Data Mining Determination of Sunlight Average Input for Solar Power Plant

Fl. Loury, P. Sablonière, C. Lamoureux, G. Magnier, Th. Gutierrez

Abstract—A method is proposed to extract faithful representative patterns from data set of observations when they are suffering from non-negligible fluctuations. Supposing time interval between measurements to be extremely small compared to observation time, it consists in defining first a subset of intermediate time intervals characterizing coherent behavior. Data projection on these intervals gives a set of curves out of which an ideally “perfect” one is constructed by taking the sup limit of them. Then comparison with average real curve in corresponding interval gives an efficiency parameter expressing the degradation consecutive to fluctuation periods of the year. The extracted information already gives plant control.

Keywords—Base Input Reconstruction, Data Mining, Efficiency Factor, Information Pattern Operator.

I. INTRODUCTION

Explain the significance of sunlight as an input for solar power plants and the challenges in predicting its availability.

II. RAW DATA ANALYSIS

Data base \( D = D(\mathbf{y}_i, \Delta) \) typically contains sunlight measurements over years \( \mathbf{y}_k \ (k=1,2,N) \) collected every time interval \( \Delta \) of the day representing \( N \times J \times v(\Delta) \) values of sun power at the location where the solar plant will be build up, where \( J \) is the average number of days in a year in the observation interval and \( v(\Delta) \) the number of daily measurements. To find a pattern over the years (following natural frequency of solar lighting), the idea has been to compare the sunlight during a typical week \( W^{k}_m \) of each trimester for each of \( k=1,2,..N \) recorded years, so \( m = n+13(t \not= I) \) with \( 0 < n < 13 \) and \( t = 1,2,3,4 \) the chosen trimester number. The choice of a week time period as a base representative
“unit” $\mathcal{F}$ is mainly motivated by the needs to have a relevant interval such that $\Delta \ll \mathcal{F} \ll Y$ in the sense that it typically represents sunlight during the trimester in which it is located in the year. One then gets a set of $4N$ curves expanding over the seven days of the considered week, i.e. with seven peaks for day time and 0 at night. Considering their envelope $E_{\mathcal{F}}$, it is quite evident that there will be larger difference between their peaks as latitude of measurements is higher. On the other hand, variations of sunlight are the result of a “perfect” sunlight modified by meteorological events which will be considered as random events (hence the choice of the week time period because it corresponds to the maximum correlation time for a heat unit coming from the sun before random dilution in atmosphere).

So to create the “theoretical” model of “perfect” sunlight at measurement location, the maximum sunlight value at each hour of observation years (i.e. over initially collected $24\times7\times52\times N$ data) has been selected. This rests upon the assumption that at measurement location, the climate will stay in the same state of repeatability determined from the average over the $N$ observation years. This allows end up with a set of only four curves representing average daily “perfect” sunlight for each trimester in the year once again constructed from observations of past $N$ years. Of course they each reflect sunlight situation during the trimester of the year with more power during summer than in winter in north hemisphere.

Next step in modeling process of sunlight variations due to meteorological events (mainly cloudiness) is to compare average day with previous “perfect” day sunlight for each trimester. By least square method it is possible to evaluate the effect of this meteorological bias as an efficiency factor $\phi$ summarizing over all the consequence of “imperfection” of local sunlight. It should be noticed that real sunlight being the product of “perfect” one by efficiency factor, it happens that even if “perfect” sunlight is larger during a trimester than in another one, the final sunlight felt on the ground may be nevertheless comparable or even larger with smaller “perfect” one due to heavy possible cloudiness considerably reducing efficiency factor in larger sunlight period. Such a result is typical of very sunny intermediate spring and autumn periods as compared to cloudy summer period which are observed in specific places, and justifies a preliminary careful choice of solar plant location for best efficiency all year round based on present analysis.

III. APPLICATION

Following previous steps, a data base has been first collected corresponding to sunlight measurements for $N = 5$ consecutive years 2000 to 2004 at each hour all representing a total of $N = 5\times 7\times 24 \times 52 = 43,680$ sunlight values. Typical week has next been determined from observation and comparison in each trimester of the year, and came out with week numbers $w_{\mathcal{F}} = 10,21,35,48$ for all $N$ years and the four trimesters in order respectively. Plotting data for each week gives the following sets of $N$ curves, see Figs. 1-4 representing sunlight measurements in W/m² for the various hours in the day.

![Fig. 1 Sunlight Measurements during Week 10 for Trimester 1](image1)

It is verified that sunlight maximum value is up to 700 W/m², but under cloudiness it can abruptly drop to 100 W/m² (during year 2001).

![Fig. 2 Sunlight Measurements during Week 21 for Trimester 2](image2)

Here the common relative peak is increased as compared to Trimester 1 and reaches 900 to 950 W/m². It can be observed that meteorological effects are more strongly influencing final observed sunlight values than during Trimester 1.

![Fig. 3 Sunlight Measurements during Week 35 for Trimester 3](image3)

The value of common relative peak is lower than for previous Trimester and stays around 800W/m². Also meteorological effects are less important and curves are smoother than for previous Trimesters.
peak is much lower around 425 W/m². Also sunlight period is
very important in strongly modifying the finally received
sunlight which could have been much more favorable in
“perfect” circumstances.

IV. CONCLUSION

Determination of realistic inputs in production systems is a
very important step in analysis of their final performances, and
is the more difficult as these inputs are of random
unmanageable nature. Solar power plants belong to this class.
In present study a method has been proposed to represent
actual sunlight inputs by an interpreter out of which analysis
of possible power plant performance receiving these inputs
and its control can be determined. It consists in defining first a
typically representative trimester reference sunlight curve by
analysis of measurement data collected over a long enough
preceding period, giving the theoretical “perfect” local
sunlight, and in correcting it in a second step by a
“meteorological” efficiency factor which reduces in
proportion expectable sunlight. This approach concentrates in
only two elements the inputs to plant system. Even if as
expectable it basically restricts good potential plant locations
to low latitude and low cloudiness ones, it also allows
compare possible locations with respect to these two aspects
and provides an interesting element of choice depending on
proposed utilization. Plant control analysis can be undertaken
next as shown elsewhere.

ACKNOWLEDGMENT

The authors are very much indebted to ECE for having
provided the necessary environment where the study has been
developed and Pr. M. Cotsaftis for his help in preparing the
manuscript.

REFERENCES

Learning: Data Mining, Inference, and Prediction, 3rd ed., Springer,
New York, 2009
[3] M. Kantardzic : Data Mining: Concepts, Models, Methods, and
L. Kurgan, P. Musilek : A Survey of Knowledge Discovery and Data
pp.1–24, 2006
Mining Applications, Acad. Press, New York, 2009
[5] Pang-Ning Tan, M. Steinbach, V. Kumar : Introduction to Data Mining,
Addison-Wesley, Reading, Mass, 2005
[6] Xin-guan Zhu, I. Davidson : Knowledge Discovery and Data Mining:
Challenges and Realities, Hershey, New York, pp.31–48, 2007
Reviews: Data Mining and Knowledge Discovery, Vol.1(5), pp. 431–
445, 2011
Scaling Algorithms, Applications and Systems, Kluwer Acad. Publ.,
Amsterdam, 1999
Acad. Press, New York, 2009
New York, 1998
Production Plant, to be published.