

Fault Detection of Drinking Water Treatment Process Using PCA and Hotelling's T^2 Chart

Joval P George, Dr. Zheng Chen, Philip Shaw

Abstract— This paper deals with the application of Principal Component Analysis (PCA) and the Hotelling's T^2 Chart, using data collected from a drinking water treatment process. PCA is applied primarily for the dimensional reduction of the collected data. The Hotelling's T^2 control chart was used for the fault detection of the process. The data was taken from a United Utilities Multistage Water Treatment Works downloaded from an Integrated Program Management (IPM) dashboard system. The analysis of the results show that Multivariate Statistical Process Control (MSPC) techniques such as PCA, and control charts such as Hotelling's T^2 , can be effectively applied for the early fault detection of continuous multivariable processes such as Drinking Water Treatment. The software package SIMCA-P was used to develop the MSPC models and Hotelling's T^2 Chart from the collected data.

Keywords— Principal Component Analysis, Hotelling's T^2 Chart, Multivariate Statistical Process Control, Drinking Water Treatment.

I. INTRODUCTION

It is well understood that Industrial Process Control applications are statistically oriented, often requiring the measurement and storage of large amounts of information. However, the analysis of such data becomes particularly difficult when large numbers of process variables are recorded, typically leading to a data rich, information poor situation.

In many cases, the collected data is highly correlated and is therefore difficult to analyse. The financial aspect of effective process control or data handling is a key issue in process industries. To overcome these problems methods of statistical process control are used to analyse the data obtained from the process. With the help of statistical process control techniques highly correlated data can be analysed and the resultant information can be plotted in a clear understandable manner.

Joval P George is with the ABB AG Power Systems Technology Division, P.O.Box 46249, Abu Dhabi, UAE. (Phone: 00971-50-8547820; fax: 00971-26318907; e-mail: jovalpg@yahoo.com).

Dr. Zheng Chen is with University of Glyndwr (formerly known as North East Wales Institute of Higher Education), Wrexham, United Kingdom, LL11 2AW, email: z.chen@glyndwr.ac.uk).

Philip Shaw is with the O&M Strategy control of United Utilities Northwest, Ellesmere Port WTW, CH2 4HZ, England, UK. (e-mail: Philip.shaw@uuplc.co.uk).

The multistage water treatment process under investigation was classified into five stages based on the parameters taken from at each stage. Usually Data taken from Water Treatment Processes are highly correlated making it difficult to discriminate between parameters and understand the relationships between them. This particular site was faced with issues such as anonymous chemical reactions occurring within the process, which if not addressed, could reduce operation efficiency and in the worst case may result in a compliance failure.

Hence the main requirement of this work was to design a suitable Multivariate Analysis Technique to analyse the set of bulk data which was collected from the processes. The objective was to explore the possibilities of applying Multivariate Analysis techniques to describe the overall state of the process. Using Principal Components Analysis (PCA), a data reduction technique, would allow the condition of a complete process to be accurately monitored, greatly improving upon the traditional signal alarming methods based on individual sensor thresholds currently employed on the site.

II. MOTIVATION

Water is a basic life sustaining requirement in the day-to-day life of a human being. Hence the quality of the water which is consumed in daily life should be ensured at the utmost level. Numerous amounts of parameters are considered in the monitoring and control of the water treatment process. As number of the parameters to be considered is more, the complexity of the process also becomes more.

Lack of proper control and monitoring of the water treatment process not only causes the financial losses but also can cause loss even to life. The latest developments in technology of process control and data acquisition techniques have resulted in a quick raise in the quantity of data coming from the process. The data handling of similar processes have become a key issue nowadays. The collected data in similar processes are highly correlated and difficult to analyze.

The paper published by De Veaux, R *et al* [1] explains the use of Statistical approaches to Fault analysis in Multivariate process control. Also the work published by Lennox, B [3], details the use of Multivariate statistical technology to improve the control system capabilities. These works clearly details the effective use of multivariate statistical techniques in the analysis of the process.

The paper published by Praus, P [6] explains the use of principal component analysis on hydrological data. According to him the hydrochemical data sets contains not only important information for quality assessment and/or treatment technology but also confusing noise. He also mentions that the measured variables are not normally distributed rather they are co-linear or auto correlated containing erroneous or nonsense values. He says that the PCA can be used to search the new conceptual orthogonal eigen vectors which explain most of the data variation in new co-ordinate system. According to him each principal component is a linear combination of the original variables and describes a different source of information. He demonstrates in the paper that PCA easily provides an unbiased view of water composition and thus can be used as a very useful tool for water quality assessment.

Hence from the above mentioned literature survey it is made clear that the multivariate statistical techniques can be effectively used in the analysis of continuous process like water treatment. The company chosen for the case study was United Utilities limited, which is one of the leading suppliers of clean water to around seven million people in the north west of England.

III. PRINCIPAL COMPONENT ANALYSIS (PCA)

A. Literature Review

Work previously published by Praus, P on water quality assessment using SVD-based principal component analysis of hydrological data, shows how PCA is able to extract latent variables/singular vectors from the noisy hydrological data. He further showed that these Singular vectors can be presented in form of the scatter and loading plots which allow mapping of water quality in different localities of the supply system. Additional work published by Ruiz, M et al illustrates how effectively MSPC techniques can be used for detecting and diagnosing events that cause a significant change in the dynamic correlation structure among the process variables. His work also establishes that dividing data into meaningful groups enables to better identify the batches with abnormal behavior and getting characterization of the types of batch process; normal behavior, atmospheric changes, etc. The work published by Klancar, G, states that the main advantage of PCA approach is that only past data records are needed and no process model is required. It also provides relatively simple realization and possibility of adding new process operation modes (faults).

McGregor et al [2] explains that monitoring of continuous process using MSPC techniques, requires a PCA model built based on data collected from various periods of plant operation when performance was good. Any periods containing variations arising from special events that one would like to detect in the future are omitted at this stage. The choice of reference set is critical to the successful application. The paper published by De Veaux *et al* [1] suggests that PCA is useful when there is sufficient correlation between the sensor readings available, but sensors

may show mostly uncorrelated noise if the process is in control. When information about these faults is available then latent variables can be developed that include the faults and to classify out of control situations to these faults, based on their distance to the mean fault in the group.

Lennox *et al* [4] recommends that PCA is a technique that can be used to identify a new set of variables that reflect the underlying characteristics that actually drive the process. The PCA model can be developed from the process data that is free from faults and this model then acts as a template which compares the operating data. The deviations from the template indicate the presence of abnormal conditions which could be possibly from the process faults.

IV. WATER TREATMENT PROCESS-CASE STUDY

The multistage water treatment process of United Utilities takes place in five different stages. At the first stage, the raw water which is being stored in the reservoir is pumped to the coagulation and flocculation chambers. During second stage, to remove the impurities from raw water, coagulants such as salts of iron or aluminium are added to form solid precipitates known as floc, which in turn adsorbs colloidal impurities. The floc thus formed is then separated using the process called the flocculation. The chemicals added at the multistage water treatment process are Lime, Aluminium Sulphate & PolyDADMAC. The coagulated water is then passed through 3x Microfloculators. A polyelectrolyte is added after the coagulation stage to minimize the coagulant loading to subsequent processes.

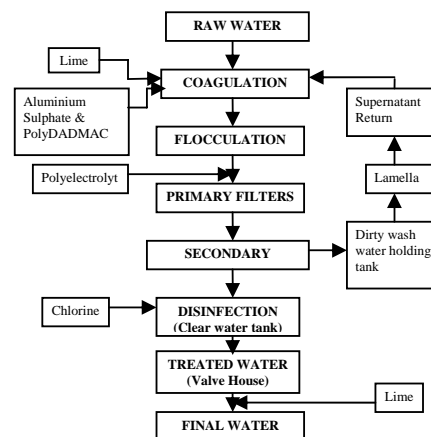


Figure 2: Water treatment process _Schematic

At the next stage Rapid gravity filters (Primary filters) and slow sand filter(Secondary filters) are employed to filter out the impurities from the later stage. At the fourth stage the filtered out water is subjected to disinfection. It is done in water with the help of addition of certain chemicals. In the discussed process site it is achieved by addition of Chlorine.

The disinfected water is then stored at the valve house tank. The valve house chlorine is measured at this stage. Lime dosing is undertaken before it is being taken for the final supply. At the final stage, the treated water is pumped to customers via three supply lines. A summary of the parameters measured at all the water treatment process stages is shown in the Appendix I.

Process data analysis presented major issues due to the extremely large amount of bulk information taken from the systems. Because water treatment processes are continuous and multivariate, methods employing univariate chart techniques are inefficient. The following sections discuss the development of the PCA model and control chart with the help of SIMCA-P package.

V. DEVELOPMENT OF PCA MODEL

The PCA model was developed from the data collected from the IPM systems installed at the Multistage United Utilities Water Treatment Works. The model consisted of 10 X variables and 13 Y variables.

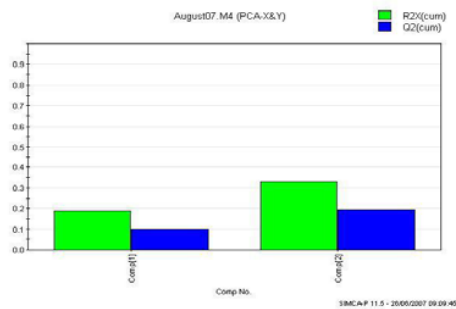


Figure 1: Goodness of Fit plot of the PCA Model

The plot displays the cumulative R2 and Q2 for the X matrix, after each component. R2X is the percent of the variation of all the X explained by the model and Q2 (Cum) is the percent of the variation of all the X that can be predicted by the model. With the component 2 of 32.9% cumulative variation it shows a good fitness of the model. The significance is shown in the list as R1.

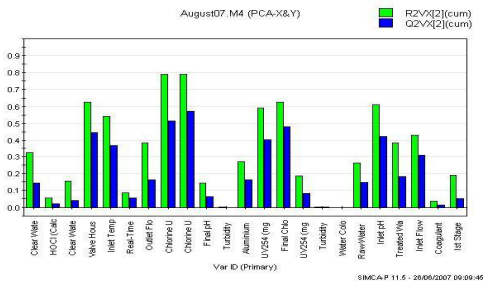


Figure 2: Goodness of Fit plot of the X and Y

This plot displays the cumulative R2VX and Q2VX for each X&Y variable. The model describes good fit for the data related to parameters like chlorine and pH. Where as the data related to parameters like turbidity and water color is showing very less fit to the model.

A. Model Results

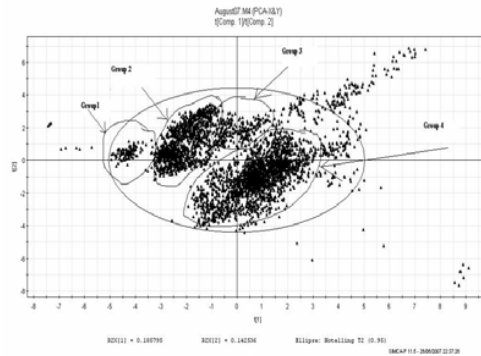


Figure 3: scores plot groups-PCA

In the scores plot mainly four groups are visible inside the ellipse. The contribution plot of the group1 shows that these are set of observation with low values of valve house chlorine, final chlorine, final dissolved organics and high values of inlet temperature, chlorine unit cost, Chlorine usage, and aluminium and inlet flow. The contribution plot of group2 shows these are groups of observations with low values of valve house chlorine, final chlorine, inlet pH and High values of Chlorine Usage and Chlorine Unit cost, inlet flow. Group3 shows that these observations are those with low values of clear water chlorine, clear water pH, Inlet temperature, inlet pH and high values of chlorine usage, chlorine unit cost and final dissolved organics. Group4 are set of observations with low values of chlorine usage, chlorine unit cost and also high values of inlet pH.

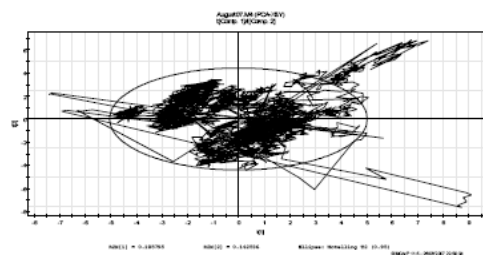


Figure 4: Analysis of outliers-Scores line plot

The scores line plot shows that there are few out of control situations happened in the process, which are not well explained by the PCA model in the scores plot.

VI. FAULT DETECTION OF THE PROCESS USING PCA

The analysis in the previous section has given a clear understanding of the outliers in the model. To have a further understanding of the process disturbances and common faults a study of Hotelling's plots has also been made.

1) Hotelling's T^2 Range Plot

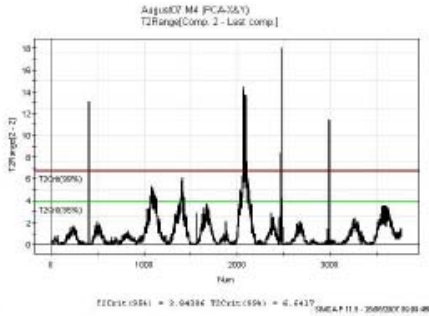


Figure 5: Hotelling's Chart-PCA

This plot summarizes the scores from the third to the last component. Hotelling T^2 is a combination of all the scores (t) in all PCA components. It is a measure of how far away an observation is from the centre of the PC model. In the above plot there are some points which cross the upper limit of Hotelling's T^2 range, which reflects there have been some disturbances in the process. There were four such disturbances found.

a) First Process Disturbance

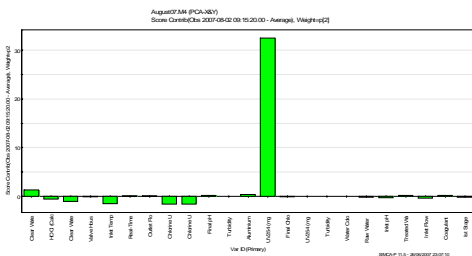


Figure 6: First process disturbance contribution plot-PCA

The plot shows that the first disturbance in the process was due to the increase in final dissolved organics.

b) Second Process disturbance

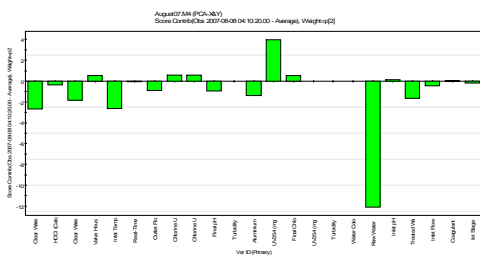


Figure 7: Second process disturbance contribution plot-PCA

The plot shows that the second disturbance in process was due to the decrease in the clear water Chlorine, clear water pH, Inlet temperature, Raw Water Inlet temperature and increase in the value of final dissolved organics.

c) Third Process Disturbance

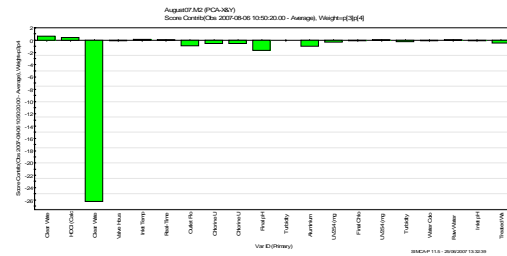


Figure 8: Third process disturbance contribution plot-PCA

The plot shows that the third disturbance in the process was due to the decrease in clear water tank chlorine.

d) Fourth Process Disturbance

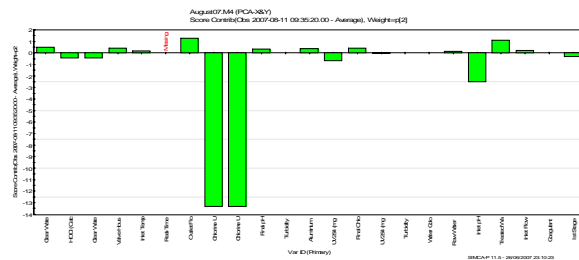


Figure 9: Fourth process disturbance contribution plot

The plot shows that the fourth process disturbance occurred due to the decrease in inlet pH, Chlorine usage, Chlorine Unit cost, Inlet pH and due to the missing values for real time contact.

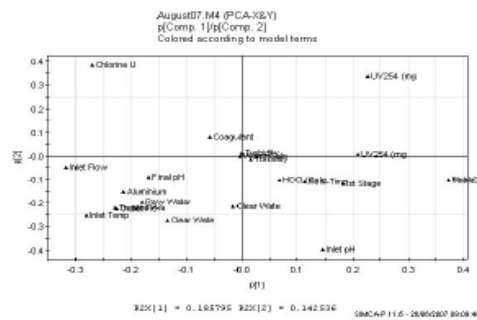


Figure 10: Loadings p[1] vs. p[2]-PCA

The scores are weighted averages of the variables with weights p1 in the first dimension and p2 in the second dimension. These weights, the loadings p's are computed from the correlation structure of the X's. Hence p1 vs. p2 display how the X variables

correlate with each other. The plot shows how the X's vary in relation to each other, which ones provide similar information, which ones are negatively correlated, or not related to each other, and which ones are not well explained by the model. In the loadings plot we can see that variables, final chlorine, final dissolved organics, inlet pH are situated right most of the plot and Chlorine unit cost, inlet temperature, treated water flow rate, outlet water flow rate, aluminum, final pH raw water inlet temperature, inlet flow are on the left of the loadings plot. This suggests that these variables are responsible for the process disturbances.

VII. CONCLUSION

Data taken from drinking water treatment process was found to be highly correlated, which made it difficult to analyze and displaying data in a form useful to operations. This research was primarily based on PCA, a multivariate technique for detection of abnormal conditions within water treatment process. The application of PCA method with the continuous multivariable water treatment process systems, allowed dimension of the data to be reduced. The fitness of the model was verified from the Goodness of fit plot, which showed a 32.9% cumulative variation for the last component of the PCA model. The PCA scores plot enabled patterns in the observations to be classified into four distinct groups representing different stages of water treatment process. The control charts allowed the overall state of system to be determined and detection of any process disturbances. Hotelling's T^2 chart detected four disturbances in observations which crossed the upper control limit of the chart, the parameters which contributed to the disturbances were unidentified by developing the contribution plot.

The common parameters which showed disturbances in the charts were identified and concluded below. The parameters which mainly contributed to the disturbances in the process was found from the Control charts of PCA analysis were,

- a) Clear water tank pH
- b) Valve house chlorine
- c) Inlet temperature
- d) Final chlorine value
- e) Real time contact
- f) Chlorine Usage

The PCA model developed using the available process data taken over a 14 day period at this multistage treatment system clearly demonstrated that greater levels of process monitoring and therefore control, can be achieved by utilizing more effective methods of analyzing bulk process data. Instrumentation and control equipment relating to the described parameters must be calibrated and fully verified to permit accurate results to be made prior to the reduction of process disturbances. It was shown that disturbances at one stage of the process can easily propagate to other stages. For example, this was the case when disturbances in valve house chlorine directly resulted in variations in chlorine usage and final chlorine. Hence the early detection, identification and rectification of process faults can maintain a tighter control of the process, with fewer disturbances.

APPENDIX

I. List of parameters measured in the process.

Inlet and Raw water
UV254 (mg DOC/l mg/l)
Turbidity (NTU)
Water Colour Inlet ($^{\circ}$ Hazen)
Raw Water Inlet Temperature ($^{\circ}$ C)
Inlet pH
Coagulant and Flocculation
1st Stage Filt Coagulant Residual
Coagulant pH
Disinfection
Clear Water Tank pH
HOCl (Calculated) (mg/l)
Clear Water Tank Chlorine (mg/l)
Valve House Chlorine (mg/l)
Inlet Temperature ($^{\circ}$ C)
Real-Time CT (Contact Time) (mg/min/l)
Outlet Flow Rate (ML/d)
Chlorine Unit Usage (kg/ML)
Chlorine Unit Cost (£/ML)
Treated Water
Treated Water Production Flow (ML/d)
Outlet Flow - Line 1 (ML/d)
Outlet Flow - Line 2 (ML/d)
Outlet Flow - Line 3 (ML/d)
Lime Unit Usage
Lime Unit Cost
Final Water
Final pH
Turbidity (NTU)-final
Aluminium (Ug/l)
UV254 (mg DOC/l mg/l)
Final Chlorine (mg/l)

ACKNOWLEDGMENT

The First Author thanks Dr. Zheng Chen, Senior Lecturer, Department of Engineering, NEWI, UK and Mr. Philip Shaw, Maintenance Analyst, United Utilities North West for providing all the guidance and support as well as for the required data collection towards the accomplishment of this paper.

REFERENCES

- [1] De Vaux R, Ungar L, Vinson J (2007, June 5), Statistical Approaches to Fault Analysis in Multivariate Process Control [Online]. Available: <http://citeseer.ist.psu.edu>
- [2] Klancar G, Fault Detection and Isolation by means of Principal Component Analysis, Faculty of Electrical Engineering, University of Ljubljana, Ljubljana, Slovenia.
- [3] Lennox B, Montague G, Marjanovic O (2007, June 19) *Detection of faults in Batch Processes: Application to an industrial fermentation and a steel making process* [Online]. Available: <http://www.forum-ira.com/pdf/4conf3.pdf>
- [4] McGregor J, Kourti T, (1994), *Process analysis, monitoring and diagnosis, using multivariate projection methods*, Chemometrics and Intelligent Laboratory Systems 28 (1995) 3-21.
- [5] Praus P, *Water quality assessment using SVD-based principal component analysis of hydrological data*, Department of Analytical Chemistry and Material Testing, VSB-Technical University Ostrava, 17. listopadu 15, 708 33 Ostrava, Czech Republic
- [6] Ruiz M, Colomer J and Mel'endez J, *Multiway Principal Component Analysis and Case Base-Reasoning approach to situation assessment in a Waste Water Treatment Plant*, Control Engineering and Intelligent Systems Group –eXiT, Department of Electronics, Computer Science and Automatic Control University of Girona, Campus Montilivi CP 17071 Building PIV, Girona – Spain
- [7] Wise B. (2007 June 30). *Adapting Multivariate Analysis for Monitoring and Modeling of Dynamic Systems*, University of Washington. *Partial Least Squares* [Online]. Available: <http://www.statsoft.com/textbook/stpls.html>, on June 30,2007



Joval P George is an Instrumentation and Control Engineer with the ABB Power Systems Technology Division, Abu Dhabi (UAE). During his career he served as a full time Lecturer of Applied Electronics and Instrumentation Department of Rajagiri School of Engineering and Technology, India. Also he served as a Part-time teaching faculty of the Engineering Department at North East Wales Institute, NEWI, UK. He completed his bachelor's degree in Instrumentation from Cochin University of Science and Technology, India. He had his Master's Degree in Electrical and Electronic Systems and Digital Technologies, from University of Wales, UK.

Dr Zheng Chen is a Senior Lecturer and Programme Leader in Aeronautical Engineering with the Glyndwr University, UK. He has been involved in the academic area of aviation electronic systems (avionics), aerodynamics, and aircraft maintenance for more than 15 years. He is actively involved in research in the area of reliability centred aircraft maintenance, robust control system design, fault detection and isolation, and intelligent process control in utility industries. He obtained his BEng and MEng in Beijing and his PhD in Southampton, UK.

Philip Shaw is a Maintenance Analyst with the O&M Strategy Control of United Utilities Plc, UK. He obtained his B.Eng(hons) and MSc.(Eng) degrees in Electronic Engineering at the University of Liverpool, England. Prior to joining UU I he worked in RF and microwave development. He joined UU in 1996 working in the "Technology Development Team" (a water/wastewater/network R&D department) within the Process Control and Instrumentation (PC&I) team. During this time he was responsible for the design and development of a wide range of novel instrumentation and other equipment for water processes. In 2001, he moved into Policy and Standards responsible to the operation of CHP (combined heat and power systems), before moving into Plant Maintenance where he was primary responsible for setting up the ICA Team. Following this he moved to Asset Management and Regulation where he was involved in the setting up and management of a range of R&D projects, particularly in the area of Process Control. He is now working on the development of diagnostic techniques and advanced control algorithms for water treatment processes.