# UC San Diego
## UC San Diego Electronic Theses and Dissertations

**Title**
Solar forecasting review

**Permalink**
https://escholarship.org/uc/item/7zv8s0w9

**Author**
Inman, Richard Headen

**Publication Date**
2012

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

**Solar Forecasting Review**

A Thesis submitted in partial satisfaction of the
requirements for the degree
Master of Science

in

Engineering Sciences (Aerospace Engineering)

by

Richard Headen Inman, Jr.

Committee in charge:

    Professor Carlos F. M. Coimbra, Chair
    Professor Jan Kleissl
    Professor Alison L. Marsden

2012

The Thesis of Richard Headen Inman, Jr. is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____
Chair

University of California, San Diego

2012

## DEDICATION

In recognition of all of their unconditional love and support; for never letting me forget that I am capable of accomplishing anything I put my mind to; and for constantly reminding me of where I came from and what is important, this thesis is lovingly dedicated to my wonderful family.

EPIGRAPH

*Chaos is inherent in all compounded things.*
*Strive on with diligence.*
—Buddha

*As far as the laws of mathematics refer to reality,*
*they are not certain,*
*and as far as they are certain,*
*they do not refer to reality.*
—Albert Einstein

*Its not a race.*
—Ryan Verne Inman

TABLE OF CONTENTS

LIST OF TABLES

ACKNOWLEDGEMENTS

Thanks to Professor Carlos F. M. Coimbra for his guidance during the writing of this thesis, Professor Sayantani Ghosh for encouraging me to pursue research as an undergraduate, Lana Jordan for sparking my initial interest in physics and engineering, my loving mother Rose Mary for a wonderful job raising my brother and I as well as contented support, my father Richard for his insightful guidance and persistence, my younger brother Ryan for bringing be back down to Earth in the way only he can, Michelle Bullock for teaching me lessons that one cannot learn in graduate coursework, and whoever deserves credit for Fridays at Scripps, Twiggs on Park and every pub in San Diego.

# VITA

| 2010 | B. S. in Mechanical Engineering, University of California, Merced. |
|------|---|
| 2010 | B. S. in Physics, University of California, Merced. |
| 2012 | M. S. in Engineering Sciences (Aerospace Engineering), University of California, San Diego. |

# PUBLICATIONS

Y. K. Verma, R. H. Inman, C. G. L. Ferri, H. Mirafzal, S. N. Ghosh, D. F. Kelley, L. S. Hirst, S. Ghosh and W. C. Chin, "Electrical modulation of static and dynamic spectroscopic properties of coupled nanoscale GaSe quantum dot assemblies", *Phys. Rev. B*, 82(16), 2010.

L. S. Hirst, J. Kirchhoff, R. Inman and S. Ghosh, "Quantum dot self-assembly in liquid crystal media", *Proceedings of SPIE*, 7618, 2010.

S. Ghosh, G. Shcherbatyuk, R. Inman and J. Clayton, "Nanostructured, luminescent solar concentrators" (invited article), *SPIE Newsroom: Solar and Alternative Energy*, 2010.

G. V. Shcherbatyuk, R. H. Inman, R. Winston and S. Ghosh, "Design and performance of nanostructure-based luminescent solar concentrators ", *Proceedings of SPIE*, 7769, 2010.

G. V. Shcherbatyuk, R. H. Inman, C. Wang, R. Winston and S. Ghosh, "Viability of using near infrared PbS quantum dots as active materials in luminescent solar concentrators ", *Applied Physics Letters*, 96, 191901, 2010.

R. H. Inman, G. V. Shcherbatyuk, D. Medvedko, A. Gopinathan and S. Ghosh, "Cylindrical luminescent solar concentrators with near-infrared quantum dots", *Optics express*, 19, 24, 24308 2011.

# FIELDS OF STUDY

Major Field: Engineering Sciences (Aerospace Engineering)

Areas of Specialization: Fluid Mechanics, Linear and Optimal Control, and Environmental Engineering

ABSTRACT OF THE THESIS

**Solar Forecasting Review**

by

Richard Headen Inman, Jr.

Master of Science in Engineering Sciences (Aerospace Engineering)

University of California, San Diego, 2012

Professor Carlos F. M. Coimbra, Chair

There has recently been a global increased interest in expanding the renewable energy portfolio of many countries. One of the most promising technologies is that of solar photovoltaics due to the high solar resource potential. However, the solar resource at ground level is highly dependent on local meteorological conditions such as aerosol content and most notably cloud fields. This renders the solar resource inherently variable which poses problems associated with the cost of dispatchable and ancillary generators and grid reliability. As a result, high accuracy forecasts are required on multiple time horizons. The theory as well as various examples of applications in the literature are given for a number of forecasting methods including stochastic techniques, artificial neural networks, clear sky models, persistence models, numerical weather predictions as well as satellite

and ground based imaging techniques. The limited range of spatial and temporal horizon is discussed for each method and the conclusion is drawn that a high fidelity solar forecast engine would need to take advantage of a number of different forecasting methods in order to span all spatial and temporal horizons of interest.

# Part I

# Fundamental Considerations

# Chapter 1

# Introduction

The increased demand for renewable energy sources has created recent interest in the economic and technical issues associated with growling levels of PhotoVoltaic (PV) penetration into the 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 [6, 7]. As a result, PV systems are outfitted with ancillary generators for periods of high variability which increases their associated cost and complexity[8]. Accurate solar forecasts are necessary if Independent System Operators (ISOs) are to realize increased levels of operational PV penetration while maintaing relatively low cost and high reliability of the grid [9, 10, 11]. Solar forecasts on multiple time horizons become increasingly important as PV penetration grows for the purposes of grid regulation, load-following generation, power scheduling and unit commitment. To date, accurate solar forecasts have been evasive and the problem proven to be complex due to the impact of weather patterns on the intensity of solar irradiance; most notably local cloud fields. Consequently, a number of promising approaches have beed developed. This paper presents an overview of these methods.

First, some fundamental considerations are addressed including clear sky modeling, clear sky and clearness indexes, and the suggestion of a standardized metric for the evaluation of forecasting quality. In part II, time series approaches are discussed. These methods take as inputs historical time series of irradiance data which are used to predict future values. Time series approaches include stochastic

methods and artificial neural networks. Part III focuses on image based and physical methods. Both ground based and satellite remote sensing are covered. Satellite models may be statistical or physical depending on the approach. The most popular physical approach, numerical weather predictions are also addressed. Finally, hybrid systems which incorporate two or more of the methods are investigated. These methods show the most promise for the future of high fidelity irradiance forecasting.

# Chapter 2

# Clear Skies

Before addressing the methods which have been developed to forecast solar irradiance at the ground level, it will be helpful to review some fundamental solar engineering concepts such as clear sky models, clear sky and clearness indexes, persistence models, and the evaluation of forecast quality.

## 2.1   Clear Sky Models

Many solar irradiance forecasting models require the knowledge of clear sky conditions. Clear sky models are used for persistence forecasts, discussed in §2.3, and for normalization in the metric which is described in §2.4. Satellite based forecasts, covered in §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 2005 [12]. Ineichen concluded that the input parameters, most notably turbidity, had the highest influence on model accuracy. In addition, and perhaps more significantly, 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 complexity is an issue, the ESRA or Ineichen model would be appropriate due to the requirement of only a single input parameter. If complexity is not an issue or if spectral resolution is required, Solis would be an appropriate choice as it gives the overall best results according to [12]. The next section briefly describes ten different clear sky models, eight of which were included in Ineichen's comprehensive comparison.

### 2.1.1  Solis model

The Solis model is a spectrally resolved clear sky transmittance model and was developed by Mueller et al. [13] in 2004 for the European Heliosat-3 project. Solis makes use of a 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 550nm.

### 2.1.2  European Solar Radiation Atlas (ESRA) Model

The European Solar Radiation Atlas (ESRA) model was developed for the Heliosat-2 project by Rigollier et al. [14] in 2000 and Geiger et al. [15] in 2002. ESRA was later adapted by Zarzalejo et al. [16] and Badescu [5]. The model is derived from the Linke turbidity factor and Kasten's Rayleigh optical depth paramererization [17]. The only required input parameter is the Linke turbidity factor at an air mass of 2.

### 2.1.3  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 including: water vapor column, aerosol optical path (at 380 and 500nm) and the ozone column. For more information see Refs. [18, 12].

### 2.1.4 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.

### 2.1.5 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 [23] to establish the Linke turbidity coefficient independently of the air mass. The model posses the attractive property that only the Linke turbidity is required as an input.

### 2.1.6 CPCR2 Model

The CPCR2 model was developed by Gueymard in 1989 [24] 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 Angstrom size coefficient, the Angstrom turbidity coefficient and the aerosol optical depth.

### 2.1.7 REST2 Model

This model is another dual-band model developed by Guymard in 2004 [25] and 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, Angstrom turbidity coefficient and reduced $NO_2$ and $O_3$ vertical path lengths as inputs.

### 2.1.8  Kasten Model

This model is based on Kasten's pyrheliometric formula developed in Ref. [26]. 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 air mass 2 as an input. A full description can be found in Kasten's 1984 paper [27].

### 2.1.9  Polynomial Fit

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

$$I_{\text{clr, poly}} = \sum_{n=0}^{N} c_n \cos(\theta_z). \tag{2.1}$$

Coimbra and Marquez [1] found that typically a third order polynomial fit ($N$=3) is sufficient. The third order polynomial can be written

$$I_{\text{clr, poly}} = c_3 \left(\cos\theta_z\right)^3 + c_2 \left(\cos\theta_z\right)^2 + c_1 \left(\cos\theta_z\right) + c_0, \tag{2.2}$$

where the coefficients are determined by fitting the polynomial in (2.2) 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.

### 2.1.10  ASCE Evapotranspiration Model

The ASCE evapotranspiration model was developed for the University of California by R. L. Snyder in 2002 and is based on the Penman-Monteith (hourly) reference Evapotranspiration Equations [28]. The model was originally developed to estimate reference evapotranspiration, however, calculation of clear sky GHI is an intermediate step. The difference in this model is that is requires no meteorological variables as inputs. In fact, only location (latitude, longitude and altitude) is required as an input.

## 2.2   Clear Sky and Clearness Indexes

Two parameters which are frequently used in solar forecasting are the clear sky index and the clearness index. They are quite similar and differ only in the normalization factor used in their definition. The previous section covered several clear sky models, which approximate the irradiance available at the ground level. However, some applications require knowledge of the available extraterrestrial irradiance which is much easier to model due to the absence of physical atmospheric fluctuations. A simple extraterrestrial irradiance model can be written as

$$I_{\text{EX}} = I_0 \cos(\theta_z), \tag{2.3}$$

where $I_0 = 1365$ W/m$^2$ is the solar constant and $\theta_z$ is the solar zenith angle. In contrast to the clear sky models, this extraterrestrial irradiance model does not involve any location specific fitting parameters except for $\theta_z$ which depends on latitude.

The clear sky index $k_t$ is defined as the ratio of the measured irradiance to the clear sky modeled irradiance

$$k_t = \frac{I_t}{I_{\text{clr}, t}}. \tag{2.4}$$

On the other hand, the clearness index $K_t$ is similarly defined except normalization is performed with respect to the extraterrestrial irradiance

$$K_t = \frac{I_t}{I_{\text{EX}, t}}. \tag{2.5}$$

Because $k_t$ and $K_t$ are dependent on $I_{\text{clr}, t}$ and $I_{\text{EX}, t}$ respectively, there can be several clear sky or clearness indexes depending on which clear sky model or extraterrestrial model is used. Figure 2.1 illustrates the relationship between the clear sky and clearness indexes.

## 2.3   Persistence Models

Persistence models are some of the most simple forecasts to implement and are often used as a baseline for the performance evaluation of other forecast engines.
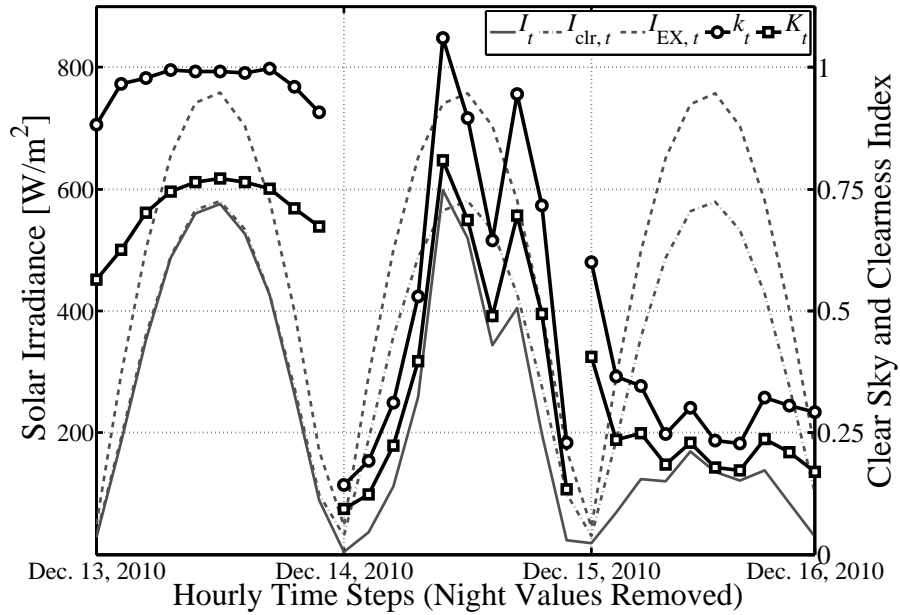
**Figure 2.1**: Time series of measured global horizontal irradiance values $I_t$, estimated clear-sky irradiance $I_{\text{clr}, t}$, estimated extraterrestrial irradiance $I_{\text{EX}, t}$ with corresponding clear sky index $k_t$ and clearness index $K_t$. Note that the clearness index is always less then the clear sky index by definition. The Ineichen model was used for $I_{\text{clr}, t}$ and equation (2.3) was used for $I_{\text{EX}, t}$. Measured data is for December 13-15, 2010 at UCSD EBUII in La Jolla, CA (Courtesy of J. Kleissl).

Persistence models, as the name implies, are defined as having the clear sky or clearness conditions persist for the next time step. For the clear sky persistence model we define the following persistence index

$$k_{t+\Delta t} = \frac{I_t}{I_{\text{clr}, \, t}} \tag{2.6}$$

Similarly, for the clearness persistence model we define a corresponding index

$$K_{t+\Delta t} = \frac{I_t}{I_{\text{EX}, \, t}} \tag{2.7}$$

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

$$\hat{I}_{\text{pers}, \, k_t} = k_{t+\Delta t} I_{\text{clr}, \, t+\Delta t} \tag{2.8}$$

and

$$\hat{I}_{\text{pers}, \, K_t} = K_{t+\Delta t} I_{\text{EX}, \, t+\Delta t}. \tag{2.9}$$

Since these models rely on the definition of a clear sky or clearness model, there can be several persistence modes depending on how $I_{\text{clr}}$ and/or $I_{\text{EX}}$ is estimated. Two noteworthy properties of the persistence models are shown in Figure 2.2. First, during cloudless periods the persistence model performs quite well, however, large forecasting errors occur during rapid changes in $I_t$ which result from the passing of opaque clouds. Secondly, the persistence model displays an obvious time delay due to the implied persistence of the clear sky conditions.

## 2.4 Evaluation of Solar Forecasting

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 [9, 1]. However, at times the solar resource exhibits a high degree of variability, see Figure 2.3 . 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 [9, 10, 11]. Despite the large number of forecasting methods described in this paper, the comparison of results and evaluation of relative
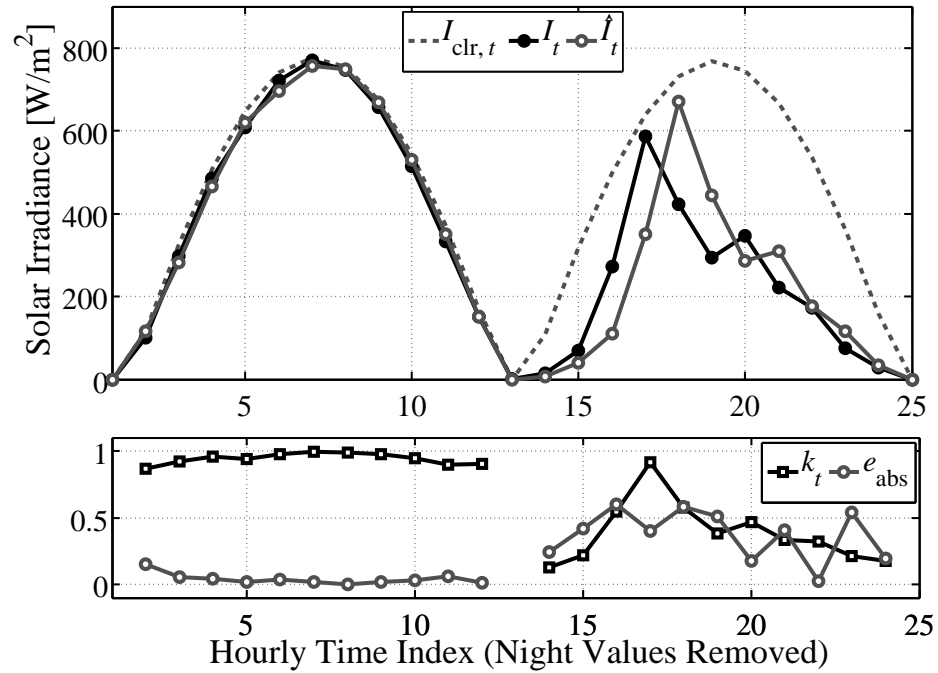
**Figure 2.2**: Example of persistence model performance for a clear and cloudy day (Oct. 11, 2010 and Oct. 13, 2010) at UCSD EBUII in La Jolla, CA (Courtesy of J. Kleissl). The persistence model performs relatively well on the clear day. However, large errors which occur during abrupt changes in measured irradiance and a 'time delay' are visible in the persistence model during the cloudy day. Clear sky index $k_t$ and absolute error $e_{\mathrm{abs}} = |I - \hat{I}|/I$ of the persistence model are shown on the lower part of the graph for the same time steps.

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. Traditional statistical metrics to characterize model quality include the coefficient of determination which compares the variance of the errors to the variance of the data which is to be modeled is given by

$$R^2 = 1 - \frac{\text{Var}(\hat{I} - I)}{\text{Var}(I)}, \tag{2.10}$$

the Root-Mean Squared Error (RMSE) which is a measure of the average spread of the errors

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^{N} \left( \hat{I}_t - I_t \right)^2}, \tag{2.11}$$

the Mean Bias Error (MBE) which is a measure of the average bias of the model

$$\text{MBE} = \frac{1}{N} \sum_{t=1}^{N} \left( \hat{I}_t - I_t \right), \tag{2.12}$$

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

$$\text{MAE} = \frac{1}{N} \sum_{t=1}^{N} \left| \hat{I}_t - I_t \right|, \tag{2.13}$$

the Mean Absolute Percentage Error (MAPE)

$$\text{MAPE} = \frac{100\%}{N} \sum_{t=1}^{N} \left| \frac{\hat{I}_t - I_t}{I_t} \right|, \tag{2.14}$$

and the correlation coefficient

$$\rho = \frac{\left( \text{Cov}(\hat{I}, I) \right)^2}{\text{Var}(\hat{I})\text{Var}(I)}. \tag{2.15}$$

However, none of these measures actually quantify the amount of variability present in the irradiance data itself.

As a solution to these obstacles, 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 the forecast uncertainty [1]. 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 §2.1, 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 [23]. 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.

### 2.4.1 Solar Variability

Variability of solar irradiance at ground level is affected by a number of factors, but most importantly solar position and cloud cover [1]. Fluctuations due to solar position are completely deterministic and are typically calculated through the use of clear sky or clearness models, see §2.1. Cloud induced fluctuations, on the other hand, are considered stochastic processes for which accurate models have been evasive. As a result, the component of solar variability which is of the most importance to accurate solar irradiance forecasting models is the stochastic cloud induced portion [11, 29, 30]. Thus, in the definition of solar irradiance variability, the authors in [1] 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

$$V = \sqrt{\frac{1}{N} \sum_{t=1}^{N} \left( \frac{I_t}{I_{\text{clr}, t}} - \frac{I_{t-1}}{I_{\text{clr}, t-1}} \right)^2} = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (\Delta k_t)^2}. \tag{2.16}$$

The definition of solar variability above is essentially synonymous with those used by Kleissl and Lave [31] and Hoff [32] except for the refinement to include the deterministic changes as was performed in [29, 30]. Figure 2.3 shows a sequence of clear and cloudy days with the associated values of $\Delta k_t$. The desired removal of deterministic variability is apparent during the clear day for which fluctuations of $\Delta k_t$ are much smaller than for the cloudy days where large ramps manifest themselves in the $\Delta k_t$ signal.
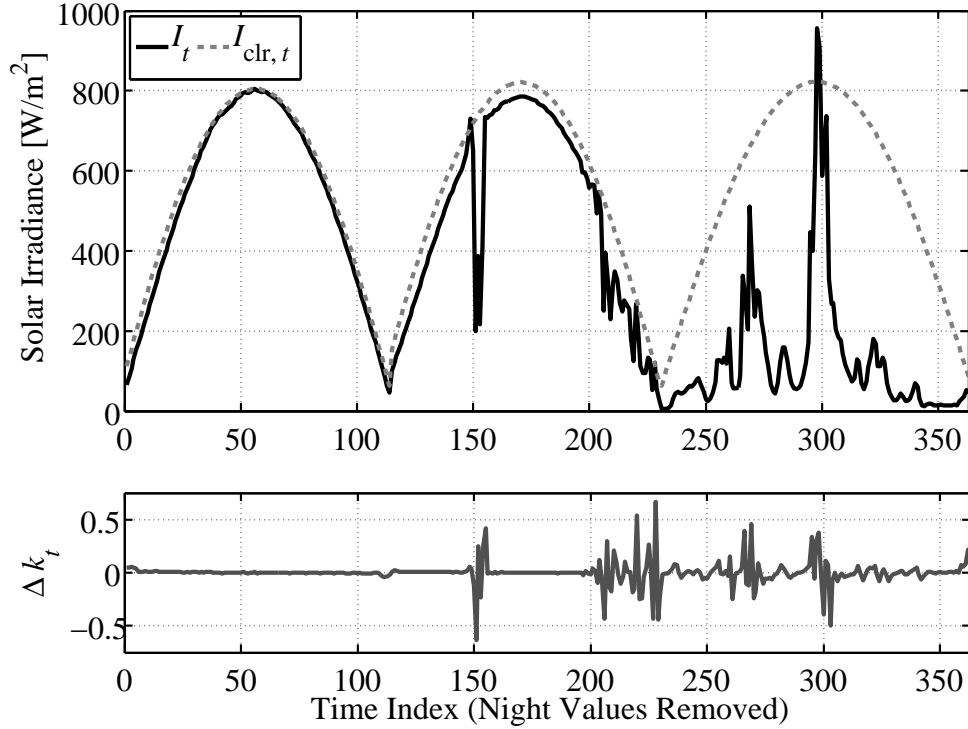
**Figure 2.3**: Measured GHI values at 5 minute resolution, Ineichen model clear-sky irradiance values and the calculated values of step-changes of the clear sky index $\Delta k_t$ (Data for Sep. 28-30, 2010 at EBUII in La Jolla, CA; Courtesy of J. Kleissl). During clear periods the solar variability is near zero. Ramps in the solar irradiance coincide with ramps in the $\Delta k_t$ signal. It should be noted that low ramps in the $\Delta k_t$ signal may also occur during periods when the measured irradiance diverges from the clear sky model (see time indexes 275 to 300 and 310 to 350). This illustrates that the use of a clear sky model for normalization does not bias the definition of variability for periods far enough removed from sunrise and sunset.
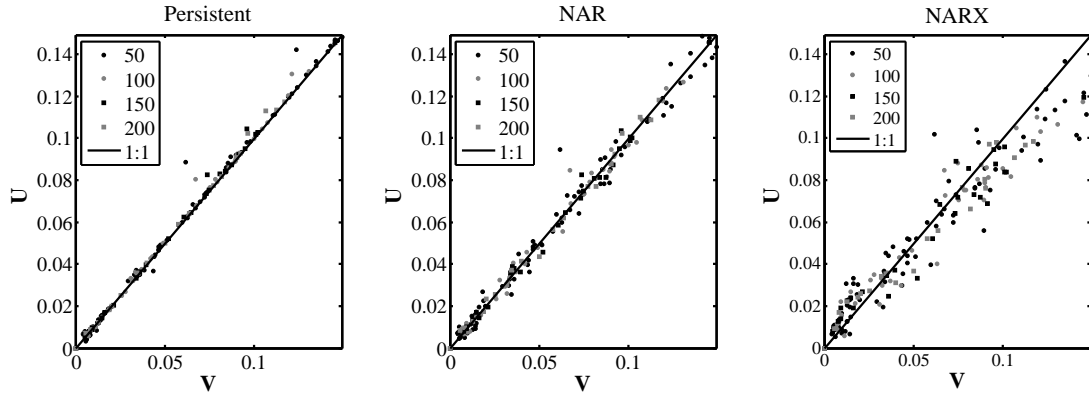
**Figure 2.4**: Scatter plot of $U$ and $V$ using ESRA clear sky model. Adapted from [1].

### 2.4.2 Forecast Uncertainty

Uncertainty of the forecast as defined in [1] is a normalized version of the RMSE used in [33] which is similar to the relative RMSE used in [34, 35]. The difference here is that normalization is performed with respect to $I_{clr}$ rather than the average irradiance as

$$U = \sqrt{\frac{1}{N} \sum_{t=1}^{N} \left( \frac{\hat{I}_t - I_t}{I_{clr,\, t}} \right)^2}. \tag{2.17}$$

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

### 2.4.3 Performance Metric

Now that solar variability $V$ and forecast uncertainty $U$ have been defined, we can write the performance metric suggested in [1] as the difference between $V$ and $U$ normalized with respect to $V$

$$s = \frac{V - U}{V} = 1 - \frac{U}{V}, \tag{2.18}$$

where it is implied that $U$ and $V$ are calculated using the same data set. The authors in [1] point out that when $s = 1$ the forcast performs perfectly due to the vanishing uncertainty, however, when $s = 0$ the forecast is dominated by solar

variability. Close inspection of Equations (2.8), (2.16) and (2.17) reveal that the persistence model, by definition, should have a performance metric value of $s = 0$, see Table 2.1. As a result, the ratio $U/V$ can be thought of as measure of the quality performance of a forecast with the persistance model as the baseline. If a developed forecast model produces a performance metric for which $s < 0$, the model performs worse than a persistance 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 [1].

As a result of $U$ and $V$ being random variables, $s$ is also a random variable. Thus, the authors in [1] 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 $V$, see Figure 2.2. The opposite is also true. Events of high variability whcih are completely predictible, such as an eclipse, will not negatively bias the value of the performance metric because uncertainly is near zero.

Marquez and Coimbra applied the preceding analysis to two forecasting models based on ANNs (NAR and NARX) for hourly data collected from Jan. 1 - Oct. 31 of 2010 and found that the metric is essentially a statictical invariant which is preserved over a wide variety of time horizons [1]. This makes the metric an attractive and robust candidate for the comparison of solar forecast model performance. Figure 2.4 shows scatter plots of $U_j$ versus $V_j$ computed for each $j$th time window for which $N_w = 50$, 100, 150 and 200 and Table 2.1 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 model's. However, it should be noted that the NARX model shows

**Table 2.1**: Forecasting quality metrics for the persistent, NAR, and NARX models on validation and training data sets from [1].

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

significant forecasting improvemnt over the persistence as many of the points fall below the 1:1 reference line.

# Part II

# Time Series Approaches

# Chapter 3

# Stochastic 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 [36] incorporated GHI data from over 300 locations into a world map of average solar energy available at the earths surface. Whillier [37] derived graphical models from correlations of measured total solar radiation in several parts of the world, at latitudes ranging from the equator to 50 degrees north and south. Swartman and Ogunlade [38, 39] derived relationships correlating solar radiation intensity with sunshine and relative humidity for tropical conditions and the Toronto area. Lund [40] 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 radiation. 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 direct radiation [41]. 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 radiation data change independently [42]. However, changes due to local weather conditions generate time series of data which are not independent but strongly dependent. The correlated nature of solar radiation 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 1920's G. Udny Yule, the influential British statistician, 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 [43]. While these methods are pervasive in almost all areas of the physical sciences, the rigid nature of strict periods limits the applications for which the methods of Fourier and Schuster 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 coined the scheme of Moving Averages (MA). The theory originates from two papers regarding the regular fluctuations of purely random series which are given, for instance, by throwing dice [44, 45]. The second of Yule's approaches, the scheme of linear Auto-Regression (AR), came in 1927 in a well known paper which discussed the periodicities of disturbed series with special reference to Wolfers sunspot numbers [46]. Shortly after, in the 1930s, the work of Yule was furthered by the Norwegian-born economist and statistician H. Wold who, while working in Sweden, applied the theory to business cycles and econometric statistics [43]. It wasnt 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 [47]. 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. 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 [43]. The use of ARMA and ARIMA provide a basis for many problems outside the realm of solar radiation forecasting including: economic and business planning, production planning, inventory and production control and optimization of industrial processes [47].

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

$$qz_t = z_{t+1} \tag{3.1}$$

The forward shift operator can be applied successively to yield

$$q^k z_t = z_{t+k} \tag{3.2}$$

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

$$\hat{q} z_t = z_{t-1} \tag{3.3}$$

which in turn yields

$$\hat{q}^k z_t = z_{t-k} \tag{3.4}$$

The backward shift operator can be used to the construct the backward difference operator $\widetilde{\nabla}$ as follows

$$\widetilde{\nabla} z_t = z_t - z_{t-1} = (1 - \hat{q}) z_t \tag{3.5}$$

which has for its inverse the summation operator $S$ written

$$S z_t = \widetilde{\nabla}^{-1} z_t = (1 - \hat{q})^{-1} z_t \tag{3.6}$$

where $\widetilde{\nabla}$ is used in order to differentiate it from the gradient operator $\nabla$. The stochastic models that follow are based on the result from Yule [46] that if a strong dependency exists between successive terms in a time series, which is, for example, a characteristic of solar radiation data, it can be effectively generated by
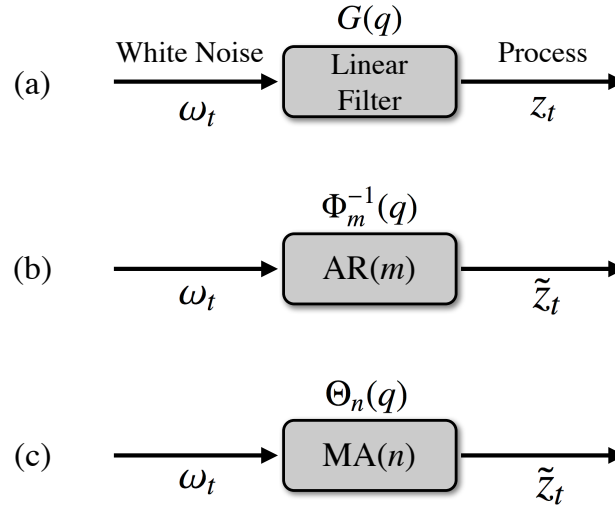
**Figure 3.1**: (a) Linear filter model of a time series generated from a white noise process with transfer function $G(q)$. (b) AR process modeled as a linear filter with transfer function $\Phi^{-1}(q)$. (c) MA process modeled as a linear filter with transfer function $\Theta(q)$.

a series of independent shocks which are assumed not to have permanent effects and are drawn at random from a stationary distribution having zero mean and variance $\sigma_\omega^2$. Any series of such random shocks $\omega_t, \omega_{t-1}, \omega_{t-2}, \dots$ is referred to as a white noise process. A linear filter can be used to transform the white noise into the stochastic process $z_t$, which is depicted in Figure 3.1(a). The linear filter is defined by the following operation

$$z_t = \mu + \omega_t + g_1\omega_{t-1} + g_2\omega_{t-2} + \dots = \mu + G(q)\omega_t \tag{3.7}$$

where $\mu$ 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, see Figure 3.2. The operator that transforms $\omega_t$ into $z_t$ is coined the transfer function $G(q)$ of the filter and is given by

$$G(q) = \sum_{k=0}^{\infty} g_k \hat{q}^k = 1 + g_1\hat{q} + g_2\hat{q}^2 + \dots \tag{3.8}$$

It should be noted that we choose the less-obvious $q$ as the argument of $G$ rather than $\hat{q}$ for clarity and to remain in agreement with the methods of Fourier. The
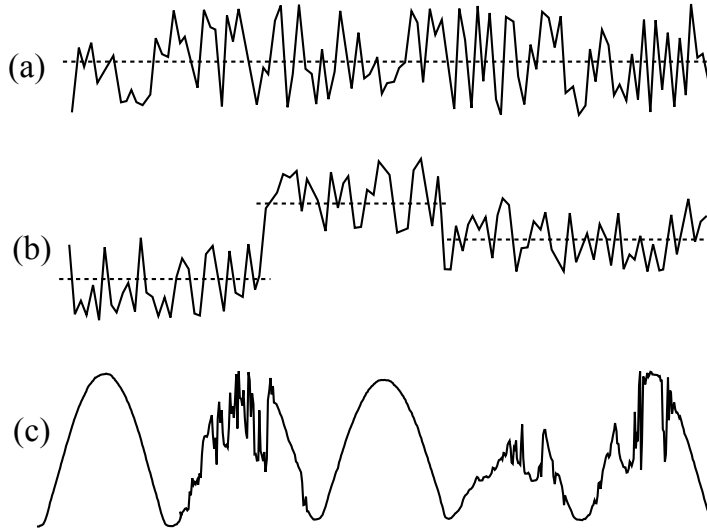
**Figure 3.2**: (a) An example stationary time series fluctuating about a static level $\mu$ which can be represented by the model $\Phi_m(q)\tilde{z}_t = \Theta_n(q)\omega_t$. (b) An example time series showing non-stationary behavior in local level can be represented by the model $\Phi_m(q)\widetilde{\nabla}z_t = \Theta_n(q)\omega_t$. (c) GHI data at 5 minute resolution with night values removed showing non-stationary behavior in local level and slope which can be represented by the model $\Phi_m(q)\widetilde{\nabla}^2 z_t = \Theta_n(q)\omega_t$.

sequence of weights $g_1, g_2, ...$ may be finite or infinite which gives us an interpretation for the level $\mu$ 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 (3.8) is finite, or infinite and convergent, the linear filter is said to be stable and the process $z_t$ to be stationary. Stationary time series are static with respect to their general shape. The fluctuations may appear ordered or completely random, nonetheless the character of the series is, on the whole, the same in different segments. In this case, the parameter $\mu$ may be interpreted as the average value about which the series fluctuates, Figure 3.2(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 [43]. In fact, the stochastic portion a solar radiation data set is often framed as a stationary process [48].

### 3.1.1 Auto-Regressive (AR) Models

The so-called auto-regressive 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 $\omega_t$. Thus, the process is said to be regressed on the previous values. If we define the stochastic portion of the time series $\tilde{z}_t, \tilde{z}_{t-1}, \tilde{z}_{t-2}, ...$ as deviations from the mean value $\mu$ as

$$\tilde{z}_t = z_t - \mu \tag{3.9}$$

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

$$\tilde{z}_t + \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + ... + \phi_m \tilde{z}_{t-m} = \omega_t \tag{3.10}$$

We can simplify the previous expression by defining the Auto-Regressive operator of order $m$, $AR(m)$, as

$$\Phi_m(q) = \sum_{k=0}^{m} \phi_k \hat{q}^k = 1 + \phi_1 \hat{q} + \phi_2 \hat{q}^2 + ... + \phi_m \hat{q}^m \tag{3.11}$$

then the $AR(m)$ model may be written conveniently as

$$\Phi_m(q) \tilde{z}_t = \omega_t \tag{3.12}$$

where 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 $\phi_1, \phi_2, ..., \phi_m, \mu$ and $\sigma_\omega^2$. Typically these are calculated from the data using the techniques covered in Section 3.1.3. It is illustrative to note that Equation (3.12) implies

$$\tilde{z}_t = \Phi_m^{-1}(q) \omega_t \tag{3.13}$$

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 $\omega_t$ as the input, see Figure 3.1(b).

In order for the $\text{AR}(m)$ process to be stationary a set of conditions must be satisfied. In [47] the authors point out that the general $\text{AR}(m)$ process has the inverse transfer function

$$\Phi_m(q) = (1 - \Gamma_1 \hat{q})(1 - \Gamma_2 \hat{q})...(1 - \Gamma_m \hat{q}) \tag{3.14}$$

which allows expansion of the process in partial fractions,

$$\tilde{z}_t = \Phi_m^{-1}(q)\omega_t = \sum_{k=1}^{m} \frac{\kappa_i}{(1 - \Gamma_i \hat{q})} \omega_t \tag{3.15}$$

where it is clear that if $\Phi_m^{-1}(q)$ is to be a convergent series for $|\hat{q}| \leq 1$, then we must have $|\Gamma| < 1$, where $k = 1, 2, ..., m$. 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 $\text{AR}(m)$ processes see [43, 47, 49].

## 3.1.2 Moving Average (MA) Models

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

$$\tilde{z}_t = \omega_t + \theta_1 \omega_{t-1} + \theta_2 \omega_{t-2} + ... + \theta_n \omega_{t-n} \tag{3.16}$$

Let us pause here and note that the terminology moving average can be a bit mis-leading due to the fact that the weights in Equation (3.16) do not, in general, need to be positive nor does their sum necessarily equal unity [47]. Nonetheless, the name is used for historic convention. The $\text{MA}(n)$ operator is defined

$$\Theta_n(q) = \sum_{k=0}^{n} \theta_k \hat{q}^k = 1 + \theta_1 \hat{q} + \theta_2 \hat{q}^2 + ... + \theta_n \hat{q}^n \tag{3.17}$$

and as a result we can write the MA model in an economic fashion

$$\tilde{z}_t = \Theta_n(q)\omega_t \tag{3.18}$$

Hence, 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 $\omega_t$ as the input, see Figure 3.1(c).

Like its counterpart, the MA model contains $n + 2$ undetermined parameters $\theta_1, ..., \theta_n, \mu, \sigma_\omega^2$ which must be determined from the data using the techniques described in the next section. Unlike AR($m$) processes, MA($n$) processes do not have a stability condition and, as a result, are unconditionally stable [43].

### 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 $\Theta$ weights and $\Phi$ weights [47]. Operating on both sides of Equation (3.12) by $\Theta(q)$ and making use of Equation (3.18), yields

$$\Theta(q)\Phi(q)\tilde{z}_t = \Theta(q)\omega_t = \tilde{z}_t \tag{3.19}$$

which implies

$$\Theta(q)\Phi(q) = 1 \tag{3.20}$$

that is

$$\Phi^{-1}(q) = \Theta(q) \tag{3.21}$$

Equation (3.21) indicates that the $\Phi$ weights may be arrived at from knowledge of the $\Theta$ weights, and vice-versa. Thus the *finite* MA process $\tilde{z}_t = \Theta(q)\omega_t$ can be written as an *infinite* AR process

$$\tilde{z}_t = -\theta_1 \tilde{z}_{t-1} - \theta_1^2 \tilde{z}_{t-2} - ... + \omega_t \tag{3.22}$$

However, if the process were really MA($n$), we would arrive at a non-parsimonious representation in terms of an AR($m$) method. By the same reasoning, an AR($m$) method could not be parsimoniously represented using a MA($n$) process. 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,

$$\tilde{z}_t + \phi_1 \tilde{z}_{t-1} + ... + \phi_m \tilde{z}_{t-m} = \omega_t + \theta_1 \omega_{t-1} + ... + \theta_n \omega_{t-n} \tag{3.23}$$
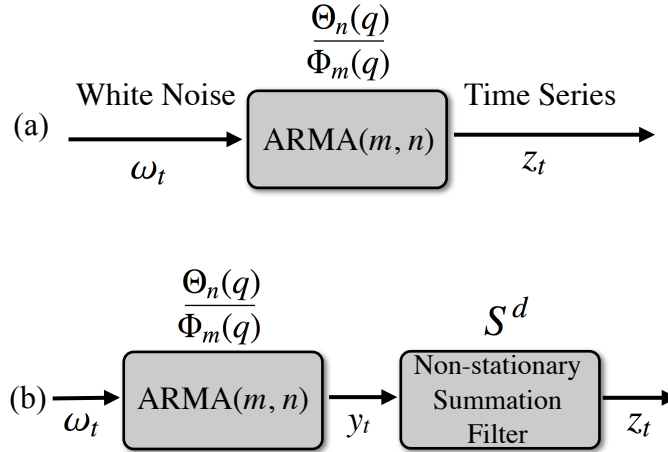
**Figure 3.3**: (a) ARMA($n$,$m$) modeled as a linear filter with transfer function $\Phi_m^{-1}(q)\Theta_n(q)$. (b) ARIMA($n$,$d$,$m$) modeled as a stationary ARMA($m$,$n$) linear filter in series with a non-stationary summation filter $S^d$.

or

$$\Phi_m(q)\tilde{z}_t = \Theta_n(q)\omega_t \tag{3.24}$$

Equation (3.24) is referred to as the mixed Auto-Regressive Moving Average (ARMA) process of order $(m, n)$. It is illustrative to note that the ARMA($m, n$) process can be written

$$\tilde{z}_t = \frac{\Theta_n(q)}{\Phi_m(q)}\omega_t = \frac{1 + \theta_1\hat{q} + ... + \theta_n\hat{q}^n}{1 + \phi_1\hat{q} + ... + \phi_m\hat{q}^m}\omega_t \tag{3.25}$$

and as a result can be thought of as the output $\tilde{z}_t$ from a linear filter, whose transfer function is the ratio of two polynomials $\Theta_n(q)$ and $\Phi_m(B)$, with white noise $\omega_t$ as the input, Figure 3.3(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 [47, 48]. 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 [50]. 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 radiation data. Boileau [51] and Guerrier [52] evaluated several stochastic models in the field of solar meteorology for 24 hour time steps. Later, Guerrier, Boileau and Bernard [53] would employ seasonal and non-seasonal stochastic techniques for solar insolation prediction which were originally developed by Box and Jenkins. Guerier [54] showed the usefulness of small correlations of consecutive day solar insolation data through an optimization of stochastic techniques. Brinkworth [55] used the simple ARMA processes to model insolation in cloudy areas of the U.K. in order to predict the output of solar thermal systems. The synthetic series produced by Brinkworth displayed characteristics equivalent to the natural insolation only for long periods of time. Benard et al. [56] estimated empirical values for stochastic models from the measurements performed at the three sites: a temperate site (Trappes, France, 49N), a Mediterranean site (Carpentras, France, 44) and an equatorial altitude site (Huallao, Peru, 13S, 3km). Subsequently, these values were used to develop ARMA models for the prediction of daily solar insolation [57]. Mustacchi et al. [58] compared ARMA models with factor analysis, Markov transition-matrices, Gaussian mapping and transmittance transition tensor techniques and concluded that the ARMA(1,0) and ARMA(2,0) models contained sufficient statistical information for hourly time series analysis. Boch et al. [59] attempted to predict solar radiation at hourly time steps using stochastic techniques for locations in France. ARMA processes have also been used to model hourly GHI for the for the calculation of the optimal control of buildings in France and Japan [60, 61, 62]. Aguiar and Collares-Pereira modeled daily sequences of hourly radiation 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 [63]. Later, Aguiar and Collares-Pereira used an ARMA(1, 0) model to generate hourly series of solar radiation values which were transformed backwards to generate synthetic sequences of clearness index $K_t$ values [64]. Al-Awahdi et al. proposed an ARMA model which used a bilinear time series to generate daily global radiation models for Kuwait [65]. Mora-Lopez et al. [66] employed a multiplicative ARMA model for global

radiation time series with regular 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 one hour and the value at the previous hour; and the relationship between the value at one hour in one day and the value at the same hour in the previous day. Moreno-Muñoz et al. [67] more recently used multiplicative ARMA models to generate instantaneous series of global irradiation in southern Spain during a four year period. Lately, however, ARMA methods typically find applications as components of a robust hybrid system architecture [68, 69, 70, 71, 72], see Section 8.

### 3.1.4 Mixed Auto-Regressive Moving Average Models with Exogenous Variables (ARMAX)

All of the linear stationary stochastic techniques discusses 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 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 models but affect its value, are referred to as exogenous variables. We can include into the $ARMA(m, n)$ models $p$ exogenous input terms which allows us to write the $ARMAX(m, n, p)$ process as

$$\tilde{z}_t + \phi_1 \tilde{z}_{t-1} + ... + \phi_m \tilde{z}_{t-m} = \omega_t + \theta_1 \omega_{t-1} + ... + \theta_n \omega_{t-n}$$
$$+ \lambda_1 e_{t-1} + ... + \lambda_p e_{t-p} \qquad (3.26)$$

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

$$\Lambda_p(q) = \sum_{k=0}^{p} \lambda_k \hat{q}^k = 1 + \lambda_1 \hat{q} + \lambda_2 \hat{q}^2 + ... + \lambda_p \hat{q}^p \qquad (3.27)$$

**Table 3.1**: Special cases of the generalized linear model (3.29). Polynomials of the shift operator which are not listed for a model are assumed to equal unity.

| Polynomials | Model |
|---|---|
| $\Lambda$ | FIR (Finite Impulse Response) |
| $\Phi, (\Lambda/\Xi = 0)$ | AR (Auto-Regressive) |
| $\Theta, (\Lambda/\Xi = 0)$ | MA (Moving Average) |
| $\Theta, \Phi, (\Lambda/\Xi = 0)$ | ARMA (Auto-Regressive Moving Average) |
| $\Phi, \Lambda$ | ARX (AR-eXogenous) |
| $\Theta, \Phi, \Lambda$ | ARMAX (ARMA-eXogenous) |
| $\Phi, \Theta, \Xi$ | ARARX (Auto-Regressive-ARX) |
| $\Phi, \Theta, \Lambda, \Xi$ | ARARMAX (Auto-Regressive ARMAX) |
| $\Theta, \Psi$ | OE (Output Error) |
| $\Theta, \Psi, \Lambda, \Xi$ | BJ (Box-Jenkins) |

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

$$\Phi_m(q)\tilde{z}_t = \Theta_n(q)\omega_t + \Lambda_p(q)e_t \tag{3.28}$$

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

$$\Phi(q)\tilde{z}_t = \frac{\Theta(q)}{\Psi(B)}\omega_t + \frac{\Lambda(q)}{\Xi(q)}e_t \tag{3.29}$$

where $\Phi(q)$, $\Theta(q)$, $\Lambda(q)$, $\Psi(q)$ and $\Xi(q)$ are polynomials of the shift operator $q$ [49, 73]. Table 3.1 summarizes some of the commonly used discrete time models which can be considered special cases of (3.29), however, the current study only focuses on 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 nonlinear behavior such as: chaos, hysteresis and saturation effects or a combination of several non-linear problems [73]. A step

towards nonlinear modeling is made by introducing the Non-linear AR-eXogenous (NARX) model as

$$\tilde{z}_t = f(\tilde{z}_{t-1}, \tilde{z}_{t-2}, ..., \tilde{z}_{t-m}, e_{t-1}, e_{t-2}, ..., e_{t-n}) + \omega_t \qquad (3.30)$$

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

$$\tilde{z}_t = f(\tilde{z}_{t-1}, \tilde{z}_{t-2}, ..., \tilde{z}_{t-m}, e_{t-1}, e_{t-2}, ..., e_{t-n},$$
$$\omega_{t-1}, \omega_{t-2}, ..., \omega_{t-p}) + \omega_t \qquad (3.31)$$

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.

## 3.3   Linear Non-Stationary Models

If the sequence of weights in Equation (3.8) 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, $\mu$ has no physical meaning except as a reference to the level of the process, Figure 3.2. 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 an normalization factor in the analysis of the evolution of a phenomenon from an initial state [43]. Several observed time series behave as if they has no specified mean about which they fluctuate, for example, daily stock prices or hourly readings from a chemical process [47].

### 3.3.1   Auto-Regressive Integrated Moving Average Models (ARIMA)

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, Figure $3.2(b)(c)$. What follows is a description of an important class of models for which it is assumed that the $d^{th}$ differene 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

$$\Phi_m(q)\tilde{z}_t = \Theta_n(q)\omega_t \tag{3.32}$$

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

$$\Phi_m(q)\tilde{z}_t = \Theta_n(q)(1 - \hat{q})^d \tilde{z}_t = \Theta_n(q)\omega_t \tag{3.33}$$

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

$$\Phi_m(q)\widetilde{\nabla}^d z_t = \Theta_n(q)\omega_t \tag{3.34}$$

Defining $y_t = \widetilde{\nabla}^d z_t$ allows one to express the model in a more illustrative way

$$\Phi_m(q)y_t = \Theta_n(q)\omega_t \tag{3.35}$$

where it is clear that the model is in agreement with the assumption that the $d^{th}$ difference of the time series can be regarded as a stationary ARMA$(p, q)$ process. If we not invert Equation (3.35) we see that

$$z_t = S^d y_t \tag{3.36}$$

which implies that the process can be arrived at by summing, or integrating, the stationary process $d$ times. Thus, we refer to (3.34) as the Auto-Regressive Integrated Moving Average (ARIMA) process. Because the AR operator $\Phi_m(q)$ is

of order $m$, the $d^{th}$ difference is taken and the MA operator $\Theta_n(q)$ is of order $n$ in (3.34) we refer to the process as ARIMA($m$, $d$, $n$). In practice, $d$ is typically 0, 1 or at most 2 [47]. 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 $\Phi_n^{-1}\tilde{\nabla}^{-d}\Theta_n$ and takes white noise $\omega_t$ as an input, see Figure 3.3.

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. [74] made use of ARIMA models to compare 10-min, 20-min, 30min and 1-hour averages of solar irradiance levels for a site in the UK. Santos et al. [75] 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 [76] 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. [77] imposed ARIMA based techniques to generate solar radiation, temperature and relative humidity forecasts for use in the control strategy of agricultural facilities. Stanhill and Cohen [78] 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 durring the $20^{th}$ century. Kärner [79] 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 [80] 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 radiation. Reikard found that, in nearly all the tests, the best results are obtained using the ARIMA in logs, with time-varying coefficients. In Ref. [81] 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 [70, 71, 72, 80].

## 3.3.2 Auto-Regressive Integrated Moving Average Models with Exogenous Variables (ARIMAX)

In a similar way to the ARMAX$(m, d, n)$ 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 $(m, d, n, p)$

$$
\begin{aligned}
\tilde{z}_t = \phi_1 \tilde{\nabla}^d z_{t-1} + ... + \phi_m \tilde{\nabla}^d z_{t-m} + \omega_t \\
+ \theta_1 \omega_{t-1} + ... + \theta_n \omega_{t-n} \\
+ \lambda_1 e_{t-1} + ... + \lambda_p e_{t-p}
\end{aligned}
\tag{3.37}
$$

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

$$
\Phi_m(q) y_t = \Theta_n(q) \omega_t + \Lambda_p(q) e_t
\tag{3.38}
$$

which again looks very similar to Equation (3.29).

# Chapter 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 [82]. Simon was consulting at the RAND Corporation when he saw a printer using customary text and punctuation symbols to produce images [83]. 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 [84]. Meanwhile, Claude Shannon demonstrated how a machine could be used to play a reasonable game of chess [85]. 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 organised by John McCarthy, Marvin Minsky, Nathaniel Rochester and Claude Shannon at Darthmouth College in New Hampshire in 1956 [82]. 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 [86].

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 [87]. However, the predominant criticism of the GPS code was that it could not learn from problems solved in advance [82]. This paved the way for McCarthy's LISP programming language to become the standard in AI development [88].

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 [82]. ELIZA used rules correlated with words like 'I', 'you', 'like' etc., which were prompted if one of these words was encountered [89]. Bobrow [90] and Winograd [91] 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 form previously solved problems [82]. Parallel to this work, Marvin Minsky and his group at MIT developed a computer program that could perform visual analogies; that is to say if two figures have some

visual relationship with each other the program could locate another set of figures with similar relationships from a larger set [92] and supported the idea that AI is fundamentally a symbol processing exercise [93].

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 $\rightarrow$ PROCESS $\rightarrow$ 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 [94]. In addition, these systems are limited by the fact that all the responses must be explicitly programmed beforehand. These symbolic approaches are refereed to as brittle do to the "crashing" that occurs when an unusual set of circumstances, which were not anticipated in the response coding, occurs. Symbol processing remained the popular motif in AI systems development well into the 1980s. Even so, there was another school of thought which defended that intelligent systems should be modeled after living organisms which posses 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" [95] was the result of a joint effort between Warren McCulloch and Walter Pitts. In this paper, which would later be included in Hebb's famous Orgization of Behavior [96], the authors purposed the first artificial neuron called the Threshold Logic Unit. In 1958 Frank Rosenblatt, a computer scientists working at Cornell Aeronautical Labaratories, published a book titled "Princilples of neurodnamics" [97]. In this book, Rosenblatt discussed machines with the ability to learn how to classify information by adapting weights, which he coined perceptrons. The Pandemonium architecture, which was proposed by Oliver Selfridge in 1959 [98] 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 [99]. This work resulted in the emergence of a new dicipline of AI techniques known as Artificial Neural Networks (ANNs). ANNs are essentially attempts to create at an electrical analogue 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" [100], Holland, motivated by Darwinian views of evolution, conceptualized an algorithm which would identify an optimal solution through the 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 durring the 1970's, however, with an increase in available computing power, the 1980s saw a renewed interest in ANNs and GAs [86, 101].

In contrast to the symbolic approach, these "evolutionary" approaches emphasize learning and adaptation over the representational power of symbolic processing. The "intelegence" of these systems is attributed to the ability to distinguish between two general classes of patterns without being directly programmed [94]. 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 [31]. Evolutionary techniques are much more systematic and it should be noted that the internal structure of the system itself is critical [94].

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 visualisation; 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 mainly on ANNs and their applications in various Hybrid Systems (HS). For a comprehensive review of AI techniques see [82, 86]. Before moving on, it should be noted that, while there have been several attempts to define AI [102, 103, 104, 105], no single definition has been universally accepted. The authors in [86] suggest that this is likely due to the various definitions references to the word "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 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
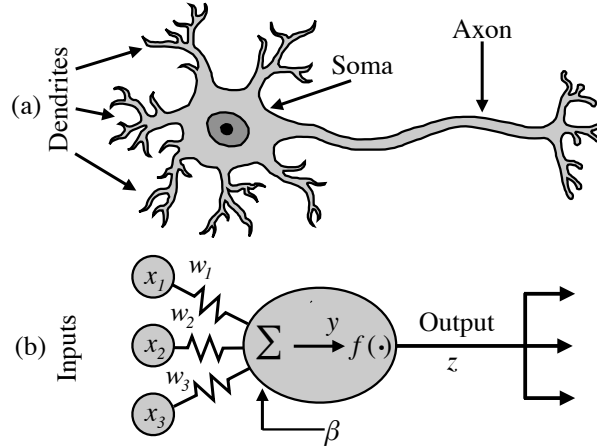
**Figure 4.1**: (a) Simplified diagram of a biological neuron. Antenna-like structures which extend from the cell body, or soma, allow the neruon to communicate with other cells. The structures which allow the neuron to accept input signals are called dendrites. The structure which carries signals away 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 $z$.

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, see Figure 4.1(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 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 [106].

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 Figure 4.1(b). The quantity $\beta$ 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

$$y = \beta + \sum_{i=1}^{3} w_i x_i \tag{4.1}$$

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 Figure 4.2. The most commonly used transfer function is the sigmoid or logistic function due to its attractive mathematical properties such as: monotonicity, continuity and differentiability [86, 106]. Employing the sigmoid transfer function with the linear net function, the output $z$ of the artificial neuron in Figure 4.1(b) can be written in a very simple manner

$$z = \frac{1}{1 + e^{\left(-\sum_i w_i x_i + \beta\right)}} \tag{4.2}$$

ANNs offer several advantages over symbolic processing. Conventional computing algorithms apply elaborate sets of equations to solve specific and well defined problems through the programer'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 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 [107]. 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.
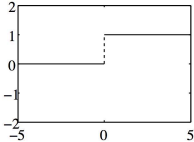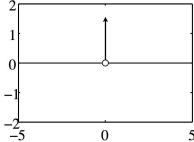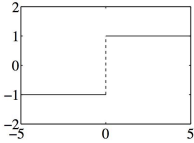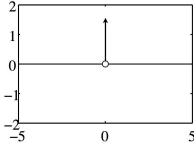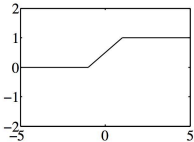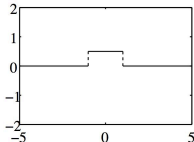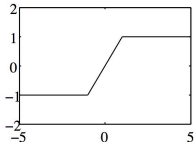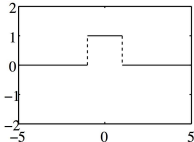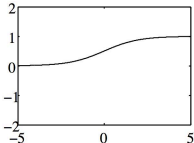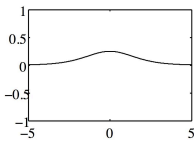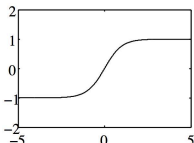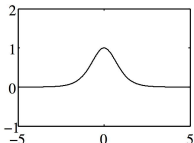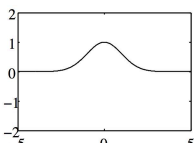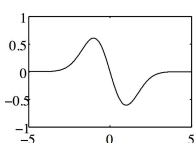
| Classification | Expression | Derivative | $f(z)$ | $f'(z)$ |
|---|---|---|---|---|
| Unipolar Step or Heavyside (Threshold) | $f(z) = H(z) = \begin{cases} 1, & \text{if } z > 0 \\ 0, & \text{if } z < 0 \end{cases}$ | $\delta(z) = \begin{cases} 0, & \text{if } z \neq 0 \\ \infty, & \text{if } z = 0 \end{cases}$ | | |
| Bipolar Step (Threshold) | $f(z) = \text{sign}(z) = 2H(z) - 1$ | $\delta(z) = \begin{cases} 0, & \text{if } z \neq 0 \\ \infty, & \text{if } z = 0 \end{cases}$ | | |
| Unipolar Linear | $f(z) = \begin{cases} 0, & \text{if } z < -1 \\ \frac{1}{2}(z+1), & \text{if } |z| < 1 \\ 1, & \text{if } z > 1 \end{cases}$ | $\frac{1}{2}[H(x+1) - H(x-1)]$ | | |
| Bipolar Linear | $f(z) = \begin{cases} -1, & \text{if } z < -1 \\ z, & \text{if } |z| < 1 \\ 1, & \text{if } z > 1 \end{cases}$ | $H(x+1) - H(x-1)$ | | |
| Unipolar Sigmoid (Logistic) | $f(z) = \dfrac{1}{1 + e^{-z}}$ | $f(z)(1 - f(z))$ | | |
| Bipolar Sigmoid (Hyperbolic Tangent) | $f(z) = \tanh(z)$ | $(1 - |f(z)^2|)$ | | |
| Gaussian Radial Basis | $f(z) = \exp(-\|x - m\|^2/\sigma^2)$ | $-2(x - m)f(z)/\sigma^2$ | | |



**Figure 4.2**: Artificial neuron transfer functions.

Now that the basic principles of artificial neurons have been established we can move on to the primary ANN structures which facilitate distributed computing. Table 4.1 summarizes the early developments in ANN structure development which is discussed in the next section. Careful examination of Table 4.1 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 XOR problem and researchers lost interest as a result [73]. 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.

## 4.1.1 Threshold Logic Unit (TLU)

As mentioned above, the earliest formulation of an artificial neuron was the TLU developed by McCulloch and Pitts in 1943. The TLU employed, as a transfer function, the threshold or Heaviside step function, see Table 4.2. In Ref. [95] 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.

**Table 4.1**: Historic Milestones in ANNs

| Number | Authors | References | Year | Structure |
|--------|---------|-----------|------|-----------|
| 1 | McCulloch & Pitts | [95] | 1942 | Threshold Logic Unit, the first Artificial Neuron (not a network). |
| 2 | Hebb & Pitts | [96] | 1942 | First learning rule; an object can be memorized by adapting the weights (not a network). |
| 3 | Rosenblatt | [97] | 1958 | Perceptron, which is the earliest ANN. |
| 4 | Lee | [108] | 1959 | Artron, which is a statistical switch closely related to Adaline. |
| 5 | Widrow and Hoff | [99] | 1960 | Adaline (ADAptive LInear NEuron), an early artificial neuron (not a network) and LMS rule for training. |
| 6 | Hopfield | [109] | 1982 | Hopfield Network. Different in many important ways, especially its recurrent feedback between nodes. |
| 7 | Widrow and Winter | [110] | 1988 | Madaline (Many Adaline) which is a network formulation of the Adaline neuron. |
| 8 | Rumelhart et al. | [111] | 1986 | Back-Propagation network, a multi-layer perceptron based ANN. |
| 9 | Hecht-Nielsen | [112] | 1987 | Counter-Propagation network uses Self-Organizing Mapping to accelerate unsupervised learning. |
| 10 | Chua & Yang | [113] | 1988 | Cellular networks in which neurons are connected to their nearest neighbors only. |

Although these assumptions are some of the earliest in ANN development, they do not necessarily apply to current ANNs, especially assumption 5 which is commonly replaced with the Hebbian Learning Law [96], allowing for changes in the structure of the network.

## 4.1.2   Simple Preceptron

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 Figure 4.3. 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 given input patterns. The input layer used a unipolar liner 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 $n$th training set

$$\Delta w_{ij}^{(n)} = \alpha(d_j^{(n)} - z_j^{(n)})x_i^{(n)} = \alpha\varepsilon^{(n)}x_i^{(n)} \tag{4.3}$$

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

$$\mathbf{W}^{(n+1)} = \mathbf{W}^{(n)} + \Delta\mathbf{W} \tag{4.4}$$

The result after $n + 1$ training sets is that each of the weights $w_{ij}$ will have been updated according to the rule outlined in Table 4.2. The perceptron learning rule can be optimized through the tuning of the learning rate coefficient $\alpha$. Small $\alpha$ corresponds to a stable and slow learning scheme. Therefore, in practice, one would like to make $\alpha$ as large as possible to ensure quick learning without introducing unstable oscillations about the desired value which result from over-relaxation.

**Table 4.2**: 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 |

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 provides guaranteed convergence of the weight matrix to values which give the correct outputs if and only if the said set of weights exits. For a detailed outline of the proof see [114].

However, there are many tasks for whcih a set of weights do not exits and, as discussed before, in 1969 Minsky and Pappert published a book concerning the strengths and limitations of single layer perceptron networks [93]. Included in this book, and the less popular earlier text by E. B. Carne [115], was a criticism of the single layer perceptron's inability to solve the simple two-state EXclusive-Or (XOR) problem which is described in Table 4.3. Solution of the XOR problem would require a non-linear partitioning of even parity points from odd parity points. This is impossible for single layer perceptrons which 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.

### 4.1.3 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, Figure 4.3(a). Unlike the unipolar step transfer function of the perceptron, ADALINEs employed a bipolar
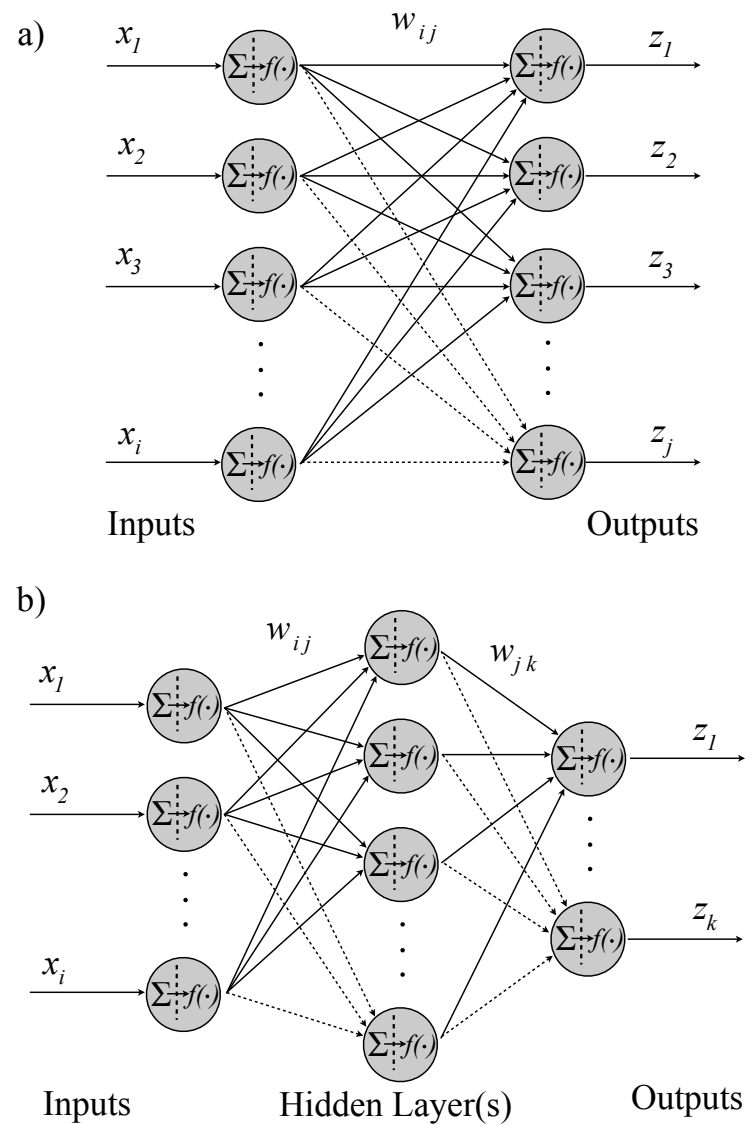
**Figure 4.3**: (a) Simple perceptron logic only capable of mapping which require linear separability. (b) Multi-Layer ANN capable of mapping which require non-linear separability.

**Table 4.3**: XOR State Table
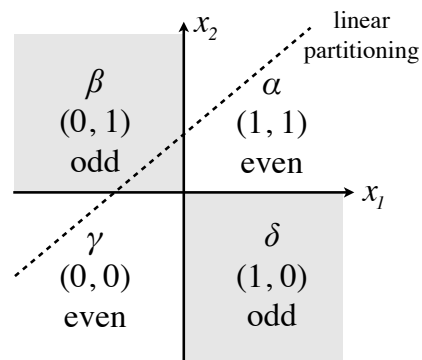


| | | inputs | | output | |
|---|---|---|---|---|---|
| state | quadrant | $x_1$ | $x_2$ | z | parity |
| $\alpha$ | 1 | 1 | 1 | 0 | even |
| $\beta$ | 2 | 0 | 1 | 1 | odd |
| $\gamma$ | 3 | 0 | 0 | 0 | even |
| $\delta$ | 4 | 1 | 0 | 1 | odd |

$(x_1 \text{ or } x_2) \bigcap (\overline{x}_1 \text{ or } \overline{x}_2)$, where $\overline{x} \equiv \text{ not } (x)$

step function, see Figure 4.2, which gave an output of -1 or +1. Inputs of ADA-LINEs are also typically bipolar. The main contribution of Widrow and Hoff was the Least Mean Square (LMS) training rule in which the weights linked with each input node are adjusted with an adaptive learning algorithm, which follows.

Given a finite number $N$ of training sets $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N$ and desired outputs $d_1, d_2, ..., d_N$, Widrow and Hoff defined a training cost given by

$$J(\mathbf{w}) = E[\varepsilon_n^2] \cong \frac{1}{N} \sum_{n=1}^{N} \varepsilon_n^2 \qquad (4.5)$$

where $E[\cdot]$ indicates the expectation and $\varepsilon_n$ denotes the training error at the $n$th set which again is simply $d_n - z_n$. Combining the definitions $E[\mathbf{x}d] = \mathbf{p}$ and $E[\langle \mathbf{x}, \mathbf{x} \rangle] = R$, where $\langle \cdot, \cdot \rangle$ denotes the inner product, with the notation from above yields

$$E[\varepsilon_n^2] = E[d_n^2] + \mathbf{w}^T R \mathbf{w} - 2\mathbf{w}^T \mathbf{p} \qquad (4.6)$$

Combining this result with (4.5), gives a gradient $\nabla J$ which is used to minimize the cost function as

$$\nabla J = \frac{\partial J(\mathbf{w})}{\partial \mathbf{w}} = 2R\mathbf{w} - 2\mathbf{p} = 0 \qquad (4.7)$$

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

$$\mathbf{w}^{LMS} = \frac{1}{R} \mathbf{p} \qquad (4.8)$$

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, see [114].

## 4.1.4   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 multilayer network which consisted of Many ADALINEs (MADALINE) [110]. In addition to the input and output layers of simple ANN structures in Figure 4.3(a), the MADALINE architecture included addition layers, see Figure 4.3(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 nonlinear 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 multilayer ANNs to mappings which require non-linear separability in fact began in 1900 with Hilbert's formulation of his famous 23 problems for the coming century [116]. 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 [117] whose reasoning was later refined by Sprecher in 1965 [118] who provided the following theorem [73].

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 $\varepsilon$, $0 < \varepsilon < \delta$, such that every real continuous function of $n$ variables $f(\mathbf{x})$, can be represented by

$$f(\mathbf{x}) = \sum_{j=1}^{2n+1} \chi \left[ \sum_{i=1}^{n} \lambda^i \psi \left( x_i + \varepsilon \left( j - 1 \right) \right) + j - 1 \right] \tag{4.9}$$

where the function $\chi$ is real and continuous and $\lambda$ is a constant which is independent of $f$.

The coupling of the previous theorem and multilayer ANNs was pointed out by Hecht-Nielsen in 1987 [118] when he demonstrated that, as a result of the Sprecher theorem, any continuous mapping $f$ can be regarded as a from of multilyer ANN with two hidden layers. The output function of the first and second layers are given by $\psi$ and $\chi$ respectively [118, 119]. In addition to the work of Sprecher, Hornik *et al.* [120], Funahashi [119] and Cybenko [121] independently

demonstrated that a general multilayer ANN with sufficient hidden layers is adequate for the approximation of a continuous nonlinear function on a finite closed interval.

Widrow and Winter also developed a training algorithm known as the MADALINE adaptaion 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 coined the *principle of minimal disturbance* [110, 114]. The MR training algorithm has beed refined several times yeilding procedures labeled MR, MRII, MRIII etc. The procedure oulined below is MRIII.

The decision wheteher or not the weights for a given layer should be adjusted requires an input vector $\mathbf{x}$ and the appropriate target output $\mathbf{d}$ in order to calculate the sum of the squared output errors. Widrow changed the input to the kth neuron by some small amount $\Delta s$ and investigated the change in the sum squared output error given by

$$\Delta \varepsilon^2 = \Delta \left( \sum_k \varepsilon_k^2 \right) \tag{4.10}$$

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

$$\nabla_k = \frac{\partial(\varepsilon_k^2)}{\partial s_k} x_k \cong \frac{\Delta(\varepsilon_k^2)}{\Delta s} x_k \tag{4.11}$$

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

$$w_{k+1} = w_k - \alpha \frac{\Delta(\varepsilon_k^2)}{\Delta s} x_k \tag{4.12}$$

Finally, defining $f' = f(1 - f)$ as the derivative of the sigmoid function (see Table 4.2), the MRIII training rule given by

$$w_{k+1} = w_k + 2\alpha \varepsilon_k f'(s_k) x_k \tag{4.13}$$

where $\alpha$ 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 [114].

### 4.1.5    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 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 Figure 4.3(b). However, MLPs differ from MADALINEs in the derivation of the popular BPL algorithm which is 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, which uses upstream variables is developed.

Several authors independently derived the BPL method in the 1970's and 1980's. Paul J. Werbos proposed the BPL method first in his 1974 Harvard University doctoral dissertation [122]. Almost ten years later, in 1985, D. B. Parker re-derived the BPL method in his MIT technical report [123]. Even so, credit for developing the BPL algorithm into an 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 [124].

In order to begin a discussion of BPL, we consider a MPL with a single hidden layer which has been modified for the BPL algorithm as shown in Figure 4.4. The results can then be generalized to a MLP with an arbitrary number of
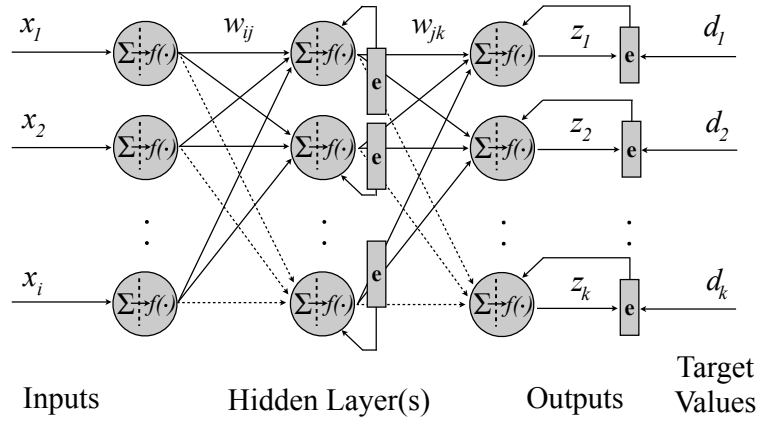
**Figure 4.4**: Modification of a MLP network for BPL.

hidden layers. As before, we adopt the following convention for weight indexing. The synaptic weight connecting input neuron $i$ to hidden layer neuron $j$ as $w_{ij}$ and the weight connecting hidden layer neuron $j$ to output neuron $k$ by $w_{jk}$. We will also define the net input to the hidden and output layer as neurons as

$$H_j = \sum_i w_{ij} x_i, \tag{4.14}$$

$$O_k = \sum_j w_{jk} z_j. \tag{4.15}$$

where $x_i$ is the input pattern and $z_j$ is the output from the hidden layer. This is similar to an unbiased version of equation 4.1. Using the above notation we can write the outputs of node $j$ of the hidden layer and unit $k$ of the output layer as

$$z_j = f(H_j) \tag{4.16}$$

$$z_k = f(O_k) \tag{4.17}$$

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 toatal system error for all training patterns through an adjustment of the weights. The total system error $E_{\text{sys}}$ is defined as the average of the output errors over all training patterns as,

$$E_{\text{sys}} = \frac{1}{N} \sum_{n=1}^{N} E_n. \tag{4.18}$$

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

$$\Delta W^{(k+1)} = -\alpha \frac{\partial E_n}{\partial \mathbf{W}^{(k)}} \tag{4.19}$$

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

$$\frac{\partial E_{\text{sys}}}{\partial \mathbf{W}} = \frac{1}{N} \sum_{n=1}^{N} \frac{\partial E_n}{\partial \mathbf{W}}. \tag{4.20}$$

Here we define $E_n$ as the mean square error,

$$E_n = \frac{1}{2} \sum_{n=1}^{N} (d_n - z_n)^2 . \tag{4.21}$$

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 $k$

$$\frac{\partial E}{\partial z_k} = -(d_k - z_k) \tag{4.22}$$

and

$$\frac{\partial z_k}{\partial I_k} = f'(I_k), \tag{4.23}$$

we can write

$$\frac{\partial E}{\partial I_k} = \frac{\partial E}{\partial z_k} \frac{\partial z_k}{\partial I_k} = -(d_k - z_k) f'(I_k). \tag{4.24}$$

By defining the previous operation as a generalized delta term,

$$\delta_k = (d_k - z_k) f'(I_k), \tag{4.25}$$

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

$$\Delta w_{jk} = -\alpha \frac{\partial E}{\partial w_{jk}} = \alpha \delta_k y_j. \tag{4.26}$$

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

$$\Delta w_{ij} = -\alpha \frac{\partial E}{\partial w_{ij}} = -\alpha \frac{\partial E}{\partial H_j} \frac{\partial H_j}{\partial w_{ij}}. \tag{4.27}$$

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

$$\frac{\partial E}{\partial H_j} = \frac{\partial E}{\partial z_j}\frac{\partial z_j}{\partial H_j} = \frac{\partial E}{\partial z_j}f'(H_j), \tag{4.28}$$

while the second term can be written explicitly from equation (4.14) as,

$$\frac{\partial H_j}{\partial w_{ij}} = \sum_i \frac{\partial}{\partial w_{ij}}(w_{ij}x_i) = x_i. \tag{4.29}$$

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

$$z_k = f(I_k) = f\left(\sum_j w_{jk}z_j\right) = f\left(\sum_j w_{jkf(H_j)}\right)$$

$$= f\left(\sum_j w_{jk}f\left(\sum_i w_{ij}x_i\right)\right) \tag{4.30}$$

along with the error defined in equation (4.21), we are able to directly differentiate the $\partial E/\partial z_j$ term in equation (4.28), obtaining

$$\frac{\partial E}{\partial z_j} = \frac{1}{2}\sum_k \frac{\partial\left(d_k - f\left(\sum_j w_{jk}z_j\right)\right)^2}{\partial z_j} \tag{4.31}$$

$$= -\sum_k (d_k - z_k)f'(I_k)w_{jk}. \tag{4.32}$$

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

$$\delta_j = f'(H_j)\sum_k \delta_k w_{jk}, \tag{4.33}$$

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

$$\Delta w_{ij} = \alpha\delta_j x_i. \tag{4.34}$$

The results for the BPL learning rule are summarized in Table 4.4.

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 reached the output layer where node $k$ calculates its output $z_k$. Subsequently,

**Table 4.4**: 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  $\delta_j = f'(H_j) \sum_k \delta_k w_{jk}$ |

errors are calculated from knowledge of the desired output pattern $d_n$ and weights $w_{jk}$ are adjusted according to Equation (4.26) as

$$w_{jk}^{(n+1)} = w_{jk}^{(n)} + \Delta w_{jk}^{(n)} \tag{4.35}$$

which is very similar to the Perceptron learning rule summarized in Table 4.2. 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 $\delta_j$

$$w_{ij}^{(n+1)} = w_{ij}^{(n)} + \Delta w_{ij} \tag{4.36}$$

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 Ref. [73, 107, 114].

## 4.2   Applications of ANNs

Since the late 1990's 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 [125]. Sfetsos and Coonick developed 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 [69]. Cao and Cao combined a recurrent ANN with wavelet

analysis for the forecast of daily solar irradiance [126, 127]. Hontoria et al. used supervised RNNs for the generation of synthetic hourly GHI time series [128]. 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 [129]. Cao and Lin combined Recurrent Neural Network (RNN) with a Wavelet Neural Network (WNN) to develop a Diagonal Recurrent Wavelet Neural Network (DRWNN) for hourly GHI forecasting [130, 131]. Yona et al. used MLPs, RNNs, and RBNNs for 24-hour-ahead generating power forecasting for PV power systems [132]. Chaouachi et al. presented the applicability of an ensemble ANN for 24 hour ahead solar power generation forecasting of a 20 kW photovoltaic system [133]. Paoli et al. recently used MLPs for the prediction of daily GHI [134, 135]. Azadeh et al. also used MLPs for forecasts, however their analysis was based on monthly average meteorological data from six cities in Iran [136]. Mellit and Pavin proposed a MLP-model to forecast the solar irradiance on a base of 24hours using the current values of the mean daily solar irradiance and air temperature for the operation of a grid connected PV plant in Italy [137]. Mellit et al. employed a Field Programmable Gate Array (FPGA) and MLP to predict daily GHI over a year [138]. Chen et al. made use of past power measurements and meteorological forecasts of relative humidity and temperature with an ANN to provide online 24 hour ahead forecasts of GHI appropriate for operational planning of transmission system operator [139]. Fiorin et al. recently presented the use of ANNs for the solar energy assessment in Brazil using data collected at SONDA sites operated by the Center for Earth System Science of the Brazilian Institute for Space Research [140]. ANN 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 [141]. For more information on ANN and solar engineering applications see [86, 33].

# Part III

# Image Based & Physical Methods

# Chapter 5

# Satellite Models

The solar forecasting methods discussed so far all require accurate measurement of the solar resource. Solar forecasts are also utilized 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 [41, 5, 142, 143, 144, 145]. 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. In this past, when no data was available for a given location it was common to extrapolate data from the nearest ground measurement station [146]. This approach is appropriate for sites in proximity to measurement stations as 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 [146, 147] thereby limiting its application. An alternative to a national or global scale ground sourced monitoring network is provided though 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" [148]. A decade later, in 1929, Goddard launched a liquid fueled rocket carrying instrumentation which included a thermometer, a barometer and a camera [149]. Advancements in rocket technology over the two decades following God-
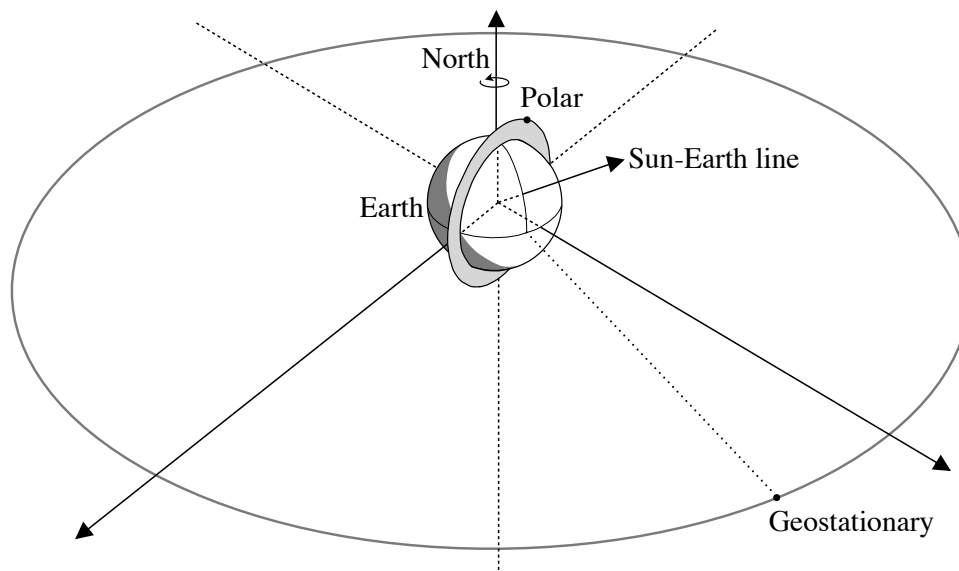
**Figure 5.1**: Illustration of a Sun-Synchronous Polar Low Earth Orbit (SSPLEO) satellite and a GeoStationary Orbit (GSO) satellite. The SSPLEO satellite has an altitude which is nearly 50 times lower than the GSO and an inclination of nearly 90° with respect to the GSO and equatorial plane. As a result, the SSPLEO 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 SSPLEO 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.

dard's work led to the first images of clouds from high altitudes [150, 151] and eventually the launch of Sputnik-1; the first artifical 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.

## 5.1 Orbits

Before discussing the assortment of satellite based methods it will be useful to address some fundamental considerations. There are two basic types of orbits which a meteorological satellite may occupy: Low Earth Orbit (LEO) and Geo-

Stationary 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 to 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 to 100 minutes, 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 a SSLEO might cross over the equator fifteen times a day each time at 11:00 a.m. local time. Satellites in a 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 the to 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 the 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 to 5 km. Figure 5.1 illustrates some of the fundamental differences between LEOs and GSOs. Once the various types of meteorological satellite orbits are understoond 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 [152]. The radiometer's third channel measured solar radiation reflected by Earth and its atmosphere in a spectral range of 0.2 to 6 $\mu$m. 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 a SSPLEO allowed the ESSA series to image a given point on the surface every time the satellite rotated about its axis. However, Radiometers were only deployed only on ESSA-7 and ESSA-9.

As a result of the sucess 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 NIM-BUS 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 monotor the Earth's water cycle and changing climate respectively [153].

Current operational LEO satellites include the U.S. National Oceanic and Atmospheric Administration (NOAA) series, also known as the Advanced TIROS-N (ATN), and the European Organization for the Exploration of Meteorological Satellites (EUMETSAT) MetOp series [153]. 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) [153]. 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 analyses. 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, however 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
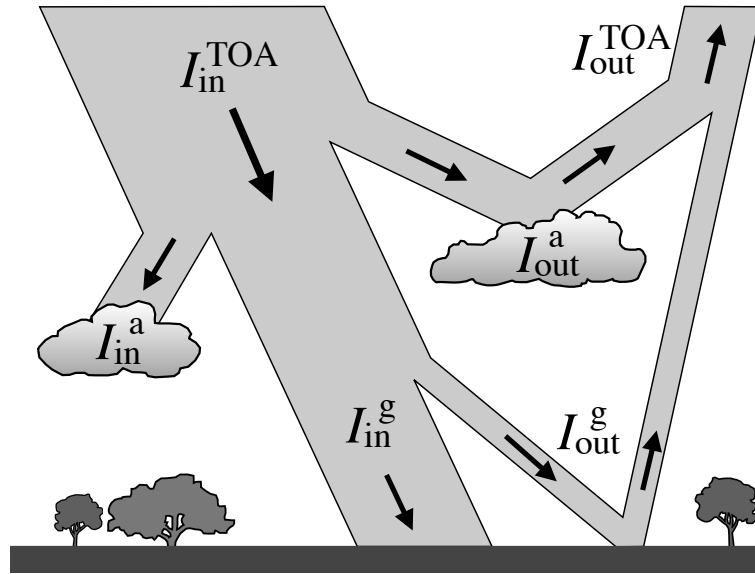
**Figure 5.2**: Radiation budget at the foundation of satellite based forecasting methods.

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) [153]. 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 minutes, although the Meteosat's Spinning Enhanced Visible and Infrared Imager (SEVIRI) possesses a nominal sampling frequency of 15 minutes [153, 154] and the GOES imager can scan small areas of interest in as little as 30 seconds [153, 155].

## 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 [147, 156]. 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 [157, 158, 159]. Other approaches involved the estimation of cloud distribution over the planet and the assignment of specific reflectivities to various could types [160, 161, 162].

Dispite these efforts, satellites equipped with radiometers allow the direct measurements of the planetary albedo. Figure 5.2 illustrates the fundamental relationship between satellite observation of planetary albedo and and ground level irradiance. By the conservation principle, the difference between the net irradiance at the top of the atmosphere $I_{in}^{TOA}$ 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,

$$I_{in}^{TOA} - I_{in}^{g} = I_{in}^{a} + I_{out}^{TOA}. \tag{5.1}$$

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

$$I_{out}^{TOA} = I_{out}^{g} + I_{out}^{a}. \tag{5.2}$$

Substution of (5.2) into (5.1) combined with normalization with respect to $I_{in}^{TOA}$ and a separation of atmospheric and terrestrial terms yields,

$$\frac{I_{in}^{g}}{I_{in}^{TOA}} + \frac{I_{out}^{g}}{I_{in}^{TOA}} = 1 - \frac{I_{in}^{a}}{I_{in}^{TOA}} - \frac{I_{out}^{a}}{I_{in}^{TOA}}. \tag{5.3}$$

The left hand side of previous equation can be written as,

$$\frac{I_{in}^{g}}{I_{in}^{TOA}} + \frac{I_{out}^{g}}{I_{in}^{TOA}} = \left(1 + \frac{I_{out}^{g}}{I_{in}^{g}}\right) \frac{I_{in}^{g}}{I_{in}^{TOA}}, \tag{5.4}$$

which given the following relationship

$$\frac{I_{\text{in}}^{\text{g}}}{I_{\text{in}}^{\text{TOA}}} = \frac{1 - \frac{I_{\text{in}}^{\text{a}}}{I_{\text{in}}^{\text{TOA}}} - \frac{I_{\text{out}}^{\text{a}}}{I_{\text{in}}^{\text{TOA}}}}{1 + \frac{I_{\text{out}}^{\text{g}}}{I_{\text{in}}^{\text{g}}}}. \tag{5.5}$$

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

$$\alpha_a = \frac{I_{\text{in}}^{\text{a}}}{I_{\text{in}}^{\text{TOA}}}; \tag{5.6}$$

the atmospheric albedo,

$$\rho_a = \frac{I_{\text{out}}^{\text{a}}}{I_{\text{in}}^{\text{TOA}}}; \tag{5.7}$$

and the terrestrial albedo,

$$\rho_t = \frac{I_{\text{out}}^{\text{g}}}{I_{\text{in}}^{\text{g}}}. \tag{5.8}$$

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,

$$K_t = \frac{1 - \alpha_a - \rho_a}{1 + \rho_g}. \tag{5.9}$$

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 (5.9).

The method used to approximate the aforementioned parameters typically varies by author. Nonetheless, the relationship between clearness index and the atmospheric and terrestrial parameters in equation (5.9) is at the foundation of all satellite based irradiance forecasting models. Measurement of $I_{\text{out}}^{\text{TOA}}$ combined with an approximation of $I_{\text{in}}^{\text{a}}$ (through the knowledge of the Linke turbidity coefficient for example) and a priori knowledge of $I_{\text{out}}^{\text{g}}$ (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 [163, 164]. 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 these statistical models are much simper due to their independence of the precise measurement of the composition of the atmosphere, 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 2.1, 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, however, 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 [164].

### 5.4.1    Gautier-Diak-Masse Model

One of the earliest physical models was developed by Gautier, Diak and Masse (GDM) in 1980 [165]. 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 [166, 3] included the T-minimum brightness determination [163] 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 [167, 168] while the absorption coefficients made use of the expressions from Paltridge [169]. Improvements to GDM model include the absorption of ozone and aerosols by Gautier and Frouin [170] in addition to multiple reflections by Tanrè *et al.* [171].

The cloudy sky RTM allowed for the simple treatment of stratiform low and middle altitude clouds [165]. 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 [165].

## 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 [172]. 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 [173] 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 and varied 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 [174], however the authors used the RTM developed by Kerschgens [175] which was more complex than the previous models. Improvements to the MR method include the addition of more parameters to accurately describe the sate 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 a 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 [176] 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 [177] together with the RTM of Tanrè [171], 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 [163].

## 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, along with current brightness, minimum brightness and maximum brightness of each pixel. As pointed out in [163] 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) and a large solid viewing angle ($2\pi$). 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 [12,20,12,16].

## 5.5.1   Hay-Hanson model

One of the simplest statistical satellite models was developed by Hay and Hanson (HH) in 1978 [178]. The model was developed for the Global Atmospheric Research Program's Atlantic Tropical Experiment to generate maps of the short-wave radiation (0.55 - 0.75 $\mu$m) reaching the surface of the ocean. The HH model is based of a statistical linear regression of the clearness index and atmospheric albedo:

$$K_t = a - b\alpha_a. \tag{5.10}$$

Substituting equation (5.10) into equation (2.3), this becomes

$$I = K_t I_0 \cos(\theta_z). \tag{5.11}$$

Hay and Hanson [178] 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 [3] to be

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

which gives a better agreement with their dataset.

It has been pointed out [163] 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 Informations Services (NESDIS) and the Great Planes Agricultural Council over the central U.S. in late

**Table 5.1**: Top: Regression coefficients for the Tarpley model as determined by Tarpley [2] and Raphael and Hay [3], in parenthesis. Bottom: Regression coefficients for the Justus-Paris-Tarpley model as determined by Justus *et al.* [4].

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

1970s [2]. 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,

$$n = \frac{0.5N_2 + N_3}{N},\tag{5.12}$$

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,

$$I = \begin{cases} a + b\cos(\theta_z) + cK_t + dn + e\left(\dfrac{B_m}{B_0}\right)^2 & \text{if } n \leq 0.4 \\[3mm] a + b\cos(\theta_z) + cn\left(\dfrac{B_{\text{cld}}}{B_n}\right)^2 & \text{if } 0.4 < n < 1, \\[3mm] a + b\cos(\theta_z) + c\left(\dfrac{B_{\text{cld}}}{B_n}\right) & \text{if } n = 1 \end{cases}$$

where $B_m$ is the mean target brightens, defined as the mean brightness of a 7 x 6 pixel array; $B_{\text{cld}}$ 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 $B_n = B_0(\theta_z = 45°, \phi_s = 105°)$ is the normalized clear brightness which

is a special case of the clear brightness $B_0$ which is obtained from the following regression,

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

Raphael and Hay [3] also estimated their own regression coefficients for this model which are different from Tarpley's treatment, see Table 5.1.

This model was later refined by Jutus, Paris and Tarpley (JPT) [4] 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,

$$I = I_0 \left(\frac{r_0}{r}\right)^2 \cos(\theta_z)\left[a + b\cos(theta_z) + c\cos^2(\theta_z)\right]$$
$$+d\left(B_m^2 - B_0^2\right), \tag{5.13}$$

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

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

As before, the authors in [4] 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 $B_0$ [163] . Regression coefficients for both the Tarpley model and the JPT model can be found in Table 5.1.

### 5.5.3  Cano-HELIOSAT Model

Cano developed a model for the French HELIOSAT project in 1982 which used visible band METEOSAT data [179]. 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,

$$K_t(i,j) = A(i,j)n_t(i,j) + B(i,j), \tag{5.14}$$

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

$$n_t(i,j) = \frac{\rho_t(i,j) - \rho_0(i,j)}{\rho_c - \rho(i,j)}, \tag{5.15}$$

where $\rho_t$ is the measured ground albedo, $\rho_0$ is the reference ground albedo and $\rho_c$ is the average albedo of the top of the clouds. The reference ground albedo was calculated using Bourges model [180] 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. [181]. These corrections were subsequently used to derive the following relationship between the cloud index $n_t$ and a clear sky index $k_t$,

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

More recent developments of the Cano-HELIOSAT model include consideration of the three dimensional structure of cloud in the determination of the cloud index [182], modification of the previous $k_t$-$n_t$ relationship to include moments of the cloud index distribution [183], corrections for non-Lambertian reflectivity and the backscattering of clouds [184] and integration of the SOLIS-RTM platform [185].

**Table 5.2**: Regression coefficients for the Perez Operational Satellite Model [5].

| 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 |

### 5.5.4   Perez Operational Model

One of the most widely used statistical satellite models is the operational model of Perez [186]. The Perez model uses a modified version of Kasten's clear sky model (see section 2.1.8) which defines a Linke turbidity coefficient independent of airmass [23]. 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 [5].

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

$$I_t = I_{\text{clr, } t} f(n_t)[aI_{\text{clr, } t} + b], \tag{5.16}$$

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

$$f(n_t) = c_5 n_t^5 + c_4 n_t^4 + c_3 n_t^3 + c_2 n_t^2 + c_1 n_t + c_0. \tag{5.17}$$

Values of the coefficients as calculated by Perez in are given in Table 5.2. 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 [187].

# Chapter 6

# Ground Based Imaging

## 6.1  Total Sky Imagers (TSI)

Both NWPs 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. One instrument which has seen increased application lately is the Total Sky Imager (TSI) manufactured by Yankee Environmental Systems [188] which is shown in Figure 6.1. An example TSI image from the University of California Merced is shown in Figure 10.2.

Typically the methodology for ground based images is similar to satellite based techniques. Projections of observed solar radiation conditions based on immediate measured history while the position and impact of clouds is deduced from their motion. In the case of TSIs the CCD image is digitally processed in order to detect locations of the sky covered by clouds. The could image is then propagated forward in time resulting in a forecast. Recently, Kleissl 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 Californa, Sand Diego [189]. Cloud fields were propagated forward in time using matching errors between the future image and the current image translated in the direction of the computed velocity vector. Kleissl's study suggests that TSI images are useful for prediction of GHI on time horizons up to 15 minutes.
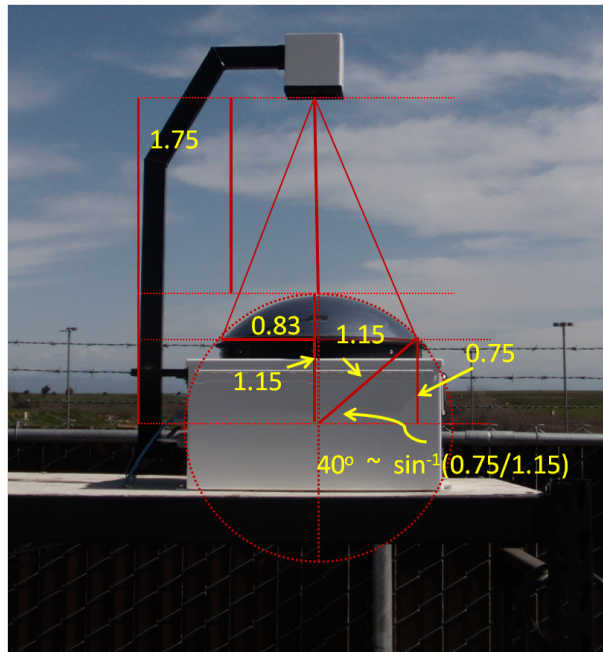
**Figure 6.1**: TSI-880 from Yankee Environmental Systems. The instrument consists of a hemispherical mirror with a CCD camera located above it. The mirror contains a sun tracking shadow-band in order to protect the camera from the sun's reflection.



**Figure 6.2**: Image from TSI taken at the University of California Merced on June 1, 2011.

Coimbra and Marquez also made use of a TSI for intra-hour forecasts, however their study focused on DNI forecasting rather than GHI [190]. The work in [190] focused on detailed image processing procedures along with Particle Image Velocimetry (PIV) to deterime grid cloud fractions in an area of interest. In agreement with the work of Kleissl, Coimbra and Marquez showed that TSIs show promise for forecasting horizons up to 15 minutes with the lowest error associated with time horizons of 5-6 minutes.

It should be noted that while these TSI based provide local meteorological information enabling intra-hour forecasts, their time horizon is restricted to approximately 30 minutes do to their limited range 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 relative cost associated with the TSI ($\sim$\$2,000) and the dynamic nature of local cloud fields which may limit the correlation of successive images poses difficulties for current ground based imaging methodologies. In addition to an upper bound on the time horizon of the TSI, a lower band is also imposed. The lower bound is a result of circum-solar scattering of light as well as limitations introduced by the shadow-band which currently renders time horizons shorter than 2 minutes inaccessible [189, 190].

## 6.2   Wireless Network Systems

None of the methods discussed so far posses intra-minute time horizons. Satellite and NWP models typically posses time horizons on the order of 30 minutes while stochastic and AI methods have not been widely applied to time horizons less than 15 minutes. TSIs are limited by the circum-solar scattering of light and the shadow-band to time horizons no longer than 3 minutes [189, 190]. Semiconductor point sensors are capable of very high sampling frequencies but fail to correctly characterize the distributed nature of an operational scale PV plant [191]. An alternative has been suggested by Coimbra and coworkers at the University of California, Merced [192]. A 1MW PV array was outfitted with with 40 TelosB
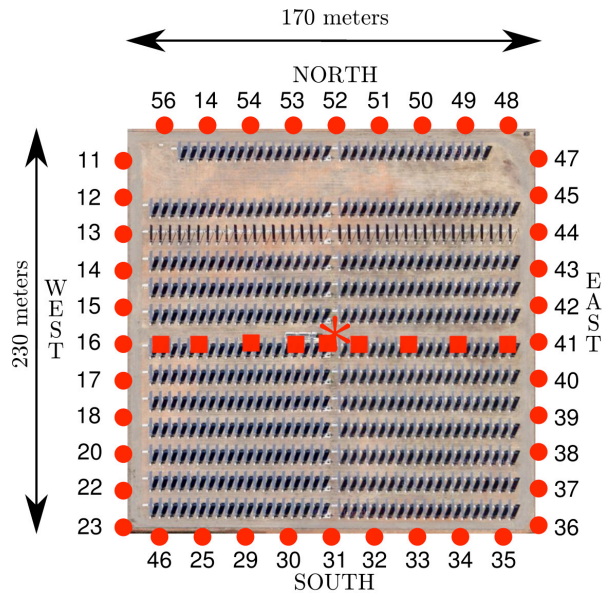
**Figure 6.3**: Birds-eye view of solar array deployment site with relative node locations. Maximum distance between successive nodes is 21 meters. The solar array covers roughly 10 acres.

nodes equipped with low cost solar irradiance sensors, see Figure 10.3. The authors in [192] proposed a forecasting algorithm which utilized multiple readings from the spatially distributed network of sensors to compute future values of the distributed power output. The forecasting approach utilized spatial cross-correlations between sensor nodes which provided forecasts in the range of 20-50 seconds. Calculated velocities agreed with TSI calculated cloud velocity field over 70% of the time [192]. This work demonstrates the potential of wireless sensor networks as low cost and highly accurate approaches for intra-minute solar forecasting.

# Chapter 7

# Numerical Weather Predictions (NWPs)

As discussed in the introduction to Chapter 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 foundation of NWP methods is illustrated in Figure 7.1. First a physical domain is chosen. The domain is subsequently discretized into a desired resolution. Physical laws of motion and thermodynamics are then numerically solved on the discrete spatial grid. It is the time marching of solutions to the physical and thermodynamic equations which provide the basis of NWP forecasts. NWP models may be divided into two main categories: global or regional, depending on the domain which is used, see figure 9.2.

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 - 20km, see Table 7.1) 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 is a 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
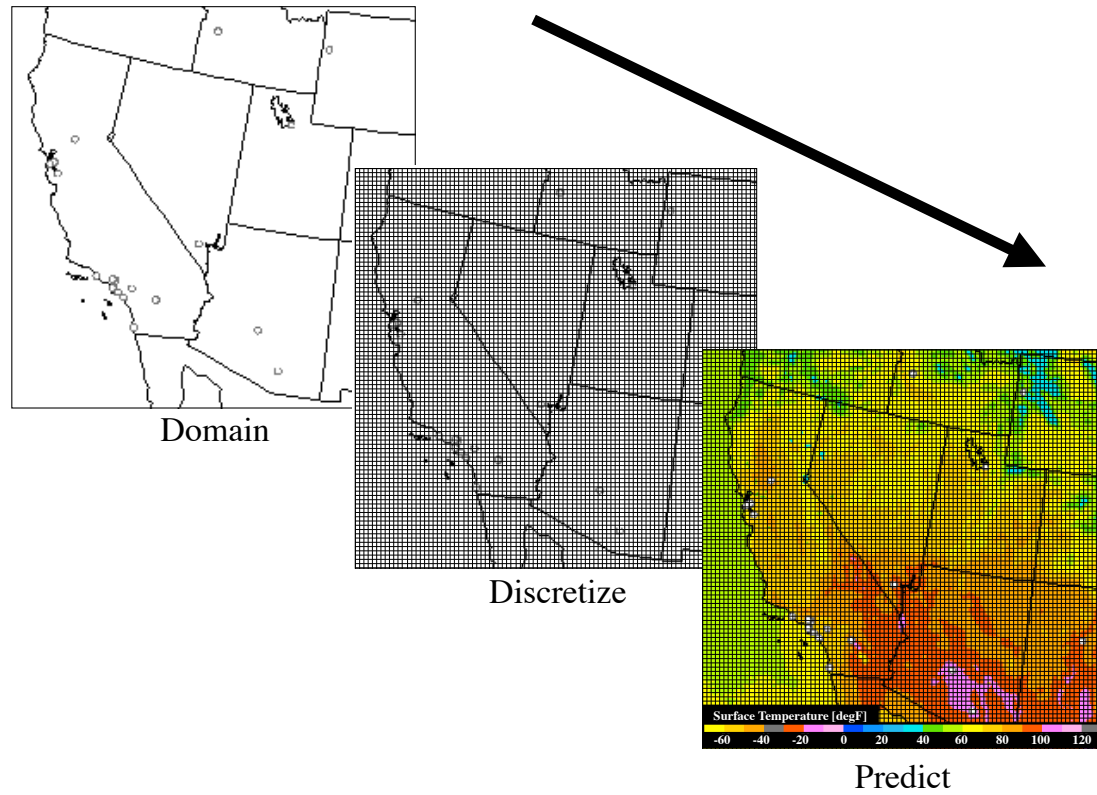
**Figure 7.1**: Basis of all Numerical Weather Prediction (NWP models). First a domain is defined. Secondly, the domain is spatially discretized to a desired resolution. Finally, the NWP predicts desired information by solving equations of motion and thermodynamic laws.

is typically required for the formation of clouds. During post-processing of the NWP's 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.

Nonetheless, NWP provide many benefits over previously discussed forecasting methods. Among these is a relatively long time horizon (15 to 240 hours, see Table 7.1). Information about the propagation of large scale weather patters is obtained though the regional and global modeling of atmospheric physics. It has been shown that NWPs provide more accurate forecasts than satellite based methods for time horizons exceeding 4 hours [193, 194]. Accordingly, NWPs provide the most attractive option for medium to long term atmospheric forecasting.

### 7.0.1   North American Mesoscale (NAM) Model

The North American Mesoscale (NAM) model is the NCEP's primary mesoscale environmental modeling tool. NAM produces 12km x 12km horizontally resolved forecasts with 60 atmospheric layers out to a time horizon of 96 hours over North America and is updated four times daily. The NAM model loop time steps are 6 hours from the time of analysis out to 84 hours (3.5 days). The NAM model used 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 atmoshpere 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.

### 7.0.2   High Resolution Rapid Refresh (HRRR) Model

The High Resolution Rapid Refresh (HRRR) model is an NOAA operated, experimental, hourly updated, 3km x 3km resolution atmospheric model. The HRRR was previously only nested over the eastern 2/3 of the continental United States, however as of June 2009 coverage was expanded to the CONUS region

similar to the former RUC, see figure 7.2.1. The RHHH models uses the 13km 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.

### 7.0.3   Weather Research and Forecasting (WRF) Model

Many of the NWPs discussed are based on a version of the WRF which was created thought 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 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.

## 7.1   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 six hours and produces forecasts up to 384 hours (16 days) in advance on a 28km x 28km grid for the global domain [195]. The GFS loop time steps are 6 hours out to 180 hours (7.5 days), then change to 12-hour time steps out to 384 hours (16 days). In addition to the 28km x 28km 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 [196]. It should be noted that the

Table **7.1**: Comparison of various NWP models.

| Name | Resolution | Layers | Update Period | Time Horizon | Time Step |
|------|-----------|--------|---------------|--------------|-----------|
| GFS | 28km | 64 | 6 hours | 180 hours | 6 hours |
| - | - | - | - | 384 hours | 12 hours |
| RUC/RAP | 13km | 50 | 1 hour | 18 hours | 1 hour |
| NAM | 12km | 60 | 6 hours | 96 hours | 6 hours |
| HRRR | 3km | 50 | 1 hour | 15 hours | 15 minutes |
| WRF | ≥1km | User Specific | User Specific | User Specific | User Specific |

radiant flux attenuation is dependent on $H_2O$ phase, temperature and particle size which makes the GFS sensitive to temperature errors.

## 7.2   Regional NWP Model

Unlike global NWP models, regional NWP model only a a sub-domain of the global space, see figure 7.2.1 . 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) model. Details of each are discussed in the following sections.

### 7.2.1   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 13km x 13km horizontally resolved forecasts with 50 atmospheric layers out to a time horizon of 18 hours. The RUC loop time steps are 1 hour from time of analysis out to 18 hours. 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 assumed Rayleigh scattering which failed to capture the inversely proportional relationship between intensity of scattering and wavelength of radiation. In addition, only downwelling irradiances were attenuated which sometimes lead to the underestimation of diffuse irradiance
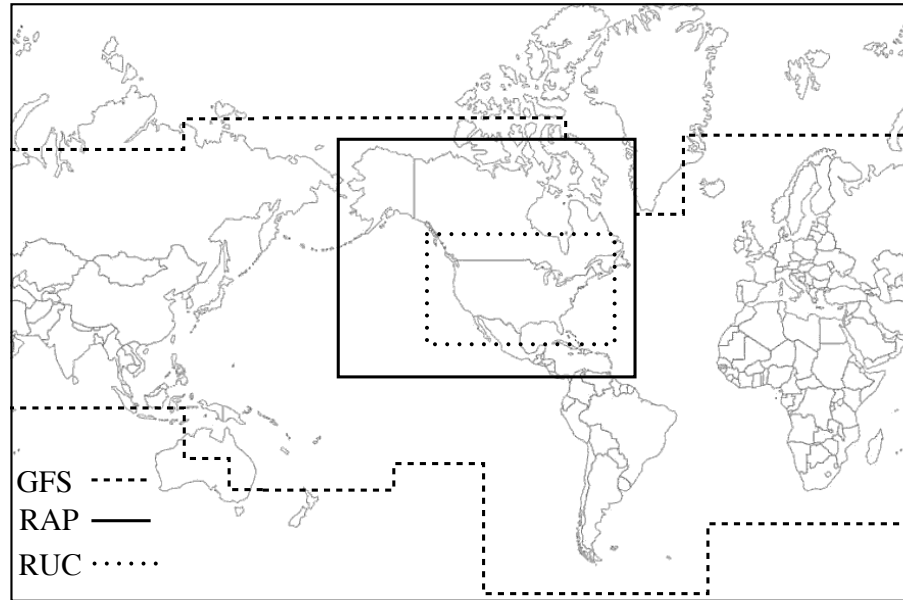
**Figure 7.2**: Domains of the Global Forecast System (GFS), RAPid refresh (RAP) and Rapid Update Cycle (RUC) models.

due to backscattering [197].

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 posses the same spatial and atmospheric resolution (12km x 12km, 50 layers) but it based on a new rapid update configuration of the WRF model (see §7.0.3). 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 Figure 9.2.

# Part IV

# Hybrid Systems & Conclusion

# Chapter 8

# HybridSystems

As of late, several HSs have been used to produce high quality forecasts. A HS is characterized by a combination of any two or more of the methods described in this paper.

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 [69]. Zarzalejo et al. used ANN in conjunction with satellite-derived cloud indexes for the forecasting of hourly mean GHI [183]. 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 [72]. 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 a 5 minutes ahead based on stochastic models and Kalman filtering [68]. Reikard preformed forecasting tests using regressions in logs, ARIMA, and unobserved components models, transfer functions, neural networks and hybrid models [80]. 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 [35]. Mellit et al. also developed an adaptive model for predicting hourly GHI and DNI using a dataset of measured air temperature, relative humidity, direct, diffuse and global horizontal irradiance

[198]. The adaptive model's performance was compared agains a MLP ANN. Ji and Chee [199] 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 Autoregressive and Moving Average (ARMA) model is used to predict the stationary residual series and a time delay neural network is applied also applied. Voyant et al. [71] proposes a study of the contribution of exogenous meteorological data as time series to their optimized MLP. comparisons with different forecasting methods including a persistence model, ARIMA reference predictor, an ANN with preprocessing using only endogenous inputs and an ANN with preprocessing using endogenous and exogenous inputs.

More recently, Marquez and Coimbra [200] developed and validated a merdium-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 in input selector to their ANN in order to select the most relevant input patters. Voyant et al. used a hybrid ARMA/ANN with NWPs to predict hourly mean GHI [201]. Marquez et al. [202] also used cloud indexes obtained from a TSI as well as cloud indexes derived from infrared radiometric measurements in order to improve result for hourly forecasts of GHI. Marquez, Pedro and Coimbra have developed a forecast based on satellite images and ANNs for time series predictions of GHI up to 2 hours ahead [203].

# Chapter 9

# Conclusion

Several methods for solar irradiance forecasting have been covered in this paper. From the description of the numerous approaches presented, it can be seen that several successful forecasts have been developed for a number of different spatial and temporal resolutions.

Stochastic 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 utilized 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 which case exogenous inputs can be included with models like the ARMAX and/or ARIMAX. These methods have been applied to a wide variety of time horizons ranging from intra-hour to yearly averages, see Chapter 3.

ANNs offer improved performance and provide an alternative approach to physical modeling of irradiance data. ANNs no not require knowledge of the internal system parameters and present compact solutions for several non-linear, stochastic and multivariate problems. Like the stochastic methods, ANNs are not typically temporally limited and successfully modeled irradiance on intra-hour to yearly time horizons, see Chapter 4.

Satellite imaging provides an alternative to expensive ground based pyra-

nometer networks and allows 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 Chapter 5.

NWPs also allow the prediction of irradiance data in locations were no data is available. NWPs are scaleable 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 offer the most attractive option for long term forecasting with time-horizons ranging from intra-hour out to a week or more, see Chapter 7.

For shorter time horizons, ground based imaging or wireless network systems have seen rather limited application, put appear very promising. The use of TSIs allows for the acquisition of local meteorological conditions which provides information on higher frequency fluctuations of irradiance with time horizons ranging from 3 - 30 minutes. Intra-minute forecasting is difficult for the TSI due to circum-solar scattering of light and the use of shadow-bands. A solution to intra-minute forecasting is provided thought the distribution of low-cost light sensors throughout the PV plant in order to determine the effect of passing local clouds fields on power output, see Chapter 6. Recently, various HSs have been presented. These offer increased coverage in the spatial and temporal coverage of forecasting techniques. Integration of NWP/satellite models with stochastic/AI techniques results in higher accuracy long-term forecasts. By the same reasoning, assimilating local meteorological information from TSIs and WNSs with stochastic/AI methods allows higher fidelity access to intra-hour time horizons.

Thus, one can see that if a high fidelity and robust solar irradiance forecast engine that span the entire spectrum of temporal and spatial horizons, from intra-minute to multiple days and from single point radiometers to continental regions, a wide variety of methods must be used. Figure 9 shows a comparison of time horizon and spatial resolution where the time horizon is plotted in seconds and the spatial resolution is given in $1/m^2$. Solid lines indicate current span of respective technologies while dashed lines indicate the future coverage based on the potential of current work.
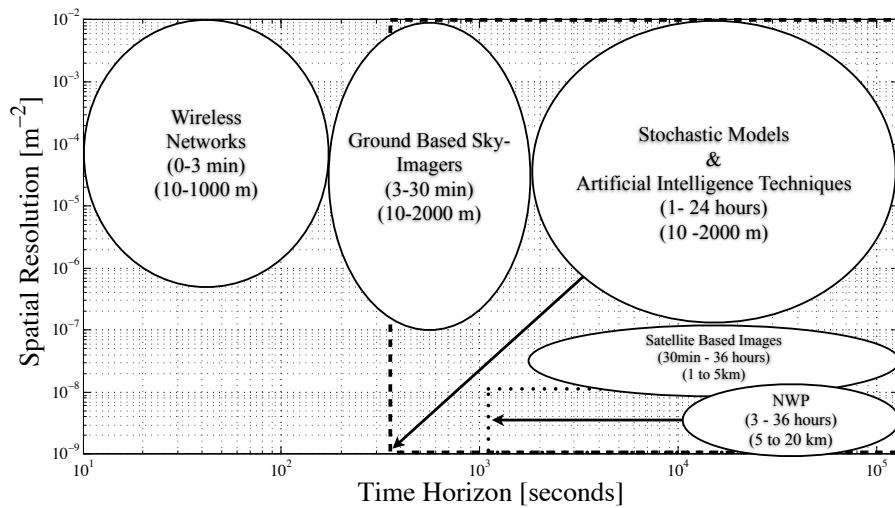
**Figure 9.1**: Comparison of time horizon and spatial resolution. Solid lines indicate current limits of techniques while the dashed lines and arrows indicate the future progress of work. AI techniques will continue to include local, mesoscale and global meteorological data which will allow for both 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.

# Bibliography

[1] R. Marquez and C. F. M. Coimbra, "A Proposed Metric for Evaluation of Solar Forecasting Models," *ASME Journal of Solar Energy Engineering (in press)*, 2012.

[2] J. D. Tarpley, "Estimating incident solar radiation at the surface from geostationary satellite data," *Journal of Applied Meteorology*, vol. 18, no. 9, pp. 1172–1181, 1979.

[3] C. Raphael and J. E. Hay, "An assessment of models which use satellite data to estimate solar irradiance at the earth's surface," *Joural of Climate and Applied Meteorology*, vol. 23, no. 5, pp. 832–844, 1984.

[4] C. G. Justus, M. V. Paris, and J. D. Tarpley, "Satellite-measured insolation in the united states, mexico, and south america," *Remote Sensing of Environment*, vol. 20, pp. 57–83, 1986.

[5] V. Badescu, *Modeling Solar Radiation at the Earth Surface.* Berlin Heidelberg: Springer-Verlag, 2008.

[6] P. Denholm and R. M. Margolis, "Evaluating the limits of solar photovoltaics (PV) in traditional electric power systems," *Energy Policy*, vol. 35, no. 5, pp. 2852 – 2861, 2007.

[7] P. Denholm and R. M. Margolis, "Evaluating the limits of solar photovoltaics (PV) in electric power systems utilizing energy storage and other enabling technologies," *Energy Policy*, vol. 35, no. 9, pp. 4424 – 4433, 2007.

[8] C. ISO, "Flexible ramping constraint." Feb. 2010.

[9] D. Lew and R. Piwko, "Western wind and solar integration study," tech. rep., National Renewable Energy Labratories, 2010.

[10] V. Sundar, "Integration of renewable resources: Operational requirements and feneration fleet capability at 20 percent rps," tech. rep., California Independent System Operator (CAISO), 2010.

[11] G. D. Rodriguez, "A utility perspective of the role of energy storage in the smart grid," in *Power and Energy Society General Meeting, IEEE*, pp. 1–2, 2010.

[12] P. Ineichen, "Comparison of eight clear sky broadband models against 16 independent data banks," *Solar Energy*, vol. 80, no. 4, pp. 468–478, 2006.

[13] R. Mueller, K. Dagestad, P. Ineichen, M. Schroedter-Homscheidt, S. Cros, D. Dumortier, R. Kuhlemann, J. Olseth, G. Piernavieja, C. Reise, L. Wald, and D. Heinemann, "Rethinking satellite-based solar irradiance modelling - The SOLIS clear-sky module," *Remote Sensing of Environment*, vol. 91, pp. 160–174, MAY 30 2004.

[14] C. Rigollier, O. Bauer, and L. Wald, "On the clear sky model of the ESRA - European Solar Radiation Atlas - With respect to the Heliosat method," *Solar Energy*, vol. 68, pp. 33–48, JAN 2000.

[15] M. Geiger, L. Diabate, L. Menard, and L. Wald, "A web service for controlling the quality of measurements of global solar irradiation," *Solar Energy*, vol. 73, no. 6, pp. 475–480, 2002.

[16] L. F. Zarzalejo, J. Polo, and L. Ramirez, "Gc_model5_irradiance," 2004. (Matlab Computer Program) CD-ROM accompanying [5].

[17] F. Kasten, "The Linke turbidity factor based on improved values of the integral Rayleigh optical thickness," *Solar Energy*, vol. 56, pp. 239–244, MAR 1996.

[18] R. E. Bird and R. L. Huldstrom, "Direct insolation models," *Trans. ASME J. Sol. Energy Eng.*, vol. 103, pp. 182–192, 1980.

[19] R. A. McClatchey and J. E. Selby, "Atmospheric transmittance from 0.25 to 28.5 $\mu$m: computer code lowtran2," environ. res. paper no. 427, Airforce Cambridge Research Labratories, 1972.

[20] B. Molineaux, P. Ineichen, and N. O'Neill, "Equivalence of pyrheliometric and monochromatic aerosol optical depths at a single key wavelength," *Applied Optics*, vol. 37, pp. 7008–7018, OCT 20 1998.

[21] A. Berk, L. S. Berntein, and D. C. Robertson, "Modtran: a moderate resolution model for lowtran 7. gl-tr-89-0122." updated and commercialized by Ontar Crp, 1996, 1989.

[22] C. Gueymard, "Smarts2, a simple model for atmospheric radiative transfer of sunshine: algorithms and performance assessment. Rep. FSEC-PF-270-95," tech. rep., Florida Solar Energy Center, Cape Canaveral, Florida, 1995.

[23] P. Ineichen and R. Perez, "A new airmass independent formulation for the Linke turbidity coefficient," *Solar Energy*, vol. 73, no. 3, pp. 151–157, 2002.

[24] C. Gueymard, "A 2-band model for the calculation of clear sky solar irradiance, illuminance, and photosynthetically active radiation at the earths surface," *Solar Energy*, vol. 43, no. 5, pp. 253–265, 1989.

[25] C. Gueymard, "High performance model for clear sky irradiance and illuminance," in *ASES Conference*, pp. 251–258, 2004.

[26] F. Kasten, "A simple parameterization of two pyrheliometric formulae for determining the linke turbidity factor," *Meteorol. Rundsch.*, vol. 33, pp. 124–127, 1980.

[27] F. Kasten, "Parametrisierung der globaslstrahlung durch bedekungsgrad und trübungsfaktor," *Annalen der Meteorologie, Neue Folge*, vol. 20, pp. 49–50, 1984.

[28] R. L. Snyder, "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] A. Mills and R. Wiser, "Implications of wide-area geographic diversity for short-term variability of solar power," tech. rep., Lawrence Berkeley National Laboratory, 2010.

[30] T. E. Hoff and R. Perez, "Modeling pv fleet output variability," *Solar Energy, in press, doi: 10.1016/j.silener.2011.11.005*, 2011.

[31] M. Lave and J. Kleissl, "Solar variability of four sites across the state of Colorado," *Renewable Energy*, vol. 35, no. 12, pp. 2867–2873, 2010.

[32] T. E. Hoff and R. Perez, "Quantifying PV power output variability," *Solar Energy*, vol. 84, pp. 1782–1793, Oct 2010.

[33] A. Mellit, "Artificial intelligence technique for modelling and forecasting of solar radiation data: a review," *International Journal of Artificial Intelligence and Soft Computing*, pp. 52–76, 2010.

[34] R. Perez, S. Kivalov, J. Schlemmer, K. Hemker, D. Renne, and T. E. Hoff, "Validation of short and medium term operational solar radiation forecasts in the US," *Solar Energy*, vol. 84, no. 5, pp. 2161–2172, 2010.

[35] L. Martin, L. F. Zarzalejo, J. Polo, A. Navarro, R. Marchante, and M. Cony, "Prediction of global solar irradiance based on time series analysis: application to solar thermal power pants energy production planning," *Solar Energy*, vol. 84, no. 10, pp. 1772–1781, 2010.

[36] H. E. Lansdberg, "Solar radiation at the earths surface," *Solar Energy*, vol. 5, no. 3, pp. 95–98, 1961.

[37] A. Whillier, "Solar radiation graphs," *Solar Energy*, vol. 9, no. 3, pp. 164–164, 1965.

[38] R. K. Swartman, "A statistical relationship between solar radiation, sunshine and relative humidity in the tropics," *Canadian Meteorological Society*, vol. 5, no. 2, pp. 25–34, 1970.

[39] R. K. Swartman, "Correlation of solar radiation with common parameters in toronto, canada," *Solar Energy*, vol. 13, no. 3, pp. 345–345, 1971.

[40] I. A. Lund, "Relationships between insolation and other surface weather observations at blue hill massachusetts," *Solar Energy*, vol. 12, no. 1, pp. 95–95, 1968.

[41] D. Yogi, F. Kreith, and J. F. Kreider, *Principles of Solar Engineering*. 270 Madison Ave, New York, NY 10016: Taylor and Francis, 2 ed., 2000.

[42] R. Walpole and R. Meyers, *Probability and statistics for engineers and scientists*. New York: Macmillian, 1972.

[43] A. Wold, *Analysis of stationary time series*. Uppsala: Almquist and Wicksell, 1938.

[44] G. U. Yule, "On the time-correlation problem, with special reference to the variate-difference correlation method," *Journal of the Royal Statistical Society*, vol. 84, pp. 497–537, July 1921.

[45] G. U. Yule, "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*, vol. 89, pp. 1–69, Jan. 1926.

[46] G. U. Yule, "On a method of investigating periodicities in disturbed series, with special reference to wofer's sunspot numbers," *Philosophical Transactions of the Royal Society of London Series A-Containing Papers of a Mathematical or Physical Character*, vol. 226, pp. 267–298, July 1927.

[47] G. E. P. Box and G. M. Jenkins, *Time series analysis, forecasting and control*. Wiley, 1998.

[48] M. Sulaiman, W. Hlaing, M. Wahab, and Z. Sulaiman, "Analysis of residuals in daily solar radiation time series," *Renewable Energy*, vol. 11, pp. 97–105, May 1997.

[49] L. Ljung, *System Identification: Theory for the User*. Prentice-Hall, 1987.

[50] T. N. Goh and K. J. Tan, "Stochastic modeling and forecasting of solar-radiation data," *Solar Energy*, vol. 19, no. 6, pp. 755–757, 1977.

[51] E. Boileau, "Discussion d'un modéle statistique en météorologie solaire," *Rev. Phys. Appl*, vol. 14, pp. 145–153, 1979.

[52] E. Boileau and B. Guerrier, "Comparaison de modéles statistiques saisonniers et non saisonniers en météorologie solaire," *La Météorologie*, vol. VI, pp. 115–130, 1979.

[53] B. Guerrier, E. Boileau, and C. Bernard, "Analyse statistique temporelle de l'irradiation solaire globale quotidienne: modélisation d'une variale réduite á l'aide de modéles statistiques a.r.m.a.," *Rev. Phys. Appl.*, vol. 15, pp. 93–102, 1980.

[54] B. Guerrier, *Modélisation appliquée á l'étude des entrées-sorties de systémes thermiques: cas des composants linéaires de l'habitat solaire passif.* PhD thesis, Paris, 1981.

[55] B. J. Brinkworth, "Autocorrelation and stochastic modeling of insolation sequences," *Solar Energy*, vol. 19, no. 4, pp. 343–347, 1977.

[56] C. Benard, A. Wirgin, Y. Body, and D. Gobin, "Caractérisation de la stabilité de l'intensité solaire par l'analyse de séries aléatoires in france et au pérou," *Le Métérologie*, vol. VI, pp. 53–71, 1978.

[57] E. Boileau, "Use of some simple statistical-models in solar meteorology," *Solar Energy*, vol. 30, no. 4, pp. 333–339, 1983.

[58] C. Mustacchi, V. Cena, and M. Rocchi, "Stochastic simulation of hourly global radiation sequences," *Solar Energy*, vol. 23, no. 1, pp. 47–51, 1979.

[59] G. Boch, E. Boileau, and C. Benard, "Search for a statistical-model of hourly integrated solar irradiation," *Revue de Physique Appliquee*, vol. 16, no. 5, pp. 237–247, 1981.

[60] C. Benard, E. Boileau, and B. Guerrier, "Modeling of the global solar radiant exposure with arma processes - application to short-time prediction (hourly), for the calculation of optimal controls in bulidings," *Revue de Physique Appliquee*, vol. 20, pp. 845–855, Dec. 1985.

[61] S. Hokoi, M. Matsumoto, and T. Ihara, "Statistical time-series models of solar-radiation and outdoor temperature - identification of seasonal models by kalman filter," *Energy and Buildings*, vol. 15, no. 3-4, pp. 373–383, 1991.

[62] H. Yoshida and T. Terai, "An arma type weather model for air-conditioning, heating and cooling load calculation," *Energy and Buildings*, vol. 16, no. 1-2, pp. 625–634, 1991.

[63] R. J. Aguiar and M. Collares-Pereira, "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* (T. Horigome, K. Kimura, and T. Takakura, eds.), vol. 3, (Kobe City, Japan), pp. 2107–2111, September 1989.

[64] R. Aguiar and M. Collares-Pereira, "Tag - a time-dependent, autoregressive, gaussian model for generating synthetic hourly radiation," *Solar Energy*, vol. 49, pp. 67–67, SEP 1992.

[65] S. Al-Awadhi and N. El-Nashar, "Stochastic modelling of global solar radiation measured in the state of kuwait," *Environmetrics*, vol. 13, pp. 751–758, NOV 2002.

[66] L. L. Mora-Lopez and M. Sidrach-De-Cardona, "Multiplicative arma models to generate hourly series of global irradiation," *Solar Energy*, vol. 63, pp. 283–291, Nov. 1998.

[67] A. Moreno-Muñoz, J. J. G. de la Rosa, R. Posadillo, and F. Bellido, "Very short term forecasting of solar radiation," in *Photovoltaic Specialists Conference, 33rd IEEE*, pp. 1–5, May 2008.

[68] M. Chaabene and M. Ben Ammar, "Neuro-fuzzy dynamic model with kalman filter to forecast irradiance and temperature for solar energy systems," *Renewable Energy*, vol. 33, pp. 1435–1443, Jul. 2008.

[69] A. Sfetsos and A. H. Coonick, "Univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques," *Solar Energy*, vol. 68, pp. 169–178, Feb. 2000.

[70] J. Wu and C. K. Chan, "Prediction of hourly solar radiation using a novel hybrid model of arma and tdnn," *Solar Energy*, vol. 85, pp. 808–817, May 2011.

[71] C. Voyant, M. Muselli, C. Paoli, and M.-L. Nivet, "Optimization of an artificial neural network dedicated to the multivariate forecasting of daily global radiation," *Energy*, vol. 36, pp. 348–359, Jan. 2011.

[72] A. Mellit, M. Benghanem, A. Arab, and A. Guessoum, "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*, vol. 79, no. 5, pp. 469–482, 2005.

[73] J. A. K. Suykens, J. P. L. Vandewalle, and B. L. R. Moor, *Artificial neural networks for modelling and control of non-linear systems*. P.O. Box 17, 3300 AA Dordrecht, The Netherlands: Kluwer Academic, 1996.

[74] C. Craggs, E. M. Conway, and N. M. Pearsall, "Statistical investigation of the optimal averaging time for solar irradiance on horizontal and vertical surfaces in the uk," *Solar Energy*, vol. 68, pp. 179–187, Feb. 2000.

[75] J. M. Santos, J. M. Pinazo, and J. Canada, "Methodology for generating daily clearness index index values k-t starting from the monthly average daily value (k)over-bar(t). determining the daily sequence using stochastic models," *Renewable Energy*, vol. 28, pp. 1523–1544, Aug. 2003.

[76] O. Karner, "On nonstationarity and antipersistency in global temperature series," *Journal of Geophysical Research-Atmospheres*, vol. 107, Sep.-Oct. 2002.

[77] K. Yüerekli, H. Simsek, B. Cemek, and S. Karaman, "Simulating climatic variables by using stochastic approach," *Building and Environment*, vol. 42, pp. 3493–3499, Oct. 2007.

[78] G. Stanhill and S. Cohen, "Solar radiation changes in japan during the 20th century: Evidence from sunshine duration measurements," *Journal of the Meteorological Society of Japan*, vol. 86, pp. 57–67, Feb. 2008.

[79] O. Kärner, "ARIMA representation for daily solar irradiance and surface air temperature time series," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 71, pp. 841–847, Jun. 2009.

[80] G. Reikard, "Predicting solar radiation at high resolutions: A comparison of time series forecasts," *Solar Energy*, vol. 83, pp. 342–349, Mar. 2009.

[81] R. Perdomo, E. Banguero, and G. Gordillo, "Statistical modeling for global solar radiation forecasting in bogotá," in *Photovoltaic Specialists Conference (PVSC), 2010 35th IEEE*, pp. 002374–002379, 2010.

[82] C. S. Krishnamoorthy and S. Rajeev, *Artificial Intelegence and Expert Systems for Engineers.* CRC Press, 1996.

[83] D. Crevier, *AI: The tumultuous search for artificial intelligence.* New York, NY: Basic Books, 1993.

[84] B. Russel and A. N. Whitehead, *Principia mathematica*, vol. 1-3. The University Press, 2 ed., 1925.

[85] C. Shannon, "Programming a computer for playing chess," *Philosophical Magazine*, pp. 256–275, 1950.

[86] A. Mellit and S. A. Kalogirou, "Artificial intelligence techniques for photovoltaic applications: A review," *Progress in Energy and Combustion Science*, vol. 34, pp. 574–632, Oct. 2008.

[87] A. Newell, J. Shaw, and H. Simon, "Report on a general problem-solving program," in *Proceedings of the International Conference on Information Processing*, pp. 256–264, 1959.

[88] J. McCarthy, "Recursive functions of symbolic expressions and their computation by machine," *Communications of the Association for Computing Machinery*, vol. 7, pp. 184–195, 1960.

[89] J. Weizenbaum, "Eliza - a computer program for the study of natural language communication between man and machine," *Communications of the Association for Computing Machinery*, vol. 9, pp. 36–45, 1966.

[90] D. G. Bobrow, "Natural language input for a computer problem solving system," in *Semantic Information Processing* (M. Minsky, ed.), (Cambridge), pp. 133–215, MIT Press, 1968.

[91] T. Winograd, *Understanding Natural Language*. New York: Academic Press, 1972.

[92] M. Minsky, "A framework for representing knowledge," *The Psychology of Computer Vision*, 1975.

[93] M. Minsky and S. Papert, *Perceptrons*. Cambridge, MA: MIT Press, 1969.

[94] J. Stender and T. Addis, eds., *Symbols versus neurons?* Amsterdam: IOS Press, 1990.

[95] W. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bulletin of Mathematical Biology*, vol. 5, pp. 115–133, 1943. 10.1007/BF02478259.

[96] D. O. Hebb, *The organization of behavior*. New York: Wiley, 1949.

[97] R. Rosenblatt, *Principles of neurodynamics*. New York: Spartan Books, 1959.

[98] O. G. Selfridge, "Pandemonium: a paradigm for learning in mechanisation of thought processes," in *Symposium on Mechanisation of Thought Process* (D. Blake and A. Uttley, eds.), (H. M. Stationary Office, London), pp. 513–526, 1959.

[99] B. Widrow and M. E. Hoff, "Adaptive switching circuits," in *IRE Wescon*, (New York), pp. 96 – 104, 1960.

[100] J. H. Holland, *Adaptation in natural and artificial systems*. The University of Michigan Press, 1975.

[101] D. E. Goldberg, *Genetic algorithms in search, optimisation and learning.* Reading, Mass.: Addison-Wesely Publishing Co., 1989.

[102] J. McCarthy, "Circumscription - a form of non-monotonic reasoning," *Artificial Intelligence*, vol. 13, no. 1-2, pp. 27–39, 1980.

[103] J. Haugeland, ed., *Artificial intelligence: the very idea.* Cambridge: MIT Press, 1985.

[104] R. Kurzweil, *The age of intelligent machines.* Cambridge: MIT Press, 1990.

[105] A. Newell and H. A. Simon, *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall, 1972.

[106] K. L. Priddy and P. E. Keller, *Artificial Neural Networks: An Introduction.* SPIE Press, 2005.

[107] D. Graupe, "Principles of artificial neural networks," in *Advanced Series in Circuts and Systems* (W. Chen and D. A. Mlynski, eds.), vol. 3, ch. 1 - 12, pp. 4–189, World Scientific, 1 ed., 1997.

[108] ACM National Meeting, *Generalization of learning in a michine*, vol. 14th, Septermber 1959.

[109] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proceeding of the National Academy of Sciences*, vol. 79, pp. 2554 – 2558, 1982.

[110] B. Widrow and R. Winter, "Neural nets for adaptive filtering and adaptive pattern recognition," *Computer*, vol. 21, pp. 25–39, 1988.

[111] D. E. Rummelhart and J. L. McClelland, "An interactive activation model of the effect of context in language learning," *Psychological Review*, vol. 89, pp. 60 – 94, 1986.

[112] R. Hecht-Nielsen, "Counter propagation networks," *Applied Optics*, vol. 26, pp. 4979 – 4984, 1987.

[113] L. O. Chua and L. Yang, "Cellular neural networks: Theory and applications," *IEEE Transactions on Circuits and Systems*, vol. 35, no. 10, pp. 1257–1290, 1988.

[114] D. W. Patterson, *Artificial neural networks: Theory and Applications.* Prentice-Hall, 1996.

[115] E. B. Carne, *Artificial Intelligence techniques.* Washington D.C.: Spartan Press, 1965.

[116] D. Hilbert, "Mathematische probleme," *Göttinger Nachrichten*, pp. 253–297, 1900.

[117] A. N. Kolmogorov, "The representation of continuous functions of many variables by superposition of continuous functions of one variable and addition," *Doklady Akademii NAUK SSSR*, vol. 114, no. 5, pp. 953–956, 1957.

[118] D. A. Sprecher, "On the structure of continuous function of several variables," *Transactions of the American Mathematical Society*, vol. 115, no. 340-355, 1965.

[119] K. I. Funahashi, "On the approximate realization of continuous mappings by neural networks," *Neural Networks*, vol. 2, pp. 183–192, 1989.

[120] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networkd are universal approcixators," *Neural Networks*, vol. 2, pp. 359–366, 1989.

[121] G. Cybenko, "Approximations by Superpositions of a Sigmoidal Function," *Mathematics of Control, Signals and Systems*, vol. 2, pp. 183–192, 1989.

[122] P. J. Werbos, *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. PhD thesis, Harvard University, 1974.

[123] D. B. Parker, "Learning logic, technical report tr-47," *Center for Computational Research in Economics and Management Science, MIT*, 1985.

[124] D. J. Rumelhart and D. Zipser, "Feature discovery by competitive learning," *Cognitive Science*, vol. 9, pp. 75–112, 1985.

[125] S. Al-Alawi and H. Al-Hinai, "An ANN-based approach for predicting global radiation in locations with no direct measurement instrumentation," *Renewable Energy*, vol. 14, pp. 199 – 204, 1998.

[126] S. Cao and J. Cao, "Forecast of solar irradiance using recurrent neural networks combined with wavelet analysis," *Applied Thermal Engineering*, vol. 25, no. 2–3, pp. 161 – 172, 2005.

[127] J. Cao and S. Cao, "Study of forecasting solar irradiance using neural networks with preprocessing sample data by wavelet analysis," *Energy*, vol. 31, no. 15, pp. 3435 – 3445, 2006.

[128] L. Hontoria, J. Aguilera, J. Riesco, and P. Zufiria, "Recurrent Neural Supervised Models for Generating Solar Radiation Synthetic Series," *Journal of Intelligent & Robotic Systems*, vol. 31, pp. 201–221, 2001.

[129] A. Sözen, E. Arcaklıoğlu, M. Özalp, and N. Çağlar, "Forecasting based on neural network approach of solar potential in Turkey," *Renewable Energy*, vol. 30, no. 7, pp. 1075 – 1090, 2005.

[130] J. Cao and X. Lin, "Study of hourly and daily solar irradiation forecast using diagonal recurrent wavelet neural networks," *Energy Conversion and Management*, vol. 49, pp. 1396–1406, JUN 2008.

[131] J. Cao and X. Lin, "Application of the diagonal recurrent wavelet neural network to solar irradiation forecast assisted with fuzzy technique," *Engineering Applications of Artificial Intelligence*, vol. 21, no. 8, pp. 1255 – 1263, 2008.

[132] A. Yona, T. Senjyu, A. Saber, T. Funabashi, H. Sekine, and C.-H. Kim, "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*, pp. 1 –6, July 2008.

[133] A. Chaouachi, R. M. Kamel, R. Ichikawa, H. Hayashi, and K. Nagasaka, "Neural Network Ensemble-based Solar Power Generation Short-Term Forecasting.," *World Academy of Science, Engineering & Technology*, vol. 54, pp. 54 – 59, 2009.

[134] C. Paoli, C. Voyant, M. Muselli, and M.-L. Nivet, "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), pp. 898–907, Springer-Verlag, 2009.

[135] C. Paoli, C. Voyant, M. Muselli, and M.-L. Nivet, "Forecasting of preprocessed daily solar radiation time series using neural networks," *Solar Energy*, vol. 84, no. 12, pp. 2146 – 2160, 2010.

[136] A. Azadeh, A. Maghsoudi, and S. Sohrabkhani, "An integrated artificial neural networks approach for predicting global radiation," *Energy Conversion and Management*, vol. 50, no. 6, pp. 1497 – 1505, 2009.

[137] A. Mellit and A. M. Pavan, "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*, vol. 84, no. 5, pp. 807 – 821, 2010.

[138] A. Mellit, H. Mekki, A. Messai, and S. Kalogirou, "FPGA-based implementation of intelligent predictor for global solar irradiation, Part I: Theory and simulation," *Expert Systems with Applications*, vol. 38, no. 3, pp. 2668 – 2685, 2011.

[139] C. Chen, S. Duan, T. Cai, and B. Liu, "Online 24-h solar power forecasting based on weather type classification using artificial neural network," *Solar Energy*, vol. 85, no. 11, pp. 2856 – 2870, 2011.

[140] D. V. Fiorin, F. R. Martins, N. J. Schuch, and E. B. Pereira, "Forecast of solar energy resource by using neural network methods," *Revista Brasileira de Ensino de Física*, vol. 33, no. 1, pp. 1–20, 2011.

[141] S. A. Kalogirou, "Applications of artificial neural-networks for energy systems," *Applied Energy*, vol. 67, no. 1–2, pp. 17 – 35, 2000.

[142] F. Kogan, L. Salazar, and L. Roytman, "Forecasting crop production using satellite-based vegetation health indices in kansas, usa," *International Journal of Remote Sensing*, vol. 33, no. 9, pp. 2798–2814, 2012.

[143] C. A. Reynolds, "Real-time hydrology operations at usda for monitoring global soil moisture and auditing national crop yield estimates," *Satellite Rainfall Applications for Surface Hydrology*, pp. 267–286, 2010.

[144] D. L. Toll, B. Doorn, E. T. Engman, and R. G. Lawford, "Using NASA products of the water cycle for improved water resources management," *American Geophysical Union, Fall Meeting*, no. H34D-08, 2010.

[145] P. Zhant, B. Anderson, M. Barlow, B. Tan, and R. B. Myneni, "Application of a satellite-based climate-variability impact index for crop yield forecasting in drought-stricken regions," *African Journal of Plant Science*, vol. 4, pp. 82–94, April 2010.

[146] R. Perez, R. Seals, R. Stewart, and A. Zelenka, "Using satellite-derived insolation data for the site/time specific simulations of solar energy systems," *Solar Energy*, vol. 53, no. 6, pp. 491–495, 1994.

[147] A. Zelenka, G. Czeplak, V. D'Agostino, W. Josefsson, E. Maxwell, and R. Perez, "Techniques for supplementing solar radiation network data," *Final Report of International Energy Agency Solar Heating and Cooling Program, Task 9, subtask 9D*, 1992.

[148] R. H. Goddardd, "A method of reaching extreme altitudes," in *Smithsonian Miscellaneous Collections*, vol. 71, 1919.

[149] W. L. Smith, W. P. Bishop, V. F. Dvorak, C. M. Hayden, J. H. McElroy, F. R. Mosher, V. J. Oliver, J. F. Purdom, and D. Q. Wark, "The meteorological satellite: Overview of 25 years of operation," *Science*, vol. 231, pp. 455–462, Jan 1986.

[150] W. W. Vaughan, "2227," in *NASA Conference Publication*, 1981.

[151] W. W. Kellogg, "2257," in *NASA Conference Publication*, 1982.

[152] R. A. Hanel, "Radiometric measurements from satellites," *Aerospace Engineering*, vol. 21, pp. 34–39, 1962.

[153] C. Kidd, L. Vincenzo, and P. Baur, "A review of satellite meteorology and climatology at the start of the twenty-first century.," *Progress in Physical Geography*, vol. 33, no. 4, pp. 474–489, 2009.

[154] T. J. Schmit, M. M. Gunshor, W. P. Menzel, J. J. Gurka, J. Li, and A. S. Bachmeier, "Introducing the Next-Generation Advanced Baseline Imager on GOES-R," *Bulletin of the American Meteorological Society*, vol. 86, no. 8, pp. 1079–1096, 2005.

[155] W. P. Menzel and J. F. W. Purdom, "Introducing GOES-I: The First of a New Generation of Geostationary Operational Environmental Satellites," *Bulletin of the American Meteorological Society*, vol. 75, no. 5, pp. 757–781, 1994.

[156] K. J. Hanson, T. H. Vonder Harr, and V. E. Suomi, "Reflection of sunlight to space and absorption by the earth and atmosphere over the United States durring spring 1962," *Monthly Weather Review*, vol. 95, no. 6, pp. 354–362, 1967.

[157] A. Danjon, "Nouvelles researches sur la photométrie de la lumière cendrée et l'albedo de la terre," *Annales de l'Observatoire de Strasbourg*, vol. 3, pp. 139–181, 1936.

[158] J. Dubois, "La variation de brilliance de la lumière cendrée de la lune au cours du dernier cycle solarie," *L'Astronomie*, vol. 69, pp. 242–246, 1955.

[159] A. Angstrom, "Atmospheric turbidity, global illuminationa and planetary albedo of the earth," *Tellus*, vol. 14, pp. 435–450, Nov 1962.

[160] G. C. Simpson, "The distribution of terrestrial radiation," *Memoirs of the Royal Meteorological Society*, vol. 3, no. 23, pp. 53–78, 1929.

[161] H. G. Houghton, "On the annual heat balance of the northern hemisphere," *Journal of Meteorology*, vol. 11, pp. 1–9, Feb 1954.

[162] J. London, "A study of the atmospheric heat balance, final report, contract no. af19(122)-165," *Department of Meteorology and Oceanography, New York University*, pp. 99–138, 2957.

[163] M. Noia, C. Ratto, and R. Festa, "Solar irradiance estimation from geostationary satellite data: I. statistical models," *Solar Energy*, vol. 51, no. 6, pp. 449 – 456, 1993.

[164] M. Noia, C. Ratto, and R. Festa, "Solar irradiance estimation from geostationary satellite data: II. Physical models," *Solar Energy*, vol. 51, no. 6, pp. 457 – 465, 1993.

[165] C. Gautier, G. Diak, and S. Masse, "A Simple Physical Model to Estimate Incident Solar Radiation at the Surface from GOES Satellite Data," *Journal of Applied Meteorology*, vol. 19, no. 8, pp. 1005–1012, 1980.

[166] C. Raphael, "Models for estimating solar irradiance at the Earth's surface from satellite data: An initial assessment, Technical Report, Atmospheric Environment Service." (Unpublished Manucript), 1983.

[167] K. Coulson, "Characteristics of the radiation emerging from the top of a rayleigh atmosphere—I: Intensity and polarization," *Planetary and Space Science*, vol. 1, no. 4, pp. 265 – 276, 1959.

[168] K. Coulson, "Characteristics of the radiation emerging from the top of a Rayleigh atmosphere—II: Total upward flux and albedo," *Planetary and Space Science*, vol. 1, no. 4, pp. 277 – 284, 1959.

[169] G. W. Paltridge, "Direct measurements of water vapour absorption of solar radiation in the three atmosphere," *Journal of Applied Meteorology*, vol. 30, pp. 156–160, 1973.

[170] C. Gautier and R. Frouin, "Downward longwave irradiance at the ocean surface using satellite data: Methodology and in situ validation," *Journal of Geophysical Research*, vol. 93, pp. 597–598, 1985.

[171] D. Tanrè, M. Herman, P. Y. Deschamps, and A. DeLeffe, "Atmospheric modeling for space measurements of fround relectances, including bidirectional properties," *Applied Optics*, vol. 18, pp. 3587–3594, 1979.

[172] S. Marullo, G. Dalu, and A. Viola, "Incident short-wave radiation at the surface from METEOSAT data," *Il Nuovo cimento della Società italiana di fisica*, vol. 10, no. 1, pp. 77 – 90, 1987.

[173] J. Schmetz, "On the parameterization of the radiative properties of broken clouds," *Tellus. Series A, Dynamic meteorology and oceanography*, vol. 36A, no. 5, pp. 417–417, 1984.

[174] W. Möser and E. Raschke, "Incident Solar Radiation over Europe Estimated from METEOSAT Data," *Journal of Climate and Applied Meteorology*, vol. 23, no. 1, pp. 166–170, 1984.

[175] M. Kerschegens, U. Pilz, and E. Raschke, "A modified two stream approcimation for computations of the solar radiation budget in a couldy atmosphere," *Tellus. Series A, Dynamic meteorology and oceanography*, vol. 30, pp. 429–429, 1978.

[176] G. Dedieu, P. Y. Deschamps, and Y. H. Kerr, "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*, vol. 26, no. 1, pp. 79–79, 1987.

[177] A. A. Lacis and J. Hansen, "A Parameterization for the Absorption of Solar Radiation in the Earth's Atmosphere," *Journal of the Atmospheric Sciences*, vol. 31, no. 1, pp. 118–133, 1974.

[178] J. E. Hay and K. J. Hanson, "A satellite-based methodology for determining solar irradiance at the ocean surface durring GATE," *Bulletin of the American Meteorological Society*, vol. 59, pp. 1549–1549, 1978.

[179] D. Cano, *Etude de l'Ennuagement par Analyse de Séquences d'Images de Satellite. Application à l'Evaluation du Rayonnement Solaire Gloibal au Sol.* PhD thesis, Ecole Nationale Supérieure des télécommunications, 1982.

[180] G. Bourges, "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.

[181] C. Rigollier, "The method HELIOSAT-2 for deriving shortwave solar radiation from satellite images," *Solar Energy*, vol. 77, no. 2, pp. 159–169, 2004.

[182] M. Girodo, R. W. Mueller, and D. Heinemann, "Influence of three-dimensional cloud effects on satellite derived solar irradiance estimation - First approaches to improve the Heliosat method," *Solar Energy*, vol. 80, no. 9, pp. 1145 – 1159, 2006.

[183] L. F. Zarzalejo, L. Ramirez, and J. Polo, "Artificial intelligence techniques applied to hourly global irradiance estimation from satellite-derived cloud index," *Energy*, vol. 30, no. 9, pp. 1685 – 1697, 2005.

[184] K.-F. Dagestad and J. A. Olseth, "A modified algorithm for calculating the cloud index," *Solar Energy*, vol. 81, no. 2, pp. 280 – 289, 2007.

[185] R. Mueller, K. Dagestad, P. Ineichen, M. Schroedter-Homscheidt, S. Cros, D. Dumortier, R. Kuhlemann, J. Olseth, G. Piernavieja, C. Reise, L. Wald, and D. Heinemann, "Rethinking satellite-based solar irradiance modelling: The SOLIS clear-sky module," *Remote Sensing of Environment*, vol. 91, no. 2, pp. 160 – 174, 2004.

[186] Richard Perez and Pierre Ineichen and Kathy Moore and Marek Kmiecik and Cyril Chain and Ray George and Frank Vignola, "A new operational model for satellite-derived irradiances: description and validation," *Solar Energy*, vol. 73, no. 5, pp. 307 – 317, 2002.

[187] R. Perez, P. Ineichen, M. Kmiecik, K. Moore, D. Renne, and R. George, "Producing satellite-derived irradiances in complex arid terrain," *Solar Energy*, vol. 77, no. 4, pp. 367 – 371, 2004.

[188] Y. E. Systems, "TSI-880 Automatic Total Sky Imager." Airport Industrial Park 101 Industrial Blvd. Turners Falls, MA 01376 USA, 2012.

[189] C. W. Chow, B. Urquhart, M. Lave, A. Dominguez, J. Kleissl, J. Shields, and B. Washom, "Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed," *Solar Energy*, vol. 85, no. 11, pp. 2881 – 2893, 2011.

[190] R. Marquez and C. F. M. Coimbra, "Intra-Hour DNI Forecasting Methodology Based on Cloud Tracking Image Analysis," *submitted to Solar Energy*, 2012.

[191] M. Lave, J. Kleissl, and E. Arias-Castro, "High-frequency irradiance fluctuations and geographic smoothing," *Solar Energy*, pp. 1–17, 2011.

[192] A. Kamthe, R. Marquez, C. F. M. Coimbra, and A. Cerpa, "Sub-Minute Solar Irradiance Forecasting Using Wireless Sensor Networks." University of California, Merced, 2011.

[193] E. Lorenz, D. Heinemann, H. Wickramarathne, H. G. Beyer, and S. Bofinger, "Forecast of ensemble power production by grid-connected PV systems," in *Proceedings of the 20th European PV Conference, Milano, Italy*, 2007.

[194] E. Lorenz, J. Hurka, D. Heinemann, and H. Beyer, "Irradiance forecasting for the power prediction of grid-connected photovoltaic systems," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 2, no. 1, pp. 2–10, 2009.

[195] F. Yang, H.-L. Pan, and S. K. Krueger, "Evaluation of the NCEP Global Forecast System at the ARM SGP Site," *Monthly Weather Review*, vol. 134, pp. 3668–3690, 2006.

[196] F. Yang, K. Mitchell, Y.-T. Hou, Y. Dai, X. Zeng, Z. Wang, and X.-Z. Liang, "Dependence of Land Surface Albedo on Solar Zenith Angle: Observations and Model Parameterization," *Journal of Applied Meteorology and Climatology*, vol. 47, no. 11, pp. 2963–2982, 2008.

[197] S. G. Benjamin, W. R. Moninger, S. S. Weygandt, M. Hu, D. Devenyi, J. M. Brown, T. Smirnova, J. Olson, C. Alexander, K. Brundage, G. Grell, S. Peckham, T. L. Smith, S. R. Sahm, and B. Jamison, "Technical Review of Rapid Refresh/RUC Project," tech. rep., NOAA/ESRL/GSD internal review, Nov. 2009.

[198] A. Mellit, H. Eleuch, M. Benghanem, C. Elaoun, and A. M. Pavan, "An adaptive model for predicting of global, direct and diffuse hourly solar irradiance," *Energy Conversion and Management*, vol. 51, no. 4, pp. 771 – 782, 2010.

[199] W. Ji and K. C. Chee, "Prediction of hourly solar radiation using a novel hybrid model of arma and tdnn," *Solar Energy*, vol. 85, no. 5, pp. 808 – 817, 2011.

[200] R. Marquez and C. F. M. Coimbra, "Forecasting of Global and Direct Solar Irradiance Using Stochastic Learning Methods, Ground Experiments and the NWS Database," *Solar Energy*, vol. 85, no. 5, pp. 746 – 756, 2011.

[201] C. Voyant, M. Muselli, C. Paoli, and M.-L. Nivet, "Numerical weather prediction (NWP) and hybrid ARMA/ANN model to predict global radiation," *Energy*, vol. 39, no. 1, pp. 341 – 355, 2012.

[202] R. Marquez, V. G. Gueorguiev, and C. F. M. Coimbra, "Forecasting of Global Horizontal Irradiance Using Sky Cover Indices," *ASME Journal of Solar Energy Engineering (in press)*, 2012.

[203] R. Marquez, H. T. C. Pedro, and C. F. M. Coimbra, "Hybrid ANN/Satellite Imaging Forecasting Method for Global Horizontal Irradiance," *submitted to Solar Energy*, 2012.