Miroslaw SWIERCZ[1], Halina MROCZKOWSKA[2]

Bialystok University of Technology (1), ENEA Cieplo Sp. z o.o. (2)

doi:10.15199/48.2019.10.43

Application of PCA for early leak detection in a pipeline system of a steam boiler

***Abstract**. The application of the Principal Component Analysis (PCA) method for early detection of leakages in the pipeline system of a steam boiler in a thermal-electrical power plant is presented and discussed. The PCA model built from historical measurements of 12 selected process variables, mapped to the reduced space of three Principal Components (PC) of the highest magnitude, was used to establish the confidence ellipsoid, i.e. the feasible region in the PC coordinates, occupied by the values of process variables related to the 'healthy' system. Changes of the current location of the process operating point in the PC space created the 'fault trajectory' and were the basis for making a decision of leakage detection.*

Streszczenie.** W artykule przedstawiono zastosowanie metody składowych głównych (PCA) do wczesnego wykrywania wycieków z rurociągów kotła parowego pracującego w elektrociepłowni miejskiej. Model PCA, zbudowany na podstawie pomiarów 12 wybranych zmiennych procesowych, przedstawiony w przestrzeni trzech składowych głównych (PC) o największych modułach, został wykorzystany do określenia tzw. elipsoidy ufności, tj. obszaru w przestrzeni PC, w którym mieszczą się wartości zmiennych odpowiadające poprawnemu działaniu systemu. Zmiany aktualnego punktu pracy kotła tworzyły tzw. trajektorię uszkodzenia w przestrzeni PC i były podstawą do podejmowania decyzji na temat ew. wycieku z rurociągów.* ***(Zastosowanie metody PCA do wczesnego wykrywania wycieków w rurociągach kotła parowego)

Keywords: fault detection, pipeline leaks, steam boiler, PCA method.
Słowa kluczowe: detekcja uszkodzeń, wycieki z rurociągów, kocioł parowy, metoda PCA.

Introduction

Continuous technological progress and increasing complexity of modern manufacturing plants require specialized solutions which can provide sufficiently high level of reliability and guarantee safety of the technical personnel. Despite increasing reliability of all elements of the plant and substantial development of computer-based plant control systems, a variety of faults of technological components, measurement equipment, control devices and staff errors can cause serious material or human losses. That is why the efficient tools for automatic (or at least semiautomatic) on-line fault detection and diagnosis of the production processes are extremely important. A major concern in implementation of a fault detection system is its capability of indentifying incipient faults, while maintaining the false alarm rate to an acceptable minimum.

Fault diagnosis methods can be divided into three main categories [1]:

1) model based, using mathematical formulation of the knowledge about the system (expressed in the form of physical, balance and chemical equations or the black-box or grey-box model) to detect and analyze faults;
2) signal processing based, using spectral analysis, principal component analysis (PCA), wavelet transforms and fast Fourier transforms (FFTs) to analyze system behavior and identify faults;
3) artificial intelligence based, using neural networks, fuzzy systems, expert systems, support vector machines (SVM) or grey correlation to develop diagnostic systems that, once trained, can identify specific faults.

Boilers are important components in power, chemical and oil refinery industries; they transform water into steam for power generation or other industrial applications. A common boiler fault is the tube leakage in the riser and downcomer sections due to aging and thermal stress. Early detection of such faults during device operation is important, as it helps in reducing possible damage to equipment and productivity loss caused by (otherwise) unscheduled boiler shutdown [2]. Early recognition of small leakages in the pipelines can also protect the plant against secondary malfunctions or damages and can provide improved safety level for process operators.

The paper presents the application of a data-driven method of statistical process monitoring to early detection of the leakage in a pipeline system of a steam boiler. As the construction and identification of a mathematical model of the plant under study is almost impossible in industrial practice (due to production plans, high costs of experiments, organizational problems, insufficient human resources, etc.), we decided to adopt and tune the Principal Component Analysis (PCA) method to process data acquired by the existing measurement and control system. In a number of numerical experiments based on our approach (described and discussed in the paper), we confirmed the usefulness of the PCA method for solving a complex engineering problem of fault detection in a thermal-electrical power plant.

Briefly, the paper is organized as follows. Section 2 presents a short review of boiler leak detection methods, currently used in industrial practice or developed and verified on laboratory equipment. Section 3 focuses on theoretical backgrounds of the Principal Component Analysis algorithm. In Section 4 some modifications of the basic PCA methods in the use to fault detection are presented, as well as some successful applications of PCA to boiler fault detection problems. In Section 5 the boiler water system working in Elektrocieplownia Bialystok is briefly presented. The problem statement of leakage detection in the steam boiler pipeline system in Elektrocieplownia Bialystok and analysis of available measurement data are the contents of Section 6. This section also presents the proposed leakage detection algorithm and results obtained on a number of leakages from the period 2010-2015, with detailed discussion of two case studies. Finally, Section 7 offers some concluding remarks and considerations about further work on improvement of fault detection accuracy and sensitivity.

Fault detection in a pipeline system of a steam boiler – current approaches

Detection and localization of leaks in a pipeline system of an industrial steam boiler is a difficult engineering and research problem. Boilers are complex nonlinear systems, which work under time-varying operating conditions (e.g. due to daily and seasonal changes of heat/electricity demands), what causes various changes in their transient responses. Also the number of process disturbances and their various characteristics (e.g. fluctuations in the combustion process), influencing the process variables and, finally, the plant characteristics, are the factors which make mathematical modeling of the boiler considerably difficult. Mathematical models, potentially useful for pipeline leak detection, have usually the form of the black-box models [3], as a small class of models based on physical relationships can be adapted to a certain industrial boiler only in a limited scope.

The tube leak is a special type of the process fault, which belongs to the class of gross error problems [2]. The development of the fault is itself a non-stationary process, which at its early stage may be compensated and 'masked' by the normal control actions, as in the standard control response to process disturbances. Due to small loses of mass and energy per time unit the pipe leaks at their early stages can be hardly detected with the use of simple methods of limitations check or statistical analysis. The above reasons cause that the steady-state based methods do not provide satisfactory performance in the leak detection problems.

The methods of leak detection in pipelines of steam boilers, which are most frequently used in industrial practice, can be roughly divided into the following groups [4]:

1) acoustic monitoring and analysis, which makes use of acoustic waves generated by the escaping steam;
2) steam/water balance testing;
3) monitoring of gas humidity in the flue;
4) other methods, based on monitoring and analysis of process variables and their relationships, sometimes supported by mathematical models of the technological process.

The methods based on acoustic monitoring and analysis are the oldest industrial implementations of the leak detection methods in steam boilers, patented from the mid 70's [5] and systematically developed through the next decades. They require installation of rather expensive devices (sensors) and careful tuning, however they cannot detect small to medium leaks (less than about 10 000 kg/h). The methods make use of high frequency pressure waves (generated by the fluid escaping from a leak in a boiler tube), which may be converted into electronic voltage signals with a variety of sensitive dynamic pressure transducers (sensors) that are in contact with the medium of interest [6]. The multi-channel signals are amplified, filtered and processed to determine energy content [7] and then continuously analyzed to detect abnormalities corresponding to leaks and, if possible, to localize the probable place where the fault occurs.

The major problems influencing the sensitivity and reliability of the acoustic system for leakage detection are changes of wave characteristics caused by the environmental factors (e.g. boiler construction, leakage localization) and physical phenomena (e.g. deflection, interference, etc.) which vary background noise level and attenuate sound within the medium. Another objective is relatively precise localization of the leakage, which can save some time and decrease the repair cost after shutting down of the installation. Several systems of multi-microphone arrays for passive source localization have been developed, which take into consideration the effect of sound propagation through the combustion temperature gradient field inside the boiler and the reverberation in the boiler enclosure on the leakage localization [8, 9]. Recent approaches to leak detection and localization employ 3D source location techniques [10], with an accelerometer used to collect structure borne noise.

The fast Fourier transform (FFT) for power spectral density estimation has been traditionally used for the purpose of leak detection – to differentiate the acoustic spectrum of a tube leak from the ambient acoustic noise due to combustion. To improve reliability of fault detection in the complex signal environment present in the boiler, the autoregressive moving average methods have been also employed to produce earlier and reliable detection of tube leaks [11]. In the last years the time-frequency decomposition tools (e.g. the wavelet transform), combined with the multi-microphone passive acoustic source localization method [12], have been applied as efficient methods to remove the effect of the background noise in boilers. The accuracy of the leak source localization can be improved effectively by applying the cepstrum method, used as a tool to reduce the reverberation effects of the leak source location.

Steam/water balance testing is quite simple, but rather time consuming and insensitive to small leaks; the maximal frequency of tests is usually too low for preventing serious damages of the pipeline system. The natural consequence of a leakage is the increase of the amount of water needed to satisfy the steam demand, what can be used in the mass balance method. The flow meters around the waterside of the boiler may be used to calculate the amount of water entering and leaving the boiler. By the combination of measurements of flow and chemical concentration, the mass balance of a specific stable and nonvolatile species (such as phosphate or molybdate) around the waterside of the boiler may be calculated. If a statistically significant loss is noticed, the water leak is suspected and an alarm is triggered to alert the operator [13]. However in typical situation the boiler does not operate in the steady-state conditions (with a known and constant heating value of the fuel), so the method is significantly affected by the process noise, i.e. variations in the individual and composite signals. Another drawback of the mass balance method is that it cannot be used for localization of the leakage.

The accuracy of the mass balance method is also decreased by the blowdown effect. The blowdown is used to control the dissolved solids in the boiler water, based on results from periodic boiler water testing. As it was noticed [14], the blowdown effect is important in that it constitutes up to 3% of the boiler mass balance equations. It can contribute to missed or false alarms if it is not compensated for, accurately, in the leak detection method. The least squares (LS) algorithm may be used for correction of the ARX model for the boiler water/steam mass balance due to the blowdown effect.

The monitoring of gas humidity method has limited specificity, as the measured changes of humidity can be caused by water added to the combustion chamber, soot blowing, etc., as well as by the steam leaks to be detected. More advanced approaches to leak detection are based on the patented Input/Loss Method (1994-2004), which computes fuel chemistry, heating value and fuel flow by integrating effluent measurements with thermodynamics [15]. The method is based on integration of system stoichiometrics with thermodynamics (i.e. the boiler efficiency and system-wide mass/energy balances). The ability to detect tube leaks and their location is highly dependent on the ability to compute fuel chemistry online, based on system stoichiometrics and to correct errors which may be present in any parameter effecting system stoichiometrics, e.g. stack CO_2, boiler or stack O_2, and generally stack H_2O, injected limestone, air heater leakage, O_2 in the ambient air, etc.

Several interesting applications of model-based methods of leak detection in the steam boilers have been presented in literature in the last few years. Most of the methods have not been implemented in industrial practice, though the authors used real data recorded in heat and electricity plants or performed the experiments on small or medium-size pilot plants equipped with measurement devices typically used in industry. As the complete boiler models based on energy, mass and volume balance and other physical and chemical relations are very difficult to be built and successfully identified, simplified modeling techniques have been employed and boiler response characteristics have been used to evaluate the quality of the models (comparing to the experiments performed in industrial conditions and/or laboratory setup). The most successful approaches use different kinds of mathematical models of the process running in normal and abnormal operating conditions and sophisticated tools for signal processing and classification (e.g. Artificial Intelligence algorithms: neural networks, fuzzy systems, genetic algorithms, etc.).

The attempts to model static and dynamic behavior of steam boilers have been presented in literature for at least two decades [16]. Some modeling techniques are based on fundamental conservation principles [17, 18], supplemented by empirical relations and performance correlations obtained from a lab-scale steam cycle setup for which dynamic measurements are available. The model is first validated quantitatively against steady state values and subsequently the dynamic validation is performed, involving some disturbances of different magnitude imposed on different parts of the plant and on the flue gas mass flow [19]. The classical model-based approach to detection of leakages and sensor biases employs the least-squares algorithm which yields an increased value of the performance index when the plant outputs do not match the modeled ones. Due to non-stationarity of the process caused by various operating conditions and development of the fault, historical values of process variables passed to the model may be used with time varying forgetting factors [20]. Such an approach can improve the reliability of detection of leaks and instrument biases under the conditions of load disturbances and time variation of the system dynamics caused by the system nonlinearity [2].

Also the observers and state estimators are used in the fault detection schemes to handle both reconstruction of immeasurable variables and estimation of process parameters with uncertainties in closed-loop stabilized systems that operate under strict nonlinear detectability conditions. The studies [21, 22] show that the state estimator is able to follow the faulty system, detecting faults by examining changes in the controlled outputs with respect to the setpoint and then probing variations in the parameters estimated. The detection mechanism may be supported by the fault identification subsystem, which uses the information provided from the state estimator about differences from normal operation trends. The observers are viewed as filters designed to track the dynamics of the errors of desirable frequency domain characteristics in the presence of a class of model uncertainties (e.g. caused by process faults), as well as persistent excitation and measurement noise. Some other approaches use the Kalman filter to improve state estimation accuracy by preserving the nonlinearities of the boiler equations [23].

In the recent years the AI tools, especially artificial neural networks (ANN), have been successively applied for building non-parametric approximate models of heat transfer and combustion processes in boilers, used in control tasks and fault detection and identification problems. The fault tolerant control schemes presented in literature [24, 25] use the ANN as universal approximators of any nonlinear input-output mappings, both as the process models and as nonlinear controllers, periodically adapted to changeable process characteristics with the use of real-time data taken from a running boiler system [26]. The most attractive is the ability of self-adaptation of a neural net-based fault tolerant control system, yielding a small stabilization error and a short time of adjusting to changeable operating conditions [27].

Neural networks are also combined with fuzzy logic, both for modeling of a boiler system and for detection of process faults [28]. The system based on the artificial neural network and the fuzzy logic, applied to leak detection in the boiler tube has been also patented by Alouani and Chang [29]. In their solution the instant detection scheme starts with identification of the boiler tube leak process variables, which are divided into universal sensitive variables, local leak sensitive variables, group leak sensitive variables, and subgroup leak sensitive variables, and which may be automatically obtained using a data driven approach and a leak sensitivity function. In the first part of the solution there is a direct mapping between appropriate leak sensitive variables and the leak behavior, provided by a trained artificial neural network (ANN). In the second design ANN are used for learning, while approximate reasoning and inference engines performed by fuzzy systems are used for decision making. Advantages of such an approach include the use of already monitored process variables, without any additional hardware and/or maintenance requirements. Systematic processing of recorded data does not require an expert system and/or a skilled operator, and the systems are portable and can be easily tailored for use on a variety of different boilers when properly ‘tuned’ to specific plant data.

Some papers [30] show a combination of fuzzy logic, neural networks and genetic algorithms (GA), employed to develop proper models for the subsystems of a steam boiler. The multilayer neuro-fuzzy models approximate the subsystems of a real plant, while the GA extracts the optimized fuzzy rules. In several approaches also the expert system reasoning and Bayesian networks are used in the boiler fault detection schemes [31]. The main benefit of these approaches, comparing with probabilistic temporal models, is that the representation based on state changes at different times instead of state values at different times, is able to represent complex systems changing over time.

Principal Component Analysis – theoretical preliminaries

Principal component analysis (PCA) is a multivariate unsupervised statistic technique, commonly used in process monitoring [32]. The main benefits in processing a big data set using the PCA method are: dimensionality reduction, decorrelation, and, to some extents, feature extraction and data denoising, saving as much information from the original variables, as possible. PCA transforms a set of correlated random variables with the zero mean value into a small number of decorrelated variables called principal components, where the first principal component (PC_1) defines the direction of the greatest variability within the original data set, with subsequent principal components explaining a decreased amount of variability. Consequently lower order principal components can be excluded without losing essential information from the original variables, as they characterize mainly the process noise [33].

Data to be decomposed by the PCA method are gathered in a matrix $\boldsymbol{X} \in \boldsymbol{R}^{(n \times m)}$, consisting of m variables and n samples (observations), as below:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & \cdots \\ \cdots & \cdots & \cdots & x_{n-1,m} \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}. \quad (1)$$

Assuming that each data vector (which constitutes a column of the matrix $\boldsymbol{X}$) is centered about its mean and scaled to unity variance, and performing a singular value decomposition of the matrix $\boldsymbol{X}$, we may express the empirical covariance matrix by:

$$S = \frac{1}{n-1} X^T X = \left(U\Sigma V^T\right)^T \times \left(U\Sigma V^T\right) = \Lambda V^T, \quad (2)$$

where $\Lambda = \Sigma^T \Sigma \in R^{m\times m}$ is a diagonal matrix with nonnegative elements of its main diagonal and $U \in R^{n\times n}$ and $V \in R^{m\times m}$ are unitary matrices.

The diagonal matrix Λ is the covariance matrix of the principal components and consists of eigenvalues of the covariance matrix $\boldsymbol{S}$. The elements on the main diagonal of the matrix Λ are real eigenvalues, arranged in not increasing order (i.e. $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq 0$) and the *i*-th eigenvalue is the square of the *i*-th singular value: $\lambda_i = \sigma_i^2$ [34].

The data matrix $\boldsymbol{X}$ is decomposed by the projection onto two orthogonal latent subspaces: the principal component subspace ($\hat{X}$), capturing the most of data variations and the residual subspace ($\tilde{X}$) that includes some uncorrelated changes and noises [35]:

$$\begin{aligned} X &= \hat{X} + \tilde{X} = TP^T + E = TP^T + \tilde{T}\tilde{P}^T = \\ &= \sum_{i=1}^{l} t_i p_i^T + \sum_{i=1}^{m-l} \tilde{t}_i \tilde{p}_i^T. \end{aligned} \quad (3)$$

where the matrices $\boldsymbol{T}$ (score matrix) and $\boldsymbol{P}$ (loading matrix) correspond to the largest singular values, $\boldsymbol{E}$ denotes the residual portion of $\boldsymbol{X}$ and the products TT^T and PP^T are orthogonal.

The number of principal components *l* usually influences the sensitivity of fault detection and must be a compromise between significant reduction of data dimensionality (*l* << *m*) and maximal retention of data variability (for bigger values of *l*). A general, simple and intuitive approach, considers the cumulative percentage of the total variation which one desires that the selected PCs contribute [33]:

$$cr_m = 100 * \frac{\sum_{k=1}^{l} \lambda_k}{\sum_{k=1}^{m} \lambda_k} [\%]. \quad (4)$$

It turned out that when the accumulate contribution rate cr_m is above 85% (or between 80% and 90%), the first *l* principal components could sufficiently reflect the main information about the system [36].

A general approach for detecting the abnormal status of the process employs the T^2 and *SPE* (*Q*) statistics for the loading vectors retained in the PCA model [37]. The T^2 statistics, which measures the variations in the score space and can detect most of the faults that produced large mean shifts in the measurement variables, can be computed by:

$$T^2 = x^T P \Sigma_l^{-2} P^T x, \quad (5)$$

where $\boldsymbol{\Sigma}_l$ contains the first *l* rows and columns of the matrix $\boldsymbol{\Sigma}$, and $\boldsymbol{x}$ is the observation vector.

The appropriate threshold for T^2 statistics, based on the level of significance α, can be determined as defined below [38]:

$$T_\alpha^2 = \frac{\left[\left(n^2-1\right)m\right]}{\left[n\left(n-m\right)\right]} F_\alpha\left(m, n-m\right), \quad (6)$$

where $F_\alpha(m, n\text{-}m)$ is the value of *F*-distribution at the significance level of α with *m* and *(n-m)* degrees of freedom.

The squared prediction error (*SPE* or *Q*) statistics measures the amount of variation not captured by the PCA model, what may be considered as the lack of fit of the PCA model to the data. The *Q* statistics is then defined on the residual vector $\tilde{x}_{k+1}$ [39]:

$$\begin{aligned} Q &= e_{k+1}{}^2 = \tilde{x}_{k+1}^T \tilde{x}_{k+1}, \\ &\quad where \quad \tilde{x} = \left(I - PP^T\right)x, \end{aligned} \quad (7)$$

i.e. on the portion of the data space which corresponds to the lowest (*m-l*) singular values.

The *Q* statistics can be monitored by using the threshold value computed as:

$$Q_\alpha = \theta_1 \left[\frac{C_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0 \left(h_0 - 1\right)}{\theta_1^2} \right]^{\frac{1}{h_0}}, \quad (8)$$

where

$$\theta_i = \sum_{j=l+1}^{m} \lambda_j^i \qquad i = 1,2,3, \quad (9)$$

$$h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2}, \quad (10)$$

and c_α is the normal deviate corresponding to the upper ($1-\alpha$) percentile.

Any value of T^2 and *Q* statistics that exceeds the threshold values defined by (6) and (8) indicates abnormal (faulty) conditions of the process under monitoring. However, when using T^2 and *SPE* statistics for process monitoring, the calculation of their control limits is made under the assumption that the latent variables are Gaussian-distributed [32]. Otherwise, the control thresholds (6) and (8) may inaccurately represent the boundary of normal operation region of the process, so they may be misleading or cause false alarms.

Application of Principal Component Analysis in fault detection

PCA and its modifications in statistical process monitoring

In the fault detection and identification problem, the effectiveness of the approaches based on quantitative model-based methods (e.g. based on state/parameter observers or methods of parameter estimation employing mathematical models constructed from the first principle) strongly depends on the precision of the mathematic models being used. The qualitative model-based methods which use cause-effect reasoning (e.g. fault trees analysis or Bayesian networks) are restricted to systems with a relatively small number of variables or states. The alternative approaches make use of data driven methods, such as those involving artificial intelligence techniques (e.g. neural networks or fuzzy systems) and statistical dimensionality reduction (e.g. PCA). The above fault detection methods find patterns (corresponding to normal operating conditions and faulty conditions) or compute meaningful statistics directly from historical data, eliminating the use of detailed models for large-scale systems which can be expensive and difficult to develop [37].

It must be however mentioned that the efficiency of data driven methods is also limited by several factors. The process monitoring of a real large scale system is difficult due to complex interactions between faults (especially multiple) and symptoms, high correlation among the measured variables and a large number of sensors, actuators and process devices to be monitored. So the large input dimension, which increases the computational requirements and the complexity of the fault detection system, is probably the main factor which decreases the effectiveness of process monitoring. Also the use of neural networks for fault classification is limited due to their black box characteristics, high computational load for complex systems and the necessity to construct (in fact) a separate neural classifier for each fault to be detected; the above drawbacks are only slightly reduced by the use of fuzzy and neuro-fuzzy systems.

That is why statistical methods for dimensionality reduction, such as Principal Component Analysis, have been widely applied to capture the variability in the data corresponding to fault occurrence and to express the relationships between process variables. Another common feature of statistical methods is their ability to reduce correlations between variables, what enables efficient extraction of the relevant information and analysis of changes in data characteristics. The PCA method can compress high dimensional and correlated process measurements into much lower dimensions while keeping the important information. The PCA has been successfully used in numerous areas including data compression, feature extraction, image processing, pattern recognition, signal analysis, and process monitoring [40]. However the level of dimensionality reduction, i.e. the ‘optimal’ number of principal component (PCs) to retain, which can provide the best sensitivity of fault detection and the lowest dimension of the decision space, is still under intensive investigations [41].

An important problem for the PCA monitoring approach is that the conventional PCA and its basic modification – the sliding window principal component analysis (SWPCA), are linear methods, while most processes are nonlinear or parameter varying [42]. That is why there are several modifications of the basic PCA method, like the “generalized PCA” (GPCA), in which the normal data set is extended to include nonlinear functions of its elements and the PCA transformation is performed on the extended data set. The recursive PCA method contains an adaptation mechanism which is especially suitable for time-dependent processes with slow changes [35, 38]. The Multi-scale principal component analysis (MSPCA), which is a combination of PCA and wavelet analysis, removes the autocorrelations of variables by means of wavelet analysis and eliminates cross-correlations between variables with PCA. The Multiway Principal Component Analysis (MPCA) is the extension of the PCA to batch processes. In that method the normal batch data are compressed and the information is extracted by projecting the data onto a low-dimensional space that summarizes both the variables and their time trajectories. Having established the normal process behavior, the process of a given batch is then monitored by comparing the time progression of the projections in the reduced space with those of the normal batch data [43]. Also the combinations of neural networks and the PCA into the nonlinear principal component analysis (NLPCA) were found to be useful for fault detection. In order to increase the sensitivity of the residuals with respect to various faults and achieve fast and robust detection, the structured residuals generated from both partial PCA and isolation enhanced PCA can be evaluated, employing the CUSUM method [44].

Principal Component Analysis for steam boiler fault detection

Several papers have been published about the use of Principal Components Analysis to detection faults in steam boilers. The authors use data obtained from the literature models or real-life steam boilers, working in normal operating conditions and in a limited set of abnormal states, caused by abrupt faults. Decisions about assessment of the considered data set to a specific class of fault is made on the basis of T^2 and Q statistics, sometimes used together with the prediction residual sum of squares (PRESS) statistic [38]. In the publications which use mathematical models of the plant the faults are typically emulated by changing values of selected model parameters. In general, the authors report satisfactory results of the PCA-based approach to boiler fault detection [45]; however the reliability may substantially differ for various changes of model parameters.

To overcome the problem of frequent changes of the plant operating point a set of principal component models (PCMs) corresponding to stable operating points can be built – each of them is built on the basis of a data subset separated (e.g. with the use of the K-mean cluster analysis) from the entire process data [46]. During the detection stage the PCM suitable for the current operating condition may be dynamically selected (e.g. based on fuzzy data partitioning) and appropriate statistics may be calculated. Also other two problems have to be solved: the first is how to pick up the modeling data and build the multi-PCM, the second one is how to implement the operating condition adaptability of the PCM during fault detection. Some other papers [36, 47] confirm that the multi-PCM method is an efficient approach to fault detection in changeable operating conditions.

The most comprehensive approach to the application of the PCA method to leak detection in a steam boiler system is presented in the work of Xi Sun et al. [48]. The authors implemented some extensions of the basic PCA method and tested the sensitivity of the detection system using the simulation model and real-life data obtained from the Syncrude Canada utility plant, which is an industrial co-generation system, utilizing a complex header system for steam distribution. The authors proposed a new data preprocessing scheme which can reduce the negative influence of noise and disturbances on detection reliability and developed the novel moving cumulative alarm (MCA) techniques combined with T^2 and *SPE*. Also a dynamic PCA model, based on changeable boiler characteristics, was developed and studied in the paper. The leak detection results using real industrial leak data demonstrate that the proposed approach significantly reduced the false alarms (which exist in the conventional PCA) and gives early fault warnings to the operators.

Boiler water system in Elektrocieplownia Bialystok

The OP-230 boiler, manufactured by RAFAKO Ltd. (Raciborz, Poland), is a part of the BC-50 thermal unit. The unit operates between 25 MWe and 55 MWe, providing 230 t of steam per hour (as a maximum continuous output) and has a back-pressure extraction turbine for district heating. The OP-230 is a one-drum and two-pass boiler with a natural water circulation. The unit is 27.0 m high, 7.5 m deep, and 8.4 m wide; the simplified furnace scheme is shown in Fig. 1. The OP-230 is tangentially fired with pulverized bituminous coal and equipped with four burner columns (six levels of swirl burners in each of them) installed in each of four corners.

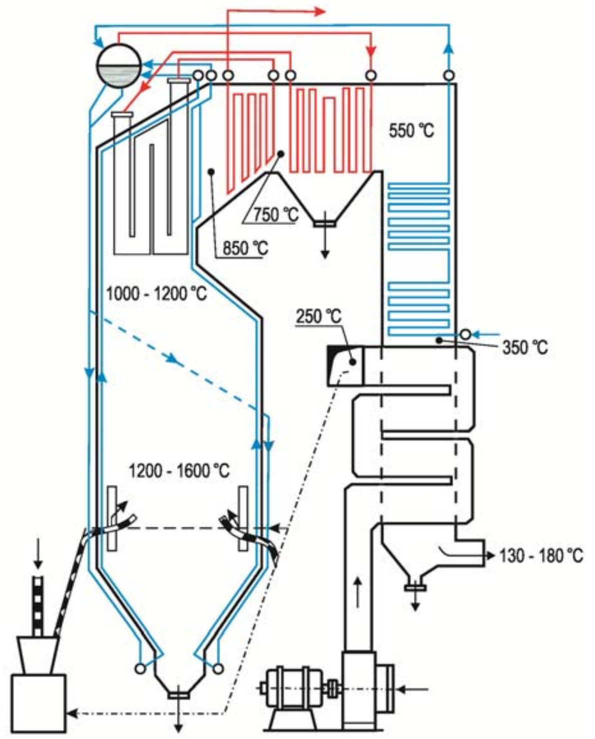

Fig.1. The schematic view of the OP-230 boiler [49]. Blue lines represent the pipeline system for circulation of water and low-pressure steam-water mixture. Red lines represent the pipeline system for high-pressure steam transportation

The main elements of the OP-230 boiler are: the drum, combustion chamber water walls, steam reheater and superheater, two attemperators, one economizer, two rotary air heaters and a supporting structure with a casing and platforms. The steam from the drum is supplied to the first stage of the convection superheater in the first pass, followed by the first steam attemperator. Then the steam flows to a platen superheater (the second stage), the second steam attemperator, the steam superheater of the third stage and the outlet collector. The air is supplied to the fans both from inside and from outside of the boiler room. A tube type economizer is located in the second pass of the boiler. The swirl burners generate short and wide flame, yielding corrosion problems on the walls and elements inside the furnace. The OP-230 boiler is equipped with antiexplosion protection.

The OP-230 boiler is equipped with 12 automatic control subsystems, which enable stabilization of main process variables to provide the proper run of fuel combustion and the steam generation process. In the period 2007-2009 in Elektrocieplownia Bialystok there was developed the Honeywell's Distributed Control System (DCS) 'Experion Process Knowledge System', which integrates the individual control loops and provides an additional level of high-integrity of process control and management. The main control and safety functions of the 'Experion PKS' system are: burner/boiler management, process safeguarding and emergency shutdown, turbine and compressor safeguarding, fire and gas detection and pipeline monitoring. The system provides the staff with complete real-time information about current operating conditions of the technological process and its individual elements and also about process faults and failures.

The following control subsystems act for maintaining a stable desired operating point of the boiler:

- underpressure in the combustion chamber (furnace draft),
- air flow, supplying the burners,
- air flow to OFA nozzle,
- air temperature after the steam air heater,
- contents of O_2 in the exhaust fumes,
- contents of NO_x in the exhaust fumes,
- temperature of the air-pulverized fuel mixture after the mills (four control loops),
- air flow delivered to the mills (four control loops),
- boiler load,
- steam temperature at the boiler outlet,
- steam pressure at the boiler outlet,
- water level in the drum.

Despite the existence of the advanced automatic control system (with sophisticated Fail Safe Control functions), each year over 30 major failures occur, which cause the unscheduled shutdowns of the boiler. The reasons of the boiler shutdowns, which happened in the last seven years, are shown in Table 1. As it can be seen the tube leak in the riser and downcomer sections (due to aging and thermal stress) is a quite common boiler fault. The faults of this kind are rather difficult to detect at their early stage by the diagnostic systems or by the process operator during on-line monitoring of the plant.

Table 1. Reasons of the shutdowns of power generating units in consecutive years

Reason of the block shutdown – year:	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
Reaction of safety systems – violation of alarm boundaries	11	19	9	15	19	16	11
Staff error	4	3	1	3	2	3	3
Boiler leak	12	10	9	11	7	7	6
Outer failure (fire, blackout in external electrical grid, etc.)	2	3	2	3	1	4	2
Total number of shutdowns	***29***	***35***	***21***	***32***	***29***	***30***	***22***

Boiler pipeline leak detection using real plant data

In the case of leakages in a boiler pipeline system the volume of a tube crack can only increase, so the failure can only expand in time. We may consider the development of a failure as a dynamic process, of unknown 'dynamics', usually with time constants changing in time, typically more rapidly after reaching a certain level of the tube damage. That is why it is very hard to determine the period from arising of the pipe leak to the moment it reaches the size when the operating staff can notice clear symptoms of the failure. The cases when one leak causes cracks of neighboring tubes, what can be regarded as failure propagation and multiplication, are also typical. The length of the pipeline section (where the leaks arise) amounts to several dozen of meters and the cracks appear in random locations, what makes it impossible to use only a single process variable for diagnostics of the pipe leak. To identify the fault with an acceptable level of certainty (as the fault effects unfold over time and this information becomes available to the personnel) it is necessary to combine various measurements acquired from inside the plant and to correlate them with typical fault patterns.

This section describes the application of the PCA method to detection of leakages in a pipeline system of a steam boiler in Elektrocieplownia Bialystok. We studied over 25 episodes of tube cracks, which took place during the period 2010-2015 and caused emergency shutdowns of the boiler. As the concept of the future advisory scheme for early leak detection consists in full cooperation with the existing distributed control system, in our trials we used the subset of the process variables which are now acquired and recorded by the measurement and control equipment. All the cases of failures we studied were well documented, i.e. we could use all the historical measurement data routinely archived in the plant database, as well as the protocols (including photographic documentation) from the inspection of the pipeline system during its repair. The number of cases we studied was too small to make any statistical conclusions about the efficiency or the accuracy of the presented approach, so the qualitative summary of our results is given later at the end of this section. Nevertheless we present and discuss two cases (in which the safety shutdown of the boiler was required), where the application of PCA would successfully detect leaks several hours earlier than there were actually noticed by the personnel.

Data analysis and preprocessing

The set of 37 quantities represents the process variables directly measured and recorded by the control equipment installed in the OP-230 boiler. Following the guidelines of the staff and careful analysis of historical data we selected the subset of 12 variables, which changes can be clearly observed when the leakage grows up to the considerable size. The number of 12 process variables includes:

- steam flow at the boiler output,
- two temperatures of steam, measured in specific locations in the boiler,
- five temperatures of fumes, measured in specific locations in a combustion chamber,
- feed water flow, measured at the boiler input,
- air flow at the boiler input,
- O_2 concentration in a selected location,
- lift in the hearth chamber.

These variables turned out to be the most sensitive to pipeline leaks, so we decided to use them in the experiments which examined the usability of the PCA method for leak detection. In another series of experiments we also examined different subsets of the above variables.

For the control purposes all the process variables are measured with the sampling period of 2 seconds and then recorded in the historical database. Such a sampling interval is definitely too short, regarding the typically observed dynamics of the tube cracking process and not useful for the functioning of the automatic system which would support leakage diagnostics. That is why we decided to aggregate the signal samples by averaging them in the period of 2 minutes (i.e. the average value was computed from 60 originally recorded samples). We also performed some experiments with the time interval equal to 5 minutes, but the shorter period gave better results of leak detection. The averaging may be also considered as a very simple operation of signal filtration, applied to remove fast fluctuations of the process variables, irrelevant from the point of view of the diagnostic procedure.

Three kinds of essential variability of heat demand affect the operating point of the steam boiler: seasonal (long-term), weekly or day-to-day (mid-term) and daily (short-term) changes. The first one can be observed as the changes of average monthly production of thermal energy, e.g. differences between February and July; however this kind of the long-term variability may be neglected in the design of the PCA model for leak detection purposes. As it will be discussed later, the PCA model which represents the 'nominal' operating conditions is created from data acquired in the last 2-4 days, so the distant historical data do not influence noticeably the construction of the principal component space. The daily profiles show the substantial changes of the demand, depending on the time of the day, i.e. higher demand during working hours and lower in the night. This type of demand variability usually requires multiple changes of setpoints of control equipment, device configuration, etc. Typical daily load of the steam boiler in the winter season (November 2011) is presented in Fig. 2.

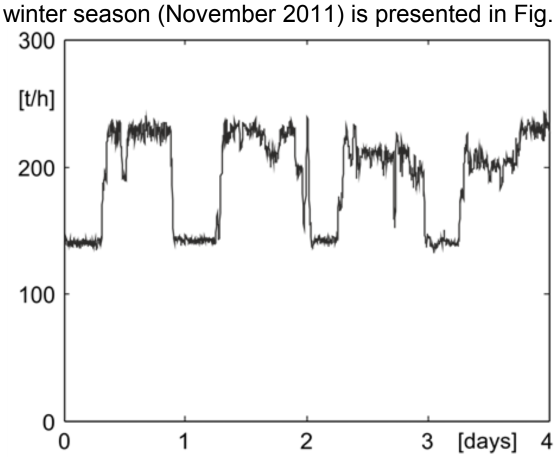

Fig.2. Typical daily variability of the steam boiler load in the winter season (November 2011). The changes in steam load during the 24-hour period are usually large (or even huge) and depend on interrelated demand on heat and electricity. It can be seen something of daily periodicity, however temporary fluctuations interfering with a periodical course are substantial

As it can be seen, there may be significant differences between demand profiles in the consecutive days, so it is not easy to establish a typical daily demand profile. Also the overall plant productivity during each 24 hours usually changes during a week, what makes the differences between demand profiles, e.g. from Sunday and Wednesday. That is why most of the signals which can be employed for leak detection do not fulfill the assumption of stationarity and the PCA model created at the moment from a set of historical data can be used for the analysis of a boiler state and fault prediction in a limited time horizon.

Solution of the leak detection problem with the use of PCA

The experimental approach to the leak detection problem, proposed in this paper, has two steps repeated iteratively every *n* day: the design (learning) phase and the monitoring (testing) phase. In the design phase, the set of historical data is used to develop the PCA model of the 'healthy' pipeline system of the boiler. Data used for model development represent the segments of process variables mentioned above, collected in the same, sufficiently long time period. Data matrix, containing the segment of one process variable in a column, is then normalized by removing its mean value and dividing by a variance of each column. Based on such historical measurements, the principal components model is then established and the principal components contribution rate is calculated [47]. The dimensionality of the principal component subspace (to be used in further steps of the algorithm) should be chosen according to the guidelines given in literature [33, 36]. In our case the three-dimensional principal component subspace has turned out to be the best one, because depending on the number of variables selected from the set of 12 original process variables, three principal components give the accumulate contribution rate cr_m (expressed by (6)) between 88% and 96%.

The points used to develop the PCA model are then mapped on the coordinate system created from three principal components of the largest contribution. The region containing 95% of such mapped points may be bounded by an ellipsoid, which is considered as the confidence region. As the mean value was removed from data before PCA decomposition, the center of the ellipsoid is located in the center of the coordinate system. The size of the ellipsoid represents the covariance of data used for model development and mapped to the principal components space. The data points located inside the confidence ellipsoid represent the 'healthy' operating conditions of the boiler pipeline system. During the monitoring phase, the newly acquired data point, which represents current values of the process variables, is normalized and mapped to the PC coordinate system. The location of the data point outside the confidence ellipsoid is the indicator of possible fault appearing in the pipeline system. Such a finding is also supported by the results of the comparison of the values of Hotelling's T^2 and *SPE* (Q) statistics (equations (5) and (7)) with their thresholds given by (6) and (8), respectively. The block diagram of the detection scheme is presented in Fig. 3.

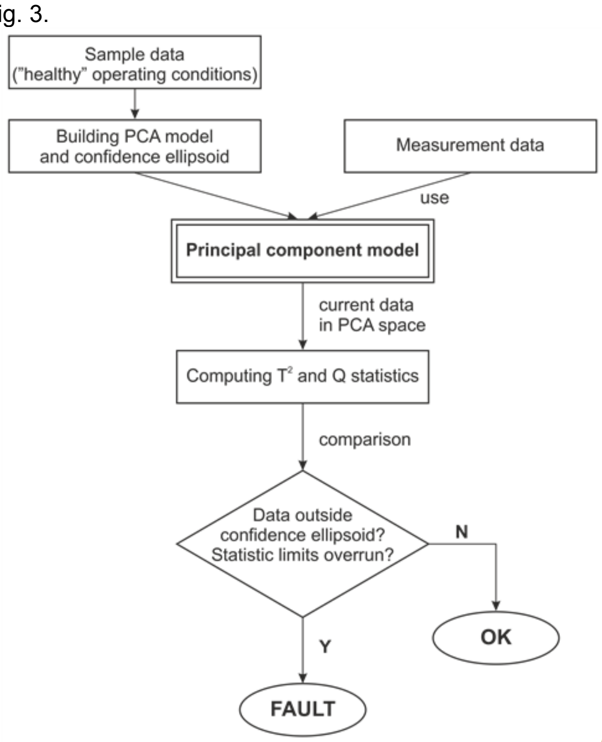

Fig.3. Block diagram of the PCA-based pipeline leak detection. In the numerical experiments such a sequence was repeated every three days, what provided adaptation of the PCA model of a 'healthy' system. The model was built from data contained in a sliding time window and used to establish the confidence region, employed (together with T^2 and Q statistics) for making a decision about leakage detection

As it was mentioned above, the PCA method has been successfully used as the alternative of model-based or AI-based approaches to fault detection. However the assessment of the current condition of the pipeline system to the 'healthy' or 'faulty' category is actually a complicated problem because of technical complexity of the plant and substantial and frequent changes of the boiler operating point. Analysis of data acquired from the examined boiler and results of preliminary simulations revealed the following problems, which proper solution turned out to be crucial for the successful application of the PCA method to leakage detection:

- The size and location of the time window used for building the PCA model of a 'healthy' system (i.e. how many historical samples should be used to develop the model and from which historical period?).
- Adaptation of the PCA model (i.e. should the model be adapted to possible changes of boiler operating conditions and how often such a procedure should be repeated?).
- Sensitivity of the PCA model to leakages in the whole pipeline system (i.e. is the model able to detect a leak in any part of the pipeline?).

The above problems cannot be solved theoretically; any practical solutions can be based only on general hints coming from heuristic approaches reported in literature and experimental work on tuning the algorithm to a specific case. Proper choice of the length of a data segment to be used for building the PCA model of a 'healthy' system is very important for the ability of a presented method to detect leakages. As it has been mentioned above, the plant operates in variable conditions, so the segments of data collected in a long time period do not fulfill the stationarity assumption. From the other hand too short data segments (e.g. corresponding to the 24-hours period) yielded the models with poor performance. So, it turned out from our numerical experiments that the best accuracy and sensitivity of the proposed detection scheme to detect a growing leak has been obtained when the PCA model of a 'healthy' system was developed using data collected in the period of 2-4 days. Such a PCA model fits to changeable operating condition in a limited time period, so it must be adapted similarly to the approach described in [35]. In our problem the best results were obtained when the PCA model of the 'healthy' system was periodically recomputed once a day, using data segments of the constant length, contained in a sliding time window. In the other words, for each process variable the 'tail' of the oldest historical daily recordings was removed from the data segment and the measurements from the new day were added at the 'head' of the segment.

We tested the proposed approach on 25 cases of leakages, which caused unscheduled shutdowns of the boiler in the years 2010-2015. As it can be seen from Table 1, there were 56 such cases in the above period, however not all of them could be used in our experiments. Except of the shutdowns caused by leakages, the boiler was also shut down due to other reasons mentioned in Table 1 and sometimes the intentional shutdowns were taking place, e.g. according to the heat production plan, plant repair schedule, etc. So, in many cases the period of boiler operation between the shutdowns was too short for building the PCA model and testing its ability to predict the leakage in a desired time horizon, assumed in our experimental setup. Table 2 shows the overall results of tests performed on 25 cases considered in numerical experiments, i.e. the number of days before the shutdown when our PCA-based approach gave undoubted indication of forthcoming failure.

In the experiments, which results are shown in Table 2 and in some details discussed below, we applied the following rules:

- The PCA model of a 'healthy' system was built using data segments of 12 process variables, recorded in the 3-days period.
- The model was then applied to process data recorded in the next 5 days after the moment of model building – the consecutive data points displayed on the confidence ellipsoid (as a background) in the PC space formed a specific 'fault trajectory'.
- The procedure described in two above rules was repeated for data with the time shift of 24 hours (the time horizon was shortened adequately, while approaching the shutdown).
- The moment when the fault trajectory was permanently leaving the confidence ellipsoid was regarded as the indication of the leakage appearance.

The last rule of making decision about the most likely appearance of a leakage needs a comment. It is quite obvious that due to process disturbances, transient states caused by changes of operating conditions of a boiler, etc., some isolated data points (or small groups of subsequent data points) corresponding to the 'healthy' state may be located outside the confidence ellipsoid. It is also quite likely that the opposite situations may appear when the fault trajectory, especially at the early stages of leakage development, may temporary enter the confidence region. So we considered that the one-hour segment of the fault trajectory lies outside the confidence region when not more than 5 data points (of 30 points analyzed each hour) and not more than 3 consecutive points are located within the ellipsoid. Such an assumption is not crucial because due to continuous development of the leakage the trajectory was always driven away from the confidence region (and finally there were no points inside the ellipsoid), but it was helpful to establish unambiguously the moment of the leak detection, as it is presented in Table 2.

Table 2. Performance of the PCA model in detection of boiler leakages

Time horizon of leakage prediction (number of days, *n*)	**≤ 1**	**(1; 2]**	**(2; 3]**	**(3, 4]**	**(4; 5]**
Number of cases, for which the leakage was predicted *n* days before its detection by the personnel	2	3	5	8	7

As it can be seen from Table 2, only two of 25 leakages (i.e. 8%) were detected by our method not earlier than 24 hours before the shutdown. It can be also noticed that the number of failures detected 3-5 days before the shutdown amounts to 15 (60% of all cases), what allows to apply the more restrictive interpretation of the results. The PCA model of the 'healthy' system, created at the beginning of the n-th day, used data from days: $n-1$, …, $n-p$ – the results shown in Table 2 were obtained for the value $p = 3$. However at the current time moment we are not sure whether the process is still running without any fault, as the measurements of the process variables are the only source of information about current conditions of the plant. We can increase the reliability that data used for model building represent the 'healthy' system if we apply the time shift of the sliding window which contain historical data used for development of the PCA model, i.e. the model is built based on the data from days: $n-1-d$, …, $n-p-d$. Of course we do not know the dynamics of crack development because the fault may rise slower or faster, depending on its location and the boiler operating conditions, but we may assume that if there are no visible symptoms of the leak at the moment, so d days ago (and earlier) the pipeline system was also 'healthy'. Taking into account these considerations and assuming the value $d = 2$, we can notice that in 60% of the analyzed cases the proposed approach can detect the leakage at least one day before the shutdown was necessary. And (applying the two-day time shift in data used for model building) only in 17% of cases the PCA method did not detect the leakages sufficiently early to provide the personnel enough time for appropriate actions before the shutdown.

The last problem mentioned above, i.e. the ability of the PCA model to detect leakages in any part of the pipeline system has not been solved in a satisfactory way. The length of the pipeline section of a boiler amounts to several dozen of meters and the measurement devices are located only in several points, neighboring to relatively small areas of the pipes. So if the crack appears in a certain location of a bended pipe, it is likely that it causes different changes in the course of the process variables than the leak arising in another point. Unfortunately we had not enough data and sometimes incomplete reports from pipeline inspection after the boiler shutdown to perform systematic analysis of the problem and make statistical conclusions. Although, in our opinion, it is impossible to work out any rules which could associate the fault trajectory in the PC space with the location of the pipe crack, the solution of the problem of more precise localization of the leak seems to be the promising direction of our future work.

Two case studies

Data recordings from the period 2010-2015 were used in numerical experiments, what allowed us to check the performance of the PCA-based approach to leakage detection for different physical characteristics of the plant. In 2011 the boiler was substantially renovated, due to creeping degradation of pipelines and an increasing number of leaks, some sections of pipelines were replaced by the new ones, also some pieces of new equipment were installed. Replacement of the pipeline sections and some other repairs of the boiler noticeably changed static and dynamic properties of the plant, so we can regard that the presented approach was tested on a very wide range of operating conditions. As a result of the renovation the number of tube cracks decreased in 2012, however in 2013 some leaks occurred in the places where the new pipes were joined to the old sections of pipelines. To illustrate the ability of the PCA model to detect leakages in the pipeline system of the boiler we have chosen two cases of faults, which required unplanned shutdowns of the boiler – the first one before the renovation (from January 2010) and the second one from the end of 2013 and beginning of 2014.

In both cases the process was shut down shortly after the personnel undoubtedly determined the cracks, mostly on the basis of acoustic and visual examination. During the inspection of the pipeline system after shutdown the single or multiple cracks of the pipes have been confirmed. The first symptoms of irregularities in the measured process variables could be observed by the operator not more than a day before auditory confirmation of the fault, however as we are showing below, the leakages could be clearly detected earlier using our approach. In both experiments presented below the full set of 12 process variables (mentioned above) has been used to prepare the PCA model of the 'healthy' system, with the principal component space reduced to three the most significant ones. No prior knowledge from the operator, whether the data represent the leak and when the leak has taken place, and whether the plant should be shut down as a result of a leak, was used in the experiments. The detection scheme followed the flowchart presented in Fig. 3 – in both cases discussed below we used the 'healthy' models constructed from three days recordings and testing periods equal to five days. The figures in the next subsections show the evolution of the 'fault trajectory' in the principal component space, until it goes outside the confidence ellipsoid at least five days before the boiler shutdown. Also the Hotelling's T^2 and *SPE* (*Q*) statistics are presented to confirm the symptoms of pipe leaks detected by the algorithm.

Case 1

The first ambiguous symptoms of the arising failure were noticed during a routine inspection at about 11:30 o'clock on 22nd January 2010, as a very loud noise coming from the area of the steam superheater on the right side of the boiler. There were no clear symptoms of escalating fault which could be noticed by the process operator, e.g. as substantial changes of characteristics of process variables. The boiler was shut down very quickly after that moment, i.e. at 14:07 on 22nd January 2010, immediately after setting in motion a reserve boiler. The inspection of the superheater area on the right side of the boiler, performed several hours later, has shown the cracks of two neighboring pipeline sections, as presented in Fig. 4.

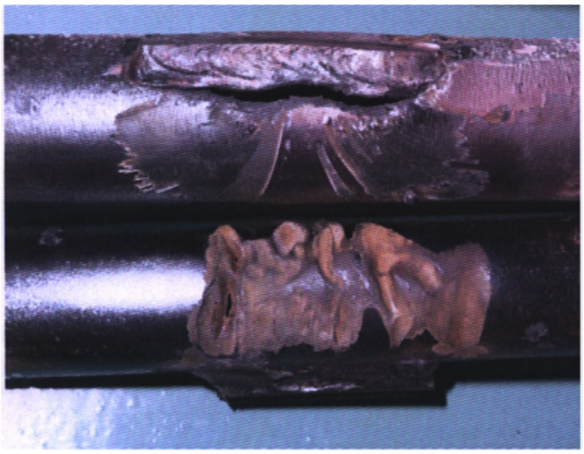

Fig.4. Cracked pipes, damaged on 22nd January 2010. The photo shows two neighboring pipes, which were damaged during that failure. The cracks caused very loud noise in the combustion chamber, what allowed the personnel to detect the fault during routine inspection

The ability of the PCA-based approach to detect the leakage at its early stage can be derived from the analysis of the evolution of 'fault trajectory' presented in Fig. 5a)-c). In all figures the green dotted lines represent the trajectory formed by data points used to build the model of a 'healthy' system (recorded within three days), while the solid dark grey lines represent the 'fault trajectory' created by data points from the next five days. The points located inside the confidence ellipsoid are marked with circles, while the points outside the ellipsoid are marked with crosses ('x' or '+' symbols). The arrows show the direction of the time flow; the straight-line segments connect the data points mapped with the time distance of four hours.

As it was mentioned in the previous section, the PCA models of a 'healthy system' were built from data sets representing 12 process variables measured in the period of the preceding three days.

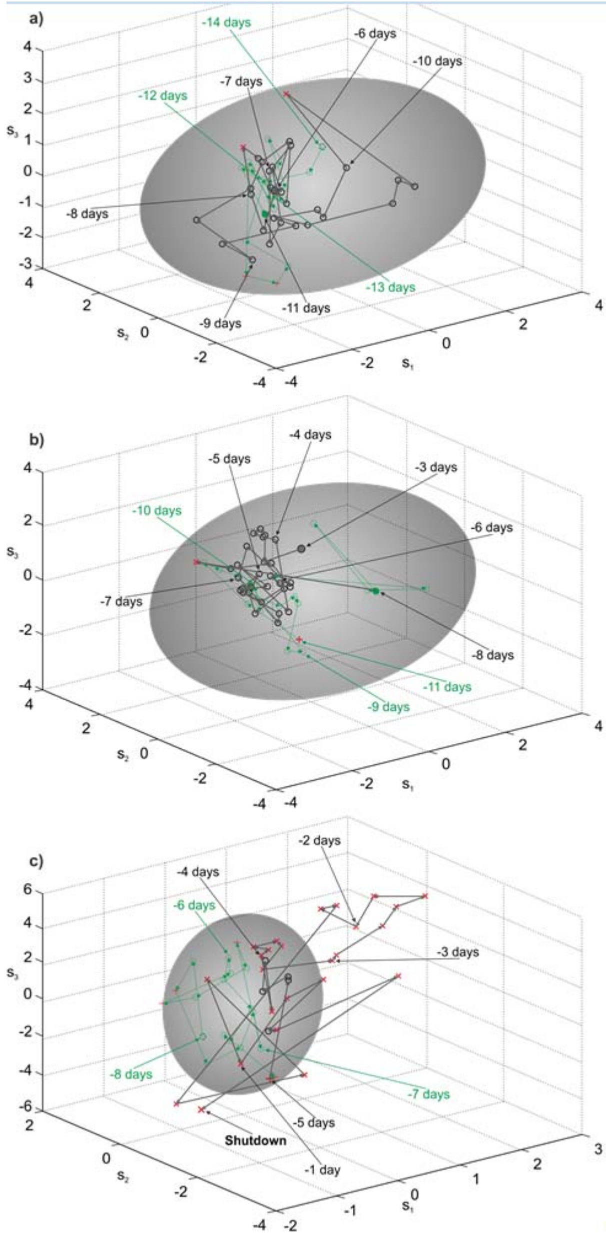

Fig.5. Fault trajectories in the 3-dimensional Principal Components space, shown in the two-week period before the shutdown on 22nd January 2010. The graphs show the location of data used for model building (green dashed lines) and measurement data collected in the next time period (black solid lines). The distance in time between the consecutive points is equal to four hours. The a), b) and c) subplots show three 'snapshots' corresponding to the evolution of the pipeline state, taken every three days in the time period preceding the shutdown. Data points located within the confidence ellipsoid are marked with circles, while the points lying outside the confidence region are marked with the (+) or (x) characters. In all cases a data segment from three days was used for building the PCA model of a 'healthy' system, which was then used to test data from the next five days. In the a) subplot the 3-day period started 14 days before the shutdown, while the next five days (starting from the 11th day before the shutdown) created the trajectory of fault evolution. In the b) subplot the 3-day period used for PCA model building started 11 days before the shutdown, and in the c) subplot it started 8 days before the shutdown. The 5-day period for testing the trajectory of fault evolution started from the 8th day before the shutdown in the b) subplot, and from the 5th day before the shutdown in the c) subplot

The reduced Principal Component space was created by three components of the highest magnitude and the analyzed 'fault trajectory' contains data from five days following the data segment used for model building. So, Fig.

5a) shows the confidence ellipsoid corresponding to the PCA model created from data collected in the 14th, 13th and 12th days before the shutdown. As it can be seen almost all the data used for PCA model building (except of one point) are located within the confidence ellipsoid; also the location of data points from the next five days does not fulfill our criteria for leakage detection, described in the previous section.

Similar observations may be made from the analysis of trajectories presented in Fig. 5b), which shows the confidence ellipsoid derived from the PCA model based on data from 11th, 10th and 9th days before the shutdown. Both the ‘model trajectory’ and the ‘fault trajectory’ are placed inside the confidence region, except of a very small number of isolated points. So, it may be concluded that the model created from data taken about 10 days before the shutdown did not give any warnings, analyzing data up to three days before the shutdown. The only symptom of presumable changes of the condition of the pipeline system is the small change of the size and position of the confidence ellipsoid; however it may be caused by the alterations of the plant operating point.

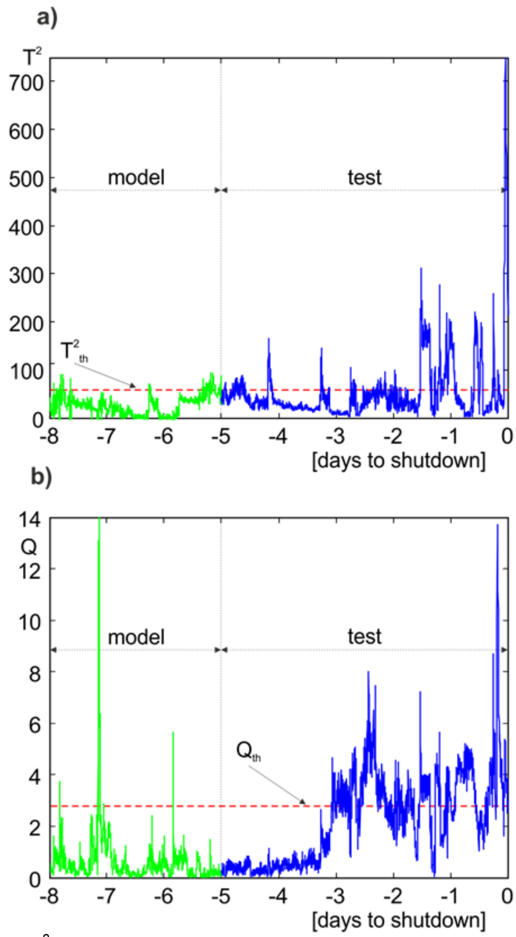

Fig.6. T^2 and Q statistics for the last eight days before the shutdown – the statistics correspond to the data set presented in Fig. 5 c). The green lines represent the statistics for data used for PCA model building, while the blue lines represent data collected in the last 5-day period just before the shutdown. The dotted red lines mark the threshold values, computed according the formula (6) for the T^2 statistic and the formula (8) for the Q statistic

Fig. 5c) shows the confidence ellipsoid corresponding to the PCA model created from data collected in the 8th, 7th and 6th days before the shutdown and the ‘fault trajectory’ obtained from data recorded in the next five days, until the shutdown. The picture shows substantially different behavior of the ‘fault trajectory’, which is only partially located within the confidence region during the fifth day before the shutdown, but after that permanently leaves the confidence ellipsoid. Also the confidence ellipsoid changes its size and location comparing to previous ones, which correspond to the ‘healthy’ pipeline system. It can be considered that the PCA-based approach generated the clear warning about the developing leakage at least 100 hours (more than 4 days) before the moment of the emergency shutdown. Even if we choose the more restrictive interpretation (presented in the previous section) with the two-day time shift of selection data to create the PCA model, in this case the method was able to detect the leakage more than two days before the shutdown.

To confirm the conclusions made from the analysis of the location of the ‘fault trajectory’ in relation to the confidence ellipsoid, we also checked the Hotelling’s T^2 and the Q statistics. The values of both statistics in the last eight days before the shutdown (corresponding to Fig. 5c)) are presented in Fig. 6 a) and b). The first segments of both statistics (depicted in green color) were computed for data used to build the PCA model. The last segments (depicted in blue) show the T^2 and the Q values for the data from the last five days directly preceding the shutdown. The thresholds (dashed red lines) indicating the faulty conditions were computed with the use of equations (6) and (8), respectively. Although we can observe qualitative changes of both statistics in the last five days before the shutdown (comparing to the three-day period representing the ‘healthy’ conditions), but clear symptoms of fault may be derived from these statistics only three days before the shutdown. So, the analysis of the ‘fault trajectory’ in the PC space seems to be the more sensitive tool for fault detection.

Case 2

The first symptoms of leakage were noticed by the process operator at about 1:30 o’clock in the night on 13th January 2014 as small disturbances of process variables measured by the control equipment. Then the personnel reports from the routine round confirmed fast increase of the leakage, but the boiler was kept working until 9:46 on 13th January 2014, when the substitute boiler was started. The inspection of the lower area of the collector chamber, performed during the boiler repair, revealed the crack presented in Fig. 7.

Fig.7. The photograph of the damage, which caused the shutdown of the boiler on 13th January 2014. The leakage occurred in the lower area of the collector chamber; the crack was revealed by the personnel during the routine inspection

Fig. 8a)-c) demonstrate the process of fault development in the PC space, using the same presentation manner as in the first case discussed above (see Fig. 5a)-c)). In the first period (Fig. 8a)) more or less all the ‘fault trajectory’ is located inside the confidence ellipsoid. In Fig.

8b) representing the period between the 8th and the 3th day before the shutdown, the ‘fault trajectory’ leaves the confidence region in the middle of the last day of testing the PCA-based detection method. It may be regarded as the first symptom of developing the pipe crack, however it should be confirmed in the next test, after adaptation of the PCA model. In the last term, presented in Fig. 8c), the ‘fault trajectory’ is located outside the confidence region at the beginning of the five-day period, thereafter in the fourth day before the shutdown the trajectory goes into the ellipsoid and then, approximately 85 hours (about 3.5 days) before the shutdown, permanently leaves the confidence region. Even if we consider the interpretation of the fault detection moment assuming the two-day time delay of selection data to create the PCA model, in this case the method was able to detect the leakage about 1.5 days before the shutdown.

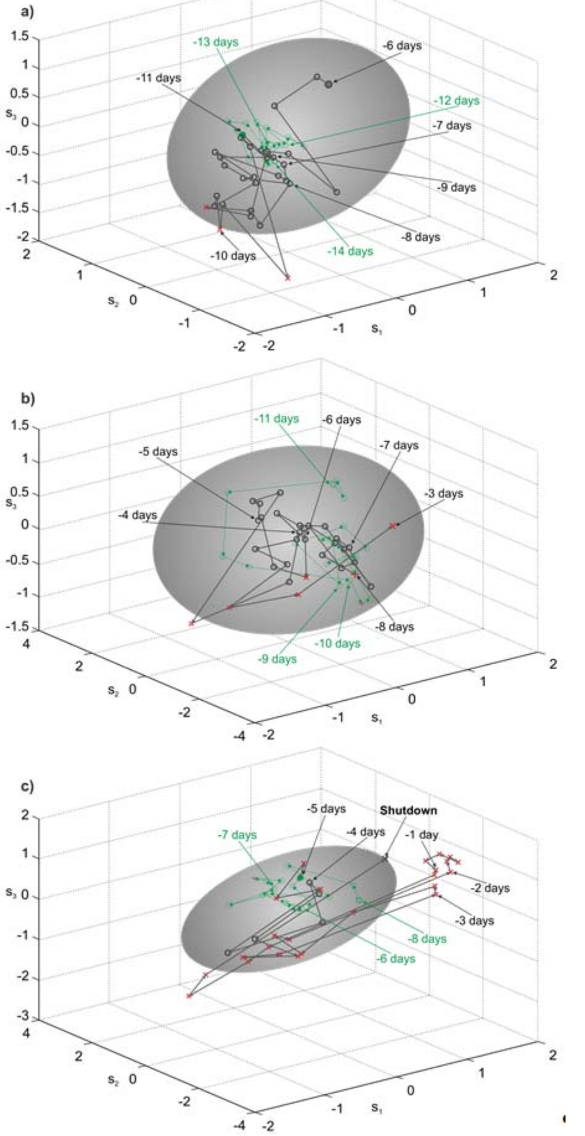

Fig.8. Fault trajectories in the 3-dimensional Principal Components space, shown in the two-week period before the shutdown on 13th January 2014. Detail description of the plots is the same as for those, presented in Fig. 5

The three pictures representing the condition of the pipeline system significantly differ: there are only slight changes of the shape and the size of the confidence ellipsoid in the first two of them (probably due to changing operating conditions of a boiler) and significant changes of the confidence region and the shape of the ‘fault trajectory’ in the third picture, what gives a clear warning about the failure. The indication that the boiler works in faulty conditions may be confirmed by the analysis of the Hotelling’s T^2 statistics in the last five-day period before the shutdown, presented in Fig. 9. In this experiment the Q statistics turned out to be visibly less sensitive to the leak development, as its value exceeds the threshold only in a small part of the last five days of the observation.

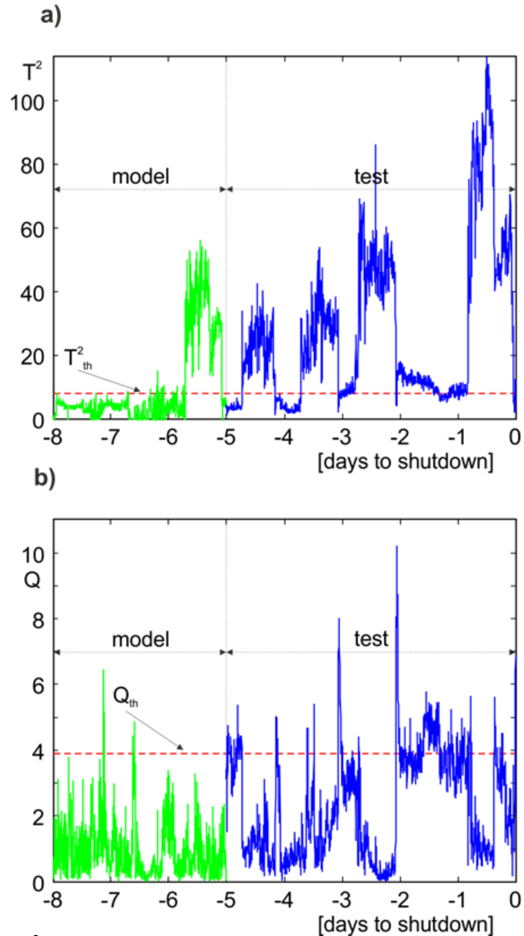

Fig.9. T^2 and Q statistics for the last eight days before the shutdown – the statistics correspond to the data set presented in Fig. 8 c). Detail description of the graphs is the same as for those, presented in Fig. 6

Conclusions

The studies presented in this paper confirmed that the PCA method with proposed improvements is a useful tool for early detection of a specific class of faults, i.e. the leakages in the pipeline system of a steam boiler. The numerical experiments reported in the paper have been performed using real data obtained from an industrial plant. It should be emphasized that the presented problem has some specific characteristics, which make it complicated and impose some difficulties or even restrictions on the application of other fault detection algorithms:

- The moment when the leak begins is unknown and impossible to be established precisely (or sometimes even approximately) by any fault detection technique.
- The dynamics of the leak growing is unknown, leak development changes in time and depends on many factors (e.g. operating conditions, location of the pipe cracks, material properties, etc.).
- The possibilities to develop and identify the process model, adequate to be used for detection of leakages (which occur in a wide number of locations) in the case of the plant considered in the paper are very limited, regardless of the reported in literature attempts to utilize simplified (e.g. linearized) models for this purpose.

- The operating conditions of a plant (resulting from energy demand) can change substantially in a relatively short period, what is a real challenge for the fault detection methods which assume steady-state conditions necessary for e.g. linearization of plant characteristics.

That is why we applied the completely data-driven approach and examined some improvements of the classical usage of the PCA scheme. The approach discussed in the paper confirmed its ability to detect real-life pipeline leakages and in our opinion replaces the necessity of the construction and identification of the process model, what is always a time consuming and expensive task. The decision making process was based on the construction of the 'confidence space' (which represents the feasible area occupied by the values of process variables for the 'healthy' system in the PC coordinates) and continuous monitoring of the 'fault trajectory' which represents the transition of the process operating point in PC coordinates. The PC coordinates represent the main features extracted from a set of process variables used by the plant control system. The fault detection is supported by the analysis of the T^2 and Q statistics, commonly used in statistical process monitoring problems.

As the changing conditions of the plant operation (yielding nonstationarity of the signals) cause the serious obstacle in the successful use of the PCA method [2], we proposed periodical, with a one-day period, adaptation of the PCA model, what substantially improved the performance of the method. We also demonstrated that the current PCA model of the 'healthy' system should be constructed not from the newly recorded data, but from the data segment shifted some days into the past, what increases the chance that the modeled system was working in the unfaulty conditions. Such time shifts, as well as the length of the data segment, which are the most suitable in the case of leak detection, were established experimentally in our studies.

The analysis of a representative number of cases observed during over five years of the plant use confirmed that method described in the paper can effectively extract fault information and reduce the effect of noises and disturbances on leak detection. In most of the analyzed historical cases the system was able to foresee the leaks earlier than the operator, typically 3-5 days before the shutdown. We also tried to extend that forecasting horizon, checking the ability of the PCA model to detect the leak symptoms 7 or more days before the shutdown, however the sensitivity to the early faults noticeably decreased, comparing to the results discussed above. Nevertheless, the implementation of our approach may be the important element of an on-line system, which would lead to significant improvement in safety and maintenance of industrial boilers in Elektrociepłownia Bialystok. The early warning can be given for plant operators, based on periodically updated PCA model of the 'healthy' system and the analysis of the fault trajectory, together with T^2 and Q statistics.

The work was supported by the Dean's Project No. S/WE/1/2016, Faculty of Electrical Engineering, Department of Automatic Control and Electronics, Bialystok University of Technology.

***Authors**: Miroslaw Swiercz, PhD, DSc (Eng), Bialystok University of Technology, Faculty of Electrical Engineering, ul. Wiejska 45D, 15-351 Bialystok, Poland, E-mail: m.swiercz@pb.edu.pl; Halina Mroczkowska, PhD (Eng), ENEA Cieplo Sp. z o.o. w Bialymstoku, ul. Warszawska 27, 15-062 Bialystok, Poland, E-mail: halina.mroczkowska@enea.pl.*

REFERENCES

[1] Zhang H., Xu L., Diagnostic System for Current-Carrying Fault: Modeling, Precaution, and Prediction, *IEEE Trans. Power Delivery*, 29 (2014), No. 3, 1318-1325, doi: 10.1109/TPWRD.2013.2295005

[2] Sun X., Chen T., Marquez H.J., Efficient model-based leak detection in boiler steam-water systems, *Computers and Chemical Engineering*, 26 (2002), No. 11, 1643-1647

[3] Sun X., Chen T., Marquez H.J., Boiler Leak Detection Using a System Identification Technique, *Industrial & Engineering Chemistry Research*, 41 (2002), No. 22, 5447-5454

[4] Jankowska A., Approach to Early Boiler Tube Leak Detection with Artificial Neural Networks. In: Jablonski R., Turkowski M., Szewczyk R. (eds.), *Recent Advances in Mechatronics*, Springer-Verlag, Berlin Heidelberg, 2007, 57-61

[5] Sato R., Yasukouchi K., Yamamoto T., Device for early detection of rupture of the pressure part of a boiler, *United States Patent 3831561*, August 27, 1974

[6] Jung G.J., Cho Y.S., Kim Y.C., Baek S.H., Sung J.H., Frequency Shifting Signal Detection and Analysis of Boiler Tube Leaks, *18th International Conference on Composite Materials*, 21-26 August 2011, ICC Jeju, Korea

[7] Studdard B., Arrington P., Rechner M., Operating Experience Using Acoustic Leak Detection at Gaston Station, *Technical Paper BR-1492, Babcock & Wilcox, a McDermott company*, presented to: Power-Gen '92, 17-19 November 1992, Orlando, USA

[8] Jiang G., Wang L., Tian J., and Pan J., Research on acoustic source positioning method for boilers tube leakage, *Proceedings of 20th International Congress on Acoustics, ICA 2010*, 23-27 August 2010, Sydney, Australia, PACS: 43.20.-f, 1-6

[9] An L., Wang P., Sarti A., Antonacci F., Shi J., Hyperbolic boiler tube leak location based on quaternary acoustic array, *Applied Thermal Engineering*, vol. 31 (2011), No. 16, 3428-3436

[10] Kim D.-H., Yang B.-S., Lee S.-B., 3D boiler tube leak detection technique using acoustic emission signals for power plant structure health monitoring, *Prognostics and System Health Management Conference* (PHM-Shenzhen), 24-25 May 2011, 1-7

[11] Jones C.S., Koch D.B., Morris K.W., Classical and parametric spectral analysis instrument using a visual programming language, *9th IEEE Instrumentation and Measurement Technology Conference, IMTC '92*, 12-14 May 1992, 533-538

[12] Jiang G., Wang L., Research on the Detection and Location of Sound Emission Caused by Tube Leakage in Boilers, *2012 Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, 27-29 March 2012, Shanghai, 1-4, doi: 10.1109/APPEEC.2012.6307413

[13] Burgmayer P.R., Durham V.E., Effective Recovery Boiler Leak Detection with Mass Balance Methods, *Proceedings of TAPPI Engineering Conference*, Atlanta (USA), 17–21 September 2000, 1011-1025

[14] Pertew A.M., Sun X., Kent Gooden R., Marquez H.J., A new blowdown compensation scheme for boiler leak detection, *2008 American Control Conference*, Seattle, USA, 11-13 June 2008, 4309-4311

[15] Lang F.D., Rodgers D.A.T., Mayer L.E., Detection of Tube Leaks and Their Location Using Input/Loss Methods, *Proceedings of the 2004 International Joint Power Generation Conference (combined Electric Power 2004 Conference), IJPGC2004*, Baltimore, USA, 30 March – 1 April 2004

[16] de Mello F.P., Boiler models for system dynamic performance studies, *IEEE Trans. Power Systems*, 6 (1991), No. 1, 66-74

[17] Colonna P., van Putten H., Dynamic modeling of steam power cycles. Part I – Modeling paradigm and validation, *Applied Thermal Engineering*, 27 (2007), No. 2-3, 467-480

[18] Liu C., Liu J., Niu Y., Jin X., Nonlinear modeling and simulation for large scale coal-fired power unit, *30th Annual Conference of the IEEE Industrial Society 2004*, 2-6 November 2004, Busan, Korea, vol. 3, 1983-1986

[19] van Putten H., Colonna P., Dynamic modeling of steam power cycles. Part II – Simulation of a small simple Rankine cycle system, *Applied Thermal Engineering*, 27 (2007), No. 14-15, 2566-2582

[20] Sun X., Chen T., Marquez H.J., Detecting leaks and sensor biases by recursive identification with forgetting factors, *Proceedings of the 40th IEEE Conference on Decision and Control 2001*, Orlando, USA, December 2001, vol. 4, 3716-3721
[21] Castillo I., Edgar T.F. and Fernández B.R., Robust model-based fault detection and isolation for nonlinear processes using sliding modes, *International Journal of Robust and Nonlinear Control*, 22 (2012), No. 1, 89-104
[22] Marquez H.J., Riaz M., Robust state observer design with application to an industrial boiler system, *Control Engineering Practice*, 13 (2005), No. 6, 713-728
[23] Lo K.L., Rathamarit Y., State estimation of a boiler using the unscented Kalman filter, *IET Generation, Transmission & Distribution*, 2 (2008), No. 6, 917-931
[24] Nazaruddin Y.Y., Nur Aziz A., Sudibjo W., Improving the Performance of Industrial Boiler Using Artificial Neural Network Modeling and Advanced Combustion Control, *International Conference on Control, Automation and Systems, ICCAS 2008*, 14-17 October 2008, Seoul, Korea, 1921-1926
[25] Yu D.L., Chang T.K., Yu D.W., Adaptive neural model-based fault tolerant control for multi-variable processes, *Engineering Applications of Artificial Intelligence*, 18 (2005), No. 4, 393-411
[26] Rostek K., Morytko L., Jankowska A., Early detection and prediction of leaks in fluidized-bed boilers using artificial neural networks, *Energy*, 89 (2015), September 2015, 914-923, doi: 10.1016/j.energy.2015.06.042
[27] Dong X.-C., Wang H.-B., Zhao X.-X., Model Reference Neural Network Control for Boiler Combustion System, *Proceedings of the Fourth International Conference on Machine Learning and Cybernetics*, Guangzhou, China, 18-21 August 2005, vol. 8, 4694-4698
[28] Ghaffari A., Moosavian S.A.A., and Chaibakhsh A., Experimental Fuzzy Modeling and Control of a Once-through Boiler, *Proceedings of the IEEE international Conference on Mechatronic & Automation*, Niagara Falls, Canada, July 2005, vol. 3, 1340-1345
[29] Alouani A.T., Chang S.-Y.P., Artificial Neural Network and Fuzzy Logic Based Boiler Tube Leak Detection Systems, *USA Patent No: 6,192,352 B1*, Feb 20, 2001.
[30] Ghaffari A., Chaibakhsh A., Lucas C., Soft computing approach for modeling power plant with a once-through boiler, *Engineering Applications of Artificial Intelligence*, 20 (2007), No. 6, 809-819
[31] Arroyo-Figueroa G., Sucar L.E., Vilaavicencio A., Probabilistic temporal reasoning and its application to fossil power plant operation, *Expert Systems with Applications*, 15 (1998), No. 3-4, 317-324
[32] Ge Z., Song Z., Gao F., Review of Recent Research on Data-Based Process Monitoring, *Industrial & Engineering Chemistry Research*, 52 (2013), No. 10, 3543-3562
[33] Jolliffe I.T., *Principal Component Analysis, Second Edition*, Springer series in statistics, Springer, New York Berlin Heidelberg, 2002
[34] Ge Z., Song Z., Multivariate Statistical Process Control: Process Monitoring Methods and Applications, *Advances in Industrial Control*, Springer, 2013th Edition, London, 2013
[35] Deng P.C., Gui W.H., Xie, Y.F., Latent space transformation based on principal component analysis for adaptive fault detection, *IET Control Theory & Applications*, 4 (2010), No. 11, 2527-2538
[36] Ma Y.-G., Zhang J., Fault Diagnosis based on PCA and D-S Evidence Theory, *Asia-Pacific Power and Energy Engineering Conference, APPEEC 2009*, 28-31 March 2009, Wuhan, China, 1-5
[37] Lau C.K., Ghosh K., Hussain M.A., Che Hassan C.R., Fault diagnosis of Tennessee Eastman process with multi-scale PCA and ANFIS, *Chemometrics and Intelligent Laboratory Systems*, 120 (2013), 1-14
[38] Ghamari A., Khaloozadeh H., Ashraf-Modarres A., Ghamari H., Application of Quantitative Data-Based Fault Detection Methods on a Drum-Type Boiler, *Proceedings of the 3rd Conference on Thermal Power Plants (CTPP)*, 2011, 1-6
[39] Jaffel I., Taouali O., Elaissi I., Messaoud H., Comparative study of PCA approaches for fault detection: Application to a chemical reactor, *IEEE - 2013 International Conference on Control, Decision and Information Technologies CoDIT'13*, 6-8 May 2013, Hammamet, Tunisia, 57-62
[40] Ding S., Zhang P., Ding E., Yin S., Naik A., Deng P., Gui W., On the Application of PCA Technique to Fault Diagnosis, *Tsinghua Science and Technology*, 15 (2010), No. 2, 138-144
[41] Tamura M., Tsujita S., A study on the number of principal components and sensitivity of fault detection using PCA, *Computers and Chemical Engineering*, 31 (2007), No. 9, 1035-1046
[42] Huang Y., Gertler J., McAvoy T.J., Sensor and actuator fault isolation by structured partial PCA with nonlinear extensions, *Journal of Process Control*, 10 (2000), No. 5, 459-469
[43] Yoo C.K., Lee J.-M., Vanrolleghem P.A., Lee I.-B., On-line monitoring of batch processes using multiway independent component analysis, *Chemometrics and Intelligent Laboratory Systems*, 71 (2004), No. 2, 151-163
[44] Cheng H., Nikus M., Jämsä-Jounela S.-L., Evaluation of PCA methods with improved fault isolation capabilities on a paper machine simulator, *Chemometrics and Intelligent Laboratory Systems*, 92 (2008), No. 2, 186-199
[45] Yu J., Yoo J., Jang J., Park J.H., Kim S., A novel plugged tube detection and identification approach for final super heater in thermal power plant using principal component analysis. *Energy*, vol. 26 (2017), 1 May 2017, 404-418
[46] Niu Z., Liu J.-Z., Niu Y.-G., Pan Y.-S, A Reformative PCA-based Fault Detection Method Suitable for Power Plant Process, *Proceedings of the Fourth International Conference on Machine Learning and Cybernetics*, Guangzhou, China, 18-21 August 2005, 2133-2138
[47] Tong P., An L.-S., Zhang J., Liu Y.-T., Research on Fault Diagnosis Method of Principal Components Analysis and D-S Evidence Theory, *Chinese Control and Decision Conference, CCDC '09*, 2009, 1601-1605
[48] Sun X., Marquez H.J., Chen T., Riaz M., An improved PCA method with application to boiler leak detection, *ISA Transactions*, 44 (2005), 379–397
[49] Pawlik M., Strzelczyk F., *Elektrownie (Power plants)* - in Polish, WNT, Warszawa, Poland, 2009