# SHORT-TERM PREDICTION OF WEATHER PARAMETERS USING ONLINE WEATHER FORECASTS

Yi Zhang[1], Vic I. Hanby[1]

[1]Institute of Energy and Sustainable Development, De Montfort University,
The Gateway, Leicester, LE1 9BH, United Kingdom

## ABSTRACT

While people need to know tomorrow’s weather to decide suitable activities and precautions, so do the “intelligent” building management systems. The accuracy of the short-term prediction of the ambient conditions is particularly import for the development of predictive control strategies. Although the short-term prediction methods for outside air temperature have been extensively studied, reliable prediction methods for solar radiation are yet to be established. This paper discusses the possibility of using meteorological forecast data with local observation data for short-term prediction of the ambient conditions. Based on the analysis of the collected data, a new method has been developed for temperature and solar radiation prediction. It is found that a linear combination of the observation and the forecast data can effectively improve the accuracy of short-term (24 hours) prediction. The method is also suitable for the prediction of humidity and wind speed/direction.

## KEYWORDS

short-term prediction; weather forecast; control; temperature; solar radiation; wind; humidity

## INTRODUCTION

The accuracy of short-term prediction of ambient conditions is crucial to the success of the latest control technologies for building system operations (Henze et al. 2004). Especially when renewable energy sources (such as solar heating) and thermal storages are involved, correct prediction of gains and loads is the basis on which control decisions are made.

A number of methods for time series prediction have been adopted in field of building system control (Chen and Athienitis, 1996; Kawashima et al. 1995; Ren and Wright, 2002; Henze et al. 2004). Most of the methods assume that the future values of the variables are correlated to immediate past values (Like-Yesterday model), past values with moving average adjustment (ARMA models), or the statistical features of historical values (Bin method). Stochastic, adaptive, and neural network-based approaches have also been used for fine-tuning the model parameters. The prediction performance of each approach, however, seems to be dependent on not only the parameters used in the models, but also the characteristics of the climate in which it is used. It is difficult to identify a universally best solution.

Applications that make use of meteorological forecast information for short-term weather prediction have also been reported (Yoshida and Inooka, 1997). Meteorological forecast are made by using large-scale atmospheric models, satellite images, multiple-points surface observations, and massive computing power to predict the trend of weather changes. However, the on-spot accuracy of the regional weather information from the meteorological sources, and therefore its feasibility of being used for building control, has not been studied.

This paper discusses the possibility for using meteorological forecast data in conjunction with local observation data to predict temperature and solar radiation in short-term.

## ONLINE RESOURCES

The study was carried out in Leicestershire, which is the central part of England. Both local observation and weather forecast data were collected online from various websites. Given the abundance of weather information on the internet, the research suggests that it is possible for a building control system to rely on external data sources, instead of installing an on-site weather station.

There are a number of national and international weather forecasting sources that provide detailed forecasts for the region of Leicestershire. We have been collecting forecast data from four websites (see Figure 1 to 4). These regional forecasts are centered at the Leicester City. However, the local weather observation data used in this research was collected from the “Oadby Weather” website (Anonymity A) since October 2004. The observatory (N 52° 36', W 1° 05') is about 3.5 miles south of the city of Leicester. Despite the seemingly short distance to the city center, this could still be significant due to the fickleness of the weather and the possible heat island effect of the city. "Oadby Weather" provides download facility for 5-minute observation including most of the meteorological parameters, e.g. temperature, humidity, wind, pressure, precipitation and solar radiation. These data were subsequently interpolated bi-hourly, therefore, a total of 15030 data points are available.


*Figure 1. BBC Weather website*

The BBC Weather website (Anonymity B) broadcasts worldwide weather forecasts made by the meteorologists in the Met Office of the UK. On the website, three-hourly forecast data for the next 24 hours can be retrieved. Data used in this paper has been collected since November 2004. Compared to other sources, the quality of the forecasts are remarkably good. About 421 days of data were collected and interpolated to be used in the analysis.

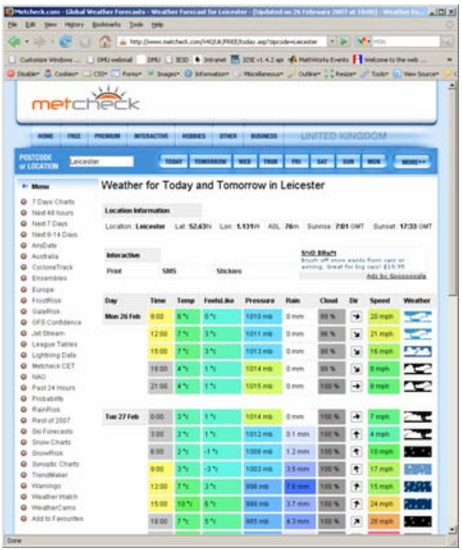


*Figure 2 METCHECK website*

Metcheck.com (Anonymity C) is an independent weather forecaster who uses observation data from the National Oceanic and Atmospheric Administration (NOAA) of the US and in-house forecasting models. It provides short-range forecasts of 25 variables in three-hour intervals. Reliable data feeds are available for a small monthly fee.

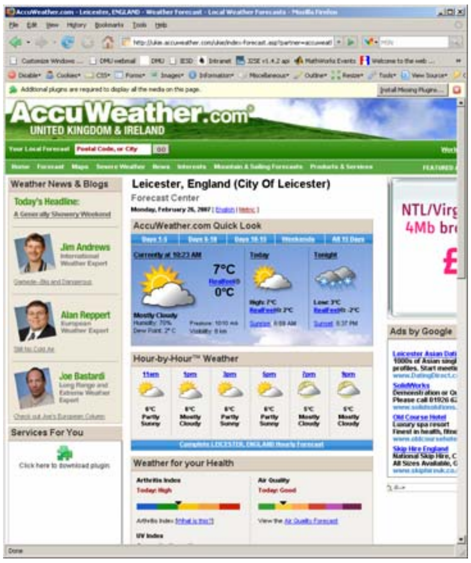


*Figure 3. AccuWeather.com*

Based in State College, Pennsylvania, USA, AccuWeather (Anonymity D) has been a weather forecast provider to the public media such as news papers, radios and websites. Hourly forecasts with worldwide coverage are available. However, the accuracy of the forecast for the Leicester city area was lower than other sources. We collected data from the website between April 2005 and November 2006.

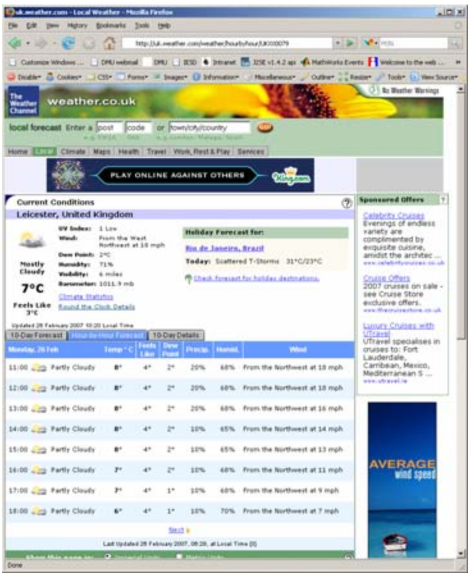


*Figure 4. The Weather Channel - UK*

We also started in September 2006, collecting forecast data from the UK branch of the Weather Channel (Anonymity E), which provide hourly forecast of weather condition, temperature, humidity, pressure, wind direction and speed. Unfortunately we were unable to include the data in the analysis at the time when this paper is written.

There are many other online sources of weather forecast data. More reliable paid services are available directly from the meteorological research centers, such as the Met Office of the UK, the European Centre for Medium Range Weather Forecasting (ECMWF), and other national institutes of meteorology around the globe. There are several issues with the use of online forecast for the purpose of building control. Firstly, most weather forecast centers use median to large-scale atmospheric models, satellite images, and local observations to predict the trend of weather changes in a wide region. Local forecast data has to be post-processed by interpolation. Depending on the forecast and interpolation models used, systematic error can exist in the forecast from a certain channel for a particular location. Secondly, due to the high dependency on the internet links, retrieval from a single forecast channel is not reliable. Multiple sources should be sought for redundancy.

In the following sections, we describe a method that combines various weather forecast sources to produce more accurate predictions on temperature and solar radiation.

## TEMPERATURE FORECAST

Linear combination of Like-Yesterday (HST), BBC (BBC), MetCheck (MCHK), and AccuWeather (ACCW) forecasts is defined as the following equation:

$$\hat{T}_t = \frac{\sum_{i=1}^{4} w_i (a_i \hat{X}_{ti} + b_i)}{\sum_{i=1}^{4} w_i} \quad (1)$$

where,

$w_i, a_i, b_i$ ($i$=1, 2, 3, 4) Model parameters

$\hat{X}_{t1} = \hat{T}_{t-24h}$ Yesterday's temperature

$\hat{X}_{t2} = \hat{T}_{t\,ACCW}$ AccuWeather forecast temperature

$\hat{X}_{t3} = \hat{T}_{t\,BBC}$ BBC forecast temperature

$\hat{X}_{t4} = \hat{T}_{t\,MCHK}$ MetCheck forecast temperature

Conversely, the Equation 1 is equivalent to the following form, in which only 5 parameters are required instead of 12.

$$\hat{T}_t = \sum_{i=1}^{4} c_i \hat{X}_{ti} + d \quad (2)$$

However, the Equation 1 is preferred because $\hat{w}$, $\hat{a}$ and $\hat{b}$ can better reflect the characteristics of each forecast source. $\hat{a}$ and $\hat{b}$ are the scalars and biases, respectively, to compensate the possible systematic error in each source, whereas the weights ( $\hat{w}$ ) indicates the relevance of each corrected source.

It is also worth noting that, due to the reliability issues in online forecast retrieval, data from each source may be intermittent. If for any missing data, "0" is given, the integrity of the Equation 1 is unaffected. The benefit is that, with the linear combination, the availability of forecast data is maximized by combining different sources.

Parameters $\hat{w}$, $\hat{a}$ and $\hat{b}$ of the model are fitted for the linear model so that minimum RMSE of the forecast is achieved for the collected data. Table 1 shows the results from data fitting. Clearly the forecast from BBC and MetCheck are more relevant to the actual observation than those from AccuWeather and Like-Yesterday predictor. Also the BBC and MetCheck both tend to slightly under estimate the temperature in the next 24-hour, while the AccuWeather often overestimate it.

*Table 1. Fitted parameters*

| SOURCE | *W* | *A* | *B* |
|---|---|---|---|
| HST | 0.075 | 0.479 | 4.698 |
| ACCW | 0.208 | 1.639 | -5.000 |
| BBC | 0.897 | 0.898 | 0.408 |
| MCHK | 0.873 | 1.072 | 0.222 |

Table 2 summarizes the accuracy of different forecast methods, including the linear model using Equation 1 (Hybrid) that combines the sources, and the model based on sinusoidal functions (Sinusoidal) and the exponential weighted moving average (EWMA) (Ren and Wright 2002), and the Bin method described by Henze et al. (2004)

*Table 2. Comparison of different methods*

| METHOD | SAMPLE SIZE | ME (K) | MAE (K) | RMS (K) | MAX (K) |
|---|---|---|---|---|---|
| HST | 12,155 | 0.09 | 2.58 | 3.36 | 13.40 |
| ACCW | 7,545 | -1.07 | 2.21 | 2.81 | 9.92 |
| BBC | 13,501 | 0.38 | 1.46 | 2.04 | 12.50 |
| MCHK | 10,189 | -0.66 | 1.56 | 2.07 | 14.40 |
| **Hybrid** | **14,535** | **0.00** | **1.26** | **1.71** | **8.97** |
| Sinusoidal* | 720 | 0.31 | 1.86 | 2.35 | 6.73 |
| EWMA* | 720 | 0.00 | 1.53 | 2.10 | 7.13 |
| BIN** | 8,760 | 0.01 | - | 2.26 | - |

* Data source: Ren and Wright, 2002
** Data source: Henze, Kalz, Felsmann and Knabe, 2004

The Hybrid model shows significantly improvement on the forecast accuracy over all three online forecast sources, as well as the Like-yesterday method. Although the RMS error of the Hybrid method is also lower than the Sinusoidal, the EWMA, and the Bin methods from literature, the results are not directly comparable due to different weather data having been used.

*Table 3. Normalized RMS Errors*

| | MEAN MEASUR -EMENT (K) | CV-RMSE | MEAN DIURNAL VAR. (K) | DIUR-CV-RMSE |
|---|---|---|---|---|
| **BBC** | 282.6 | 0.0072 | 8.6 | 0.2384 |
| **HYBRID** | | **0.0061** | | 0.2001 |
| **SINUSOI.** | 281.5 | 0.0083 | 8.2 | 0.2860 |
| **EWMA** | | 0.0075 | | 0.2556 |
| **BIN** | 295.7 | 0.0076 | 14.5 | **0.1555** |

Table 3 provides two different methods to normalize the root mean square error of the forecast in order to compare the accuracy of the methods using different weather data. Firstly, the coefficient of variance (CV) is calculated with the mean of all samples used in each method. This comparison shows the Hybrid method is a clear winner; while even without any correction, the BBC forecast is more accurate than the other method. The second method is to use diurnal variation of temperature as the base for normalization. In this case, the relative error in the Bin method is much lower than the other methods, due to the larger average daily temperature fluctuation. However, as further proven in the discussions on solar forecast, the characteristics of climate has significant impact on the accuracy of short-term weather prediction methods; therefore it is possible that, if applied in Phoenix, AZ, USA, the Hybrid method will outperform the Bin method, too.

## SOLAR FORECAST

Short-term prediction of the availability of solar radiation from online weather forecast sources is more complex than predicting temperature. Primarily, unlike temperature, the value of solar radiation (e.g. global horizontal radiation) is not directly available from the weather forecast channels. Instead, references such as weather condition summary and cloud cover, to less-direct references like visibility, precipitation, cloud base, humidity and pressure, are provided. The challenge is how to use the information to predict hourly radiation values.

In this research, a two-step approach is adopted. Firstly, the descriptive weather condition summaries from the forecast channels are converted into a measure of attenuation to the global horizontal radiation based on a clear sky condition. With the theoretical profile of clear sky solar radiation of the location and the predicted attenuation, hourly prediction can be calculated. The predictions from different sources are subsequently fused using the linear combination model described in Equation 1.

### Conversion

The first step is to convert the descriptive information from weather forecast channels to global horizontal radiation measure. A stochastic model is used for this purpose.

Firstly, the weather condition summary reports are arbitrarily divided into 11 categories corresponding to 11 levels of attenuation (observed global horizontal radiation versus theoretical clear sky profile), see Table 4. The distributions of the categories in the collected weather forecast data are shown in Figure 5. Comparing to the distribution in the observation in Leicester (Figure 6), the forecasts are inconsistent, and the correlation is unclear.

*Table 4. Definition of solar (attenuation) scale*

| SOLAR SCALE | WEATHER SUMMARY | OBSERV. VS. CLEAR SKY |
|---|---|---|
| 11 | Clear / fair / sunny | >0.95 |
| 10 | Sunny intervals | 0.85~0.95 |
| 9 | Partly cloudy | 0.75~0.85 |
| 8 | Median-level cloud | 0.65~0.75 |
| 7 | Low-level cloud | 0.55~0.65 |
| 6 | Fog | 0.45~0.55 |
| 5 | Mist / drizzle | 0.35~0.45 |
| 4 | Light rain / light snow | 0.25~0.35 |
| 3 | Sleet / shower | 0.15~0.25 |
| 2 | Heavy rain / heavy snow | 0.05~0.15 |
| 1 | (Night time) | <0.05 |

The following stochastic model that converts weather summary to attenuation of global horizontal radiation is used:

$$y = R(1) \times 2\sigma_{CAT(x)} + \mu_{CAT(x)} \qquad (3)$$

Where, $y$ is the probable attenuation of global horisontal radition; $R(1)$ is a random number with uniform distribution within [0, 1]; $CAT(x)$ is the category in which an input weather summary belongs; $\sigma_i$ and $\mu_i$ are model parameters for the category of each input. By fitting the parameters $\hat{\sigma}$ and $\hat{\mu}$, it has been possible to achieve similar distribution of solar attenuation as in the Oadby observation data. (See Figure 7) The fitted model parameters can be found in Figure 8.

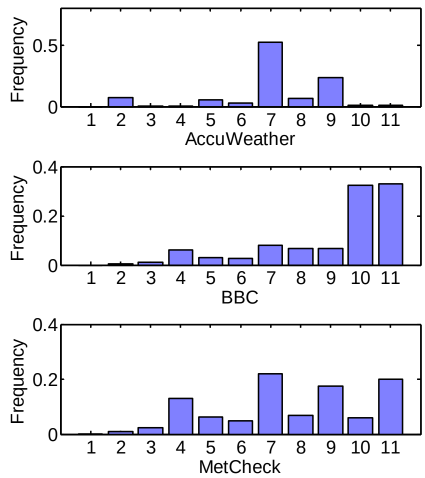


*Figure 5. Distribution of solar scales from weather forecast channels*

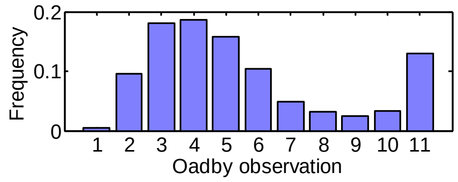


*Figure 6. Distribution of solar scales in observation data from Oadby, Leicester, UK*

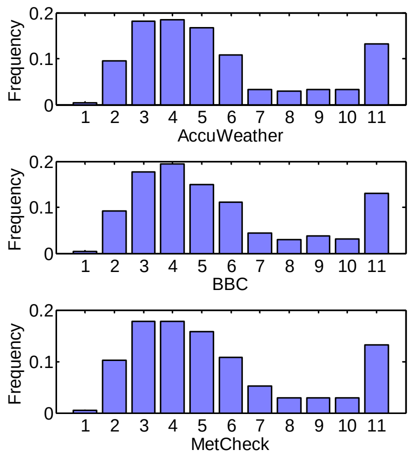


*Figure 7. Distribution of solar scales in the output of the stochastic conversion model*

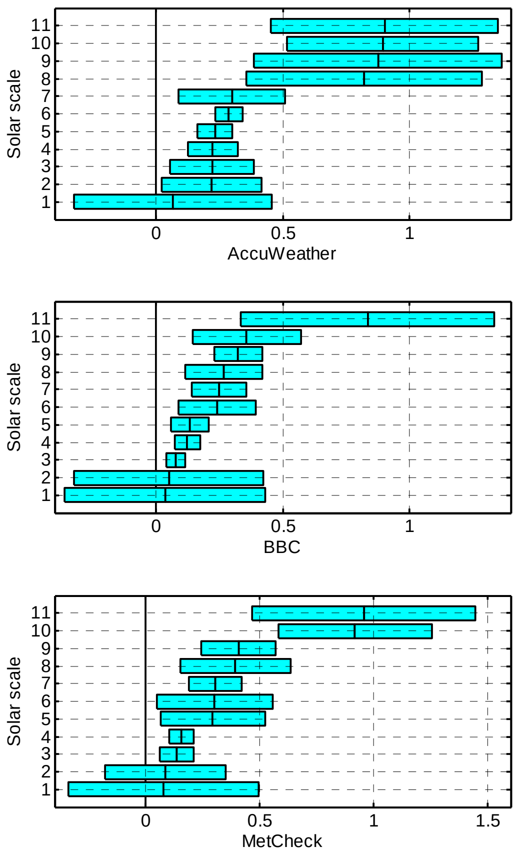


*Figure 8. Stochastic conversion model parameters*

**Data fusion**

Similar to the method for temperature prediction described in the previous section, the linear combination model (Equation 1) is used for solar prediction, too. Theoretical hourly global horizontal solar radiation in clear sky condition can be calculated using methods described by Page and Sharples (1998). Predicted solar radiation data can then be calculated by multiplying the forecasted attenuation to the clear sky data. Parameters of the Equation 1 are fitted by minimizing the RMS error in predicted solar radiation. The identified parameter values are listed in Table 5.

*Table 5. Fitted parameters*

| SOURCE | *W* | *A* | *B* |
|---|---|---|---|
| HST | 0.531 | 0.937 | 55.36 |
| ACCW | 0.425 | 0.648 | 159.2 |
| BBC | 1.000 | 0.965 | 21.39 |
| MCHK | 0.347 | 0.491 | -53.78 |

Table 6 sumarizes the accuracy of each forecast source and the Hybrid model. In terms of RMS error, the Hybrid method produces significant improvement over any single source. The results of two other solar prediction methods from the literature is listed in Table 6, too. However, due to the large difference in sample size, the accuracy results are not readily comparable. It is worthwhile to mention that all samples used in this research for solar prediction are for daytime hours; whereas in the literature, it was not clearly stated.

*Table 6. Comparison of different methods*

| METHOD | SAMPLE SIZE | ME ($W/M^2$) | MAE ($W/M^2$) | RMSE ($W/M^2$) | MAXE ($W/M^2$) |
|---|---|---|---|---|---|
| HST | 7,660 | -5.7 | 113.1 | 196.7 | 1017.6 |
| ACCU | 6,305 | -20.8 | 147.6 | 232.3 | 1001.9 |
| BBC | 8,624 | -12.2 | 97.8 | 162.1 | 1002.4 |
| MCHK | 5,238 | -9.7 | 99.8 | 168.8 | 1041.2 |
| **Hybrid** | **9,309** | **-0.8** | **92.8** | **146.6** | **1006.6** |
| Sinusoidal* | 720 | 29.1 | 111.4 | 155.0 | 554.2 |
| EWMA* | 720 | 3.5 | 68.7 | 116.1 | 473.4 |

* Data source: Ren and Wright, 2002

One useful discussion is how characteristics of a climate type impact on the accuracy of solar prediction. The study reported in this paper concerns the climate type of England, which is a humid temperate climate with cool summers. Prediction of solar radiation is particularly difficult due to probability of rapid change of the atmospheric conditions in the region. Table 7 shows a comparison between the predications made for Leicester, UK and that for Miami, Florida, USA. The CV-RMS measure, which is the RMS error normalized against mean global horizontal solar radiation, and the correlation coefficient between prediction and measurement are used. It is clear that, by using the Like-yesterday method, the prediction in Florida is much more accurate than that anything we can achieve in Leicester.

*Table 7. Impact of climate type on the accuracy of forecasts*

| | MEAN SOLAR RAD. (W/M$^2$) | RMSE (W/M$^2$) | CV-RMSE | R |
|---|---|---|---|---|
| HST (LEICS. UK) | 178.3 | 196.7 | 1.103 | 0.575 |
| HYBRID (LEICS. UK) | | 146.6 | 0.822 | 0.715 |
| HST (MIAMI, US) | 385.1 | 164.7 | 0.428 | 0.826 |

## CONCLUSION

Short-term prediction of ambient conditions is an important part of the latest control technology for the operation of building systems, especially when solar energy and thermal storages are involved. This paper discusses the possibility of using meteorological forecast data with local observation data for short-term prediction of temperature and solar radiation.

Meteorologists and research centers around the world are using large-scale atmospheric models, satellite images, and surface observations to predict the trend of weather changes. The resources available to the meteorological researchers are almost imaginable for building control engineers. The attempts to use the outputs of meteorological forecast information to assist short-term weather prediction seem to be a natural solution.

A large number of online weather forecast channels are freely available. Depending on the forecaster and the interpreter, systematic error may exist in the local forecast from a certain channel. Also, redundant retrieval from several channels may increase the reliability of the data. In this paper, we use the forecast and local observation data collected during the past two years to develop a new method that combines various weather forecast sources to produce more accurate predictions on temperature and solar radiation. The new method, which uses a linear combination (Hybrid) model of the last-24-hour observation and the forecast data from various sources, can significantly improve the accuracy of short-term prediction for both temperature and solar radiation. Future research may be needed to develop a dynamic adaptation of the model parameters in order to account for changes in climate and/or forecast models.

## ACKNOWLEDGMENT

The authors would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council (EPSRC) under grant GR/S30467, the collaboration of Trend Control Systems Ltd and the Oadby and Wigston District Council.

## REFERENCES

Anonymity A http://www.carthage.nildram.co.uk/

Anonymity B http://www.bbc.com/weather/

Anonymity C http://www.metcheck.com/

Anonymity D http://ukie.accuweather.com/

Anonymity E http://uk.weather.com/

Chen, T Y; Athienitis A K. 1996. "Ambient Temperature and Solar Radiation Prediction for Predictive Control of HVAC Systems and a Methodology for Optimal Building Heating Dynamic Operation." ASHRAE Transactions Vol.102(1):26-36.

Henze, G P; Kalz, D E; Felsmann, C; Knabe, G. 2004. "Impact of forecasting accuracy on predictive optimal control of active and passive building thermal storage inventory," HVAC & R Res. Vol. 10(2):153-178

Kemmoku Y; Orita S; Nakagawa S; Sakakibara T; 1999. "Daily insolation forecasting using a multi-stage neural network", Solar Energy, Vol.66(3):193–199

Kawashima M; Dorgan C E; Mitchell J W; 1995. "Hourly Thermal Load Prediction for the Next 24 Hours by ARIMA, EWMA, LR, and an Artificial Neural Network." ASHRAE Transactions, Vol.101(1):186-200.

Page, J K; Sharples S; 1988. "The SERC Meteorological Data Base, Volume II: Algorithm Manual (2nd Edition)", Department of Building Science, University of Sheffield, Sheffield, S10 2TN, United Kingdom

Ren M J; Wright J A; 2002. "Adaptive diurnal prediction of ambient dry-bulb temperature and solar radiation", HVAC&R Research, Vol.8(4):381-399

Yoshida, H; Terai, T. 1992. "Modeling of Weather Data by Time Series Analysis for Air Conditioning Load Calculations." ASHRAE Transactions, Vol.98(1):328-345

Yoshida H; Inooka T; 1997. "Rational operation of a thermal storage tank with load prediction scheme by ARX model approach", Procedings of Building Simulation, 97 II (1997):79–86