Mansoura Engineering Journal

Volume 25 | Issue 3 Article 9

2-5-2021

On Forecasting with Blind Data.

Hassan Soltan
Production and Mechanical Design Engineering Department.,. Faculty of Engineering., El-Mansoura University., Mansoura 35510, Egypt., hasolan@hotmail.com

Recommended Citation
Soltan, Hassan (2021) "On Forecasting with Blind Data.," *Mansoura Engineering Journal*: Vol. 25 : Iss. 3 , Article 9.
Available at: https://doi.org/10.21608/bfemu.2021.146696

Mansoura Engineering Journal, (MEJ), Vol. 25, No. 3, September 2000. M. 28

ON FORECASTING WITH BLIND DATA

عن التنبؤ في ظل بيانات غير مؤهلة

Hassan Ali Soltan
Prod. and Mech. Design Engng. Dept., Faculty of Engineering,
Mansoura University, Mansoura 35516, Egypt
Email: hasoltan@hotmail.com

خلاصة :

يستخدم المؤلف لفظ "بيانات غير مؤهلة" للتعبير عن البيانات التي لا يمكن استخدامها مباشرة للاستدلال. و توجد هذه البيانات غالبا في صورتين، البيانات المشوهة و البيانات متعددة الطيات. و تتواجد الصورة الأولى عندما يتعرض النظام لظروف غير مستقرة بينما تتواجد الثانية في الحالات الدورية المماثلة للطلب على الطاقة الكهربية. و يقدم هذا البحث مساهمة لكلا الصورتين حيث أنه يتفحص طريقة تنبؤ خاصة بالبيانات المشوهة والتي قدمها ذات المؤلف من قبل بالإضافة إلي أنه يقترح طريقة تنبؤ أخرى خاصة بالطلب على الطاقة الكهربية خلال أربعة و عشرين ساعة. و قد تم توضيح الطريقة الثانية على منحنى بيانات موجود و هذه الطريقة بدورها يمكن تطبيقها على البيانات المماثلة.

ABSTRACT

The author used the term blind data to express the data which can't be directly extrapolated or managed in other ways. Such data almost exist in two forms—the distorted data and multi-folded data. The first appears when the system is subjected to uncontrolled environment and the second appears when the output curve consists of folds—seem to be cycles—having equal periods along the time horizon such as the electricity demand. This paper investigates, by using simulation, a forecasting method developed earlier by the author for distorted data and proposed another forecasting method can be applied for hourly and half-hourly forecasting of electricity demand. An existing demand curve is used to explain the logic of the latter method, which in turn can be extended to the cases that have a behavior similar to the electricity demand.

INTRODUCTION

Earlier studies have shown that time series analysis is the most commonly quantitative technique used to model forecasting in business and production (Gardner and McKenzie 1985, Soltan 1998, Ragsdale and Plane 2000.) Accurate forecasts or at least forecasts enable a company to build a production or service plan with minimum expected interruption. Concerning the blind data, electricity demand forecasting has received a considerable attention (see Taylor and Majithia 2000 and the papers therein) but forecasting with distorted data seems to be neglected (Soltan 1998.)

Accepted September 30, 2000

Soltan (1998) has developed a heuristic method to solve the forecasting problems that have distorted data—he meant the data that are found impossible to follow any trend or can be smoothed by using the known techniques. This method will be successful when it is necessary to forecast under such critical conditions. Under a few assumptions, he stated that the problem can be solved using the time series analysis after combining the original data into several groups having equal number of points along the time horizon. All groups were assumed to follow a specific probability distribution with estimable averages and standard deviations. Moreover, he stated that at least one of the known averages (mean, mode, etc.) can be used as an estimator offers a good trend leading to an expected forecast for the next group average. After that, an expected forecast for a point can be easily obtained under an estimated probability. The term random quantities was used in the sense of each consecutive number of data but not for a distribution at each point; in other words it is different from forecasting stochastic demand or demand variation (review Bell 2000 for more details about this point.) The first contribution of this paper is to investigate the method developed by Soltan (1998) for distorted data.

The electricity demand forecasting was invoked since the industrial change of the structure of electricity supply all over the world. Forecasting the electricity demand is often based on the short-term. Day-ahead half-hourly or hourly demand forecasts are necessary for time scheduling of power and for estimating the daily electricity pool price. Bunn (1982) reviews many methods developed for short-term forecasting in the electricity supply industry. A popular heuristic approach, based on a past demand curve, predicts 10 or 11 points on the demand curve—cardinal points—and then forecasts half-hourly demand by a procedure called profiling which fits a curve to the cardinal points (Baker 1985, Taylor and Majithia 2000) The past demand curve is often a previous day demand which is chosen subjectively according to similar conditions while it was suggested that profiling half-hourly forecasts of the next day should be extended to more than one past curve. For that concern, Ramanathan et al. (1997) obtained forecasts by using separate models for each hour of the day and suggested the application of switching and combination procedures to extend their approach. Taylor and Majithia (2000) investigates possible profiling improvements by using a controlled weighting mechanism. However, there is no statement for a best approach to short-term electricity demand forecasting. The second contribution of this paper represents a method allows the use of unlimited number of past demand curves to forecast half-hourly or hourly electricity demand.

THE ANALYSIS

Distorted Data

The main assumption in the method developed by Soltan (1998) was to have a sufficient number of periodical data to enable the combining procedure to set manageable groups. Therefore, it is basically designed for the systems which respond quantities in short-periods. The objective here is to investigate the proposition that the group averages approach a trend that can be processed by using the time series analysis. In addition, it is required to differentiate between those averages from the view of preference.

A simulation procedure is followed to generate uniformly distributed data in groups having the same number of data Ten consecutive groups were generated each time with group sizes 30, 50, 60, 70, 80, and 120 observations respectively. The statistics calculated were standard deviation, arithmetic mean, median, geometric mean, harmonic mean, and root mean square. The plots of those statistics are depicted in Fig 1 (size 30), Fig. 2 (size 50), Fig. 3 (size 60), Fig 4 (size 70), Fig. 5 (size 80), and Fig 6 (size 120) respectively. Note that the mode is not calculated in this analysis to avoid the subtle of its existence, but it could be examined if real data are available The procedure can be simply carried out using electronic spreadsheets—*EXCEL 97* for instance—which become common software tools managers use to analyze data and model quantitative problems (Ragsdale and Plane 2000.)

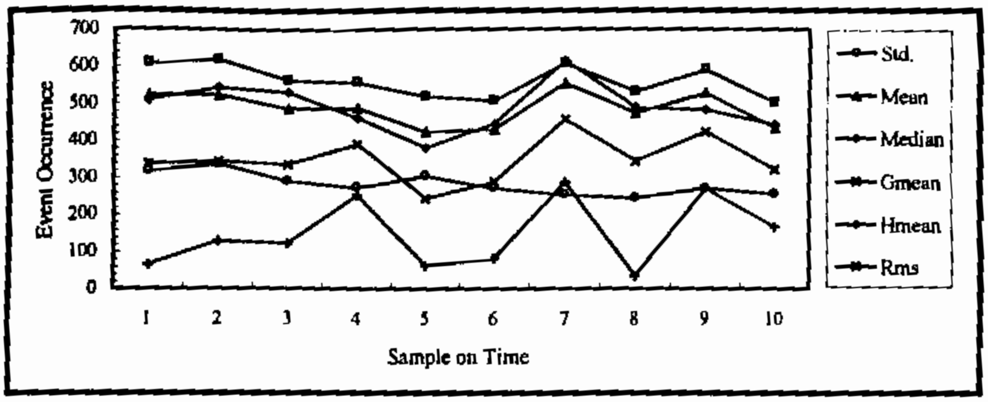

Fig. 1 Sample of 30 uniform random observations.

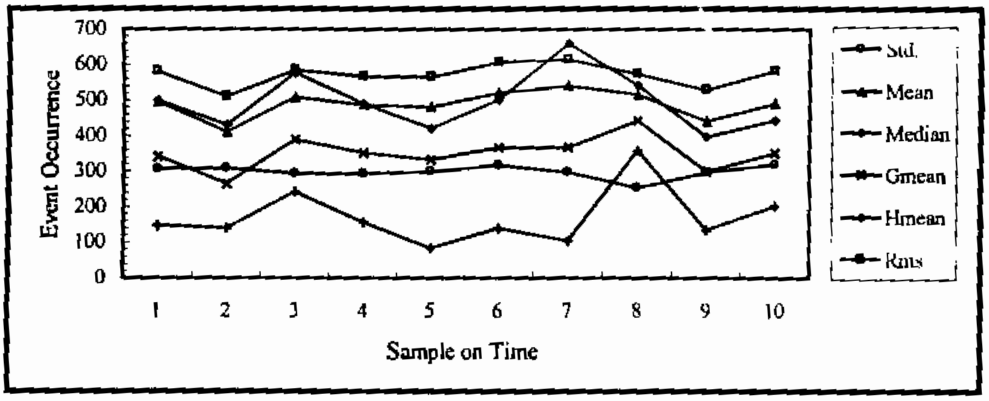

Fig. 2 Sample of 50 uniform random observations.

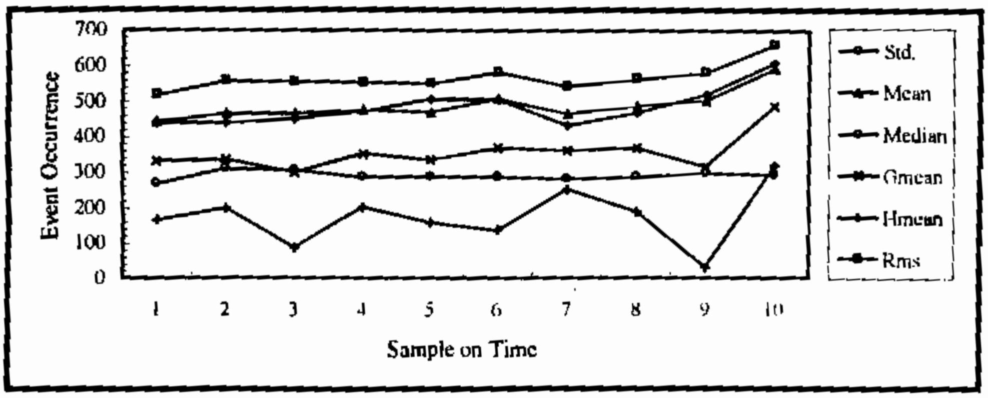

Fig. 3 Sample of 60 uniform random observations.

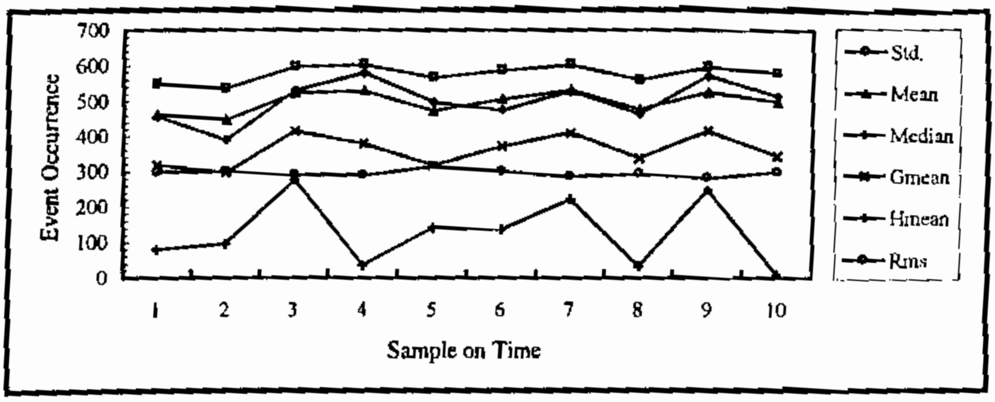

Fig. 4 Sample of 70 uniform random observations.

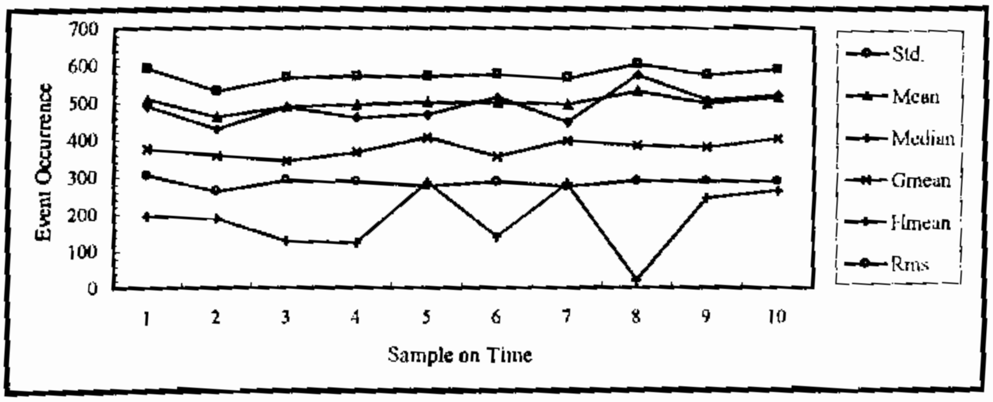

Fig. 5 Sample of 80 uniform random observations.

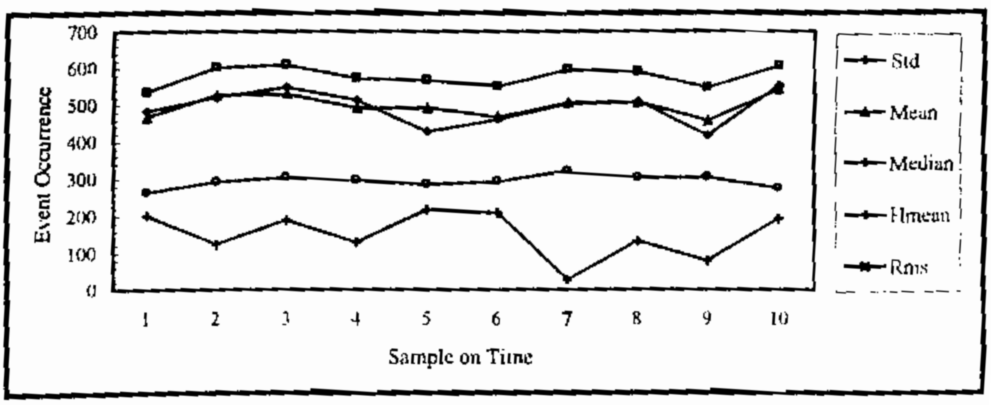

Fig. 6 Sample of 120 uniform random observations.

It seems from Figs. 1 to 6 that the harmonic mean is the worst trend followed by the geometric mean and all trends become better as the group size increases. The standard deviation fits a very good trend, which makes the related calculations accurate. Note that the geometric mean is not clearly defined for the bigger sample sizes, therefore it does not appear in Fig. 6 (size 120). However, the arithmetic mean, median, and root mean square behave in a similar fashion with trends having high determination factors. Thus verifying the statement of Soltan (1998) and making those three averages more reliable for group average forecasting. There is no statement for the best group size whereas it can be selected using experimentation. There are two trade offs—smaller the group size, the more accurate forecast for the included points and vice versa for a larger group size. The advantage of a small group size is always achieved if a trend is discovered. But the planner should be aware about the nature of data and the required forecasts. The analysis proves that the method introduced by Soltan (1998) can be fed as a complementary part to the time series analysis to deal with unstable and unconventional systems.

Another bit of work is necessary if the data distribution is not known in advance—Test of Hypothesis about the group-original data probability distributions. If the experiment fails to define a probability distribution, the planner may split the group forecasted average by the number of data included in each group but he should be conservative about the quantity declaration (Refer to the equations in Soltan 1998.)

However, the distorted data in production and business are still in need for extensive work taking in consideration the simple statistical techniques. This because the reliability of the forecasts may disappear with the complicated techniques especially if they involve many validation assumptions.

Multi-Folded Data

As declared before, the most known form of multi-folded data is the hourly or half-hourly electricity demand which may cycle every day (see Fig. 7) with change in the curve parameters. The profiling approach has a salient drawback that it restricts the number of used past curves. Other drawbacks exist; it needs to extensive analysis and its application is limited to the short-term forecasts. Although the short-term forecasting was received a considerable attention, the profiling is still the most popular approach. Moreover there is a statement that there is no consensus as to the best approach to short-term electricity demand forecasting (Taylor and Majithia 2000.) Therefore, this paper has tried to discover a simple method makes the available data managed by known approaches to the time series analysis. It will be at least a guide to the profiling approach if it is necessary. It gives an initial forecasting solution which may be more accurate than profiling.

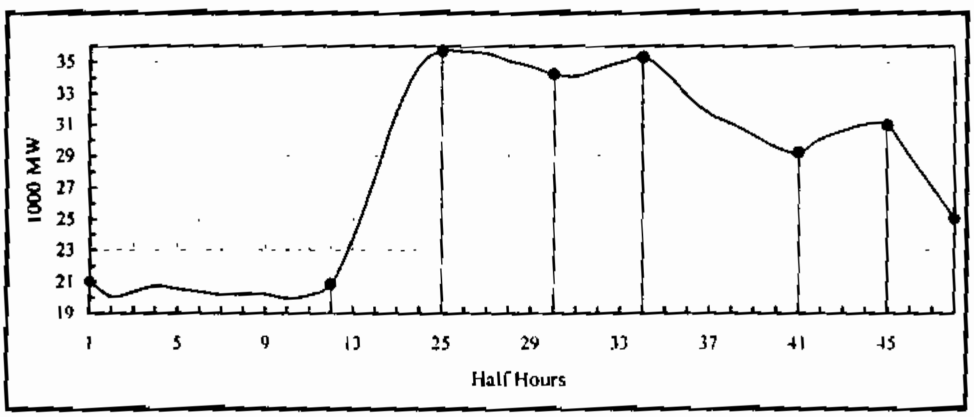

Fig. 7 A day actual demand of a company's power in MW: An approximation to the case curve found in Taylor and Majithia (2000), J. Opl. Res. Soc., Vol. 51, No. 1, p. 74.

The method will be explained guided by an actual data curve, shown in Fig. 7, at the National Grid Company in England, which is responsible for the transmission of electricity. The curve is divided into parts. Each part should include a group of points having a specific trend. It is preferred that the planner looks to more than one past curve to settle those divisiuns subjectively or quantitatively by using some statistical techniques. The objective here becomes to forecast the parameters of each division trend by using a number of past data curves on the short- or long-term. After that the forecasts of any hour or half-hour can be obtained assuming that it is located in a specific division (see Fig. 7.)

According to the given data curve, we obtain half-hourly seven divisions—1-12, 12-25, 25-30, 30-34, 34-41, 41-45, and 45-48 respectively. The first division fits approximately an exponential curve whereas the other divisions can be approximated by straight line fittings. Hence, the parameters of those trends can be forecasted easily if other past data curves are available. Points 1, 12, 25, 30, 34, 41, 45, and 48 respectively replace the cardinal points in the profiling approach and can be used to adjust the forecasted trends because they represent the most critical points on the curve.

As clear, this procedure makes the data relent to the simplest time series analysis; therefure the objective of the work is achieved at this point. Also the techniques of making the curve division are beyond the topic of this paper because they depend on the nature of dnta and the importance uf analysis. However, it can be manipulated subjectively.

CONCLUSION

This paper has addressed the problem of forecasting, which deals with unconventional data forms called by the author, blind data—ineluding distorted data and multi-folded data. Those data are not available directly to the time series analysis. An earlier approach for distorted data is investigated by the experiment, which exhibited its effectiveness and validation of its statements. Another approach is introduced for the multi-folded data (electricity demand is considered as an example) forecasting. The approaches seem to be intuitively appealing because they were construeted with simple statistieal techniques. Moreover, they ean be easily understood and computerized by the practitioners.

Moreover, the first approach can be applied to the high variability multi-folded data by eonsidering each fold as a single group, but in theses cases it fails to expect forecasts for an individual point especially in the ease of hourly electrieity demand. However, it ean supply an aceurate forecast for the total eleetricity demand during the whole day.

Profiling is the most popular approach for foreeasting electricity demand, nevertheless there is no reliability about its forecasts, although it is subjected to several improvements aod trials to include more than one past data curve (refer to Taylor and Majithia 2000.) However, there is no assurance that the current approaches, including our approach, obtain the optimum short-term forecasts for hourly or half-hourly electricity demand. Therefore, as a future work the author proposes the integration between profiling and the approach presented here or another similar one to eliminate extensive statistical work, minimize the foreeasting error, and deal with the long-term forecastiog.

REFERENCES

Baker, A.B., "Load forecasting for scheduling generation on a large intereonnected system," In: Bunn and Farmer (eds). Comparative Models for Electrical Load Forecasting, 1985. Wiley, NY, pp 57-67.

Bell, P.C., "Forecasting demand variation when there are stockouts." J. Opl. Res. Soc., Vol. 51, 2000, pp 358-363.

Bernard, P., "Three simple techniques for improving statistical forecasts," Prod. and Inventory Mgmt. J., Fourth Quarter, 1997, pp 65-69.

Bunn, D.W., "Short-term forecasting: a review of proeedures in the electrieity supply industry," J. Opl. Res. Soc., Vol. 33, 1982, pp 533-545.

Gardner, E. and E. McKenzie, "Forecasting trends in time series," Mgmt. Sci., Vol 31, 1985, pp 1237-1246.

Ragsdale, C.T. and D.R. Plane, "On modeling time series data usiug spreadsheets," Omega, Vol. 28, 2000, pp 215-221.

Ramanathan, R. et al., "Short-run forecasts of electricity loads and peaks," Int. J. Forecast, Vol. 13, 1997, pp 161-174.

Soltan, H.A., "Forecasting method for weak trend quautities," Computers and Ind. Engng., Vol. 35, 1998, pp 407-410.

Taylor, J.W. and S. Majithia, "Using combined forecasts with changing weights for electricity demand profiling," J. Opl. Res. Soc., Vol. 51, 2000, pp 72-82.