regression techniques may be used in conjunction with satellite data to provide irradiance forecasts on time horizons ranging from intra-hour to intra-day (see Section 5). These methods suffer from temporal and spatial limitations due to satellite sampling frequency and limits on spatial resolution of the satellite images.

Numerical Weather Prediction also allows for modeling of the temporal progression of cloud cover in locations were no extensive ground networks are available. NWP models are scalable due to their discrete nature, and can model regional or global atmospheric

evolution. As a result of the large-scale modeling of the atmosphere, NWP models offer the most attractive option for long-term forecasting with time horizons ranging from a few hours out to a couple of days or more as described in Section 6.

For short time horizons, ground based imaging (and recently wireless network systems) appears to be very promising. The use of sky imagers allows for the acquisition of local cloud cover thus providing information on higher frequency fluctuations of irradiance with time horizons ranging from 2 to 20 min in most environments. Intra-minute forecasting is difficult with sky imagers due to image processing time and due to circumsolar glare on optics, but these limitations may be addressable with sky imager designs that are dedicated to solar forecasting. A solution to intra-minute forecasting is provided through the distribution of low-cost irradiance sensors throughout the edges of the solar plant and at spatially correlated locations in order to determine the effect of passing local clouds fields on power output as seen in Section 7.

Recently, various hybrid methods incorporating two or more of the techniques discussed in this work have been presented. These methods offer many advantages over traditional approaches, including the ability to combine stochastic and deterministic forecast. Integration of NWP/satellite models with stochastic learning methods results in higher accuracy forecasts for several time horizons of interest. By the same reasoning, assimilating local meteorological information from sky imagers and wireless sensor networks through stochastic learning methods should result in higher fidelity coverage of intra-hour time horizons. Thus, a complete system in the form of a high fidelity solar forecast engine that spans the entire spectrum of temporal and spatial horizons, from intra-minute to multiple days, and from single point radiometers to continental regions, is possible when multiple data inputs are combined with stochastic learning.

References

[1] Denholm P, Margolis RM. Evaluating the limits of solar photovoltaics (PV) in traditional electric power systems. Energy Policy 2007;35(5):2852–61.
[2] Denholm P, Margolis RM. Evaluating the limits of solar photovoltaics (PV) in electric power systems utilizing energy storage and other enabling technologies. Energy Policy 2007;35(9):4424–33.
[3] Lew D, Piwko R. Western wind and solar integration study. Tech. rep.. National Renewable Energy Laboratories; 2010. Technical Report No. NREL/SR-550–47781
[4] Venkataraman S, Jordan G, Piwko R, Freeman L, Helman U, Loutan C, et al. Integration of renewable resources: operational requirements and generation fleet capability at 20% rps. Tech. rep.. California: ISO; 2010
[5] Rodriguez GD. A utility perspective of the role of energy storage in the smart grid. In: Power and energy society general meeting. IEEE; 2010. p. 1–2.
[6] Kopp G, Lean JL. A new, lower value of total solar irradiance: evidence and climate significance. Geophysical Research Letters 2011;38(L01706):1–7.
[7] Goody RM, Yung YL. Atmospheric radiation: theoretical basis. 2nd ed. Oxford University Press; 1995.
[8] Reno MJ, Hansen CW, Stein JS. Global horizontal irradiance clear sky models: implementation and analysis. Tech. Rep. SAND2012–2389. Sandia National Laboratories; 2012.
[9] Kasten F, Young T. Revised optical air mass tables and approximation formula. Applied Optics 1989;28:4735–8.
[10] Ineichen P, Perez R. A new airmass independent formulation for the Linke turbidity coefficient. Solar Energy 2002;73(3):151–7.
[11] Ineichen P. Comparison of eight clear sky broadband models against 16 independent data banks. Solar Energy 2006;80(4):468–78.
[12] Mueller R, Dagestad K, Ineichen P, Schroedter-Homscheidt M, Cros S, Dumortier D, et al. Rethinking satellite-based solar irradiance modeling: the SOLIS clear-sky module. Remote Sensing of Environment 2004;91(2):160–74.
[13] Rigollier C, Bauer O, Wald L. On the clear sky model of the ESRA – European Solar Radiation Atlas – with respect to the Heliosat method. Solar Energy 2000;68(1):33–48.
[14] Geiger M, Diabate L, Menard L, Wald L. A web service for controlling the quality of measurements of global solar irradiation. Solar Energy 2002;73(6):475–80.
[15] Zarzalejo LF, Polo J, Ramirez L. Gc_model5_irradiance, (Matlab computer program) CD-ROM accompanying [16] 2004.
[16] Badescu V. Modeling solar radiation at the earth surface. Berlin Heidelberg: Springer–Verlag; 2008.
[17] Kasten F. The Linke turbidity factor based on improved values of the integral Rayleigh optical thickness. Solar Energy 1996;56(3):239–44.
[18] Bird RE, Huldstrom RL. Direct insolation models. Transactions of the ASME Journal of Solar Energy Engineering 1980;103:182–92.
[19] McClatchey RA, Selby JE. Atmospheric transmittance from 0.25 to 28.5 μm: computer code LOWTRAN2. Tech. Rep. Environ. Res. Paper No. 427. Airforce Cambridge Research Laboratories; 1972.
[20] Molineaux B, Ineichen P, O'Neill N. Equivalence of pyrheliometric and monochromatic aerosol optical depths at a single key wavelength. Applied Optics 1998;37(30):7008–18.
[21] Berk A, Bernstein LS, Robertson DC. Modtran: a moderate resolution model for lowtran. Tech. rep.. DTIC Document; 1987
[22] Gueymard C. Smarts2, a simple model for atmospheric radiative transfer of sunshine: algorithms and performance assessment. Rep. FSEC-PF-270–95, Tech. rep.. Cape Canaveral, Florida: Florida Solar Energy Center; 1995
[23] Gueymard C. A 2-band model for the calculation of clear sky solar irradiance, illuminance, and photosynthetically active radiation at the earths surface. Solar Energy 1989;43(5):253–65.
[24] Gueymard CA. High performance model for clear-sky irradiance and illuminance. In: Proceedings of the solar, annual ASES conference. Portland, OR: American Solar Energy Society; 2004. p. 251–8.
[25] Kasten F. A simple parameterization of two pyrheliometric formulae for determining the Linke turbidity factor. Meteorologische Rundschau 1980;33:124–7.
[26] Kasten F. Parametrisierung der globaslstrahlung durch bedekungsgrad und trübungsfaktor. Annalen der Meteorologie 1984;20:49–50.
[27] Marquez R, Coimbra CFM. Proposed metric for evaluation of solar forecasting models. ASME Journal of Solar Energy Engineering 2013;135:0110161–9.
[28] Snyder RL. Penman-Monteith (hourly) reference evapotranspiration equations for estimating ET_{os} and ET_{rs} with hourly weather data. Regents of the University of California; 2002.
[29] Duffie JA, Beckman WA. Solar engineering of thermal processes. New York: John Wiley and Sons; 1980.
[30] Allen RG, Smith M, Pereira LS, Perrier A. An update for the calculation of reference evapotranspiration. ICID Bulletin 1994;43(2):35–92.
[31] MESOR. Management and exploitation of solar resource knowledge; European Commission's community research Sixth Framework Programme. Tech. rep.. Sandia National Laboratories; 2002–10
[32] IEA. Solar heating and cooling programme, task 36, subtask a – standard qualification for solar resource products. Tech. rep.. IEA; 2012
[33] Beyer GG, Martinez JP, Suri M, Torres JL, Lorenz E, Muller S, et al. Benchmarking of radiation products. Tech. Rep. 3. Mesor Report D; 2009.
[34] Espinar B, Ramírez L, Drews A, Beyer HG, Zarzalejo LF, Polo J, et al. Analysis of different comparison parameters applied to solar radiation data from satellite and German radiometric stations. Solar Energy 2009;83(1):118–25.
[35] Hoff TE, Perez R, Kleissl J, Renne D, Stein J. Reporting of irradiance modeling relative prediction errors. Progress in Photovoltaics: Research and Applications. http://dx.doi.org/10.1002/pip.2225.
[36] Perez R, Beauharnois M, Hemker K, Kivalov S, Lorenz E, Pelland S, et al. Evaluation of numerical weather prediction solar irradiance forecasts in the US. In: Proc. solar 2011, American solar energy society's annual conference 2011.
[37] Massey Jr FJ. The Kolmogorov-Smirnov test for goodness of fit. Journal of the American Statistical Association 1951;46(253):68–78.
[38] Mills A, Wiser R. Implications of wide-area geographic diversity for short-term variability of solar power. Tech. Rep. LBNL-3884E. Lawrence Berkeley National Laboratory; 2010.
[39] Hoff TE, Perez R. Modeling PV fleet output variability. Solar Energy 2012;86(8):2177–89.
[40] Lave M, Kleissl J. Solar variability of four sites across the state of Colorado. Renewable Energy 2010;35(12):2867–73.
[41] Hoff TE, Perez R. Quantifying PV power output variability. Solar Energy 2010;84(10):1782–93.
[42] Mellit A. Artificial intelligence technique for modeling and forecasting of solar radiation data: a review. International Journal of Artificial Intelligence and Soft Computing 2010;1(1):52–76.
[43] Perez R, Kivalov S, Schlemmer J, Hemker K, Renne D, Hoff TE. Validation of short and medium term operational solar radiation forecasts in the US. Solar Energy 2010;84(5):2161–72.
[44] Martín L, Zarzalejo LF, Polo J, Navarro A, Marchante R, Cony M. Prediction of global solar irradiance based on time series analysis: application to solar thermal power plants energy production planning. Solar Energy 2010;84(10):1772–81.
[45] Landsberg H. Solar radiation at the earth's surface. Solar Energy 1961;5(3):95–8.
[46] Whillier A. Solar radiation graphs. Solar Energy 1965;9(3):164–5.
[47] Swartman R, Ogunlade O. A statistical relationship between solar radiation, sunshine and relative humidity in the tropics. Atmosphere 1967;5(2):25–34.
[48] Swartman RK. Correlation of solar radiation with common parameters in Toronto, Canada. Solar Energy 1971;13(3):345–7.
[49] Lund IA. Relationships between insolation and other surface weather observations at Blue Hill Massachusetts. Solar Energy 1968;12(1):95–106.
[50] Yogi D, Kreith F, Kreider JF. Principles of solar engineering. 2nd ed. Taylor and Francis; 2000.
[51] Walpole R, Meyers R. Probability and statistics for engineers and scientists. New York: Macmillan; 1972.

[52] Wold A. Analysis of stationary time series. Uppsala: Almquist and Wicksell; 1938.
[53] Yule GU. On the time-correlation problem, with special reference to the variate-difference correlation method. Journal of the Royal Statistical Society 1921;84:497–537.
[54] Yule GU. Why do we sometimes get nonsense-correlations between time-series? – A study in sampling and the nature of time-series. Journal of the Royal Statistical Society 1926;89:1–69.
[55] Yule GU. On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. Philosophical Transactions of the Royal Society of London Series A-Containing Papers of a Mathematical or Physical Character 1927;226:267–98.
[56] Box GEP, Jenkins GM, Reinsel GC. Time series analysis: forecasting and control. 4th ed. Wiley; 2008.
[57] Sulaiman M, Hlaing W, Wahab M, Sulaiman Z. Analysis of residuals in daily solar radiation time series. Renewable Energy 1997;11(1):97–105.
[58] Ljung L. System identification: theory for the user. 2nd ed. Prentice-Hall; 1998.
[59] Goh TN, Tan KJ. Stochastic modeling and forecasting of solar-radiation data. Solar Energy 1977;19(6):755–7.
[60] Boileau E. Discussion d'un modéle statistique en météorologie solaire. Revue de Physique Appliquée 1979;14:145–53.
[61] Boileau E, Guerrier B. Comparaison de modéles statistiques saisonniers et non saisonniers en météorologie solaire. La Météorologie 1979;VI:115–30.
[62] Guerrier B, Boileau E, Bernard C. Analyse statistique temporelle de l'irradiation solaire globale quotidienne: modélisation d'une variale réduite á l'aide de modéles statistiques A.R.M.A. Revue de Physique Appliquée 1980;15:93–102.
[63] Guerrier B. Modélisation appliquée á l'étude des entrées-sorties de systémes thermiques: cas des composants linéaires de l'habitat solaire passif [Ph.D. thesis] 1981. Paris.
[64] Brinkworth BJ. Autocorrelation and stochastic modeling of insolation sequences. Solar Energy 1977;19(4):343–7.
[65] Benard C, Wirgin A, Body Y, Gobin D. Caractérisation de la stabilité de l'intensité solaire par l'analyse de séries aléatoires in France et au Pérou. Le Météorologie 1978;VI:53–71.
[66] Boileau E. Use of some simple statistical-models in solar meteorology. Solar Energy 1983;30(4):333–9.
[67] Mustacchi C, Cena V, Rocchi M. Stochastic simulation of hourly global radiation sequences. Solar Energy 1979;23(1):47–51.
[68] Boch G, Boileau E, Benard C. Search for a statistical-model of hourly integrated solar irradiation. Revue de Physique Appliquée 1981;16(5):237–47.
[69] Benard C, Boileau E, Guerrier B. Modeling of the global solar radiant exposure with ARMA processes – application to short-time prediction (hourly), for the calculation of optimal controls in buildings. Revue de Physique Appliquée 1985;20(12):845–55.
[70] Hokoi S, Matsumoto M, Ihara T. Statistical time-series models of solar-radiation and outdoor temperature – identification of seasonal models by Kalman filter. Energy and Buildings 1991;15(3–4):373–83.
[71] Yoshida H, Terai T. An ARMA type weather model for air-conditioning, heating and cooling load calculation. Energy and Buildings 1991;16(1–2):625–34.
[72] Aguiar RJ, Collares-Pereira M. The modeling of daily sequences of hourly radiation through autoregressive methods. In: Clean and safe energy forever. Proceedings of the 1989 congress of the international solar energy society 1989;vol. 3. p. 2107–11. Kobe City, Japan.
[73] Aguiar R, Collares-Pereira M. TAG – a time-dependent, autoregressive, Gaussian model for generating synthetic hourly radiation. Solar Energy 1992;49(3):67–8.
[74] Al-Awadhi S, El-Nashar N. Stochastic modeling of global solar radiation measured in the state of Kuwait. Environmetrics 2002;13(7):751–8.
[75] Mora-Lopez LL, Sidrach-De-Cardona M. Multiplicative ARMA models to generate hourly series of global irradiation. Solar Energy 1998;63(5):283–91.
[76] Moreno-Muñoz A, de la Rosa JJG, Posadillo R, Bellido F. Very short term forecasting of solar radiation. In: Photovoltaic specialists conference, 33rd IEEE 2008. p. 1–5.
[77] Chaabene M, Ben Ammar M. Neuro-fuzzy dynamic model with Kalman filter to forecast irradiance and temperature for solar energy systems. Renewable Energy 2008;33(7):1435–43.
[78] Sfetsos A, Coonick AH. Univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques. Solar Energy 2000;68(2):169–78.
[79] Ji W, Chee KC. Prediction of hourly solar radiation using a novel hybrid model of ARMA and TDNN. Solar Energy 2011;85(5):808–17.
[80] Voyant C, Muselli M, Paoli C, Nivet M-L. Optimization of an artificial neural network dedicated to the multivariate forecasting of daily global radiation. Energy 2011;36(1):348–59.
[81] Mellit A, Benghanem M, Arab AH, Guessoum A. A simplified model for generating sequences of global solar radiation data for isolated sites: using artificial neural network and a library of Markov transition matrices approach. Solar Energy 2005;79(5):469–82.
[82] Suykens JAK, Vandewalle JPL, Moor BLR. Artificial neural networks for modeling and control of non-linear systems. Kluwer Academic; 1996.
[83] Craggs C, Conway EM, Pearsall NM. Statistical investigation of the optimal averaging time for solar irradiance on horizontal and vertical surfaces in the UK. Solar Energy 2000;68(2):179–87.
[84] Santos JM, Pinazo JM, Canada J. Methodology for generating daily clearness index values K_t starting from the monthly average daily value $\overline{K}_t$. Determining the daily sequence using stochastic models. Renewable Energy 2003;28(10):1523–44.
[85] Kärner O. On nonstationarity and antipersistency in global temperature series. Journal of Geophysical Research 2002;107(D20):4415.
[86] Yüerekli K, Simsek H, Cemek B, Karaman S. Simulating climatic variables by using stochastic approach. Building and Environment 2007;42(10):3493–9.
[87] Stanhill G, Cohen S. Solar radiation changes in Japan during the 20th century: evidence from sunshine duration measurements. Journal of the Meteorological Society of Japan 2008;86(1):57–67.
[88] Kärner O. ARIMA representation for daily solar irradiance and surface air temperature time series. Journal of Atmospheric and Solar-Terrestrial Physics 2009;71(8–9):841–7.
[89] Gordon R. Predicting solar radiation at high resolutions: a comparison of time series forecasts. Solar Energy 2009;83(3):342–9.
[90] Perdomo R, Banguero E, Gordillo G. Statistical modeling for global solar radiation forecasting in Bogotá. In: Photovoltaic specialists conference (PVSC), 2010 35th IEEE 2010. p. 2374–9.
[91] Krishnamoorthy CS, Rajeev S. Artificial intelligence and expert systems for engineers. CRC Press; 1996.
[92] Crevier D. AI: the tumultuous search for artificial intelligence. New York, NY: Basic Books; 1993.
[93] Russel B, Whitehead AN. Principia mathematica. 2nd ed., vols. 1–3. The University Press; 1925.
[94] Shannon C. Programming a computer for playing chess. Philosophical Magazine 1950:256–75.
[95] Mellit A, Kalogirou SA. Artificial intelligence techniques for photovoltaic applications: a review. Progress in Energy and Combustion Science 2008;34(5):574–632.
[96] Newell A, Shaw J, Simon H. Report on a general problem-solving program. In: Proceedings of the international conference on information processing 1959. p. 256–64.
[97] McCarthy J. Recursive functions of symbolic expressions and their computation by machine. Communications of the Association for Computing Machinery 1960;7:184–95.
[98] Weizenbaum J. ELIZA – a computer program for the study of natural language communication between man and machine. Communications of the Association for Computing Machinery 1966;9:36–45.
[99] Bobrow DG. Natural language input for a computer problem solving system. In: Minsky M, editor. Semantic information processing. Cambridge: MIT Press; 1968. p. 133–215.
[100] Winograd T. Understanding natural language. New York: Academic Press; 1972.
[101] Minsky M, Papert S. Perceptrons. Cambridge, MA: MIT Press; 1969.
[102] Minsky M. A framework for representing knowledge. The Psychology of Computer Vision 1975;3:1–81.
[103] Stender J, Addis T, editors. Symbols versus neurons? Amsterdam: IOS Press; 1990.
[104] McCulloch W, Pitts W. A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biology 1943;5:115–33.
[105] Hebb DO. The organization of behavior. New York: Wiley; 1949.
[106] Rosenblatt R. Principles of neurodynamics. New York: Spartan Books; 1959.
[107] Selfridge OG. Pandemonium: a paradigm for learning in mechanisation of thought processes. In: Blake D, Uttley A, editors. Symposium on mechanisation of thought process. London: H. M. Stationary Office; 1959. p. 513–26.
[108] Widrow B, Hoff ME. Adaptive switching circuits. In: IRE Wescon 1960. p. 96–104. New York.
[109] Holland JH. Adaptation in natural and artificial systems. The University of Michigan Press; 1975.
[110] Goldberg DE. Genetic algorithms in search, optimisation and learning. Reading, Mass: Addison-Wesely Publishing Co; 1989.
[111] McCarthy J. Circumscription – a form of non-monotonic reasoning. Artificial Intelligence 1980;13(1–2):27–39.
[112] Haugeland J, editor. Artificial intelligence: the very idea. Cambridge: MIT Press; 1985.
[113] Kurzweil R. The age of intelligent machines. Cambridge: MIT Press; 1990.
[114] Newell A, Simon HA. Human problem solving. Englewood Cliffs, NJ: Prentice-Hall; 1972.
[115] Priddy KL, Keller PE. Artificial neural networks: an introduction. SPIE Press; 2005.
[116] Graupe D. Principles of artificial neural networks. In: Chen W, Mlynski DA, editors. Advanced series in circuits and systems. 1st ed., vol. 3. World Scientific; 1997. p. 4–189. Ch. 1-12.
[117] Patterson DW. Artificial neural networks: theory and applications. Prentice-Hall; 1996.
[118] Carne EB. Artificial intelligence techniques. Washington D.C: Spartan Press; 1965.
[119] Widrow B, Winter R. Neural nets for adaptive filtering and adaptive pattern recognition. Computer 1988;21:25–39.
[120] Hilbert D. Mathematische probleme. Göttinger Nachrichten 1900;1:253–97.
[121] Kolmogorov AN. On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition. Transactions of the American Mathematical Society 1963;28:55–9.

[122] Sprecher DA. On the structure of continuous functions of several variables. Transactions of the American Mathematical Society 1965;115:340–55.
[123] Funahashi KI. On the approximate realization of continuous mappings by neural networks. Neural Networks 1989;2:183–92.
[124] Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. Neural Networks 1989;2(5):359–66.
[125] Cybenko G. Approximations by superpositions of a sigmoidal function. Mathematics of Control, Signals and Systems 1989;2:183–92.
[126] Werbos PJ. Beyond regression: new tools for prediction and analysis in the behavioral sciences [Ph.D. thesis]. Harvard University; 1974.
[127] Parker DB. Learning logic. Tech. Rep. TR-47. Center for Computational Research in Economics and Management Science, MIT; 1985.
[128] Rumelhart DJ, Zipser D. Feature discovery by competitive learning. Cognitive Science 1985;9:75–112.
[129] Al-Alawi S, Al-Hinai H. An ANN-based approach for predicting global radiation in locations with no direct measurement instrumentation. Renewable Energy 1998;14:199–204.
[130] Cao S, Cao J. Forecast of solar irradiance using recurrent neural networks combined with wavelet analysis. Applied Thermal Engineering 2005;25(2–3):161–72.
[131] Cao J, Cao S. Study of forecasting solar irradiance using neural networks with preprocessing sample data by wavelet analysis. Energy 2006;31(15):3435–45.
[132] Hontoria L, Aguilera J, Riesco J, Zufiria P. Recurrent neural supervised models for generating solar radiation synthetic series. Journal of Intelligent and Robotic Systems 2001;31:201–21.
[133] Sözen A, Arcaklıoğlu E, Özalp M, Çağlar N. Forecasting based on neural network approach of solar potential in Turkey. Renewable Energy 2005;30(7):1075–90.
[134] Cao J, Lin X. Study of hourly and daily solar irradiation forecast using diagonal recurrent wavelet neural networks. Energy Conversion and Management 2008;49(6):1396–406.
[135] Cao J, Lin X. Application of the diagonal recurrent wavelet neural network to solar irradiation forecast assisted with fuzzy technique. Engineering Applications of Artificial Intelligence 2008;21(8):1255–63.
[136] Yona A, Senjyu T, Saber A, Funabashi T, Sekine H, Kim C-H. Application of neural network to 24-hour-ahead generating power forecasting for PV system. In: Power and energy society general meeting – conversion and delivery of electrical energy in the 21st century, 2008 IEEE 2008. p. 1–6.
[137] Paoli C, Voyant C, Muselli M, Nivet M-L. Solar radiation forecasting using ad-hoc time series preprocessing and neural networks. In: Proceedings of the 5th international conference on emerging intelligent computing technology and applications, ICIC'09. Berlin, Heidelberg: Springer–Verlag; 2009. p. 898–907.
[138] Paoli C, Voyant C, Muselli M, Nivet M-L. Forecasting of preprocessed daily solar radiation time series using neural networks. Solar Energy 2010;84(12):2146–60.
[139] Azadeh A, Maghsoudi A, Sohrabkhani S. An integrated artificial neural networks approach for predicting global radiation. Energy Conversion and Management 2009;50(6):1497–505.
[140] Mellit A, Pavan AM. A 24-h forecast of solar irradiance using artificial neural network: application for performance prediction of a grid-connected PV plant at Trieste, Italy. Solar Energy 2010;84(5):807–21.
[141] Mellit A, Mekki H, Messai A, Kalogirou S. FPGA-based implementation of intelligent predictor for global solar irradiation, part I: theory and simulation. Expert Systems with Applications 2011;38(3):2668–85.
[142] Chen C, Duan S, Cai T, Liu B. Online 24-h solar power forecasting based on weather type classification using artificial neural network. Solar Energy 2011;85(11):2856–70.
[143] Kalogirou SA. Applications of artificial neural-networks for energy systems. Applied Energy 2000;67(1–2):17–35.
[144] Duda RO, Hart PE, Stork DG. Pattern classification. Wiley; 2001.
[145] Ripley BD. Pattern recognition and neural networks. Cambridge University Press; 1996.
[146] Silverman BW, Jones MC, Fix E, Hodges JL. (1951): An important contribution to nonparametric discriminant analysis and density estimation: commentary on Fix and Hodges (1951). International Statistical Review/Revue Internationale de Statistique 1989;57(3):233–8.
[147] Yakowitz S. Nearest-neighbour methods for time series analysis. Journal of Time Series Analysis 1987;8(2):235–47.
[148] Pedro HTC, Coimbra CFM. Assessment of forecasting techniques for solar power production with no exogenous inputs. Solar Energy 2012;86(7):2017–28.
[149] Lora A, Santos J, Santos J, Ramos J, Exposito A. Electricity market price forecasting: neural networks versus weighted-distance k nearest neighbours. In: Hameurlain A, Cicchetti R, Traunmüller R, editors. Database and expert systems applications. Lecture notes in computer science, vol. 2453. Berlin/Heidelberg: Springer; 2002. p. 157–211.
[150] Lora AT, Riquelme JC, Ramos JLM, Santos JMR, Exposito AG. Influence of kNN-based load forecasting errors on optimal energy production. In: Pires FM, Abreu S, editors. Progress in artificial intelligence, vol. 2902. Berlin: Springer-Verlag Berlin; 2003. p. 189–203.
[151] St-Hilaire A, Ouarda TBMJ, Bargaoui Z, Daigle A, Bilodeau L. Daily river water temperature forecast model with a k-nearest neighbour approach. Hydrological Processes 2012;26(9):1302–10.
[152] Akbari M, Overloop P, Afshar A. Clustered k nearest neighbor algorithm for daily inflow forecasting. Water Resources Management 2011;25(5):1341–57.
[153] Bannayan M, Hoogenboom G. Predicting realizations of daily weather data for climate forecasts using the non-parametric nearest-neighbour re-sampling technique. International Journal of Climatology 2008;28(10):1357–68.
[154] Zorita E, von Storch H. The analog method as a simple statistical downscaling technique: comparison with more complicated methods. Journal of Climate 1999;12(8):2474–89.
[155] Singh D, Dimri AP, Ganju A. An analogue method for simultaneous prediction of surface weather parameters at a specific location in the western Himalaya in India. Meteorological Applications 2008;15(4):491–6.
[156] Xavier PK, Goswami BN. An analog method for real-time forecasting of summer monsoon subseasonal variability. Monthly Weather Review 2007;135(12):4149–60.
[157] Kogan F, Salazar L, Roytman L. Forecasting crop production using satellite-based vegetation health indices in Kansas, USA. International Journal of Remote Sensing 2012;33(9):2798–814.
[158] Reynolds CA. Real-time hydrology operations at USDA for monitoring global soil moisture and auditing national crop yield estimates. In: Satellite rainfall applications for surface hydrology. Springer; 2010. p. 267–93.
[159] Toll DL, Doorn B, Engman ET, Lawford RG. Using NASA products of the water cycle for improved water resources management. In: AGU fall meeting abstracts 2010;vol. 1. p. 8.
[160] Zhant P, Anderson B, Barlow M, Tan B, Myneni RB. Application of a satellite-based climate-variability impact index for crop yield forecasting in drought-stricken regions. African Journal of Plant Science 2010;4(4):82–94.
[161] Perez R, Seals R, Stewart R, Zelenka A. Using satellite-derived insolation data for the site/time specific simulations of solar energy systems. Solar Energy 1994;53(6):491–5.
[162] Goddardd RH. A method of reaching extreme altitudes. In: Smithsonian miscellaneous collections 1919;vol. 71. p. 1–69.
[163] Smith WL, Bishop WP, Dvorak VF, Hayden CM, McElroy JH, Mosher FR, et al. The meteorological satellite: overview of 25 years of operation. Science 1986;231(4737):455–62.
[164] 2227. In: Vaughan WW, editor. NASA conference publication 1981. p. 1–2.
[165] 2257. In: Kellogg WW, editor. NASA conference publication 1982. p. 1–5.
[166] Hanel RA. Radiometric measurements from satellites. Aerospace Engineering 1962;21:34–9.
[167] Kidd C, Vincenzo L, Baur P. A review of satellite meteorology and climatology at the start of the twenty-first century. Progress in Physical Geography 2009;33(4):474–89.
[168] Schmit TJ, Gunshor MM, Menzel WP, Gurka JJ, Li J, Bachmeier AS. Introducing the next-generation advanced baseline imager on GOES-R. Bulletin of the American Meteorological Society 2005;86(8):1079–96.
[169] Menzel WP, Purdom JFW. Introducing GOES-I: the first of a new generation of geostationary operational environmental satellites. Bulletin of the American Meteorological Society 1994;75(5):757–81.
[170] Hanson KJ, Vonder Harr TH, Suomi VE. Reflection of sunlight to space and absorption by the earth and atmosphere over the United States during spring 1962. Monthly Weather Review 1967;95(6):354–62.
[171] Danjon A. Nouvelles researches sur la photométrie de la lumière cendrée et l'albedo de la terre. Annales de l'Observatoire de Strasbourg 1936;3:139–81.
[172] Dubois J. La variation de brilliance de la lumière cendrée de la lune au cours du dernier cycle solarie. L'Astronomie 1955;69:242–6.
[173] Angstrom A. Atmospheric turbidity, global illumination and planetary albedo of the earth. Tellus 1962;14(4):435–50.
[174] Simpson GC. The distribution of terrestrial radiation. Memoirs of the Royal Meteorological Society 1929;3(23):53–78.
[175] Houghton HG. On the annual heat balance of the northern hemisphere. Journal of Meteorology 1954;11(1):1–9.
[176] London J. A study of the atmospheric heat balance. Research Division, College of Engineering, Department of Meteorology and Oceanography, New York University; 1962.
[177] Noia M, Ratto C, Festa R. Solar irradiance estimation from geostationary satellite data: I. statistical models. Solar Energy 1993;51(6):449–56.
[178] Noia M, Ratto C, Festa R. Solar irradiance estimation from geostationary satellite data: II. Physical models. Solar Energy 1993;51(6):457–65.
[179] Gautier C, Diak G, Masse S. A simple physical model to estimate incident solar radiation at the surface from GOES satellite data. Journal of Applied Meteorology 1980;19(8):1005–12.
[180] Raphael C, Hay JE. An assessment of models which use satellite data to estimate solar irradiance at the earth's surface. Journal of Climate and Applied Meteorology 1984;23(5):832–44.
[181] Coulson K. Characteristics of the radiation emerging from the top of a Rayleigh atmosphere—I: intensity and polarization. Planetary and Space Science 1959;1(4):265–76.
[182] Coulson K. Characteristics of the radiation emerging from the top of a Rayleigh atmosphere—II: total upward flux and albedo. Planetary and Space Science 1959;1(4):277–84.
[183] Paltridge GW. Direct measurements of water vapour absorption of solar radiation in the free atmosphere. Journal of Applied Meteorology 1973;30:156–60.
[184] Gautier C, Frouin R. Downward longwave irradiance at the ocean surface using satellite data: methodology and in situ validation. Journal of Geophysical Research 1985;93:597–8.

[185] Tanre D, Herman M, Deschamps PY, DeLeffe A. Atmospheric modeling for space measurements of ground reflectances, including bidirectional properties. Applied Optics 1979;18:3587–94.
[186] Marullo S, Dalu G, Viola A. Incident short-wave radiation at the surface from METEOSAT data. Il Nuovo cimento della Società italiana di fisica 1987;10(1):77–90.
[187] Schmetz J. On the parameterization of the radiative properties of broken clouds. Tellus. Series A, Dynamic Meteorology and Oceanography 1984;36A(5):417–8.
[188] Möser W, Raschke E. Incident solar radiation over Europe estimated from METEOSAT data. Journal of Climate and Applied Meteorology 1984;23(1):166–70.
[189] Kerschegens M, Pilz U, Raschke E. A modified two stream approximation for computations of the solar radiation budget in a cloudy atmosphere. Tellus. Series A, Dynamic Meteorology and Oceanography 1978;30:429–30.
[190] Dedieu G, Deschamps PY, Kerr YH. Satellite estimation of solar irradiance at the surface of the earth and of surface albedo using a physical model applied to METEOSAT data. Journal of Climate and Applied Meteorology 1987;26(1):79–80.
[191] Lacis AA, Hansen J. A parameterization for the absorption of solar radiation in the earth's atmosphere. Journal of the Atmospheric Sciences 1974;31(1):118–33.
[192] Hay JE, Hanson KJ. A satellite-based methodology for determining solar irradiance at the ocean surface during GATE. Bulletin of the American Meteorological Society 1978;59:1549–50.
[193] Tarpley JD. Estimating incident solar radiation at the surface from geostationary satellite data. Journal of Applied Meteorology 1979;18(9):1172–81.
[194] Justus CG, Paris MV, Tarpley JD. Satellite-measured insolation in the United States, Mexico, and South America. Remote Sensing of Environment 1986;20:57–83.
[195] Cano D. Etude de l'Ennuagement par Analyse de Séquences d'Images de Satellite. Application à l'Evaluation du Rayonnement Solaire Gloibal au Sol [Ph.D. thesis]. Ecole Nationale Supérieure des télécommunications; 1982.
[196] Bourges G. Courbes de Frequence Cumulees de l'Irradiation Solaire Globale Horaire Recue par une Surface Plane. Tech. rep.. Centre d'Energetique de l'Ecole National Superieur des Mines de Paris; 1979
[197] Rigollier C. The method HELIOSAT-2 for deriving shortwave solar radiation from satellite images. Solar Energy 2004;77(2):159–69.
[198] Girodo M, Mueller RW, Heinemann D. Influence of three-dimensional cloud effects on satellite derived solar irradiance estimation – first approaches to improve the Heliosat method. Solar Energy 2006;80(9):1145–59.
[199] Zarzalejo LF, Ramirez L, Polo J. Artificial intelligence techniques applied to hourly global irradiance estimation from satellite-derived cloud index. Energy 2005;30(9):1685–97.
[200] Dagestad K-F, Olseth JA. A modified algorithm for calculating the cloud index. Solar Energy 2007;81(2):280–9.
[201] Perez R, Ineichen P, Moore K, Kmiecik M, Chain C, George R, et al. A new operational model for satellite-derived irradiances: description and validation. Solar Energy 2002;73(5):307–17.
[202] Perez R, Ineichen P, Kmiecik M, Moore K, Renne D, George R. Producing satellite-derived irradiances in complex arid terrain. Solar Energy 2004;77(4):367–71.
[203] Bjerknes V. Das problem der Wettervorhersage, betrachtet vom Standpunkte der Mechanik und der Physik. Meteorologische Zeitschrift 1904;21:1–7.
[204] Exner FM, Trabert WT. Dynamische meteorologie. In: Encyklopadie der Mathematischen Wissenschaften mit Einschluss ihrer Anwendungen – Des Sechsten Bandes Erster Teil – Geodasie und Geophysik, vol. 6. Goettingen State and University Library; 1912. p. 179–234.
[205] Richardson LF. Weather prediction by numerical process. 2nd ed. Cambridge University Press; 2007.
[206] Ehrenforfer M. Spectral numerical weather prediction models. Society for Industrial and Applied Mathematics 2012.
[207] Lynch P. Richardson's barotropic forecast: a reappraisal. Bulletin of the American Meteorological Society 1992;73(1):35–47.
[208] Lynch P. Weather forecasting from woolly art to solid science. International Geophysics 2002;83:106–19.
[209] Courant R, Friedrichs K, Lewy H. Uber die partiellen Differenzengleichungen der mathematischen Physik. Mathematische Annalen 1928;100:32–74.
[210] Courant R, Friedrichs K, Lewy H. On the partial difference equations of mathematical physics. IBM Journal of Research and Development 1967;11(2):215–34.
[211] Boyd JP. Chebyshev and Fourier spectral methods. 2nd ed. 2001. Dover.
[212] Durran DR. Numerical methods for fluid dynamics with applications to geophysics. 2nd ed., vol. 32. Springer; 2010.
[213] Charney JG. On the scale of atmospheric motions, vol. 17. Geof. Publ; 1948. p. 251–65.
[214] Charney JG, Fjortoft R, von Neumann J. Numerical integration of the barotropic vorticity equation. Tellus 1950;2:237–54.
[215] Lynch P. The emergence of numerical weather prediction – Richardson's dream. Cambridge University Press; 2006.
[216] Charney JG. The dynamics of long waves in a baroclinic westerly current. Journal of Meteorology 1947;4:135–62.
[217] Eady ET. Long waves and cyclone waves. Tellus 1949;1:33–52.
[218] Phillips NA. Energy transformations and meridional circulations associated with simple baroclinic waves in a two-level, quasi-geotropic model. Tellus 1954;6:273–86.
[219] Smagorinsky J. On the numerical integration of the primitive equations of motion for baroclinic flow in a closed region. Monthly Weather Review 1958;86:457–66.
[220] Hinkelmann K. Ein numerisches Experiment mit den primitiven Gleichungen. In: Bolin B, editor. The atmosphere and sea in motion: scientific contributions to the Rossby memorial volume. The Rockefeller Institute Press; 1959. p. 486–500.
[221] Hinkelmann K. Primitive equations. In: Davies DA, editor. Lectures on numerical short-range weather prediction WMO regional training seminar. Moscow: World Meteorological Organization; 1969. p. 306–75.
[222] Phillips NA. The general circulation of the atmosphere. Quarterly Journal of the Royal Meteorological Society 1956;82:123–64.
[223] Lewis JM. Clarifying the dynamics of the general circulation. Bulletin of the American Meteorological Society 1998;79:39–60.
[224] Silberman I. Planetary waves in the atmosphere. Journal of Meteorology 1954;11:27–34.
[225] Platzman GW. The spectral form of the vorticity equation. Journal of Meteorology 1960;17:635–44.
[226] Baer F, Platzman GW. A procedure for numerical integration of the spectral vorticity equation. Journal of Meteorology 1961;18:393–401.
[227] Orszag SA. Transform method for the calculation of vector-coupled sums: application to the spectral form of the vorticity equation. Journal of the Atmospheric Sciences 1970;27:890–5.
[228] Eliasen E, Machenhauer B, Rasmussen E. On a numerical method for integration of the hydrodynamical equations with a spectral representation of the horizontal fields. Kobenhavns Universitet, Institut for Teoretisk Meteorologi; 1970.
[229] Robert A. The integration of a spectral model of the atmosphere by the implicit method. In: WMO/IUGG symposium on numerical weather prediction. Tokyo, Japan: Japan Meteorological Agency; 1969. p. 19–24.
[230] Robert A, Henderson J, Turnball C. An implicit time integration scheme for baroclinic models of the atmosphere. Monthly Weather Review 1972;100:329–35.
[231] Robert A. The semi-implicit method. In: Numerical methods used in atmospheric models, vol. II. World Meteorological Organization; 1979. p. 417–37. vol. 17.
[232] Hinkelmann K. Der Mechanismus des meteorologischen Larmes. Tellus 1951;3:285–96.
[233] Machenhauer B. On the dynamics of gravity oscillations in a shallow water model, with applications to normal mode initialization. Contributions to Atmospheric Physics 1977;50:253–71.
[234] Baer F, Tribbia JJ. On complete filtering of gravity modes through nonlinear initialization. Monthly Weather Review 1977;105:1536–9.
[235] Daley R. Normal mode initialization. Reviews of Geophysics and Space Physics 1981;19:450–68.
[236] Tribbia JJ. On variational normal mode initialization. Monthly Weather Review 1982;110:455–70.
[237] Errico RM. Convergence properties of Machenhauer's initialization scheme. Monthly Weather Review 1983;111:2214–23.
[238] Coiffier J. Fundamentals of numerical weather prediction 2011. Cambridge.
[239] Harper KC. Weather by the numbers 2008. Cambridge.
[240] Lorenz E, Heinemann D, Wickramarathne H, Beyer HG, Bofinger S. Forecast of ensemble power production by grid-connected PV systems. In: Proceedings of the 20th European PV conference, Milano, Italy 2007. p. 1–2.
[241] Lorenz E, Hurka J, Heinemann D, Beyer H. Irradiance forecasting for the power prediction of grid-connected photovoltaic systems, selected topics in applied earth observations and remote sensing. IEEE Journal 2009;2(1):2–10.
[242] Yang F, Pan H-L, Krueger SK. Evaluation of the NCEP global forecast system at the ARM SGP site. Monthly Weather Review 2006;134:3668–90.
[243] Yang F, Mitchell K, Hou Y-T, Dai Y, Zeng X, Wang Z, et al. Dependence of land surface albedo on solar zenith angle: observations and model parameterization. Journal of Applied Meteorology and Climatology 2008;47(11):2963–82.
[244] Benjamin SG, Moninger WR, Weygandt SS, Hu M, Devenyi D, Brown JM, et al. Technical review of rapid refresh/RUC project. Tech. rep.. NOAA/ESRL/GSD Internal Review; Nov. 2009
[245] Kalisch J, Macke A. Estimation of the total cloud cover with high temporal resolution and parametrization of short-term fluctuations of sea surface insolation. Meteorologische Zeitschrift 2008;17(5):603–11.
[246] Cazorla A, Olmo FJ, Alados-Arboledas L. Development of a sky imager for cloud cover assessment. Journal of the Optical Society of America A 2008;25(1):29–39.
[247] Martínez-Chico M, Batlles F, Bosch J. Cloud classification in a mediterranean location using radiation data and sky images. Energy 2011;36(7):4055–62.
[248] Antón M, Gil J, Cazorla A, Fernández-Gálvez J, Foyo-Moreno I, Olmo F, et al. Short-term variability of experimental ultraviolet and total solar irradiance in Southeastern Spain. Atmospheric Environment 2011;45(28):4815–21.
[249] Ghonima MS, Urquhart B, Chow CW, Shields JE, Cazorla A, Kleissl J. A method for cloud detection and opacity classification based on ground based sky imagery. Atmospheric Measurement Techniques 2012;5(4):4535–69.
[250] Calbó J, Sabburg J. Feature extraction from whole-sky ground-based images for cloud-type recognition. Journal of Atmospheric and Oceanic Technology 2010;25:3–14.

[251] Chow CW, Urquhart B, Lave M, Dominguez A, Kleissl J, Shields J, et al. Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed. Solar Energy 2011;85(11):2881–93.
[252] Crispim EM, Ferreira PM, Ruano AE. Prediction of the solar radiation evolution using computational intelligence techniques and cloudiness indices. International Journal of Innovative Computing Information and Control 2008;4(5):1121–33.
[253] Ferreira P, Gomes J, Martins I, Ruano A. A neural network based intelligent predictive sensor for cloudiness, solar radiation and air temperature. Sensors 2012;12:15750–77.
[254] Marquez R, Coimbra CFM. Intra-hour DNI forecasting methodology based on cloud tracking image analysis. Solar Energy 2013;91:327–36.
[255] Lave M, Kleissl J, Arias-Castro E. High-frequency irradiance fluctuations and geographic smoothing. Solar Energy 2012;86(8):2190–9.
[256] Kamthe A, Marquez R, Coimbra CFM, Cerpa A. Sub-minute solar irradiance forecasting using wireless sensor networks. Merced: University of California; 2011.
[257] Bosh JL, Zheng Y, Kleissl J. Deriving cloud velocity from an array of solar radiation measurements. Solar Energy 2013;87:196–203.
[258] Durre I, Vose RS, Wuertz DB. Overview of the integrated global radiosonde archive. Journal of Climate 2006;19(1):53–68.
[259] National climatic data center [online] 2012 [cited May 2013].
[260] Marquez R, Pedro HTC, Coimbra CFM. Hybrid solar forecasting method uses satellite imaging and ground telemetry as inputs to ANNs. Solar Energy 2013;92:176–88.
[261] Mellit A, Eleuch H, Benghanem M, Elaoun C, Pavan AM. An adaptive model for predicting of global, direct and diffuse hourly solar irradiance. Energy Conversion and Management 2010;51(4):771–82.
[262] Marquez R, Coimbra CFM. Forecasting of global and direct solar irradiance using stochastic learning methods, ground experiments and the NWS database. Solar Energy 2011;85(5):746–56.
[263] Voyant C, Muselli M, Paoli C, Nivet M-L. Numerical weather prediction (NWP) and hybrid ARMA/ANN model to predict global radiation. Energy 2012;39(1):341–55.
[264] Marquez R, Gueorguiev VG, Coimbra CFM. Forecasting of global horizontal irradiance using sky cover indices. ASME Journal of Solar Energy Engineering 2013;135:0110161–9.
[265] Mathiesen P, Kleissl J. Evaluation of numerical weather prediction for intra-day solar forecasting in the continental United States. Solar Energy 2011;85(5):967–77.
[266] Mathiesen P, Collier C, Kleissl J. A high-resolution, cloud-assimilating numerical weather prediction model for solar irradiance forecasting. Solar Energy 2013;92:47–61.
[267] Pelland S, Galanis G, Kallos G. Solar and photovoltaic forecasting through post-processing of the global environmental multiscale numerical weather prediction model. Progress in Photovoltaics: Research and Applications 2013;21(3):284–96.
[268] Breitkreuz H, Schroedter-Homscheidt M, Holzer-Popp T, Dech S. 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 2009;48(9):1766–79.
[269] Lara-Fanego V, Ruiz-Arias J, Pozo-Vázquez D, Santos-Alamillos F, Tovar-Pescador J. Evaluation of the WRF model solar irradiance forecasts in Andalusia (southern Spain). Solar Energy 2012;86(8):2200–17.
[270] Bacher P, Madsen H, Nielsen HA. Online short-term solar power forecasting. Solar Energy 2009;83(10):1772–83.
[271] Yang D, Jirutitijaroen P, Walsh WM. Hourly solar irradiance time series forecasting using cloud cover index. Solar Energy 2012;86(12):3531–43.
[272] National solar radiation data base [online] 2008 [cited May 2013].
[273] Mandal P, Madhira STS, haque AU, Meng J, Pineda RL. Forecasting power output of solar photovoltaic system using wavelet transform and artificial intelligence techniques. Procedia Computer Science 2012;12:332–7.
[274] Hammer A, Heinemann D, Hoyer C, Kuhlemann R, Lorenz E, Müller R, et al. Solar energy assessment using remote sensing technologies. Remote Sensing of Environment 2003;86(3):423–32.
[275] Hammer A, Heinemann D, Lorenz E, Luckehe B. Short-term forecasting of solar radiation: a statistical approach using satellite data. Solar Energy 1999;67(1–3):139–50.
[276] Boyle G. Renewable electricity and the grid: the challenge of variability. Routledge; 2012.
[277] Kiliccote S, Price P, Piette MA, Bell G, Pierson S, Koch E, et al. Field testing of automated demand response for integration of renewable resources in california's ancillary services market for regulation products. Tech. rep.. Lawrence Berkeley National Laboratory; 2012
[278] Hart E, Stoutenburg E, Jacobson M. The potential of intermittent renewables to meet electric power demand: current methods and emerging analytical techniques. Proceedings of the IEEE 2012;100(2):322–34.
[279] Crabtree G, Misewich J, Ambrosio R, Clay K, DeMartini P, James R, et al. Integrating renewable electricity on the grid 2011;vol. 1401. p. 387–405.
[280] Walawalkar R, Fernands S, Thakur N, Chevva KR. Evolution and current status of demand response (DR) in electricity markets: insights from PJM and NYISO. Energy 2010;35(4):1553–60.
[281] Lee R. Generalization of learning in a machine. In: ACM annual conference/annual meeting: preprints of papers presented at the 14th national meeting of the association for computing machinery: Cambridge, Massachusetts 1959;vol. 1.
[282] Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences 1982;79:2554–8.
[283] Rummelhart DE, McClelland JL. An interactive activation model of the effect of context in language learning. Psychological Review 1986;89:60–94.
[284] Hecht-Nielsen R. Counter propagation networks. Applied Optics 1987;26:4979–84.
[285] Chua LO, Yang L. Cellular neural networks: theory and applications. IEEE Transactions on Circuits and Systems 1988;35(10):1257–90.
[286] Hocaoglu FO, Gerek ON, Kurban M. Hourly solar radiation forecasting using optimal coefficient 2-d linear filters and feed-forward neural networks. Solar Energy 2008;82(8):714–26.
[287] Capizzi G, Napoli C, Bonanno F. Innovative second-generation wavelets construction with recurrent neural networks for solar radiation forecasting. IEEE Transactions on Neural Networks and Learning Systems 2012;23(11):1805–15.
[288] Chow SK, Lee EW, Li DH. Short-term prediction of photovoltaic energy generation by intelligent approach. Energy and Buildings 2012;55:660–7.
[289] Kean Yap W, Karri V. Comparative study in predicting the global solar radiation for Darwin, Australia. Journal of Solar Energy Engineering 2012;134(3). 034501–034501.
[290] Voyant C, Randimbivololona P, Nivet ML, Paoli C, Muselli M. Twenty four hours ahead global irradiation forecasting using multi-layer perceptron, Meteorological Applications. http://dx.doi.org/10.1002/met.1387.
[291] Lorenz E, Scheidsteger T, Hurka J, Heinemann D, Kurz C. Regional PV power prediction for improved grid integration. Progress in Photovoltaics: Research and Applications 2011;19(7):757–71.
[292] Yona A, Senjyu T, Funabashi T, Kim C-H. Determination method of insolation prediction with fuzzy and applying neural network for long-term ahead PV power output correction. IEEE Transactions on Sustainable Energy 2013;4(2):527–33.
[293] Jafarzadeh S, Fadali M, Evrenosoglu C. Solar power prediction using interval type-2 TSK modeling. IEEE Transactions on Sustainable Energy 2013;4(2):333–9.

Abbreviations

ADALINE: ADAptive LInear NEuron
AI: Artificial Intelligence
AM: Airmass
ANN: Artificial Neural Network
AR: Auto-Regressive
ARIMA: Auto-Regressive Integrated Moving Average
ARIMAX: ARIMA-eXogenous
ARMA: Auto-Regressive Moving Average
ARMAX: ARMA-eXogenous
ASCE: American Society of Civil Engineers
autoDR: automated Demand Response
AVHRR: Advanced Very High Resolution Radiometer
BPL: Back Propagation-Learning
BVE: Barotropic Vorticity Equation
CERES: Clouds and the Earths Radiant Energy System
CFL: Courant, Friedrichs and Lewy
CPV: Concentrated PV
CSP: Concentrated Solar Power
DHI: Diffuse Horizontal Irradiance
DNI: Direct Normal Irradiance
ECMWF: European Centre for Medium-Range Weather Forecasts
ESRA: European Solar Radiation Atlas
ESSA: Environmental Science Services Administration
FPGA: Field Programmable Gate Array
GA: Genetic Algorithm
GFS: Global Forecast System
GHI: Global Horizontal Irradiance
GOES: Geostationary Operational Environmental Satellites
GPS: General Problem Solver
GSO: GeoStationary Orbit
HRRR: High Resolution Rapid Refresh
HS: Hybrid System
IEA: International Energy Agency
ISO: Independent System Operator
kNN: kNearest Neighbor
KSI: Kolmogorov–Smirnov Integral
LEO: Low Earth Orbit
LMS: Least Means Square
MA: Moving Average
MADALINE: Many ADALINEs
MAE: Mean Absolute Error
MAPE: Mean Absolute Percent Error
MBE: Mean Bias Error
MLP: Multi-Layer Perceptron
MOS: Model Output Statistics
MR: MADALINE adaptation Rule

NAM: North American Mesoscale
NAR: Non-linear AR
NARMAX: Non-linear ARMA-eXogenous
NARX: Non-linear AR-eXogenous
NASA: National Aeronautics and Space Administration
NCAR: National Center for Atmospheric Research
NCEP: National Centers for Environmental Prediction
NOAA: National Oceanic and Atmospheric Administration
NWP: Numerical Weather Prediction
PIV: Particle Image Velocimetry
PLEO: Polar LEO
PSSLEO: Polar SSLEO
PV: PhotoVoltaic
RMSE: Root Mean Square Error
RTM: Radiative Transfer Model
RUC: Rapid Update Cycle
SSLEO: Sun-Synchronous LEO
TIROS: Television InfraRed Observation Satellite
TLU: Threshold Logic Unit
TSI: Total Sky Imager
WRF: Weather Research and Forecasting
XOR: eXclusive-OR