International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:7, No:9, 2013

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 effect. The method is applied to sunlight data collected in a specific place, where ideal sunlight is the one resulting from direct exposure at location latitude over the year, and efficiency is resulting from action of meteorological parameters, mainly cloudiness, at different periods of the year. The extracted information already gives interesting element of decision, before being used for analysis of plant control.

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

I. Introduction

EXPLICIT calculation of system dynamics is very often requiring the knowledge of elements belonging to system environment, because of intricate interactions rendering system isolation more difficult, if sometimes not possible. They play the role of system inputs, and a real problem is to find their correct representation to deal with so that system dynamics can be handled by adequate control in compatibility with general constraints on time and space imposed by solution granularity [1]. For instance, when analyzing piston dynamics in a cylinder under gas molecules collisions on one side, it is completely acceptable to represent their effect globally by a pressure term once Knusen number $\mathcal{K} < 1$ (easily satisfied at usual pressure), and adequate control in this case is faithfully based on pressure. More generally, the problem exists anytime the raw inputs affecting system dynamics have a space-time granularity much smaller than characteristic one for system under study [2]-[5]. In this case simple averaging moments to adequate order are known to be sufficient for providing correct approximation to system dynamics with error evaluation. Situation is much more difficult when input fluctuations are comparable to system space-time granularity. In this case, it is necessary to find faithful enough representation for these raw data to become coherent and acceptable inputs [6], [7]. Many different processes have been worked out to set up corresponding transformation, depending on correlation degree exhibited by fluctuations [8]. Worst case occurs when inputs are very regular phenomena randomly modified by "large" perturbations, ie able to change significantly regular input amplitude.

This is typically the case where the system is a solar plant, regular input is sunlight which is with clear sky very easily predictable over the year at any location on the Earth, and perturbation is a meteorological event such as cloudiness reducing initial sunlight by sometimes extremely large factor. To correctly analyze such an inevitable situation with the development of alternative energies, simple filtering is not sufficient and more elaborated data mining methods have to be used [9]. The idea is to find a pattern $\mathcal{P}$ which will structure raw data $\mathcal{D}$ for being interpretable. This is done via an interface $\mathcal{I}$ which extracts organized information from data [10], [11]. The process is on-line for controlled real time systems and may be very demanding. Another easier case is test case off-line analysis where it is intended to evaluate system response performance $\mathcal{R}$ and sensitivity $\mathcal{S}$ with different control laws $\mathcal{L}$ in order to get most appropriate robustness ball $\mathcal{B}$ within which system dynamics remain under control against a class of possible fluctuating inputs $\mathcal{F}$. In this last situation, the task is to define a representative averaged input class $\langle\mathcal{F}\rangle = \mathcal{I}\mathcal{F}$ from the set of possible ones $\mathcal{F}$ such that $\mathcal{R}(\langle\mathcal{F}\rangle, \mathcal{L}) \subset \mathcal{B} \Rightarrow \mathcal{R}(\mathcal{F}, \mathcal{L}) \subset \mathcal{B}$. The problem has been already addressed elsewhere [12], and only the definition of interface $\mathcal{I}$ will be discussed here.

II. Raw Data Analysis

Data base $\mathcal{D} = \mathcal{D}(\mathcal{Y},\Delta)$ typically contains sunlight measurements over years $\mathcal{Y}_k$ ($k=1,2,N$) collected every time interval Δ of the day representing $\mathcal{N} = N\times J\times\nu(\Delta)$ 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 $\nu(\Delta)$ 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_m^k of each trimester for each of $k=1,2,..N$ recorded years, so $m = n+13(t-1)$ 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

Fl. Loury is with the ECE Graduate School of Engineering, Paris, 75015, France (phone: +33613858627; e-mail: loury@ece.fr).
P. Sablonière is with the ECE Graduate School of Engineering, Paris, 75015, France (phone: +3325046960; e-mail: sablonie@ece.fr).
C. Lamoureux is with the ECE Graduate School of Engineering, Paris, 75015, France (e-mail: lamoureu@ece.fr).
G. Magnier and Th. Gutierrez are with the ECE Graduate School of Engineering, Paris, 75015, France (e-mail: magnier@ece.fr, gutierre@ece.fr).

"unit" $\mathcal{T}$ is mainly motivated by the needs to have a relevant interval such that $\Delta << \mathcal{T} << \mathcal{Y}$ 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 $\mathcal{E}_m$ 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 $24\times7\times52\times\mathcal{N}$ 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 ϕ 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 $\mathcal{N} = 5\times7\times24\times52 = 43{,}680$ sunlight values. Typical week has next been determined from observation and comparison in each trimester of the year, and came out with week numbers $W_m^k = 10,21,35,48$ 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/m^2 for the various hours in the day.

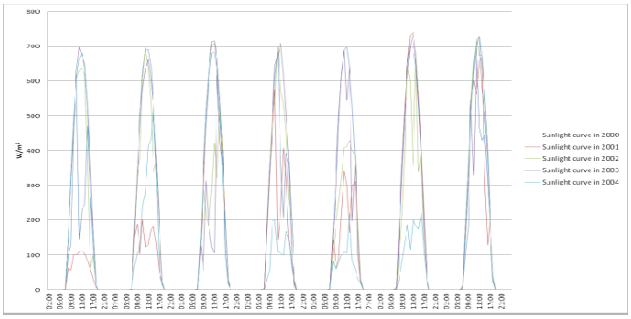

Fig. 1 Sunlight Measurements during Week 10 for Trimester 1

It is verified that sunlight maximum value is up to *700* W/m^2, but under cloudiness it can abruptly drop to 100 W/m^2 (during year 2001).

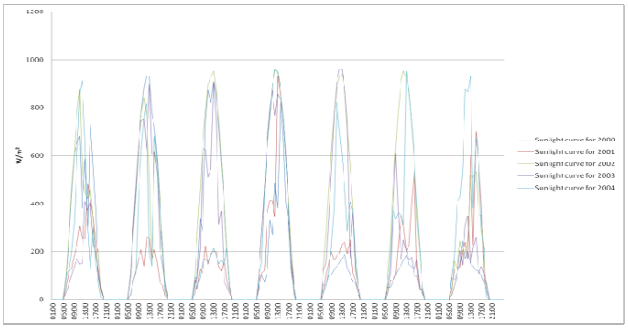

Fig. 2 Sunlight Measurements during Week 21 for Trimester 2

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.

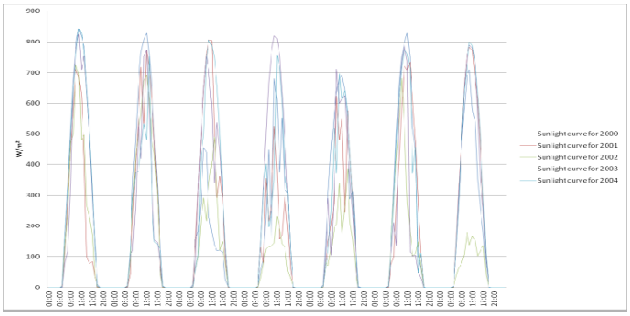

Fig. 3 Sunlight Measurements during Week 35 for Trimester 3

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.

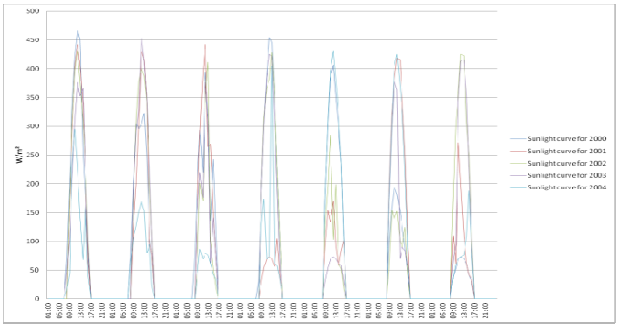

Fig. 4 Sunlight Measurements during Week 48 for Trimester 4

As expectable for winter time the value of common relative peak is much lower around 425 W/m². Also sunlight period is much shorter during the days.

From all data and following the procedure explained in previous paragraph, the four "perfect" sunlight curves corresponding to each Trimester (for the interval of N selected observation years) are obtained as shown on Fig. 5.

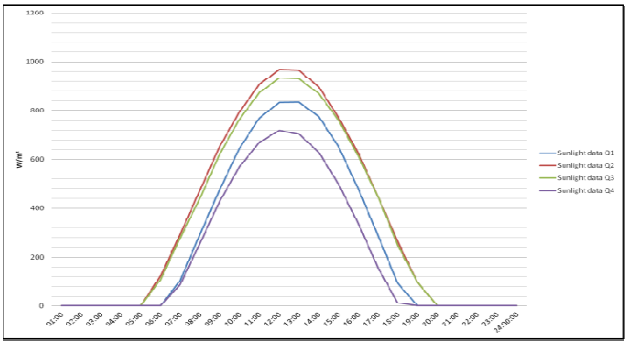

Fig. 5 2000 to 2005 Theoretical "Perfect" Sunlight Quarterly Curves

Evidently the curves are reflecting the season difference due to latitude. Comparison with daily average "real" ones gives the efficiency coefficient ϕ by the percentage of hatched part underneath "perfect" previous curves, and which will be retained in final sunlight trimester description.

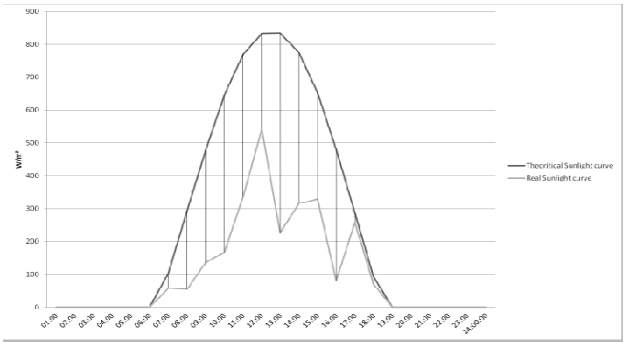

Fig. 6 Comparison of Theoretical ''Perfect'' Sunlight with Real one for Efficiency Evaluation

As already indicated, the influence of local climate can be 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

[1] J.A. O'Brien, G.M. Marakas : *Management Information Systems*, McGraw-Hill, New York, 2011

[2] T. Hastie, R. Tibshirani, J. Friedman : *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 3rd ed., Springer, New York, 2009

[3] M. Kantardzic : *Data Mining: Concepts, Models, Methods, and Algorithms*, J. Wiley & Sons, New York, 2003
L. Kurgan, P. Musilek : A Survey of Knowledge Discovery and Data Mining Process Models, *The Knowledge Engineering Review*, Vol.21(1), pp.1–24, 2006

[4] R. Nisbet, J. Elder, G. Miner : *Handbook of Statistical Analysis & Data 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] Xingquan Zhu, I. Davidson : *Knowledge Discovery and Data Mining: Challenges and Realities*, Hershey, New York, pp.31–48, 2007

[7] R. Mikut, M. Reischl : Data Mining Tools, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Vol.1(5), pp. 431–445, 2011

[8] S. Chandrasekhar: Stochastic Problems in Physics and Astronomy, Rev. Mod. Phys., Vol.15(1), pp.1-89, 1943

[9] Y. Guo, R. Grossman, (editors) : *High Performance Data Mining: Scaling Algorithms, Applications and Systems*, Kluwer Acad. Publ., Amsterdam, 1999

[10] S. Theodoridis, K. Koutroumbas : *Pattern Recognition*, 4th Edition, Acad. Press, New York, 2009

[11] S.M. Weiss, N. Indurkhya : *Predictive Data Mining*, Morgan Kaufmann, New York, 1998

[12] Fl. Loury, P. Sablonière : *Profit Optimization for Solar Electricity Production Plant*, to be published.