Time Series Data Analysis Technology Based on the Chaos Theory

Takanori Hayashi,
Tomoyuki Yomogida

Keywords Data analysis, Fault detection, Trend analysis

Abstract

We developed technology that analyze waveforms of time series data obtained from measurements in plant facilities and also detect waveforms online that are different from ordinary ones. With these technologies, we can numerically grasp changes in trends from past data in order to use the obtained data for fault detection.

For the verification of these technologies, we investigated power demand in order to numerically detect changes in trends between weekdays and Saturdays, Sundays, and holidays. Simultaneously, we reconfirmed that we must achieve the theorization of respective data in order to grasp the contents of our investigation.

By analyzing a variety of time series data obtained from measurements in actual facilities, we aim to commercialize it so that we can utilize our established technologies for abnormality detection and fault detection technology in the future.

1 Preface

Many facilities such as electrical facilities to support social infrastructures are required to assure high reliability while operating economically. Plant facilities and equipment in industrial fields also are required to perform adequate facility management to secure the continuous assured sustained reliability. In particular, prompt detection of an abnormality or failure in facilities and fast trouble-shooting are indispensable not only for reducing costs, but also for securing safety and assured confidence of users.

In these facilities, many kinds of information are being recorded in monitoring and control systems in the form of time series data obtained through measurements. Parts of these data are being utilized for typical abnormality detection. Due to the present progress of computer performance a large volume of time series data has become analyzable. It is therefore, expected that such a background allows for detecting potential abnormality, the predicting of occurrences of failure through data analysis in a new approach, and the discovery of trends concealed in time series data.

In order to accomplish efficient detection and prediction of any abnormality in equipment, we have been developing basic technologies based on the chaos theory[(1)(2)]. Recently by modifying a conventional Trajectory Parallel Measure Method (TPMM) into a system applicable to on-line detection, we have developed a Successive Trajectory Parallel Measure Method[(3)] (STPMM) with higher serviceability.

We applied the STPMM to trend analysis of power demand and confirmed its detection performance[(4)].

This paper introduces STPMM and the result of trend analysis for power demand which apply the STPMM.

2 Abnormality Detection Approach

Formerly, we developed the TPMM as an abnormality detection method based on the chaos theory. This time, we have developed the improved STPMM for online detection.

2.1 TPMM

The TPMM used for abnormality detection approach based on the chaos theory. This approach evaluates whether a locus (attractor) created by embedded time series data in a higher dimensional space has deterministic regularity, or whether it has a stochastic characteristic. For this evaluation, we use a dispersion in the tangent vector of the locus, in which the locus passes through the neighborhood in the attractor.

The Trajectory Parallel Measure (TPM) calculated as an evaluation value stays in the range between 0 and 1. The more the direction of the tangent vectors is aligned, the more the evaluation value comes closer to 0, which indicates that the deterministic property is strong. On the contrary, the more the direction of the tangent vectors is dispersed, the more the evaluation value comes closer to 0.5, where the stochastic characteristic becomes strong. If this value exceeds 0.5 and comes closer to 1, such data are considered abnormal. In the case of detection of abnormality based on the TPMM, it is assumed that normal data have a deterministic regularity and abnormal data are considered to have a stochastic characteristic. In this method, it is possible to perform abnormality detection based on the TPM value.

2.2 STPMM

The TPMM is used to evaluate the objective time series data with an average value of TPM at each evaluation point by selecting multiple evaluation points stochastically from the attractor where time series data are embedded. This approach is not always sufficient for abnormality detection in online systems where the result is requested each time data are collected successively. Fig. 1 shows an image of the STPMM. By this approach, successively changing data can be analyzed because the evaluation point of the former TPMM has been changed into the forefront embedding point. In this method, a forefront point is chosen as an evaluation value from the attractor of time series data and the tangent vector is calculated in a neighborhood space.

Such a method is called the STPMM. This is because TPM calculation is performed by successively updating the evaluation point and tangent vectors at multiple points in its neighborhood space changing with time.

3 Trend Analysis for Power Demand

The TPM value comes closer to 0 when the attractor locus is more coherent, and it becomes greater when the locus is dispersed. Consequently, this value changes in the vicinity of 0 if the periodicity of time series data is high. In this connection, it is presumed that this value becomes greater if there is a periodic turbulence due to a factor like any malfunction.

In power demand, there is generally a daily periodicity due to the activities of people, however, trends can change on weekdays, Saturdays, Sundays, and holidays. We have analyzed how the TPM is changed by such a difference in trends.

3.1 Objective Analytical Data

Fig. 2 shows power demand data (2008～2011) obtained from the service area of Tokyo Electric Power Company, Incorporated (TEPCO). This analysis was carried out based on 4 years of power demand data collected at the intervals of one hour in the service area of TEPCO as disclosed to the public by this power company. The data suggest that there are seasonal differences among seasons in maximum power demand. It is, however, difficult to grasp any daily trends from this graph.

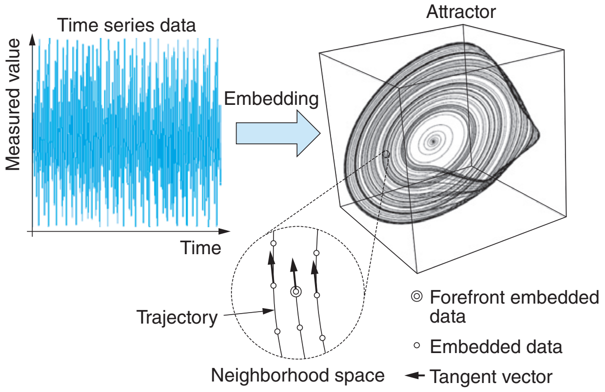

Fig. 1 STPMM Image

Time series data are successively embedded to establish an attractor so that dispersion of tangent vectors can be examined in the neighborhood space of forefront data.

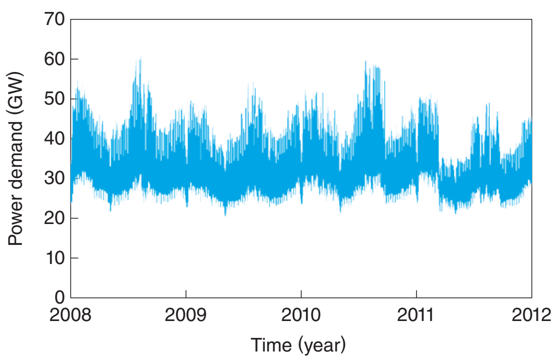

Fig. 2 Power Demand Data in the Service Area of TEPCO (2008～2011)

This graph was created from the power demand data for each hour of the service area of TEPCO that is published on the TEPCO Web site.

3.2 Conditions for Calculation

The STPMM involves three types of parameters; the embedded dimension, the lag time at the making attractor, and the number of neighborhood spaces browsed during TPM calculation. In this case, for the purpose of examining the result of preliminary evaluation and daily periodicity, we decided to adopt the embedded dimension 24, the lag time 1, and the number of neighborhood 10. During the analysis, the TPM value at 11PM when daily variation is mostly reflected is used as a typical value for daily trends.

3.3 Analysis

Fig. 3 shows a graph of changes in the TPM for 4 years. At the beginning of the period, the value appears in a large value. This is because neighborhood is selected in terms of the number of items, and even items in a far place tend to be chosen, thus causing dispersion. After that, values are mostly kept below 0.1 and variation is observed within a period of approximately one week. Values seem to lower year after year and this is presumed to result from an increase in similar patterns that can attain the nearby points. A sharp increase in value in 2011 was caused as the result of the Great East Japan Earthquake.

The description below is for the analysis done based mainly on data obtained in 2010. Fig. 4 shows changes in the TPM extracted from one-month data obtained in June 2010. The diagram suggests that the TPM value tends to increase on Saturday and Sunday. Fig. 5 shows the mean values of the TPM recorded daily. Throughout one year in 2010, values were high on Saturdays and Sundays. This is considered to be due to the differences in trends from those of the weekdays, when similar trends of changes were reflected every day on the TPM value. Among weekdays, values on Monday tend to be high for reasons that trends of holidays remained until daybreak.

In order to examine the influence of holidays, we established a graph as shown in Fig. 6 based on

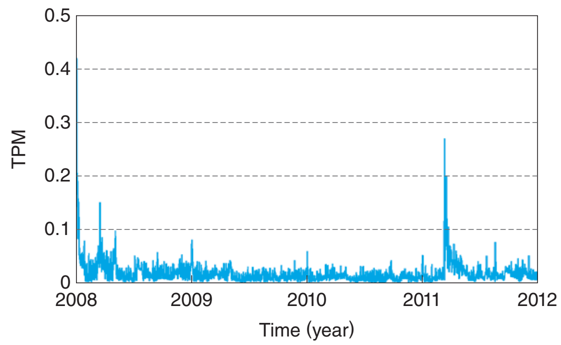

Fig. 3 Changes in TPM for 4 Years (Fiscal 2008～2011)

The graph shown is based on the extracted values from data in Fig. 2 obtained at 11PM every day through TPM calculation by STPMM.

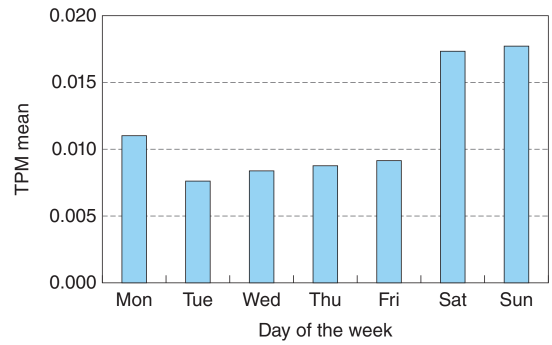

Fig. 5 TPM Mean Values on Each Day of the Week

The graph shows the TPM values in 2010 averaged on each day of the week. Values on Saturdays and Sundays were noticeably high. The value on Monday was high despite Monday being a weekday.

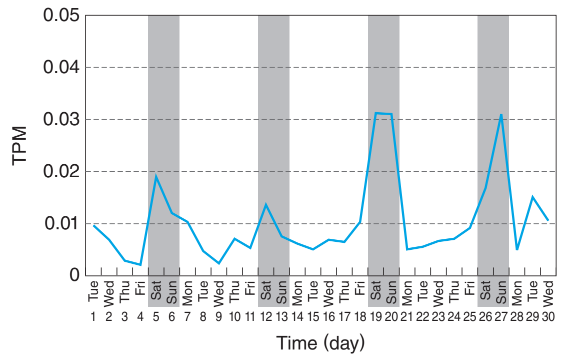

Fig. 4 Changes in TPM for 1 Month (June 2010)

Data for June 2010 were extracted from the graph in Fig. 3 and Saturdays and Sundays are shaded. TPM values appear high on Saturdays and Sundays.

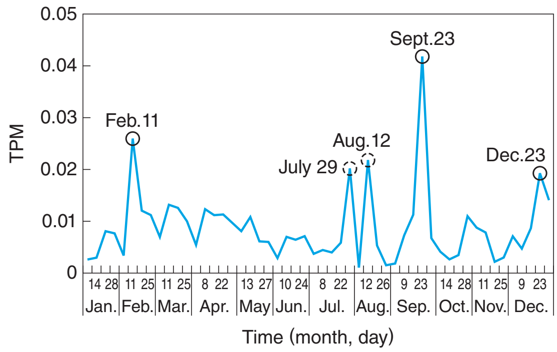

Fig. 6 Changes in TPM on a Weekday (Thursday)

The graph shows the TPM values extracted only for Thursdays in 2010. Values in holidays are noticeably high. There are also days other than holidays when the values are high.

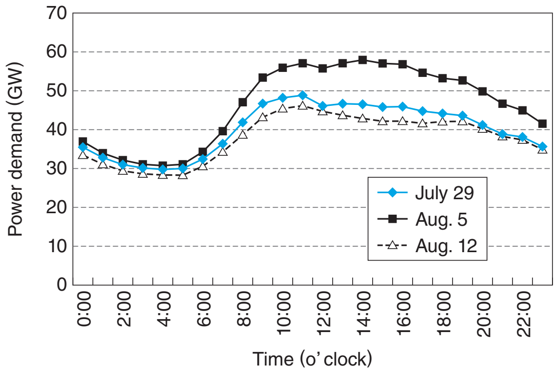

Fig. 7 **Power Demand in Weekdays when TPM is High**

The graph shows one-day power demand where the TPM values were high on July 29 and Aug. 12 irrespective of weekdays as shown in **Fig. 6** and also the TPM value on Aug. 8 was lower than the foregoing two days. This is considered to be due to the TPM being influenced by the difference in demand between daytime and nighttime.

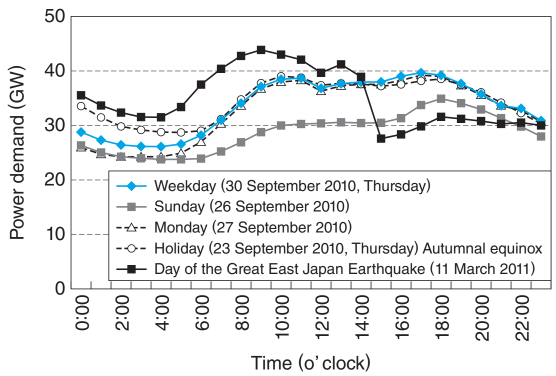

Fig. 8 **Power Demand in Weekday, Sunday, Monday, Holiday, etc.**

The graph shows power demand on a weekday, Sunday, Monday, and holiday in September 2010, and on the day of the Great East Japan Earthquake in 2011. The difference in trends can be identified.

an idea where TPM values were extracted only the weekday, Thursday. It is possible to confirm that the TPM values were high on February 11, September 23, and December 23, which were all holidays, and high values were indicated even on holidays compared with weekdays. Ironically, TPM values were also high on both July 29 and August 12, which were not holidays, though reasons are not yet clearly explained. **Fig. 7** shows power demand on weekdays when TPM is high. Compared with August 5 when TPM was low, we recognize that power demand in the daytime is greatly lowered. In order to explain the reason why the demand was lower, we have to use other information like the weather.

Fig. 8 shows several days of power demand on a weekday, Sunday, Monday, etc. Compared with weekdays, power demand on Sunday was generally low. Monday was close to Sunday until daybreak. After daybreak, the situation closer to resembled a weekday. On the autumnal equinox, which was a holiday, power demand was higher than that of any weekday until daybreak. After that, trends were similar to weekdays. On the day of the Great East Japan Earthquake, power demand tended to change at a higher level irrespective of it being a different season. In fact, the demand level was lower after the occurrence of this great disaster before 3PM.

4 Postscript

This paper introduced the features of STPMM that we developed. It also provides some examples of power demand analysis. It indicates that the difference in trends of power demand is obviously reflected on TPM values and that trends at a lower frequency were detected in a form of abnormality. High TPM values were presented due to many reasons. As such, at the time of application to abnormality detection, we must apply adequate theoretical reasoning based on real data.

Going forward, we will continue to analyze a variety of time series data obtained from measurements in actual facilities and will aim to complete our abnormality and fault detection technologies to commercialize them as applicable technologies.

《**References**》

(1) Iokibe, Sugaya, Fujimoto, Suzuki: "Local Fuzzy Reconstruction Method for Short-Term Prediction of Chaotic Time Series," Journal of Japan Society for Fuzzy Theory and Intelligent Informatics, Vol.7, No.1, 1995, pp.186-194

(2) Fujimoto, Iokibe, Tanimura: "Trajectory Parallel Measure Method to Measure the Deterministic Property of Observed Time Series Data," Journal of Japan Society for Fuzzy Theory and Intelligent Informatics, Vol.9, No.4, 1997, pp.580-588

(3) Yomogida, Hayashi: "Successive Trajectory Parallel Measure Method for Abnormality Detection from Time Series Data," Transactions of National Lecture Meeting for The Institute of Electrical Engineers of Japan, Vol.1, 2012, pp.146-147

(4) Yomogida, Hayashi: "Power Demand Analysis by the Successive Trajectory Parallel Measure Method," Transactions of National Lecture Meeting for The Institute of Electrical Engineers of Japan, Vol.1, 2013, pp.121-122