

# Understanding Power System Behavior Through Mining Archived Operational Data

Sarasij Das, P S Nagendra Rao

**Abstract**—This paper is the outcome of an attempt in mining recorded power system operational data in order to get new insight to practical power system behavior. Data mining, in general, is essentially finding new relations between data sets by analyzing well known or recorded data. In this effort we make use of the recorded data of the Southern regional grid of India. Some interesting relations at the total system level between frequency, total MW/MVAR generation and average system voltage have been obtained. The aim of this work is to highlight the potential of data mining for power system applications and also some of the concerns that need to be addressed to make such efforts more useful.

## I. INTRODUCTION

Advances in electronics, computer and information technology are fueling major changes in the area of power system instrumentation. More and more microprocessor based digital instruments are replacing analog meters. Data logging is becoming automatic and frequent. Vast quantities of data generated by extensive deployments of digital instruments are creating information pressure on Utilities. The legacy SCADA based data management systems do not support management of such huge data. The present practice is to store the acquired data in SCADA for only a few months and then delete. In few cases after removing from the SCADA system, these data are stored in compact discs. At present the usefulness of historical data is not fully explored. So, utilities do not give importance to store such data efficiently.

The traditional integrated power industry is going through a deregulation process. The market principle is bound to force competition between power utilities, which in turn demands a higher focus on profit. To optimize system operation and planning utilities need better decision-making processes that depend on the availability of reliable system information. It is expected that in this context historical data is going to be a vital asset. In [1] some possible applications of historical power system data is presented.

Apart from a business perspective, historical data is important from another point of view also. Electric power system is a very complex system. Most of the mathematical models used for analyzing/predicting system behaviors are based on several assumptions. The availability of detailed measurements of power system parameters could provide an opportunity to

validate many of such models. The importance of data is being recognized widely in the recent times.

Power system data management has been discussed in several works[1][2][3]. Data warehousing technology is being proposed to meet the future requirement of power systems. In [4] data mining as a feature of power system data warehouse is mentioned. In [5] it has been mentioned that power systems operation can be greatly improved through data analysis and/or assimilation

In our work real system data is analyzed to find interrelation between several system parameters. This analysis can be viewed as a small attempt of data mining. Data mining is essentially an analysis of data sets in order to discover new relations between various quantities which is not obvious from the recorded data in its normal form.

For our investigation, data of five system parameters- voltage, frequency, MW and VAR generation, system demand - of the southern regional grid of India, collected from Southern Regional Load Despatch Center has been used.

This paper is organized as follows. Section II outlines some features of the Southern Regional Grid. In Section III a brief description of the data set used is given. Section IV presents the results of data analysis. The paper is concluded in Section V.

## II. SOUTHERN REGIONAL GRID FEATURES

The southern regional grid of India covers an area of 6,51,000 Sq. km encompasses four states namely Andhra Pradesh, Karnataka, Kerala, Tamilnadu and one Union Territory of Pondicherry. This region comprises of several central and state owned generating stations, independent power producers, distribution companies and state transmission utilities. In India, including SRLDC, there are five major regional grids. After August 2006 four of the five regions excluding the southern region are (synchronously) interconnected. The Southern region is connected with the other regional grids only in an asynchronous manner. Southern region is connected to Western region through HVDC back to back at Ramagundam-Chandrapur and to Eastern region at Jepore-Gazuwaka back-to-back and point to point HVDC line between KolarTalcher. The total installed capacity of southern region as in the beginning of 2007 is about 37370 MW. Some other salient features of the southern regional grid are [6] :

- Covers approximately 19% of the geographical area, 22% of population and 29% of the installed capacity of the country.

Sarasij Das obtained M.Sc(Engg) in Electrical Engineering from Indian Institute of Science, Bangalore. He is currently with Power Research Development Consultant Pvt. Ltd. (e-mail: sarasijdas@gmail.com).

P S Nagendra Rao is with the Electrical Engineering Department, Indian Institute of Science, Bangalore.(e-mail: nagendra@ee.iisc.ernet.in).

- 30-70% hydro-thermal mix
- 3300 MW wind generating plants
- 8000 MW capacity independent power producers.
- 2000 MW capacity HVDC Talcher-Kolar double circuit interconnection with the Eastern Region.
- 400/220 KV transmission system

### III. DATA DESCRIPTION

The control center at Bangalore of SRLDC is equipped with a computerized load despatch and communication facilities. Around 320 Remote Terminal Units(RTU) are used for real time system monitoring and grid management. Through SCADA these RTUs communicate with the control center. From the collected data only five system parameters- voltage, MW generation, MVA<sub>r</sub> generation, frequency, system demand - are made available to us for use in this work. These data were in the form of Microsoft Excel files stored in Compact Discs. Data, starting from Jan 2004 to June 2006, is collected. The data logging interval is 1 minute for all parameters. Following are some salient features of the data of the five parameters are chosen.

#### A. Voltage

Voltage data consists of measurements at the 26 buses of the 400 KV grid. All voltages have been stored (in EXCEL file) as integers with a least count of 1 KV. Among the 26 bus voltages, the Kolar bus voltage remains constant at 400 KV all the time.

#### B. Frequency

Frequency data consists of frequency of the region. Precision of frequency data is 0.01216 Hz.

#### C. MW Generation

MW generation data consists of information from 60 generation buses. The precision of stored (in EXCEL file) data is 1 MW. The 60 outputs at generation buses represent either unit outputs or the total station output. Types of plants are hydro, thermal or nuclear.

#### D. MVA<sub>r</sub> generation

MVA<sub>r</sub> generation data corresponds to 88 generation units. Precision of stored(in EXCEL file) data is 1 MVA<sub>r</sub>. The 88 generation units include hydro, thermal and nuclear power units. In this case the data corresponds to individual units in all cases.

#### E. System demand

System demand data consists of system demand of four states and the total demand of the region. The precision of stored(in EXCEL file) data is 1 MW. One of the major challenges in handling real data is the issue of the outliers. Outliers are basically records whose features are distinct from the other records of the group. Outliers are classified into four classes. The first class arises from data entry error. For power systems data this type of error is generated due to malfunction of SCADA or communication

channels, etc. The second class of outlier is the outcome of some extraordinary events. In the context of power systems faults, switching operations etc could be the cause for such events. This type of outliers should be retained in the data set. The third class of outlier comprises of extraordinary observations for which there is no explanation. These outliers must be retained to capture some characteristics of the system (not known/explained). The fourth and final class of outlier are observations that fall within the range of each of the variables but are unique in their combination of values across the variables. In the data made available to us, it appears that no attempt had been made to identify and replace outliers. Investigated data set contains all the four types of outliers. The outliers identified in the data set are:

- Considerable change in value for in one or two consecutive instances
- Occurrence of 0 values for a considerable time period
- Occurrence of values not possible from system point of view

Identifiable outliers are substituted with average value of previous and next non-outlier values.

### IV. ANALYSIS OF PARAMETER INTERRELATIONS

An overview of some features of the collected data set has been given in the last section. In this section we investigate the interrelation of some of the measured system variables. In Figures 1 and 2 system frequency vs. average system voltage is plotted for 06/06/2006 and 16/4/2006. By average system voltage we mean average voltage of all 400 KV buses. It is seen that the points are clustered around one of the diagonals. It can be seen that to some extent higher voltages correspond to higher frequency and lower voltage corresponds to lower frequency. This correlation is relatively well defined in Figure 2.

In Figures 3 and 4 total system demand vs. system frequency is presented for 6/6/2006 and 16/4/2006. In Figures 3 and 4 by and large higher system demand corresponds to low system frequency and at lower demands system frequency is high.

Figure 5 presents the total system demand vs. average system

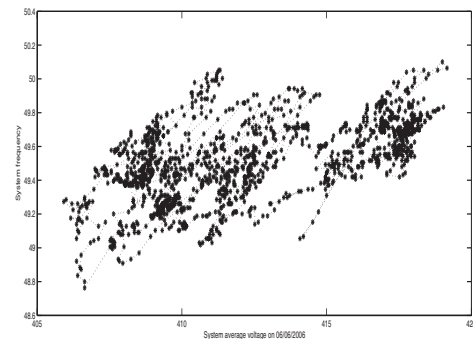


Fig. 1. System frequency vs. Average system voltage on 06/06/2006

voltage plot for 19/1/2006. Time instants corresponding to

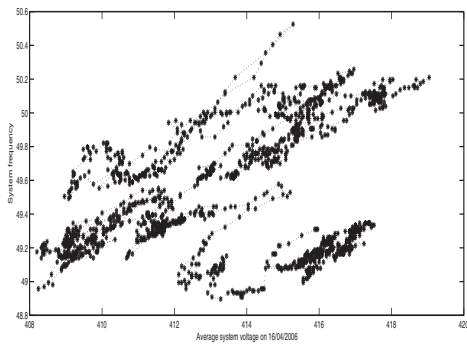


Fig. 2. System frequency vs. Average system voltage on 16/04/2006

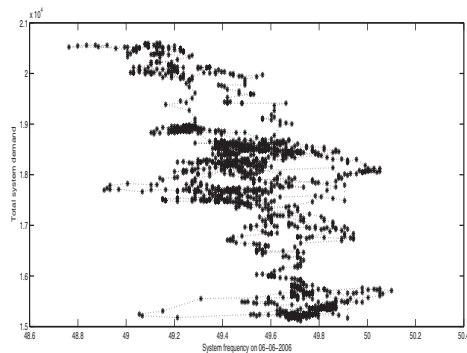


Fig. 3. Total system demand vs. System frequency on 06/06/2006

some data points are also indicated in the figure. It can be seen that from 02:00 a.m., the average system voltage gradually decreases as the system demand increases with time. After 6:20 a.m the plot shows random changes. But, after 7:00 p.m till the end of the day the average system voltage increases with decrease in total system demand. In Figure 6 the same plot of Figure 5 is shown but without the random portion of the graph. It can be seen that the plot takes two distinct paths at the start and end of the day. For the same system demand average system voltage is lower for the morning portion of the day and higher during the night. This corresponds to a 'Hysteresis' type of variation.

In Figure 7 total system demand vs. total MVar generation is presented for the same day. It is interesting to see that this plot also shows a 'Hysteresis' loop when the random portions are excluded. The Figure 7 is similar to the Figure 5 with the only difference being that in Figure 7, with time the plot moves anti-clockwise where as in Figure 5 it moves clockwise. In Figure 8 the same Figure 7 is shown but with the random variation region excluded.

In Figure 9 total MVar generation vs. average system voltage is plotted. From the Figure it can be seen that almost linear relationship exists between total MVar generation and average system voltage. As average system voltage increases MVar injection drops and generators absorb MVar at high average

system voltages. In Figure 10 average generator power factor vs. total system demand is plotted for 19/1/2006. It can be seen that as system demand increases in the morning the power factor starts improving. After 6 a.m the power factor enters a random fluctuation zone but remains high in value. At night, as system demand starts decreasing the power factor starts to decrease.

In Figure 11 total system demand vs. average system voltage

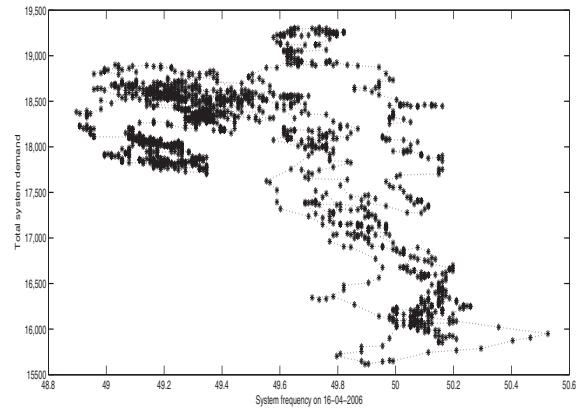


Fig. 4. Total system demand vs. System frequency on 16/04/2006

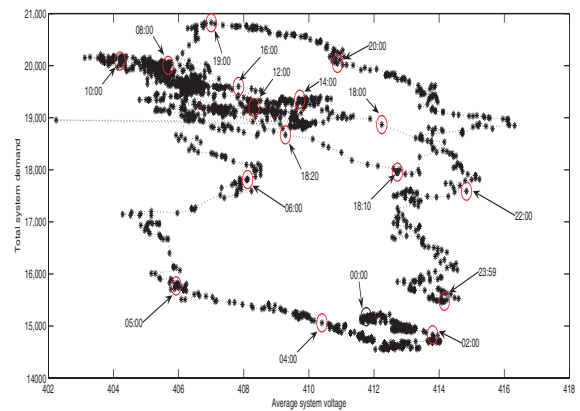


Fig. 5. Total system demand vs. Average system voltage on 19/01/2006

is plotted for 6/6/2006. In this case also random movement is seen during the middle of the day while the remaining part exhibits hysteresis type of behavior. We have seen this hysteresis across several days taken from several months. More careful investigations are necessary to identify specific performance patterns. What is evident is that there is some interesting behavior seen in these plots. Further study could help to understand this in a better way.

In Figure 12 total system MVar generation vs. average system voltage is presented for 6/6/2006. The relationship between MVar and average system voltage is almost linear (in an average sense). As average system voltage increases large

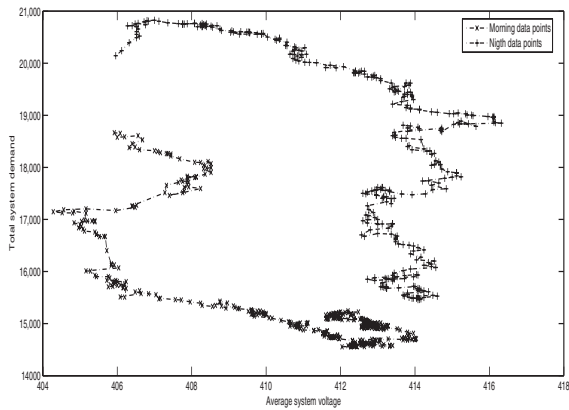


Fig. 6. Total system demand vs. Average system voltage plot (after removing random portion) of on 19/01/2006

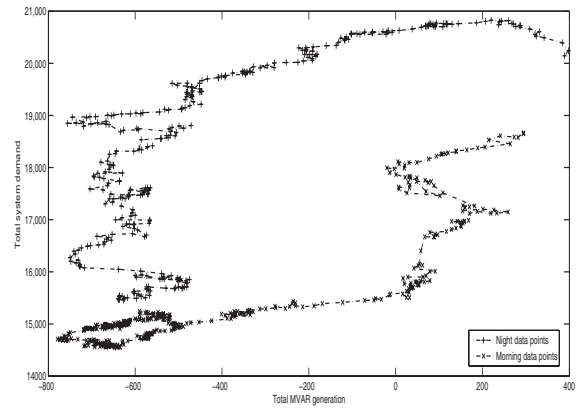


Fig. 8. Total system demand vs. total MVAR generation plot (after removing random portion) on 19/01/2006

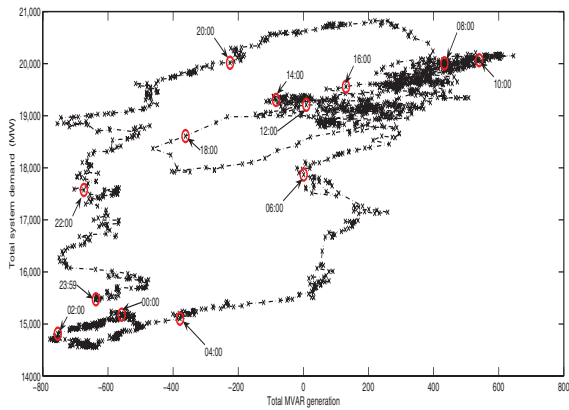


Fig. 7. Total system demand vs. total MVAR generation plot on 19/01/2006

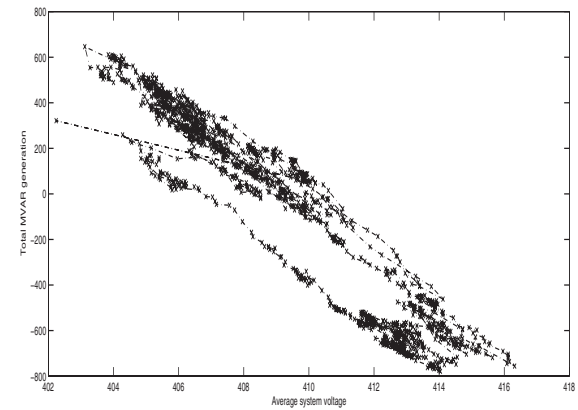


Fig. 9. Total MVAR generation vs. average system voltage plot on 19/01/2006

MVAR is consumed by generators and as average system voltage becomes low generators inject large MVAR into the grid. Near linear relationship between the parameters is also evident for other days also.

Till now we have discussed the inter-parameter relationship considering the whole system. Parameter relationships are also investigated for individual generation buses. In Figure 13 scatter plot of MVAR and voltage at different generator buses are presented for the day 6/6/2006. It can be seen that the MVAR vs. voltage scatter plot at RGM generator bus is different from other plots. For each voltage at the bus two distinct MVAR values are seen. Scatter plots (for all the 8 generating stations) indicate a near linear relationship between bus voltage and MVAR output of the units. Figure 14 shows the plot of average of MVAR generation at each voltage value vs. bus voltage value of generating stations (corresponding to Figure 13). The slope of the MVAR-voltage plot is different for different generating stations in Figure 14. The nominal voltage at VTS bus is 220 KV. As the voltage is increasing the MVAR injection at VTS bus is also increasing. On the other

hand MAP and MTPS are injecting less MVAR with increasing voltage. KAI, SHVT and SIM are absorbing large MVAR at the higher voltage levels while injecting a small amount at lower voltages. RTPS is absorbing smaller MVAR at higher voltage as compared to the absorption at lower voltages. In Figure 15 plot of voltage vs. MVAR (averaged for each voltage) is shown for the day 13/3/2006. In this case RGM is absorbing large MVAR at the higher side of the voltage range while injecting small MVAR at lower side of the voltage range. The nature of variations at other generating stations except RTPS in the figure remain same as on 6/6/2006. In Figure 15 RTPS is injecting a small MVAR at high voltage while injecting large MVAR at low voltage. In Figure 16 Voltage vs. VAR injection (averaged for each voltage) is shown for the whole month of March 2006. In this case also a near linear relationship can be seen between VAR and voltage. For SHVT and MAP the plots are not linear at the extremities. Actually the number of sample points are very small in this range. In Figure 17 total system demand (averaged for each frequency) vs. system frequency is plotted for the month of December 2005. Except at higher

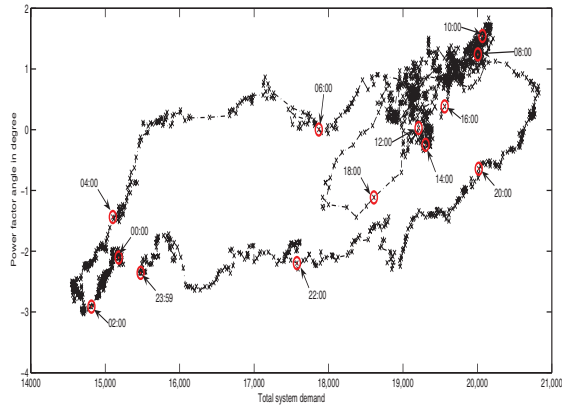


Fig. 10. Total system demand vs. power factor plot on 19/01/2006

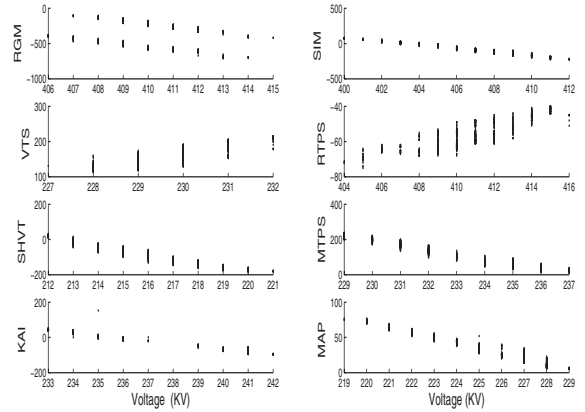


Fig. 13. Scatter plot of Voltage vs. MVAR injection at different bus on 06/06/2006

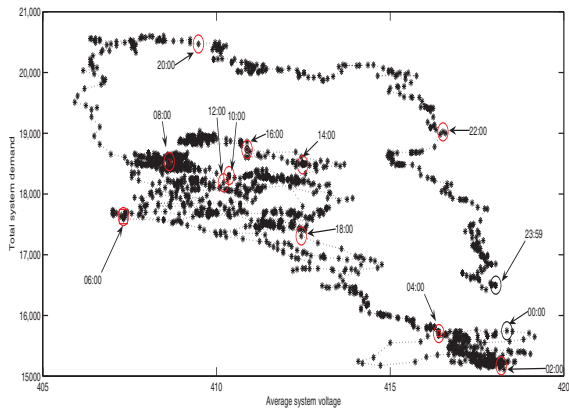


Fig. 11. Total system demand vs. Average system voltage on 06/06/2006

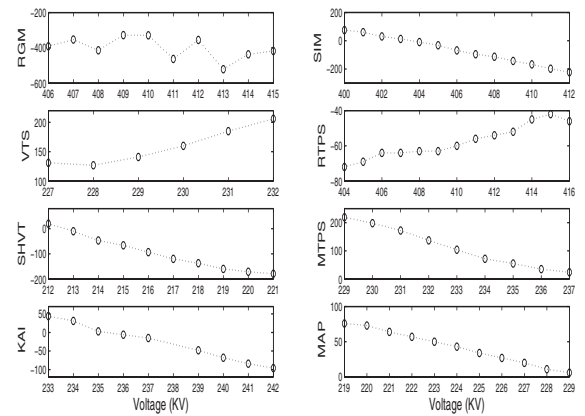


Fig. 14. Plot of Voltage vs. VAR injection (Averaged for each voltage) at different bus on 06/06/2006

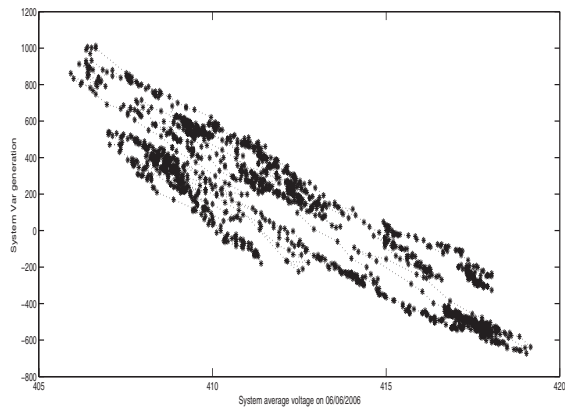


Fig. 12. Total system MVAR generation vs. Average system voltage on 06/06/2006

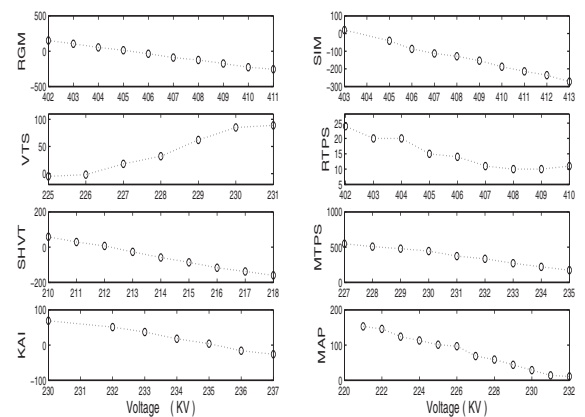


Fig. 15. Plot of Voltage vs. VAR injection (Averaged for each voltage) at different bus on 13/03/2006

frequencies the graph appears to be nearly a quadratic. As system demand increases system frequency decreases and vice versa. At the higher end of frequencies the graph is irregular

due to insufficient number of sample points. In Figure 18 the same plot is shown for the month of April 2006. Here also



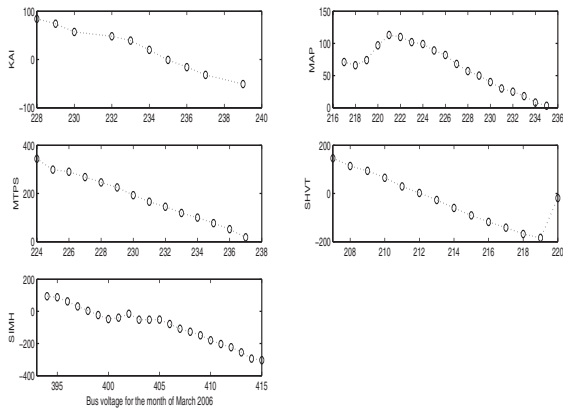


Fig. 16. Plot of Voltage vs. VAr injection (Averaged for each voltage) for the month of March 2006

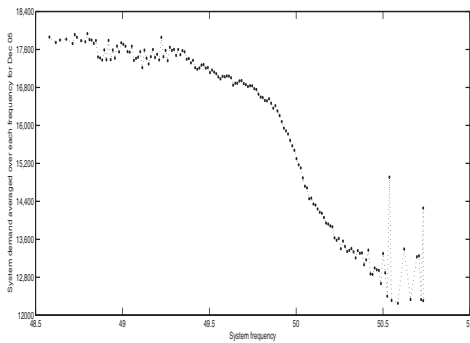


Fig. 17. Plot of System demand averaged at each frequency vs. system frequency for the month of December 2005

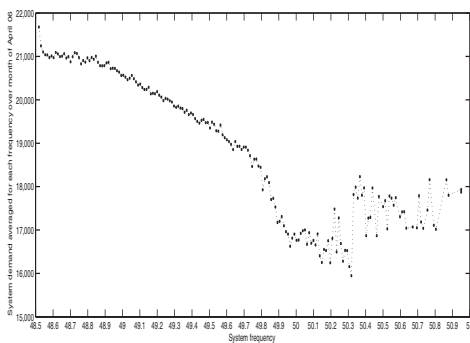


Fig. 18. Plot of System demand averaged at each frequency vs. system frequency for the month of April 2006

the graph tends to show a similar nature except at the higher frequency range.

V. CONCLUSION

The simple analysis attempted in the work brings out several interesting characteristics of the overall system behavior that are not readily available anywhere for the southern regional grid

of India. By performing similar analysis on other systems the similarities/differences in their behavior can be found. It must also be pointed out that the primary aim of the present work is to emphasize that new relations/insights can be obtained by analyzing operational data. The results presented are incidental. The present investigation had some constraints beyond our control. For example, all the parameters contained outliers. Erroneous outliers are to be identified and eliminated for the sake of meaningful analysis at the time of archiving. It is difficult to do it later. We have identified some of the erroneous values and substituted them with reasonable values. With more system information it would have been possible that more outliers are identified. Much work is needed to find suitable ways to identify and replace erroneous values at the time of recording/archiving. Application of statistical methods for outlier identification and replacement can be an interesting issue for further research.

To get the real benefit from data analysis, utilities have to focus on eliminating the limitations of the present data storage practices. The limitations observed are :

- 1) The data set has many errors
- 2) The data is not complete
- 3) the data set acquisition/storage was not motivated by data mining consideration

The aim of this investigation is to argue that if some of these limitations are overcome (it can be in fact done fairly easily), then mining such data could be extremely profitable from the point of view of efficient planning and operation of power systems.

ACKNOWLEDGMENT

The authors would like to thank Southern Regional Load Despatch Center, Bangalore for providing system data for this work.

REFERENCES

- [1] M.Werner and U.Hermansson, "Integrated Utility Data Warehousing - A Prerequisite To Keep Up With Competition On Electricity Markets", Power Systems Management and Control.2002.Fifth International Conference on(Conf.Publ.No.488), 17-19 April 2002, pp 130-135
- [2] D Shi, Y Lee, X Duan, Q.H.Wu, "Power systems data warehouse", IEEE Computer Applications in Power, July 2001, Vol. 14, No. 3, pp 49-55
- [3] Lin Xu, "Data modelling and processing in deregulated power systems", Ph.D Theses, Washington State University, May 2005
- [4] Xiaofeng He, Gang Wang and Jiancang Zhao, "Research on the SCADA /EMS System Data Warehouse Technology", IEEE/PES Transmission and Distribution Conference and Exhibition: Asia and Pacific, Dalian, China, 2005, pp 1-6
- [5] E.H.Abed, N.S.Namachchivaya, T.J.Overbye, M.A.Pai, P.W.Sauer and A.Sussman, "Data-Driven power systems Operations", Lecture Notes In Computer Science, 2006, NUMB 3993, pp 448-455
- [6] <http://www.srlc.org/Downloads/Srldc>