# ADAPTIVE RULE GENERATION FOR SOLAR FORECASTING: INTERFACING WITH A KNOWLEDGE-BASE LIBRARY

J. Nummikoski, Y. S. Manjili, R. Vega, and H. Krishnaswami

Texas Sustainable Energy Research Institute and Department of Electrical and Computer Engineering,
University of Texas at San Antonio, San Antonio, TX 78249, USA

***Abstract* — This paper covers the development of an adaptive, interactive rule generation interface applied to the full scope of solar forecasting techniques, both current and forthcoming. The interface provides a user-friendly platform for detecting patterns and correlations between elements in a database of solar irradiance, weather and photovoltaic generation information. The database consists of 10 years of data obtained from the National Renewable Energy Laboratory (NREL) data acquisition systems and the Automated Surface Observing System (ASOS). This report discusses how such an interface can be used to improve existing forecasting algorithms and also be used to create new forecasting techniques.**

***Index Terms* — Solar energy, Solar power generation, Solar forecasting, Renewable energy operations**

## I. INTRODUCTION

Solar forecasting is an area of research that is gaining more and more visibility due to the increasing penetration of solar power into electricity grids. Once the amount of solar penetration reaches a certain threshold, the variability of the solar power production becomes a problem for both grid stability and reliable bidding of electricity by utility companies. Solar forecasting solutions provide utility companies with predictions of power output from large-scale solar installations or from distributed solar generation with a time scale ranging from the next few minutes up to several days ahead.

Databases offer a wealth of knowledge to users/operators but are usually so restrictively large that information cannot be extracted, analyzed and synthetized quickly enough for daily, hourly and minute decision-making. This is where processes such as data analytics and data mining can be leveraged. Data analytics seeks to discover relevant data features and patterns through data mining processes and communicate that information in the most effective way to the end-user; either a human or a computerized process [1, 2].

This paper will discuss the current state of the art and introduce a novel approach to solar forecasting through implementation of a knowledge-base library that is ever growing and increasing in intelligence.

### A. *Current State of the Art in Solar Forecasting*

A report performed for the Regents of the University of California entitled "Current State of the Art in Solar Forecasting" [3] best summarizes the current status of this field of research. In this report Glassley et al. state that satellite and numerical weather prediction (NWP) are the preferred methods for longer duration solar forecasting (one hour to a few days) [3]. As for intra-hour forecasts, sky imagery based forecasting methods are most common, but, unfortunately, several basic assumptions regarding cloud shape and linear cloud movement vectors reduce the potential accuracy of this type of forecasting [4, 5].

Making intra-hourly or site-specific forecasts through the use of satellite data is also not very common due to the infrequent sampling interval (30 minutes) and the low image resolution [3]. This problem is only exaggerated in the NWP method due to a larger sampling interval and lower cloud imaging precision. The limitations of NWP models do not allow for shorter time-scale and smaller spatial sampling to be accounted for [6]. Some of the error associated with the aforementioned problems can be corrected through a method known as Modeled Output Statistics (MOS), which determines statistical correlations in observed weather data and related imagery (satellite, sky imagery, etc.) [3,6-8].

In order to apply MOS to a solar forecasting approach, regardless of the time scale, a database of instrumentation data is necessary [7]. The more information that is available in this database, the more useful it is; this applies to both the breadth and depth of the data. Unfortunately, this amount of data can be difficult to work with and finding accurate correlations between elements must be delegated to a statistical model, which often includes outlier data that does not improve forecasting results [3, 8]. Furthermore, a statistical model is not learning from the decision-making and rule policies learned by its users/operators over time.

### B. *Improvements in Solar Forecasting Using a Knowledge-Base Library*

The problem mentioned above can be addressed by designing an interface to work with the large knowledge-base database (herein also called "library") which would allow a user to interact with the data in a way that is not possible by statistical means alone. This program will allow users to identify and extract sub-datasets, aggregate imagery data, and couple them in a multivariate environment with other weather-related and irradiance data, all through a convenient user interface. Furthermore, the users can utilize the aggregated imagery data to perform batch image-processing techniques common to image-based solar forecasting. With this interface users could step through the data in any fashion and view the patterns with their own eyes, which would then allow more accurate adaptations to their existing forecasting algorithms

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 'Date-Time' | 'Image Data' | 'File Path' | 'Year' | 'Month' | 'Day' | 'Hour... | 'Minute' | 'Research F3' | 'Research F2' | 'Global L |
| 2 | '20120921060000' | <352x288x3 uint8> | 'C:/Users/... | 2012 | 9 | 21 | 6 | 0 | 18.2810 | 14.4830 | 13.1720 |
| 3 | '20120921061000' | <352x288x3 uint8> | 'C:/Users/... | 2012 | 9 | 21 | 6 | 10 | 30.5800 | 29.6730 | 24.9180 |
| 4 | '20120921062000' | <352x288x3 uint8> | 'C:/Users/... | 2012 | 9 | 21 | 6 | 20 | 51.8000 | 48.7380 | 47.2370 |
| 5 | '20120921063000' | <352x288x3 uint8> | 'C:/Users/... | 2012 | 9 | 21 | 6 | 30 | 75.3210 | 69.6770 | 70.1090 |
| 6 | '20120921064000' | <352x288x3 uint8> | 'C:/Users/... | 2012 | 9 | 21 | 6 | 40 | 98.0840 | 93.2510 | 94.5770 |
| 7 | '20120921065000' | <352x288x3 uint8> | 'C:/Users/... | 2012 | 9 | 21 | 6 | 50 | 126.3900 | 121.6100 | 123.9400 |

Fig. 1. Ten Minute Database (Partial Screenshot)

and capabilities by creating necessary rule policies to improve model accuracy.

The ability to intelligently segment sub-datasets from this large database would allow more meaningful analytics to be performed, such as visual analytics. Visual analytics is a mathematical procedure that enables visual data to be transformed into numerical data sets of multi-level depth. These data sets allow a computer algorithm to recognize the same patterns that can be seen by the human eye and possibly even some patterns that cannot [9]. This method of translating large sets of image data into larger sets of numerically relevant data is a key component to developing short-term and long-term solar forecasting abilities.

## II. Approach and Findings

To compile the database, a “more data is better approach” was used. The library consists of meteorological data and sky images dating back to 2004 downloaded from the National Renewable Energy Laboratory (NREL) website and the Automated Surface Observing System (ASOS) database for ceilometer cloud base height recordings. Data from more than 250 instruments, variables and readings were collected, including: global horizontal irradiance, diffuse horizontal irradiance, direct normal irradiance, solar zenith angle, solar azimuth angle, atmospheric temperature, relative humidity, cloud cover %, cloud height, wind speed, wind direction, and many more.

Throughout the process of building the library it was necessary to acquire relevant knowledge in several areas: meteorological phenomena, irradiance and weather instrumentation, sky-imaging cameras, image processing, physics, power engineering, and artificial intelligence. The database contains a large collection of weather and irradiance instrument readings which in some cases were erroneous or incomputable. In order to determine whether these readings were actually incorrect it was necessary to gather knowledge on all the instrument types and the meteorological phenomena that they measure. Learning about the types of available sky imaging cameras and the processing of their images enabled the research to produce meaningful data from the image files. The knowledge obtained from researching the current state of the art in solar forecasting provided the understanding of what data needed to be gathered for the various prediction methods. And lastly, without sufficient knowledge of physics and power engineering, a thorough understanding of the correlation between irradiance, atmospheric conditions and photovoltaic array energy output could not be obtained. Building this library required a multi-disciplinary approach.

The knowledge-base library will function as an experiential step to quantify probabilities and patterns in the historical records that can, in turn, inform real-time decisions and day-ahead models from measured lessons learned. It is a practical, decision-enhancing, rule-making tool that supports the development of more precise solar forecasting tools and acts as a memory-bank for the operators to draw their conclusions. The ultimate objective of this knowledge-base library is to enable a better scientific quantification of the confidence levels (uncertainty) in the amount of solar power generation that is expected from high-fidelity intra-hour and day-ahead models.

### *A. Building the Knowledge-Base Library*

The library was built in two separate formats: at one-minute intervals and 10-minute intervals, the latter consisting of sky imagery data and other data points. In the one-minute library all available data is given at one minute intervals for the entire length of the day. For the few instruments that do not record readings at such a small interval, either linear interpolation was used, if relevant, or a default value, “NaN” for “off”, was assumed. The 10-minute library consists of sky images, downloaded from the NREL website, dating back to 2004. The sky images are taken by a Yankee Environmental Systems Total Sky Imager 880, known as TSI-880. The images are sampled at ten minute intervals during daylight hours; this equates to roughly 60 compressed JPEG images on an average day. The 10-minute library consists of these images and all the data from the one-minute library at these given time instants, as shown in Figure 1.

The knowledge-base library can also be updated real-time and additional database information can be added to the historical database as desired. The existing database is open ended and allows new instrument (e.g. satellite imaging for solar-terrestrial energy and parallax research) measurements to be added without effecting its functionality and searchability. Also, when analyzing and processing the database through the interface it is likely that the user will uncover new variables that would be beneficial to include in the entire historical database. These new variables can be processed and added to the historical record simply through the interface. During the forecasting process, both short-term and long-term, new

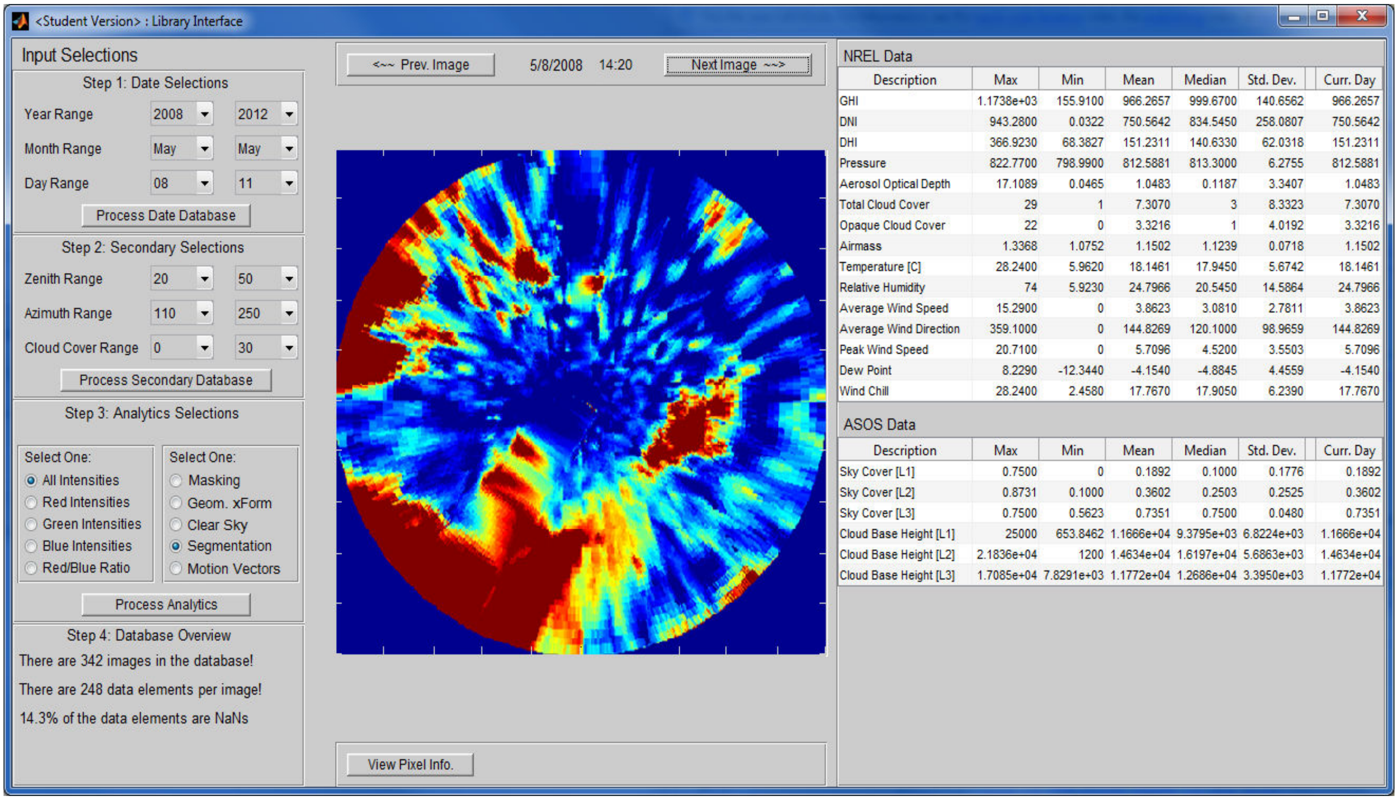


Fig. 2. Interface to the Knowledge-Base Library

variables such as cloud cover, cloud segmentation gradients and various weather related derivatives are being created and added to the database in a continuous feedback loop.

## *B. Interface Layout*

The goal of the interface is to interactively connect the users with the large libraries, allowing them to develop intelligence from coupled optical and electrical sensor data acquisition systems. A sample snapshot of the interface is shown in Figure 2.

The interface is read in a left to right, top to bottom fashion. In the top left section, labeled Step 1, the user selects a date range to analyze and processes this range to reveal the other range variables available in Step 2. Step 2 will only display the available Zenith, Azimuth and Cloud Cover Percent ranges based on the sub-database created in Step 1. After making Step 2 selections and processing to further reduce the sub-database, the user is presented with various analytics choices in Step 3, and a general overview of the sub-database characteristics in Step 4. From the information presented in Step 4 the user can determine the relevance of the current search conditions by looking at the number of images or number of NaN entries in the new sub-database.

Finally, the interface allows the user to export the intelligently selected sub-datasets and raw or processed images for further separate analyses or to other forecasting modules as desired. For example, currently in development are plans to display in this interface the displaced clouds (and their shades) to their future locations considering the weather and cloud genesis evolution, and then to export the processed images to Geographic Information Systems (GIS) for use by operators in their distributed and large-scale solar energy systems.

## *C. Developing Rules Through the Interface*

Once the users have become familiar with the function of the knowledge-base library interface, they can begin to recognize patterns in the data that would not be accessible without an intelligent approach to data aggregation. Data mining and analytics, often troublesome when working with large sets of data, are drastically simplified through the use of this interface, as it allows the user to perform all these complex tasks quickly by connecting real-time data with user/operator real-time intelligence. Visual analytics can then be performed on the smaller subsets of imagery, presenting image data such as color, shape, spatial similarities and other image qualities in numerical values as described in [9 ,10].

The insights gathered through the use of this interface can aid in the development of adaptive image processing algorithms. For instance, when computer codes have difficulty in synthetizing results that otherwise would be easier for a human to detect, the interface allows the user/operator to identify features on the images by flagging or painting areas on the image. This will allow the cloud segmentation algorithm to learn from its user by keeping the associations made. Recognizing that a particular type of cloud cover exists during a given time of year and using that knowledge-base imagery to enhance the cloud detection algorithm used on a real-time image with similar characteristics significantly improves the state of the art in data analytics, with an immediate impact to applications such as solar forecasting.

Important statistical information can also be acquired from the Library interface about the meteorological data related to a given sky image. This information can be used as the basis for training and developing rules for an Adaptive Artificial Neural Network (AANN). AANNs use a training data set to build a predictive network. From this network, simulations of future data can be performed. Neural Networks are very sensitive to the training data set they receive: the more accurate and relevant the data input, the better the ability to make accurate simulations and reliable forecasts. The intelligent coupling of imagery and meteorological data is an example of a more relevant AANN input. In addition, as image processing algorithms continually refine the properties associated with the clouds in an image, such as: type, roughness and depth; the set of AANN training data is adjusted to more closely match the expected cloud conditions.

## III. Contribution to Solar Forecasting

As discussed in Section I-A, there are many different forecasting methods currently in use over the wide range of forecasting time horizons. Virtually all methods can benefit from the additional rule generation and pattern recognition made possible through the use of the knowledge-base library interface.

### *A. Sample Adaptive Rules*

For utility companies, solar forecasting has been gaining popularity as the number and size of solar farms has been increasing. Utilities have been primarily concerned with day-ahead forecasting due to the financial impact it represents in the potential bidding of solar energy to the regional ISO (Independent System Operator). Power output in day-ahead forecasting approaches tends to be over-predicted on cloudy days and under-predicted on sunny days. In most geographic locations both errors tend to happen frequently, which is why it is extremely important to have a quantification of the uncertainty related to each day-ahead forecast. Through the use of the knowledge-base library and interface, real-time forecasting uncertainty can be quantified by searching for highly correlated days in the historical database. By knowing the uncertainty attached to each day-ahead forecast, utilities can make better decisions about the amount of generation capacity that needs to be available on stand-by.

One of the emerging methods of day-ahead forecasting is the Adaptive Artificial Neural Network (AANN) approach [3]. Through the use of the knowledge-base library and interface, intelligent adaptations can be made to the input data being fed into an AANN. The user will be able to apply knowledge of the science behind cloud formation to the database and search for patterns and correlations in atmospheric data. These correlated data points can then be used as the inputs to an AANN, resulting in a more finely tuned high-fidelity solar forecasting algorithm.

From day-ahead forecasts, utilities will learn about the expected ramp rates of their solar generation for each coming day, but atmospheric conditions are not always completely predictable and therefore, timing related to day-ahead predictions could be several hours off. This scenario is where intra-hour prediction becomes extremely important to utilities, as it allows them to make the necessary adjustments in generation availability to match with the new expected power ramps. Intra-hour forecasting can benefit in many ways from the rule-generation and pattern-recognition capabilities provided by the knowledge-base library and interface. For example, in Section II-C it was mentioned that cloud detection algorithms could be improved through an adaptive rule generation approach. In digital image processing one of the more difficult tasks has always been object detection, especially when the objects are as complex and dynamic as clouds. Object detection can be simplified if sample objects can be used as a starting point in the detection algorithm. Through the use of the library interface these “sample objects” can be determined and applied to said detection algorithm. As a result of improved detection of the clouds within an image, the high-fidelity forecasts which rely on these images can be made more reliable.

### *B. Knowledge-Base Library: Flexible Data Collection*

One problem mentioned in Glassley et al. [3] regarding the acquisition of data for use in MOS applications was the lack of availability of the instrumentation data at different locations on the planet, where cloud formations vary. Using the interface, users will be able to retrieve the nearest possible data to their location or input into the database the data that they do have. Most solar forecasting applications only require simple inexpensive irradiance sensors to be purchased, and the remainder of the data can be retrieved from nearby locations. The interface can support the use of other imaging data with which to predict patterns such as satellite or NWP images.

## IV. Conclusions

The work performed in this project was two-fold, the development of a large data and image library and the development of the connecting interface. The database contains hundreds of relevant data elements and as more discoveries are made through the use of the interface, more elements can be added to the database as necessary. The goal of the knowledge-base library and interface is to improve upon existing solar forecasting approaches and discover new methods of forecasting. Since the interface is a combination of automation and human interaction, it is possible for different users to gain insights that others may not acquire, though they are using the same tool. The tool has very significant potential in several applications and disciplines.

## V. Acknowledgements

The authors would like to thank their research sponsor CPS Energy for their generous support. This research was made possible by the Research Alliance between the San Antonio electric utility CPS Energy and the Texas Sustainable Energy Research Institute of The University of Texas at San Antonio. We would also like to thank Drs. Mo Jamshidi and Sos Agaian for their helpful advice on this work.

## REFERENCES


[1] R. Kohavi, N. Rothleder, and E. Simoudis, "Emerging Trends in Business Analytics," *Communications of the ACM, Volume 45, Issue 8*, 2002, pp. 45–48.

[2] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, “From Data Mining to Knowledge Discovery in Databases,” *AI Magazine, Volume 17,* 1996, pp. 37-54.

[3] W.Glassley, J. Kleissl, H. Shiu, J. Huang, G. Braun, and R. Holland, “Current state of the art in solar forecasting,” *Final report, Appendix A, California Renewable Energy Forecasting, Resource Data and Mapping*, California Institute for Energy and Environment, 2010.

[4] J. Sabburg and C. N. Long, “Improved sky imaging for studies of enhanced UV irradiance,” *Atmospheric Chemistry and Physics, Volume 4*, 2004, pp. 2543–2552.

[5] P. Bacher, H. Madsen, and H.A. Nielsen, “Online short-term solar power forecasting,” *Solar Energy, Volume 83, Issue 10,* October 2009, pp. 1772-1783.

[6] P. Mathiesen and J. Kleissl, “Evaluation of numerical weather prediction for intra-day solar forecasting in the continental United States,” *Solar Energy, Volume 85, Issue 5*, May 2011, pp. 967-977.

[7] H.M. Diagne, M. David, P. Lauret, and J. Boland, “Solar irradiation forecasting: state-of-the-art and proposition for future developments for small-scale insular grids,” in *Proceedings of the World Renewable Energy Forum 2012* (WREF 2012), Denver, USA, May 2012.

[8] R. Perez, K. Moore, S. Wilcox, D. Renné, and A. Zelenka, “Forecasting solar radiation – Preliminary evaluation of an approach based upon the national forecast database,” *Solar Energy, Volume 81, Issue 6,* June 2007, pp. 809-812.

[9] Chincor N, Thomas J, Wong P, Christel M, Ribarsky W. “Multimedia Analysis + Visual Analytics = Multimedia Analytics,” NSF Grant No. IIS-0705491.

[10] Aigrain P, Zhang H, and Petkovic D. “Content-Based Representation and Retrieval of Visual Media: A State-of-the-Art Review,” *Multimedia Tools And Applications, Volume 3*, 1996, pp. 179-202.