

Data pre-processing methods for NPP equipment diagnostics algorithms: an overview

Iurii D. Katser^{1,2}, Vyacheslav O. Kozitsin^{1,2}, Ivan V. Maksimov²,
Denis A. Larionov², Konstantin I. Kotsoev³

¹ Skolkovo Institute of Science and Technology, Bolshoy Boulevard 30 Bld. 1, Skolkovo Innovation Center, Moscow 121205, Russia

² Cifrum – Nuclear Industry Digitalization Support, Private Enterprise, Warszawskoye Highway 46 Floor 5 Office 529b, Moscow 115230, Russia

³ Bauman Moscow State Technical University; 2-ya Baumanskaya Street 5 Bld. 1, Moscow 105005, Russia

Corresponding author: Iurii D. Katser (Iurii.Katser@skoltech.ru)

Academic editor: Georgy Tikhomirov ♦ **Received** 27 January 2021 ♦ **Accepted** 12 May 2021 ♦ **Published** 22 June 2021

Citation: Katser ID, Kozitsin VO, Maksimov IV, Larionov DA, Kotsoev KI (2021) Data pre-processing methods for NPP equipment diagnostics algorithms: an overview. Nuclear Energy and Technology 7(2): 111–125. <https://doi.org/10.3897/nucet.7.63675>

Abstract

The main tasks of diagnostics at nuclear power plants are detection, localization, diagnosis, and prognosis of the development of malfunctions. Analytical algorithms of varying degrees of complexity are used to solve these tasks. Many of these algorithms require pre-processed input data for high-quality and efficient operation. The pre-processing stage can help to reduce the volume of the analyzed data, generate additional informative diagnostic features, find complex dependencies and hidden patterns, discard uninformative source signals and remove noise. Finally, it can produce an improvement in detection, localization and prognosis quality. This overview briefly describes the data collected at nuclear power plants and provides methods for their preliminary processing. The pre-processing techniques are systematized according to the tasks performed. Their advantages and disadvantages are presented and the requirements for the initial raw data are considered. The references include both fundamental scientific works and applied industrial research on the methods applied. The paper also indicates the mechanisms for applying the methods of signal pre-processing in real-time. The overview of the data pre-processing methods in application to nuclear power plants is obtained, their classification and characteristics are given, and the comparative analysis of the methods is presented.

Keywords

advanced analytics, data analysis, data pre-processing, diagnostics, NPP, machine learning, raw data

Introduction

Modern nuclear power plants (NPP) generate large amounts of data. The methods of intellectual analysis make it possible to apply the generated data for the purpose of detecting malfunctions, determining the operating lifetime of equipment and solving other urgent problems in NPP operation.

Such data contain valuable information about incipient faults, but it can be extremely difficult to use the so-called raw or unprocessed data in analytical algorithms. The algorithms of fault detection, pattern recognition, fault localization, prognosis of fault development, etc. require signal pre-processing for high-quality output. The pre-processing techniques include both machine learning methods (Bishop 2006, Hastie et al. 2009) and classical

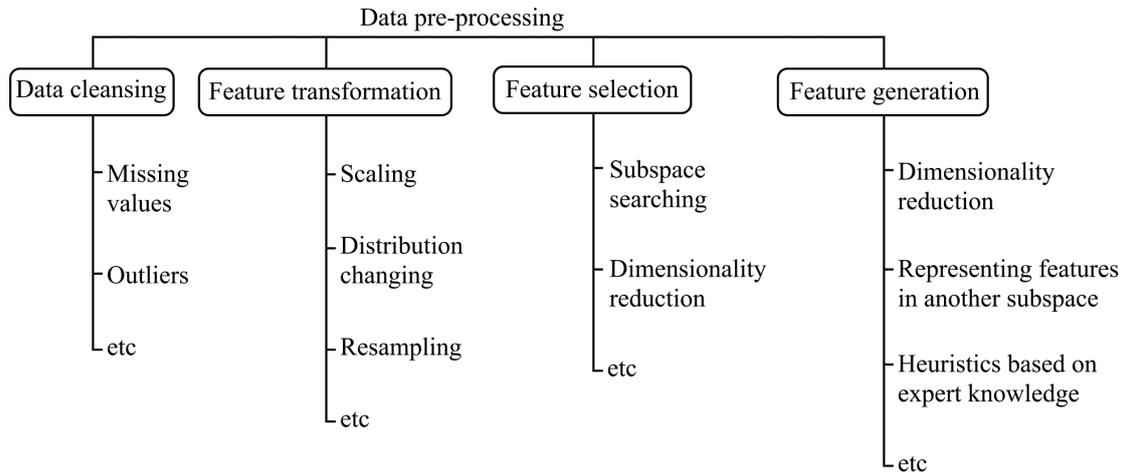


Figure 1. Taxonomy of data pre-processing methods.

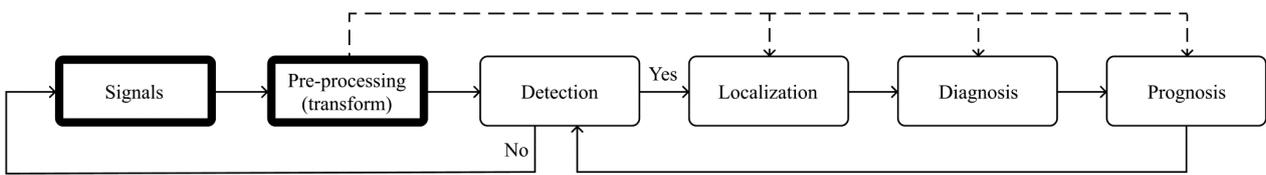


Figure 2. Equipment diagnostics loop.

signal processing methods (Chiang et al. 2001, Sergienko 2011). Modern diagnostic systems at NPPs use such pre-processing methods as spectral analysis, filtering, moving averages, generation of diagnostic features from recorded signals, and others. The academic literature on technical diagnostics has described the application of such methods for NPPs (Arkadov et al. 2004, 2019, 2020).

The pre-processing stage is very important in detection algorithms. Its relevance seems rather evident since it is an integral part to the overwhelming majority of the methods mentioned in this overview and other reviews of data processing methods (Venkatasubramanian et al. 2003a, 2003b, Qin 2009, Ma and Jiang 2011, Si et al. 2011, An et al. 2013, Dai and Gao 2013, Patel and Shah 2018). In Fig. 1 we propose the taxonomy of data pre-processing methods, which summarizes many such works.

Fig. 2 shows the flow diagram of equipment diagnostics according to GOST R ISO 13381-1-2016 (2017).

The main path of the equipment diagnostics is the sequential execution of all stages, starting with data acquisition, followed by pre-processing, fault detection, localization, diagnosis or root cause identification and prognosis of how the detected faults may develop. The dashed line indicates an auxiliary path of equipment diagnostics, in which the stages do not follow from one another. The auxiliary path can be taken either in deferred analysis when any stage is considered separately from the others; or when using the original data in its unprocessed form or adding new data at any stage; or in other pre-processing methods to prepare the original data and thus ensure algorithm operation.

It is necessary here to clarify some of the terms used in this article. The *offline* mode will refer to working with the full data sample; in this case full realization of the sig-

nals is available for analysis. The *online* mode will mean working in real time; in this case, the full data sample is unavailable for analysis, data objects (vectors) can arrive one after another as streaming data – hence, the analysis is called the *pointwise* analysis – or there can be a buffer with batch data – hence the analysis is called the *batch* analysis.

Supervised learning refers to tasks in which all the operating modes of equipment are known and the data classes are marked; in other words, the data on both the normal mode of operation and the abnormal mode of operation (preferably also on all types of abnormalities) are available. *Semi-supervised learning* refers to tasks in which only the data on normal mode of operation is available; this means that only the part of data describing normal operation of equipment has a class mark. *Unsupervised learning* refers to tasks in which there is no data on either normal or abnormal operation and no class marks for any data.

This article focuses on the Data and Pre-Processing stages, traced with heavy line in Fig. 2. It discusses the methods of signal pre-processing that help cleanse time series data and transform, isolate and select data features with respect to NPPs and other complex technical systems.

Data

An NPP may have tens of thousands of instrument channels (Akimov et al. 2015, Arkadov et al. 2019). These include approximately 3,000 temperature signals, 450 electrical signals, 4,700 binary input signals, and 3,200 pressure, level, consumption and other signals. In addition, monitoring, control and diagnostics systems generate a large amount of useful data and, in most cases, transmit only aggregated information to the Supervisory Control

And Data Acquisition (SCADA) system. Arkadov et al. (2020) distinguished the following main groups of raw data parameters:

- geometric quantities (measurements of length, position, angle of inclination, etc.);
- thermotechnical quantities (temperature, pressure, flow rate, volume of working fluid);
- electrical quantities (current, voltage, power, frequency, induction, etc.);
- mechanical quantities (deformation, forces, torques, vibration, noise level, etc.);
- chemical composition (concentration, chemical properties, etc.);
- physical properties (humidity, electrical conductivity, viscosity, radioactivity);
- parameters of ionizing radiation (radiation fields inside and outside of zoned fluxes of neutrons and gamma radiation);
- other parameters.

Most of the generated and aggregated signals relate to the raw data and represent time-series type of data. Asynchronous generation and acquisition of data present a problem in data analysis. Malfunctions of measurement channels result in data omissions, inaccurate readings and noise contamination. Moreover, self-monitoring or self-diagnostic systems of measuring equipment can either detect invalid values or skip them. However, various pre-processing methods make it possible to minimize the impact of such factors on the quality of technical diagnostics.

Data Pre-Processing

In general, the Pre-Processing stage consists of the four main steps shown in Fig. 1: Data Cleansing, Feature Transformation, Feature Engineering and Feature Selection. The following sub-sections give a more detailed account of each step.

Data cleansing

The Data Cleansing helps eliminate invalid values and outliers by removing or correcting them. At this stage, either the missing data are filled in, or the data objects containing such gaps are deleted if their share is small. The features with a large number of data gaps or invalid values can also be excluded from further analysis.

All measurements affecting NPP safety should be promptly diagnosed and marked by a validity indicator (Arkadov et al. 2019) that shows the degree of information reliability. It allows eliminating invalid data in the SCADA. However, not all measurements come with reliable self-monitoring. There is a growing body of studies that aim at solving the problem of diagnosing the measuring equipment and controlling the reliability of measure-

ments, for example (Zavaljevski and Gross 2000, Li et al. 2018a, 2018b, 2018c, 2019, Arkadov et al. 2020).

Data gaps appear due to the imperfection of modern measuring systems, communication channels and other infrastructure. This poses a problem when working with anomaly detection methods and other techniques. The simplest approaches here are to ignore features with gaps or replace the gaps with specially assigned values, for example, 0 or -1 . Also, missing values can be filled in by standard methods, such as the moving average or median over the selected window; the average (quantitative characteristic), mode (categorical characteristic) or median value over the entire time series; and the last value obtained before the gap. Alternatively, there are advanced methods to fill in missing data, for example, the machine-learning methods (for regression, see Honghai et al. (2005); for nearest neighbor method, see Batista and Monard (2002), Jonsson and Wohlin (2004); for neural networks, see Gupta and Lam (1998); for k-means and fuzzy k-means method, see Li et al. (2004), etc.) Batista and Monard (2003) and Wohlrab and Fürnkranz (2009) compared different gap filling procedures. Zagoruyko (1999) and Marlin (2008) gave reviews of gap-filling techniques with different approaches.

To tackle the problem of outliers, one can either apply conventional methods, for example, remove values that contradict the laws of physics or fail to meet the standard deviation of a feature, or resort to modern methods of data mining and machine learning. However, in most cases, the problem of finding anomalies in data is an unsupervised learning task and hence it is suggested to use the class of unsupervised learning methods. In his textbook on models for detecting outliers and anomalies, Aggarwal (2015) identified six main approaches, each corresponding to a class of models:

1. extreme value analysis;
2. clustering;
3. distance-based models;
4. density-based models;
5. probabilistic models;
6. information-theoretical models.

Zhao et al. (2019a) described the PyOD library, which includes twenty outlier detection methods, for the Python Programming Language.

Another approach to solving the problem of outlier detection is the use of ensembles (Aggarwal 2013, Aggarwal and Sathe 2015, Aggarwal and Sathe 2017, Zhao et al. 2019b). Ensembles are based on sequential or parallel application of a single base algorithm or a set of base algorithms to data subsamples or feature subspaces, with the following evaluation of the resulting response sets. Gradient boosting, random forest, bagging and some other common methods are founded on building such ensembles.

Turning now to support vector machines (SVM), there are two principal SVM-based methods for detecting anomalies in data (Scholkopf et al. 2000). The first

one, One-Class Support Vector Machine, is used to detect novelties (Scholkopf et al. 2000) and anomalies (Amer et al. 2013) in data. The idea behind this method is to apply such a transformation of the feature space that in the new space all the objects and the hyperplane, separating them from the origin of coordinates, lie as far as possible from the origin. Zhang et al. (2009) presented the online application of One-Class Support Vector Machine for outlier detection. The second one is Support Vector Data Description (Tax and Duin 2004). It transforms the feature space and then draws a boundary sphere around the data, pulling the maximum number of objects inside the sphere and keeping its radius as small as possible. Note that Support Vector Data Description is sometimes referred to as the SVM-based one-class classifier, and it causes confusion of the two methods. These methods are computationally complex and often show weak results, though the advantage is its clear mathematical and statistical base.

Isolation Forest, or iForest, identifies outliers by the low depth of outlying values in the constructed tree (Liu et al. 2018). The method cannot be applied to streaming data in real time, since building a tree and selecting outlying values require data sample. Tan et al. (2011) and Ding and Fei 2013 gave examples of the algorithm operation in the online mode with a buffer. The advantage of the method is low computational complexity and the ability to work with heterogeneous input data. The disadvantage is the inability to work with data as with a time series – they are perceived as a non-temporal set of states or instances.

Cluster analysis is the process of categorizing a set of objects into groups (clusters) so that objects in one group are similar by some of the attributes. The study by Jiang et al. (2001) was one of the first to employ cluster analysis to detect outliers in data. Breunig et al. (2000) examined the degree of being an outlier, called the Local Outlier Factor (LOF), depending on the point density. In a follow-up study, He et al. (2003) presented the Cluster-Based Local Outlier Factor and an outlier detection algorithm based on cluster analysis. Such algorithms as ROCK (Guha et al. 2000) and DBSCAN (Ester et al. 1996, Duan et al. 2009) are able to detect outliers, but these algorithms regard noise, i.e. objects that are not assigned to any selected cluster, as outliers. Loureiro et al. (2004), Pachgade and Dhande (2012) and many other studies also report attempted approaches and algorithms for detecting outliers in data using cluster analysis. As for the initial data for clustering algorithms, both initial signals and diagnostic features generated from them can be used. In addition, generally, clustering algorithms have no requirements for input data, which is one of the advantages of this method. A disadvantage is the use of heuristics in most clustering methods at various stages of solving the problem.

Katser et al. (2019) give a more detailed description of One class Support Vector Machine, Isolation Forest, and cluster analysis in terms of detecting data anomalies and equipment faults.

Let us now consider minimum covariance determinant (MCD), another method to control outliers in data

(Rousseeuw 1984). Its objective is to find the data subsample whose covariance matrix has the lowest determinant. Thus, when calculating the covariance matrix, the values that are considered to be outliers, get excluded. It improves the quality of problem solving by finding the covariance matrix (Principal Component Analysis, Independent Component Analysis, etc.). The FAST-MCD algorithm, developed for the purpose of quick search of the exact subsample, selects at least half of the observations from the total pool, making an acceptable number of operations, which allows using the method in practice (Rousseeuw and Driessen 1999).

Hubert and Debruyne (2009) presented the advantages, disadvantages, limitations and examples of MCD application in various fields. Similarly, Hardin and Rocke (2004), Fauconnier and Haesbroeck (2009) and Leys et al. (2018) examine the application of this outlier detection method to solve some practical problems. It can also be used for fault detection, for example, in conjunction with Independent Component Analysis (Cai and Tian 2014).

Feature transformation

At the Feature Transformation stage, the transformation affects the features values (scaling, change in the sampling rate), their type (categorization of discrete and continuous values), modality (videos are converted into a sequence of pictures, pictures into tables of numerical data), etc.

Most of the pre-processing algorithms require input data, the features of which are on the same scale, since the mean value and variance of features impact their significance for algorithms (Bishop 2006, Hastie et al. 2009). Among numerous scaling methods, the most common ones are the following (Shalabi et al. 2006):

- Linear:
 - Z-Normalization/Standardization: normalizing the mean to 0, standardizing the variance at 1;
 - Min-Max Normalization: rescaling the range of features to bring data to the 0 to 1 scale, with zero corresponding to the minimum value before normalization and one corresponding to the maximum;
 - Normalization by Decimal Scaling: to moving the decimal point of values of feature, take i digits of the maximum value of a time series and divide each value of the series by 10^i ;
 - MaxAbs scaling: normalizing each time series value to the maximum absolute value of the entire series;
- Non-linear:
 - Hyperbolic tangent: scaling the values to $[-1 \dots 1]$;
 - Logistic (sigmoid) function: scaling the values to $[1 \dots 0]$.

In addition to scaling, the Box-Cox transformation (taking of logarithm) is often applied to features (Sakia 1992) to make the distribution of features similar to nor-

mal. The transformation can be applied multiple times but only to positive values.

Another important problem is to bring signals with different sampling rates to a single one. In their monograph, Arkadov et al. (2020) described the main approaches in its chapter Combining Measurement Information of Different Systems:

- reducing the sampling rate of all processes to the minimum;
- increasing the sampling rate of all processes to the maximum;
- converting to an intermediate or any other sampling rate.

The choice of a specific rate, which all signals must be converted to, should be based on the characteristic rate of the analyzed process and be consistent with the subsequent stages of diagnostics. A significant decrease in the rate can lead to the loss of information in the signals while an unreasonable increase in the rate can affect the computational complexity of subsequent data analysis processes.

Arkadov et al. (2020) outlined the conditions of applicability, advantages and disadvantages of the approaches but only to the extent of spectral analysis. It is worth supplementing the chapter with several observations:

Firstly, now that the machine learning methods are gaining popularity, including due to the ability to work with Big Data, sometimes it pays to bring signals to a low frequency to reduce the total computational complexity of the problem. It also may be necessary to reduce the sampling rate if the set of sequentially applied methods is large, to be able to solve problems in real time.

Secondly, the monograph missed an important point of applying the above approaches in the real time mode. Since interpolation is not applicable in real time mode (in the pointwise analysis) and extrapolation is complex and rarely used, simpler methods can deliver the reduction to a single sampling rate, namely:

- increasing the sampling rate by filling the current range in with the last received value with subsequent sampling;
- increasing the sampling rate by filling in the average or median value at the last range with subsequent sampling;
- decreasing the sampling rate by selecting extrema, mean or median values in the range.

Feature selection and generation

Feature selection can be generally understood as declining in the number of features, for example, by searching for a subspace of a lower dimension using dimensionality reduction methods or by simply discarding a part of uninformative features. Feature selection simplifies models, reduces the complexity of the models problem training, and helps avoid the curse of dimensionality.

Zagoruyko (1999), Bishop (2006) and Hastie et al. (2009) reflected on the problem of selecting a system of informative features and the variety of methods for that purpose. According to these authors, the most common algorithms are as follows:

- complete rummage of all the feature sets;
- sequential feature selection of features (Add);
- sequential feature elimination (Del);
- genetic algorithm;
- random search;
- clustering of features.

Well-known extensions of some of these algorithms like SHAP (Lipovetsky et al. 2001) and LIME (Ribeiro et al. 2016) are successfully used nowadays for interpreting machine learning model predictions measuring feature importance. A variety of such methods are shown by Lundberg et al. (2017) in their work and references therein.

Regularization, which imposes a penalty the complexity of the model, is often applied to machine learning problems (Bishop 2006, Hastie et al. 2009). The L1 regularization and the least absolute shrinkage and selection operator (LASSO; see Tibshirani (1996)) solves the problem of feature selection, by excluding some of the original uninformative features from the subsample used for training and operation of the model.

Feature generation is possible if based on the logic and physics of the process or on standard transformations, i.e. raising to the polynomial power or performing multiplication on feature values. Engineering of new diagnostic features is also the acquisition of signal auto-features by using a sliding buffer and all kinds of correlating pairs, and other rather trivial transformations. In respect to NPPs, they are discussed in the monographs by Arkadov et al. (2004, 2018, 2020).

Most techniques of dimensionality reduction solve both the problem of reducing the number of features and the problem of engineering new diagnostic features. The techniques of dimensionality reduction project data into a lower-dimensional space and, unlike selection methods, considers all the original information, thus making it possible to simplify and improve the procedure for monitoring and searching for anomalies in signals. The dimensionality reduction problem has many applications (Chiang et al. 2001). A notable example of using the dimensionality reduction is visualization, i.e. representing a dataset in a two- or three-dimensional space.

Principle Component Analysis (PCA) is a widely used technique for reducing the dimensionality of datasets. The idea of the method is to search for a hyperplane of a given dimensionality in the original space with the subsequent projection of the data onto the found hyperplane. The axes of the new space are a linear combination of the original ones and get selected based on the variance of the original features. The transformation of the measurement space into a new orthogonal space is performed by bringing the covariance (correlation) matrix to a diagonal form;

for this reason, the original features in the new space are uncorrelated. Li et al. (2018a, b, c, 2019) and Ayodeji et al. (2018) studied applications of Principal Component Analysis for signal pre-processing and feature generation in problems of diagnosing equipment and sensors.

Independent Component Analysis (ICA), unlike Principle Component Analysis, finds a space in which the original features are not only uncorrelated, but also independent in terms of statistical moments of a higher order. In other words, Independent Component Analysis solves the problem of finding any, including non-orthogonal, space where the axes are a linear combination of the original ones. The goal is to transform the original signals so that in the new space they would be statistically independent from each other as much as possible (Kano et al. 2003, Lee et al. 2004a).

Both PCA and ICA build transformations into a new space only based on the matrix of features, without taking into account the response vector. This solves the problem of the mutual dependence of features, but fails to tackle the presence of features that do not affect the target variable (response vector). That is why such features are used in further analysis.

Compared to PCA where the axes of the new space are selected based on the variance of the original features, the Partial Least Squares (PLS) method, or Projection to Latent Structures, selects the axes of the new space proceeding from the maximization of the covariance between the matrix of features and the matrix of responses. At that, new spaces are found for both matrices. The new axes for the feature space are calculated to provide the maximum variance along the axes in the new space for the matrix of responses. Using the data on equipment faults as responses, one can obtain a lower-dimensional space for the matrix of feature and hence more accurately determine various faults (MacGregor and Kourti 1995, Chiang et al. 2001, Wang et al. 2003, Ma and Jiang 2011).

The application of the PLS method is limited due to the need to know the classes of events (faults) when training the model. For that reason, the method is often used at the pre-processing stage when solving the problem of making a diagnosis or determining the causes.

The wide applicability of these techniques is explained by the fact that they can tame multidimensional, noisy data with correlated parameters by translating the data into a lower-dimensional space that contains most of the Cumulative Percentage Variance of the original data (Chiang et al. 2001, Jiang and Yan 2014, Xu et al. 2017). However, the standard PCA, ICA and PLS methods can only find linear relationships of features and sometimes fail to solve problems efficiently enough. Hence, there appeared a number of modifications improving them:

- kernel methods: for PCA, see Lee et al. (2004a) and Choi and Lee (2004); for ICA, see Zhang and Qin (2007); for PLS, see Zhang et al. (2010), Zhang and Hu (2011), Jiao et al. (2017). Unlike the linear methods of dimensionality reduction, the non-linear

ones produce an effective dimensionality reduction due to the creation of a non-linear combination of features to create a new lower-dimensional space;

- dynamic methods: for PCA, see Ku et al. 1995, Russell et al. (2000); for ICA, see Lee et al. (2004b); for PLS, see Chen and Liu (2002). The dynamic methods, used for analysis of transient phenomena, supplement the studied sample with a certain number of previous observations and factor in autocorrelations and cross-correlations with displacements in time;
- probabilistic methods: for PCA, see Tipping and Bishop (1999), Kim and Lee (2003); for ICA, see Zhu et al. (2017); for PLS, see Li et al. (2011). The probabilistic methods model the data distribution as a multivariate Gaussian distribution. With PPCA, it is possible to construct a PPCA mixture model, which consists of several local PPCAs and detects faults in data with multimodal or complex non-Gaussian distributions (Ge and Song 2010, Raveendran and Huang 2016, Raveendran and Huang 2017);
- Sparse Principal Component Method (Sparse PCA), which has appeared only recently, takes only a part of the original features to construct a new lower-dimensional space. Gajjar et al. (2018) presented its application for fault detection;
- dynamic kernel PLS technique and a brief overview of works on PLS modifications were presented by Jia and Zhang (2016).

Linear Discriminant Analysis (LDA), or Fisher Discriminant Analysis (FDA), is a statistical analysis method that searches for a linear combination of features able to separate events from different classes (determining different faults) in the best way possible (Chiang et al. 2001). It is used for the problems of classification and dimensionality reduction of the original feature space. de Lazaro et al. (2015) demonstrated that the kernel LDA (FDA with kernels in Mika et al. (1999)) showed better results as compared to the kernel PCA. By analogy with the above methods, the probabilistic version of LDA was developed and presented by Prince and Elder (2007). The method has proven itself well in many fields, including nuclear industry (Garcia-Allende et al. 2008, da Silva Soares and Galvao 2010, Jamil et al. 2016, Cho and Jiang 2018), but it has the same limitation as the PLS method: it requires the vector of responses that often does not exist in practice.

Canonical Correlation Analysis (CCA), or Canonical Variate Analysis (CVA) is a technique of searching for lower-dimensional spaces for two sets of variables (features and responses) when projecting the data in which the cross-correlations between the two sets of variables are maximal among all possible variants of spaces (Chiang et al. 2001, Hardoon et al. 2004, Manly and Alberto 2016). The basis of the variables in the new space is a linear combination of the original variables. CCA is used as a method of dimensionality reduction but it can also be applied to informative feature selection (Kaya et

al. 2014). Chen et al. (2016b, 2016c) used CCA to monitor industrial processes, and Chen et al. (2018b) applied a modification of this technique for monitoring processes with a non-Gaussian distribution. CCA is similar to PLS and LDA by the need to resort to a response vector (Chiang et al. 2001).

Factor Analysis is a multivariate statistical analysis that serves to determine the relationship between variables and reduce their number (Harman 1976, Kim 1989, Warne and Larsen 2014, Manly and Alberto 2016). It is based on the assumption that known variables depend on fewer unknown variables and random error. This allows using Factor Analysis to replace correlated measurements with a smaller number of new variables (factors), although losing a small amount of information contained in the original data. Another requirement is to represent the factors in terms of the original variables. The factor itself is interpreted as the cause of the joint variability of several original variables. The main difficulty in Factor Analysis is the selection and interpretation of the principal factors.

Feature bagging, or bootstrap aggregation, is a learning method that searches through randomly selected feature subsamples from $n/2$ to $n - 1$ from the number of original n features and uses the basic algorithm on each subsample, and after that all results are aggregated by summation or another method (Breiman 1996). Feature bagging allows improving the performance of algorithms, for example, classification accuracy (Bryll et al. 2003). Lazarevic and Kumar (2005) provided an algorithm to solve the problem of detecting outliers in data with examples. Aggarwal and Sathe (2015) proposed a modification of the algorithm that reduces the dependence of the basic algorithms on themselves.

Bagging in combination with basic algorithms turns the problem solution into an ensemble of algorithms, increasing the computational complexity of the basic algorithms but improving the accuracy and robustness of the results. If all features are independent and important, bagging often degrades the quality of responses as each algorithm has an insufficiently informative subsample to learn.

Neural networks are also used for data processing and dimensionality reduction. Today, one of the most effective methods for the latter purpose is an autoencoder – a type of artificial neural network applied to encode data, usually in unsupervised learning (Bourlard and Kamp 1988, Sakurada and Yairi 2014, Chen et al. 2016a, Chalapathy et al. 2017). Each subsequent layer of the autoencoder up to the middle layer – the bottleneck – nearly always has fewer neurons than the previous one. Time series can be input to the network, and the main requirement to them is preliminary data normalization. An autoencoder aims to learn a representation for data in another subspace, usually for a dimensionality reduction problem. An autoencoder learns to reduce the dimensionality of the feature space of the data, received at the network input, to a specified number of features, and then to decode the compressed data back to a representation that most closely matches

the original data. Thus, the original data is supplied to the input and output of the neural network, and at each training iteration (epoch), the error between the original data and the output data is minimized.

In addition to feed-forward networks, there are a large number of modernized architectures; some of them are as follows:

- convolutional autoencoders whose architecture includes a convolutional layer that creates a convolutional kernel for the convolution of input data by one feature. It is used for data noise removal (Grais and Plumbley 2017), clustering (Chen 2015, Ghasedi et al. 2017), fault detection (Chen et al. 2018a) and other purposes;
- Recurrent Neural Network (RNN) based Autoencoders and their varieties (Elman 1990, Chung et al. 2016), such as Long Short-Term Memory (Hochreiter and Schmidhuber 1997) and Gated Recurrent Units (Chung et al. 2014);
- Variational Autoencoders (VAE), by studying the probability distributions that simulate the input data, allow the hidden-variables model to learn (Everett 2013). For more details on VAE architecture and applications, refer to Kingma and Welling (2013), Doersch (2016).

Autoencoders can be used jointly with standard fault detection methods, for example, with statistical detection criteria (Yang et al. 2015, Xiao et al. 2017). A high degree of compression of the original data, due to finding complex non-linear dependencies, and the possibility of architecture upgrade, for example, in order to remove noise (Vincent et al. 2008), are the advantages of the above neural networks, but it is worth noting the computational complexity of the algorithms and the complexity of the models tuning. Generally, neural networks, especially deep ones, are considered as techniques that extract useful features automatically. And sometimes, it is an advantage over classical machine learning and other approaches, where feature extraction is often a manual and laborious part of work. Even though this advantage of neural networks increases the quality of the model and final results by extracting more complex nonlinear features, it can also be considered a disadvantage due to a lack of knowledge of how the feature is extracted. So, data scientists mostly can't reproduce the logic of how the feature is pulled out from the original subset and what intuition and physics are behind. The popularity of this field of knowledge has grown recently. Here we recommend selecting either the problem solving quality is important or the transparency in the feature extraction and modeling processes is important.

Spectral Analysis includes time series processing associated with obtaining a representation of signals in the frequency domain. The main application of Spectral Analysis is to assess the vibration of equipment. The most popular techniques of spectral processing are the Fourier transform, the Laplace transform, the Hilbert transform

and the Hilbert-Huang transform. The results of Spectral Analysis are rather easy to interpret, and it is possible to detect faults, determine the nature of their occurrence and make a diagnosis on their basis. Arkadov et al. (2004, 2018, 2020) described the application Spectral Analysis to NPP diagnostics in detail. As for non-stationary time series, time-frequency analysis is widely used to detect malfunctions in rotary equipment under time-varying operating conditions. Kim et al. (2007) provided a comparative analysis of the windowed Fourier transform, the Wigner-Ville distribution, and the wavelet transform.

Another tool of fault detection can be to generate diagnostic features that serve as equipment health indicators. Such diagnostic features that characterize the system condition, are identified by an expert based on their experience for a clear and effective understanding of the state of a technical system and, accordingly, for detecting anomalies in operation (Leskin et al. 2011, Costa et al. 2015, Baraldi et al. 2018, Arkadov et al. 2020). In effect, principal components in PCA, bottleneck features in an autoencoder, and the Fourier spectrum in a signal are the diagnostic features, but the main distinction of equipment health indicators is that they are formulated in a purely heuristic way. An expert builds the equipment health indicators upon processing and formalization of a pattern of regularities that are not described by known physical and mathematical models of equipment.

The advantages of the diagnostic features approach include the possibility of creating a rational solution that accumulates experts' experience, and the ease of health indicator implementation. The disadvantages are the lack of physical or mathematical models that could form the foundation of the method, and its limitations for, as a rule, an indicator points only to malfunctions of the same kind in one unit of equipment.

Time series data augmentation

The problem of lacking time series data leads to the inapplicability of deep learning algorithms in some applications. In such cases, augmentation or data generation is used for adding more synthetic data for better training and working of machine learning algorithms. Though quite a bit of attention is paid to this field of knowledge, the surveys by Ivana et al. (2020) and Wen et al. (2021) highlight the state of this research field. The latter work provides the following taxonomy for time series data augmentation:

1. Basic approaches:
 - a. Time domain;
 - b. Frequency domain;
 - c. Time-frequency domain.
2. Advanced approaches:
 - a. Decomposition Methods;
 - b. Statistical Generative Models;
 - c. Learning Methods (including Embedding Space, Deep Generative Models, and Automated Data Augmentation).

Although data augmentation is quite a useful tool for improving the quality of various models, it mainly relates to the training stage. Data augmentation almost never is being a part of the equipment diagnostics pipeline. Moreover, time series data augmentation methods are not appropriately researched for real-world industrial data with noise and possible various statistical changes happening all the time.

Online application of pre-processing methods

Each pre-processing method has its own distinctive nature in relation to the original data: some are capable of working with one data object while others require the calculation of values based on a learning sample or a buffer. Moreover, real-time pre-processing must match the diagnostics model selected for learning; otherwise, the models may give incorrect results. For such cases, it is worth discussing the mechanisms for applying pre-processing methods:

- The pointwise transformation in learning and operation. This mechanism is used when the applied pre-processing methods require a state vector only at the current time. Examples of such transformations are deleting data exceeding a certain (for example, physically justified) threshold, raising a feature to the polynomial power, performing multiplication on feature values, etc.
- Complete or *batch* transformation during learning, pointwise transformation during operation. This mechanism is used when the transformation requires the calculation of values, for example, the mean or the variance of a learning sample. The values obtained at the learning stage are saved and applied in real-time operation for each new state vector. Examples of such transformations are One-Class SVM, iForest, MCD, PCA and all linear methods for reducing features to a single scale mentioned in this article.
- *Batch* transformation. It refers to the transformation of features based on the calculation of characteristics using a sliding window or a batch. An example here is calculating a moving average of a signal per a window or obtaining auto-characteristics of signals using a sliding buffer and all kinds of correlated pairs.

Let us demonstrate how methods are applied in real-time mode, assuming that our preprocessing pipeline consists of the following steps:

1. Moving average for gaps filling;
2. Z-Normalization;
3. PCA applying;

1. Selecting the first principal component for further comparison with the threshold for anomaly detection.

First of all, the new point for multivariate time series is received. Then the average value for the window with previous points is calculated if some of the values in the novel vector are missing. Into the gaps, calculated points are inserted. After that, Z-normalization is applied using previously (during the training stage, commonly, for fault-free mode) defined mean and standard deviation values. Afterward, PCA is applied using a transformation matrix calculated for the train set. Finally, the value over the first principal axis is selected for further comparison.

Conclusion

This overview has described the peculiarities of the data collected at NPPs and its pre-processing in real time. Table 1 summarizes the methods of data pre-processing, carried out before solving the main problem of diagnostics.

The problems encountered in data are not unique to the nuclear industry, but the outstanding aspect of NPPs

is the large amount of generated information, the variety of its sources and data types. Pre-processing is necessary to prepare the data for input to the diagnostic algorithms, since many of them either have requirements that rule out the input of data with gaps, outliers, signals with different sampling rates, or produce incorrect results when working with unscaled data. Another reason for using pre-processing methods is the possibility of improving the quality of the diagnostic algorithms and reducing the computational complexity of the problem, for example, by reducing the dimensionality of the initial data or lowering the sampling frequency of signals.

We find it necessary to give a summary with providing our opinion on which methods are commonly used, which are not, and why:

- When filling in gaps, the most intuitive way is to use specially assigned values to avoid generating false information about the data. But not all machine learning methods can process such values properly. That is why the most common techniques fill the gaps with some data characteristics from moving windows or over the whole signal realization.

Table 1. Characteristics of data pre-processing methods

| Item | Method | Data input limitation | Problem type | Univariate/Multivariate | Online | References |
|----------------------------------|------------------|----------------------------|-----------------|-------------------------|--------|--|
| Data Cleansing | | | | | | |
| 1 | One-class SVM | Normalization | Unsupervised | +/+ | + | Scholkopf et al. 2000 Tax and Duin 2004 Zhang et al. 2009 Amer et al. 2013 |
| 2 | iForest | Normalization | Unsupervised | +/+ | –* | Tan et al. 2011 Ding and Fei 2013 Liu et al. 2018 |
| 3 | Cluster analysis | Equidistant* | Unsupervised | +/+ | + | Ester et al. 1996 Breunig et al. 2000 Guha et al. 2000 Jiang et al. 2001 He et al. 2003 Loureiro et al. 2004 Duan et al. 2009 Pachgade and Dhande 2012 |
| 4 | MCD | Normalization, equidistant | Unsupervised | –/+ | + | Rousseeuw 1984 Rousseeuw and Driessen 1999 Hardin and Roche 2004 Fauconnier and Haesbroeck 2009 Hubert and Debryne 2009 Leys et al. 2008 Cai and Tian 2014 |
| Feature Selection and Generation | | | | | | |
| 5 | PCA | Normalization, equidistant | Semi-supervised | –/+ | + | Ku et al. 1995 Tipping Bishop 1999 Russell et al. 2000 Kim and Lee 2003 Choi and Lee 2004 Lee et al. 2004a Ge and Song 2010 Raveendran and Huang 2016, 2017 Ayodeji et al. 2018 Gajjar et al. 2018 Li et al. 2018a, 2018b, 2018c, 2019 |
| 6 | ICA | Normalization, equidistant | Semi-supervised | –/+ | + | Kano et al. 2003 Lee et al. 2004b, 2004c Zhang and Qin 2007 Zhu et al. 2017 |

| Item | Method | Data input limitation | Problem type | Univariate/Multivariate | Online | References |
|------|-------------------|----------------------------|-----------------|-------------------------|--------|---|
| 7 | PLS | Normalization, equidistant | Unsupervised | -/+ | + | MacGregor and Kourti 1995 Chiang et al. 2001 Chen and Liu 2002 Wang et al. 2003 Zhang et al. 2010 Li et al. 2011 Ma and Jiang 2011 Zhang and Hu 2011 Jiao et al. 2017 |
| 8 | LDA, FDA | Normalization, equidistant | Unsupervised | -/+ | + | Mika et al. 1999 Chiang et al. 2001 Prince and Elder 2007 Garcia-Allende et al. 2008 da Silva Soares and Galvao 2010 de Lazaro et al. 2015 Jamil et al. 2016 Cho and Jiang 2018 |
| 9 | CCA, CVA | Normalization, equidistant | Unsupervised | -/+ | + | Chiang et al. 2001 Hardoon et al. 2004 Kaya et al. 2014 Chen et al. 2016b, 2016c, 2018b Manly and Alberto 2016 |
| 10 | Factor analysis | Normalization, equidistant | Semi-supervised | -/+ | + | Harman 1976 Kim JO 1989 Warne and Larsen 2014 Manly and Alberto 2016 |
| 11 | Spectral analysis | Stationarity, equidistant* | Unsupervised | +/- | + | Arkadov et al. 2004, 2018, 2020 Kim et al. 2007 |
| 12 | Bagging | - | Unsupervised | -/+ | - | Harman 1976 Kim JO 1989 Breiman 1996 Bryll et al. 2003 Lazarevic and Kumar 2005 Warne and Larsen 2014 Manly and Alberto 2016 |
| 13 | Autoencoder | Normalization | Semi-supervised | +/+ | +* | Bourlard and Kamp 1988 Elman 1990 Hochreiter and Schmidhuber 1997 Vincent et al. 2008 Everett 2013 Kingma and Welling 2013 Chung et al. 2014, 2016 Sakurada and Yairi 2014 Chen 2015 Yang et al. 2015 Chen et al. 2016a Doersch 2016 Chalapathy et al. 2017 Ghasedi Dizaji et al. 2017 Grais and Plumbley 2017 Xiao et al. 2017 Chen et al. 2018a |
| 14 | Health indicators | -* | Unsupervised* | +/+ | + | Leskin et al. 2011 Costa et al. 2015 Baraldi et al. 2018 Arkadov et al. 2020 |

Note: * cannot be applied to some models in the method.

Machine learning techniques are quite rare and situational for such problems.

- As for outliers and impossible values detection, the most straightforward approaches to detecting values that contradict the laws of physics are the most popular ones due to the transparency of such rules for engineering personnel. Searching for deviation from some statistical characteristics, even utilizing machine learning techniques, is still fighting for attention. They are primarily used in retrospective analysis or in diagnostic systems that provide rec-

ommendations for operating personnel but not in critical safety systems.

- When transforming the data, Z-Normalization and Min-Max scaling are the most common scaling techniques because in the overwhelming majority of cases they show better results. Moreover, other methods are used when they are required for some specific reason for further analysis. Box-Cox transformation and other techniques like deriving the data are situational and used when further research requires working with normally distributed data or stationary time-series.

- A lack of sample rate for the signal or various sample rates is a frequent problem for industrial data. When selecting a unified sample rate, achieving a trade-off between the loss of information and computational complexity is vital. At the same time, the choice of a specific rate should be based on the characteristic rate of the analyzed process. When increasing the sample rate in the real-time mode, filling the current range with the last received value is the most common technique. When decreasing, both extrema and mean/median values are commonly used.
- For feature selection, a thorough analysis combining with various mentioned algorithms works the best. Analysis may also include finding dependencies of target vector from features when the problem is supervised. One of the most common ways is fitting some simple model, calculating feature importance for this model, and then selecting the most important features for fitting a more complex model. Regularisation is also commonly used when applicable. Among dimensionality reduction techniques, PCA is the most popular since it is unsupervised and provides linear transformation easy-to-understand and transparent for personnel. Although nonlin-

ear techniques, including neural networks, show state-of-the-art results, they lack interpretability of how transformation is constructed, making the approaches not popular in industrial applications.

- Feature generation in real-world applications is primarily based on the logic and physics of the process resulting in heuristical health indicators and various meaningful characteristics from spectral analysis.

The methods described in this work have already successfully proven themselves in industrial application, including at NPPs. At the same time, these methods continue to develop, and there appear supplements that improve their operation or expand their field of application. This overview, together with Katser et al. (2019), gives a sufficiently complete understanding of how the process at an NPP can be monitored from the moment of pre-processing of the collected data to the moment of solving the first diagnostic problem, i.e. detecting equipment malfunction.

Further research can be focused on over-viewing the methods used to solve such diagnostic problems at NPPs as arriving at the correct diagnosis, fault localization, and prognosis of the malfunction development.

References

- Aggarwal CC (2013) Outlier ensembles. *ACM SIGKDD Explorations Newsletter* 14: 49–58. <https://doi.org/10.1145/2481244.2481252>
- Aggarwal CC (2015) Outlier Analysis. In: *Data Mining: The Textbook*. Springer, Cham, 237–263. https://doi.org/10.1007/978-3-319-14142-8_8
- Aggarwal CC, Sathe S (2015) Theoretical foundations and algorithms for outlier ensembles. *ACM SIGKDD Explorations Newsletter* 17(1): 24–47. <https://doi.org/10.1145/2830544.2830549>
- Aggarwal CC, Sathe S (2017) *Outlier Ensembles: An Introduction*. Springer, 275 pp. <https://doi.org/10.1007/978-3-319-54765-7>
- Akimov NN, Bibikov VV, Koltsov VA, Lotov VN (2015) Automated process control system of the Belarusian NPP. *Doklady BGUIR* 2: 9–12. [in Russian]
- Amer M, Goldstein M, Abdennadher S (2013) Enhancing one-class support vector machines for unsupervised anomaly detection. In: *ODD '13: Proceedings of the ACM SIGKDD Workshop on Outlier Detection and Description*. Nineteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, IL (USA), August 2013. Association for Computing Machinery, New York, 8–15. <https://doi.org/10.1145/2500853.2500857>
- An D, Choi JH, Kim NH (2013) Options for prognostics methods: A review of data-driven and physics-based prognostics. In: *Proceedings 54th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, Boston, MA (USA), April 2013. American Institute of Aeronautics and Astronautics, 1940. <https://doi.org/10.2514/6.2013-1940>
- Arkadov GV, Egorov SV, Katser ID, Kovalev EV, Kozitsin VO, Maksimov IV, Pavelko VI, Slepov MT (2019) Predictive Analytics and Diagnostics of NPPs. *Diaprom Publ.*, Moscow, 72 pp. [in Russian]
- Arkadov GV, Pavelko VI, Finkel BM (2020) *Diagnostic Systems of WWER*. Jenergoatomizdat Publ., Moscow, 391 pp. [in Russian]
- Arkadov GV, Pavelko VI, Slepov MT (2018) *Vibroacoustics as Applied to the VVER-1200 Reactor*. Nauka Publ., Moscow, 391 pp. [in Russian]
- Arkadov GV, Pavelko VI, Usanov AI (2004) *VVER vibration diagnostics*. Jenergoatomizdat Publ., Moscow, 344 pp. [in Russian]
- Ayodeji A, Liu Y-k, Xia H (2018) Knowledge base operator support system for nuclear power plant fault diagnosis. *Progress in Nuclear Energy* 105: 42–50. <https://doi.org/10.1016/j.pnucene.2017.12.013>
- Baraldi P, Bonfanti G, Zio E (2018) Differential evolution-based multi-objective optimization for the definition of a health indicator for fault diagnostics and prognostics. *Mechanical Systems and Signal Processing* 102: 382–400. <https://doi.org/10.1007/s00500-020-04790-2>
- Batista GE, Monard MC (2002) A study of k-nearest neighbour as an imputation method. *HIS* 87(251–260): 48.
- Batista GE, Monard MC (2003) An analysis of four missing data treatment methods for supervised learning. *Applied Artificial Intelligence* 17(5–6): 519–533. <https://doi.org/10.1080/713827181>
- Bishop CM (2006) *Pattern Recognition and Machine Learning*. Springer-Verlag, New York, NY, 738 pp.
- Boursard H, Kamp Y (1988) Auto-association by multilayer perceptrons and singular value decomposition. *Biological Cybernetics* 59: 291–294. <https://doi.org/10.1007/BF00332918>
- Breiman L (1996) Bagging predictors. *Machine Learning* 24(2): 123–140. <https://doi.org/10.1007/BF00058655>
- Breunig MM, Kriegel H-P, Ng RT, Sander J (2000). LOF: Identifying Density-Based Local Outliers. *ACM SIGMOD Record* 29(2): 93–104. <https://doi.org/10.1145/335191.335388>

- Bryll R, Gutierrez-Osuna R, Quek F (2003) Attribute bagging: improving accuracy of classifier ensembles by using random feature subsets. *Pattern recognition* 36(6): 1291–1302. [https://doi.org/10.1016/S0031-3203\(02\)00121-8](https://doi.org/10.1016/S0031-3203(02)00121-8)
- Cai L, Tian X (2014) A new fault detection method for non-Gaussian process based on robust independent component analysis. *Process Safety and Environmental Protection* 92(6): 645–658. <https://doi.org/10.1016/j.psep.2013.11.003>
- Chalapathy R, Menon AK, Chawla S (2017) Robust, deep and inductive anomaly detection. In: Ceci M, Hollmén J, Todorovski L, Vens C, Džeroski S (Eds) *Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2017. Lecture Notes in Computer Science*, vol 10534. Springer, Cham, 36–51. https://doi.org/10.1007/978-3-319-71249-9_3
- Chen G (2015) Deep learning with nonparametric clustering. arXiv.
- Chen J, Liu K-C (2002) On-line batch process monitoring using dynamic PCA and dynamic PLS models. *Chemical Engineering Science* 57(1): 63–75. [https://doi.org/10.1016/S0009-2509\(01\)00366-9](https://doi.org/10.1016/S0009-2509(01)00366-9)
- Chen K, Hu J, He J (2016a) Detection and classification of transmission line faults based on unsupervised feature learning and convolutional sparse autoencoder. *IEEE Transactions on Smart Grid* 9(3): 1748–1758. <https://doi.org/10.1109/TSG.2016.2598881>
- Chen K, Hu J, He J (2018a) Detection and classification of transmission line faults based on unsupervised feature learning and convolutional sparse autoencoder. *IEEE Transactions on Smart Grid* 9(3): 1748–1758. <https://doi.org/10.1109/TSG.2016.2598881>
- Chen Z, Ding SX, Peng T, Yang C, Gui W (2018b) Fault detection for non-Gaussian processes using generalized canonical correlation analysis and randomized algorithms. *IEEE Transactions on Industrial Electronics* 65(2): 1559–1567. <https://doi.org/10.1109/TIE.2017.2733501>
- Chen Z, Ding SX, Zhang K, Li Z, Hu Z (2016b) Canonical correlation analysis-based fault detection methods with application to alumina evaporation process. *Control Engineering Practice* 46: 51–58. <https://doi.org/10.1016/j.conengprac.2015.10.006>
- Chen Z, Zhang K, Ding SX, Shardt YAW, Hu Z (2016c) Improved canonical correlation analysis-based fault detection methods for industrial processes. *Journal of Process Control* 41: 26–34. <https://doi.org/10.1016/j.jprocont.2016.02.006>
- Chiang LH, Russell EL, Braatz RD (2001) *Fault Detection and Diagnosis in Industrial Systems*. Springer-Verlag, London, 279 pp. <https://doi.org/10.1007/978-1-4471-0347-9>
- Cho S, Jiang J (2018) Optimal fault classification using fisher discriminant analysis in the parity space for applications to NPPs. *IEEE Transactions on Nuclear Science* 65(3): 856–865. <https://doi.org/10.1109/TNS.2018.2803658>
- Choi SW, Lee I-B (2004) Nonlinear dynamic process monitoring based on dynamic kernel PCA. *Chemical Engineering Science* 59(24): 5897–5908. <https://doi.org/10.1016/j.ces.2004.07.019>
- Chung J, Gulcehre C, Cho K, Bengio Y (2014) Empirical evaluation of gated recurrent neural networks on sequence modeling. In: *NIPS 2014 Workshop on Deep Learning*, December 2014.
- Chung Y-A, Wu C-C, Shen C-H, Lee H-Y, Lee L-S (2016) Audio Word2Vec: Unsupervised learning of audio segment representations using sequence-to-sequence autoencoder. In: *INTERSPEECH '16, San Francisco (USA)*, September 2016. ISCA, 765–769. <https://doi.org/10.21437/Interspeech.2016-82>
- Costa B, Angelov P, Guedes L (2015) Fully unsupervised fault detection and identification based on recursive density estimation and self-evolving cloud-based classifier. *Neurocomputing* 150: 289–303. <https://doi.org/10.1016/j.neucom.2014.05.086>
- da Silva Soares A, Galvao RKH (2010) Fault detection using Linear Discriminant Analysis with selection of process variables and time lags. In: *Proceedings of 2010 IEEE International Conference on Industrial Technology, Vina del Mar (Chile)*, March 2010. IEEE, 217–222. <https://doi.org/10.1109/ICIT.2010.5472682>
- Dai X, Gao Z (2013) From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis. *IEEE Transactions on Industrial Informatics* 9(4): 2226–2238. <https://doi.org/10.1109/TII.2013.2243743>
- de Lazaro JMB, Moreno AP, Santiago OL, da Silva Neto AJ (2015) Optimizing kernel methods to reduce dimensionality in fault diagnosis of industrial systems. *Computers & Industrial Engineering* 87: 140–149. <https://doi.org/10.1016/j.cie.2015.05.012>
- Ding Z, Fei M (2013) An anomaly detection approach based on isolation forest algorithm for streaming data using sliding window. *IFAC Proceedings Volumes* 46(20): 12–17. <https://doi.org/10.3182/20130902-3-CN-3020.00044>
- Doersch C (2016) Tutorial on variational autoencoders. arXiv.
- Duan L, Xu L, Liu Y, Lee J (2009) Cluster-based outlier detection. *Annals of Operations Research* 168(1): 151–168. <https://doi.org/10.1007/s10479-008-0371-9>
- Elman JL (1990) Finding structure in time. *Cognitive science* 14(2): 179–211. https://doi.org/10.1207/s15516709cog1402_1
- Ester M, Krieger H-P, Sander J, Xu X (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. *KDD-96 Proceedings* 96, 226–231.
- Everett B (2013) *An Introduction to Latent Variable Models*. Springer Science & Business Media, 106 pp.
- Fauconnier C, Haesbroeck G (2009) Outliers detection with the minimum covariance determinant estimator in practice. *Statistical Methodology* 6(4): 363–379. <https://doi.org/10.1016/j.stamet.2008.12.005>
- Gajjar S, Kulahci M, Palazoglu A (2018) Real-time fault detection and diagnosis using sparse principal component analysis. *Journal of Process Control* 67: 112–128. <https://doi.org/10.1016/j.jprocont.2017.03.005>
- Garcia-Allende PB, Conde OM, Mirapeix J, Cobo A, Lopez-Higuera JM (2008) Quality control of industrial processes by combining a hyperspectral sensor and Fisher's linear discriminant analysis. *Sensors and Actuators B: Chemical* 129(2): 977–984. <https://doi.org/10.1016/j.snb.2007.09.036>
- Ge Z, Song Z (2010) Mixture Bayesian regularization method of PPCA for multimode process monitoring. *AIChE Journal* 56(11): 2838–2849. <https://doi.org/10.1002/aic.12200>
- Ghasedi Dizaji K, Herandi A, Deng C, Cai W, Huang H (2017) Deep clustering via joint convolutional autoencoder embedding and relative entropy minimization. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Venice (Italy), October 2017. IEEE, 5747–5756. <https://doi.org/10.1109/ICCV.2017.612>
- GOST R ISO 13381-1-2016 (2017) *Condition monitoring and diagnostics of machines – Prognostics – Part 1: General guidelines*, IDT. Standartinform Publ., Moscow, 24 pp. [in Russian]
- Grais EM, Plumbley MD (2017) Single channel audio source separation using convolutional denoising autoencoders. In: *Proceedings of 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Montreal (Canada), November 2017. IEEE, 1265–1269. <https://doi.org/10.1109/GlobalSIP.2017.8309164>

- Guha S, Rastogi R, Shim K (2000) ROCK: A robust clustering algorithm for categorical attributes. *Information Systems* 25(5): 345–366. [https://doi.org/10.1016/S0306-4379\(00\)00022-3](https://doi.org/10.1016/S0306-4379(00)00022-3)
- Gupta A, Lam M (1998) The weight decay backpropagation for generalizations with missing values. *Annals of Operations Research* 78: 165–187. <https://doi.org/10.1023/A:1018945915940>
- Hardin J, Rocke DM (2004) Outlier detection in the multiple cluster setting using the minimum covariance determinant estimator. *Computational Statistics & Data Analysis* 44(4): 625–638. [https://doi.org/10.1016/S0167-9473\(02\)00280-3](https://doi.org/10.1016/S0167-9473(02)00280-3)
- Hardoon DR, Szedmak S, Shawe-Taylor J (2004) Canonical correlation analysis: An overview with application to learning methods. *Neural Computation* 16(12): 2639–2664. <https://doi.org/10.1162/0899766042321814>
- Harman HH (1976) *Modern Factor Analysis*. 3rd edn. The University of Chicago Press, Chicago, IL, 522 pp.
- Hastie T, Tibshirani R, Friedman J (2009) *The Elements of Statistical Learning*. Springer-Verlag, New York, NY, 745 pp. <https://doi.org/10.1007/978-0-387-84858-7>
- He Z, Xu X, Deng S (2003) Discovering cluster-based local outliers. *Pattern Recognition Letters* 24(9–10): 1641–1650. [https://doi.org/10.1016/S0167-8655\(03\)00003-5](https://doi.org/10.1016/S0167-8655(03)00003-5)
- Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural computation* 9(8): 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Honghai F, Guoshun C, Cheng Y, Bingru Y, Yumei C (2005) A SVM regression based approach to filling in missing values. In: Khosla R, Howlett RJ, Jain LC (Eds) *Knowledge-Based Intelligent Information and Engineering Systems. Ninth International Conference, KES 2005, Melbourne (Australia), September 2005*. Springer, Berlin, Heidelberg, Part III, 581–587. https://doi.org/10.1007/11553939_83
- Hubert M, Debruyne M (2009) Minimum covariance determinant. *Wiley Interdisciplinary Reviews: Computational Statistics* 2(1): 36–43. <https://doi.org/10.1002/wics.61>
- Iwana BK, Uchida S (2020) An empirical survey of data augmentation for time series classification with neural networks. arXiv preprint arXiv: 2007.15951.
- Jamil F, Abid M, Haq I, Khan AQ, Iqbal M (2016) Fault diagnosis of Pakistan Research Reactor-2 with data-driven techniques. *Annals of Nuclear Energy* 90: 433–440. <https://doi.org/10.1016/j.anucene.2015.12.023>
- Jia Q, Zhang Y (2016) Quality-related fault detection approach based on dynamic kernel partial least squares. *Chemical Engineering Research and Design* 106: 242–252. <https://doi.org/10.1016/j.cherd.2015.12.015>
- Jiang M-F, Tseng S-S, Su C-M (2001) Two-phase clustering process for outliers detection. *Pattern Recognition Letters* 22(6–7): 691–700. [https://doi.org/10.1016/S0167-8655\(00\)00131-8](https://doi.org/10.1016/S0167-8655(00)00131-8)
- Jiang Q, Yan X (2014) Just-in-time reorganized PCA integrated with SVDD for chemical process monitoring. *AIChE Journal* 60(3): 949–965. <https://doi.org/10.1002/aic.14335>
- Jiao J, Zhao N, Wang G, Yin S (2017) A nonlinear quality-related fault detection approach based on modified kernel partial least squares. *ISA Transactions* 66: 275–283. <https://doi.org/10.1016/j.isatra.2016.10.015>
- Jonsson P, Wohlin C (2004) An evaluation of k-nearest neighbour imputation using likert data. *Proceedings Tenth International Symposium on Software Metrics, Chicago, IL (USA), September 2004*, 108–118. <https://doi.org/10.1109/METRIC.2004.1357895>
- Kano M, Tanaka S, Hasebe S, Hashimoto I, Ohno H (2003) Monitoring independent components for fault detection. *AIChE Journal* 49(4): 969–976. <https://doi.org/10.1002/aic.690490414>
- Katser ID, Kozitsin VO, Maksimov IV (2019) NPP equipment fault detection methods. *Izvestiya Vuzov. Yadernaya Energetika* 4: 5–27. <https://doi.org/10.26583/npe.2019.4.01> [in Russian]
- Kaya H, Eyben F, Salah AA, Schuller B (2014) CCA based feature selection with application to continuous depression recognition from acoustic speech features. In: *Proceedings of 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence (Italy), May 2014*. IEEE, 3729–3733. <https://doi.org/10.1109/ICASSP.2014.6854298>
- Kim B, Lee S, Lee M, Ni J, Song JY, Lee CW (2007) A comparative study on damage detection in speed-up and coast-down process of grinding spindle-typed rotor-bearing system. *Journal of Materials Processing Technology* 187–188: 30–36. <https://doi.org/10.1016/j.jmatprotec.2006.11.222>
- Kim D, Lee I-B (2003) Process monitoring based on probabilistic PCA. *Chemometrics and Intelligent Laboratory Systems* 67(2): 109–123. [https://doi.org/10.1016/S0169-7439\(03\)00063-7](https://doi.org/10.1016/S0169-7439(03)00063-7)
- Kim JO (1989) *Factor, Discriminant and Cluster Analysis*. Ripol Classic, Moscow, 215 pp. [in Russian]
- Kingma DP, Welling M (2013) Auto-encoding variational bayes. arXiv.
- Ku W, Storer RH, Georgakis C (1995) Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems* 30(1): 179–196. [https://doi.org/10.1016/0169-7439\(95\)00076-3](https://doi.org/10.1016/0169-7439(95)00076-3)
- Lazarevic A, Kumar V (2005) Feature bagging for outlier detection. In: *Proceedings of the Eleventh Acm Sigkdd International Conference on Knowledge Discovery in Data Mining (KDD '05), Chicago IL (USA), August 2005*. Association for Computing Machinery, New York, NY, 157–166. <https://doi.org/10.1145/1081870.1081891>
- Lee J-M, Yoo C, Choi SW, Vanrolleghem PA, Lee I-B (2004a) Non-linear process monitoring using kernel principal component analysis. *Chemical Engineering Science* 59(1): 223–234. <https://doi.org/10.1016/j.ces.2003.09.012>
- Lee J-M, Yoo C, Lee I-B (2004b) Statistical monitoring of dynamic processes based on dynamic independent component analysis. *Chemical Engineering Science* 59(14): 2995–3006. <https://doi.org/10.1016/j.ces.2004.04.031>
- Lee J-M, Yoo C, Lee I-B (2004c) Statistical process monitoring with independent component analysis. *Journal of Process Control* 14(5): 467–485. <https://doi.org/10.1016/j.jprocont.2003.09.004>
- Leskin ST, Slobodchuk VI, Shelegov AS, Lapshin MR (2011) WWER-1000 main circulation pumps diagnostics based on their technological testing results. In: *Proceedings of the Seventh International Scientific and Technical Conference “Safety Assurance of NPP with VVER”*, Podolsk (Russia), May 2011. [in Russian]
- Leys C, Klein O, Dominicy Y, Ley C (2018) Detecting multivariate outliers: Use a robust variant of the Mahalanobis distance. *Journal of Experimental Social Psychology* 74: 150–156. <https://doi.org/10.1016/j.jesp.2017.09.011>
- Li D, Deogun J, Spaulding W, Shuart B (2004) Towards missing data imputation: a study of fuzzy k-means clustering method. In: Tsumoto S, Słowiński R, Komorowski J, Grzymała-Busse JW (Eds) *Rough Sets and Current Trends in Computing. Fourth International Conference, RSCTC 2004, Uppsala (Sweden), June 2004*. Springer, Berlin, Heidelberg, 573–579. https://doi.org/10.1007/978-3-540-25929-9_70

- Li S, Gao J, Nyagilo JO, Dave DP (2011) Probabilistic partial least square regression: A robust model for quantitative analysis of Raman spectroscopy data. In: Proceedings of 2011 IEEE International Conference on Bioinformatics and Biomedicine, Atlanta, GA (USA), November 2011. IEEE, 526–531. <https://doi.org/10.1109/BIBM.2011.94>
- Li W, Peng M, Wang Q (2018a) Fault detectability analysis in PCA method during condition monitoring of sensors in a nuclear power plant. *Annals of Nuclear Energy* 119: 342–351. <https://doi.org/10.1016/j.anucene.2018.05.024>
- Li W, Peng M, Wang Q (2018b) False alarm reducing in PCA method for sensor fault detection in a nuclear power plant. *Annals of Nuclear Energy* 118: 131–139. <https://doi.org/10.1016/j.anucene.2018.04.012>
- Li W, Peng M, Liu Y, Jiang N, Wang H, Duan Z (2018c) Fault detection, identification and reconstruction of sensors in nuclear power plant with optimized PCA method. *Annals of Nuclear Energy* 113: 105–117. <https://doi.org/10.1016/j.anucene.2017.11.009>
- Li W, Peng M, Wang Q (2019) Improved PCA method for sensor fault detection and isolation in a nuclear power plant. *Nuclear Engineering and Technology* 51(1): 146–154. <https://doi.org/10.1016/j.net.2018.08.020>
- Lipovetsky S, Conklin M (2001) Analysis of regression in game theory approach. *Applied Stochastic Models in Business and Industry*, 17(4): 319–330. <https://doi.org/10.1002/asm.446>
- Liu FT, Ting KM, Zhou Z-H (2018) Isolation Forest. In: Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, Pisa (Italy), December 2008, 413–422. <https://doi.org/10.1109/ICDM.2008.17>
- Loureiro A, Torgo L, Soares C (2004) Outlier detection using clustering methods: a data cleaning application. In: Proceedings of KD-Net Symposium on Knowledge-based Systems for the Public Sector, Bonn (Germany), June 2004.
- Lundberg S, Lee SI (2017) A unified approach to interpreting model predictions. arXiv preprint arXiv: 1705.07874.
- Ma J, Jiang J (2011) Applications of fault detection and diagnosis methods in nuclear power plants: A review. *Progress in Nuclear Energy* 53(3): 255–266. <https://doi.org/10.1016/j.pnucene.2010.12.001>
- MacGregor J, Kourti T (1995) Statistical process control of multivariate processes. *Control Engineering Practice* 3(3): 403–414. [https://doi.org/10.1016/0967-0661\(95\)00014-L](https://doi.org/10.1016/0967-0661(95)00014-L)
- Manly BFJ, Alberto JAN (2016) *Multivariate Statistical Methods: A Primer*. 4th edn. Taylor & Francis, Chapman and Hall/CRC, New York, NY, 253 pp. <https://doi.org/10.1201/9781315382135>
- Marlin B (2008) Missing data problems in machine learning. PhD Thesis. University of Toronto, Toronto.
- Mika S, Ratsch G, Weston J, Scholkopf B, Mullers K-R (1999) Fisher discriminant analysis with kernels. In: *Neural Networks for Signal Processing IX: Proceedings of the 1999 IEEE Signal Processing Society Workshop* (cat. no. 98th8468), Madison, WI (USA), August 1999. IEEE, 41–48. <https://doi.org/10.1109/NNSP.1999.788121>
- Pachgade MS, Dhande MS (2012) Outlier detection over data set using cluster-based and distance-based approach. *International Journal of Advanced Research in Computer Science and Software Engineering* 2(6): 12–16.
- Patel HR, Shah VA (2018) Fault detection and diagnosis methods in power generation plants – The Indian power generation sector perspective: An introductory review. *PDP Journal of Energy and Management* 2(2): 31–49.
- Prince SJ, Elder JH (2007) Probabilistic linear discriminant analysis for inferences about identity. In: *Proceedings of 2007 IEEE 11th International Conference on Computer Vision, Rio de Janeiro (Brazil), October 2007*. IEEE, 1–8. <https://doi.org/10.1109/ICCV.2007.4409052>
- Qin SJ (2009) Data-driven Fault Detection and Diagnosis for Complex Industrial Processes. *IFAC Proceedings Volumes* 42(8): 1115–1125. <https://doi.org/10.3182/20090630-4-ES-2003.00184>
- Raveendran R, Huang B (2016) Mixture probabilistic PCA for process monitoring – collapsed variational Bayesian approach. *IFAC-PapersOnLine* 49(7): 1032–1037. <https://doi.org/10.1016/j.ifacol.2016.07.338>
- Raveendran R, Huang B (2017) Two layered mixture Bayesian probabilistic PCA for dynamic process monitoring. *Journal of Process Control* 57: 148–163. <https://doi.org/10.1016/j.jprocont.2017.06.009>
- Ribeiro MT, Singh S, Guestrin C (2016) Why should i trust you?: Explaining the predictions of any classifier. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1135–1144. <https://doi.org/10.1145/2939672.2939778>
- Rousseeuw PJ (1984) Least median of squares regression. *Journal of the American Statistical Association* 79(388): 871–880. <https://doi.org/10.1080/01621459.1984.10477105>
- Rousseeuw PJ, Driessen KV (1999) A fast algorithm for the minimum covariance determinant estimator. *Technometrics* 41(3): 212–223. <https://doi.org/10.1080/00401706.1999.10485670>
- Russell EL, Chiang LH, Braatz RD (2000) Fault detection in industrial processes using canonical variate analysis and dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems* 51(1): 81–93. [https://doi.org/10.1016/S0169-7439\(00\)00058-7](https://doi.org/10.1016/S0169-7439(00)00058-7)
- Sakia RM (1992) The Box-Cox Transformation Technique: A Review. *Journal of the Royal Statistical Society. Series D (The Statistician)* 41(2): 169–178. <https://doi.org/10.2307/2348250>
- Sakurada M, Yairi T (2014) Anomaly detection using autoencoders with nonlinear dimensionality reduction. In: *Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis (MLSDA '14)*, Gold Coast (Australia), December 2014. Association for Computing Machinery, New York, NY, 4–11. <https://doi.org/10.1145/2689746.2689747>
- Scholkopf B, Williamson RC, Smola AJ, Shawe-Taylor J, Platt JC (2000) Support vector method for novelty detection. In: Solla SA, Leen TK, Muller K-R (Eds) *Advances in neural information processing systems*. Thirteenth Annual Neural Information Processing Systems Conference (NIPS 1999), Denver, CO (USA), June 2000. MIT Press, 582–588.
- Sergienko AB (2011) *Digital Signal Processing*. BVH-Peterburg Publ., 768 pp. [in Russian]
- Shalabi LA, Shaaban Z, Kasasbeh B (2006) Data mining: A preprocessing engine. *Journal of Computer Science* 2(9): 735–739. <https://doi.org/10.3844/jcssp.2006.735.739>
- Si X-S, Wang W, Hu C-H, Zhou D-H (2011) Remaining useful life estimation: A review on the statistical data driven approaches. *European Journal of Operational Research* 213(1): 1–14. <https://doi.org/10.1016/j.ejor.2010.11.018>
- Tan SC, Ting KM, Liu TF (2011) Fast anomaly detection for streaming data. In: Walsh T (Ed.) *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI 2011)*. International Joint Conference on Artificial Intelligence 2011,

- Barcelona (Spain), July 2011. Association for the Advancement of Artificial Intelligence (AAAI), Menlo Park, CA, 1511–1516.
- Tax DM, Duijn RP (2004) Support vector data description. *Machine Learning* 54(1): 45–66. <https://doi.org/10.1023/B:MACH.000008084.60811.49>
 - Tibshirani R (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 58(1): 267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
 - Tipping ME, Bishop CM (1999) Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61(3): 611–622. <https://doi.org/10.1111/1467-9868.00196>
 - Venkatasubramanian V, Rengaswamy R, Kavuri SN (2003a) A review of process fault detection and diagnosis: Part I: Quantitative model-based methods. *Computers & Chemical Engineering* 27(3): 313–326. [https://doi.org/10.1016/S0098-1354\(02\)00161-8](https://doi.org/10.1016/S0098-1354(02)00161-8)
 - Venkatasubramanian V, Rengaswamy R, Kavuri SN, Surya N, Yin K (2003b) A review of process fault detection and diagnosis: Part III: Process history based methods. *Computers & Chemical Engineering* 27(3): 327–346. [https://doi.org/10.1016/S0098-1354\(02\)00162-X](https://doi.org/10.1016/S0098-1354(02)00162-X)
 - Vincent P, Larochelle H, Bengio Y, Manzagol P (2008) Extracting and composing robust features with denoising autoencoders. In: *Proceedings of the 25th International Conference on Machine Learning, Helsinki (Finland), July 2008*. Association for Computing Machinery, New York, NY, 1096–1103. <https://doi.org/10.1145/1390156.1390294>
 - Wang X, Kruger U, Lennox B (2003) Recursive partial least squares algorithms for monitoring complex industrial processes. *Control Engineering Practice* 11(6): 613–632. [https://doi.org/10.1016/S0967-0661\(02\)00096-5](https://doi.org/10.1016/S0967-0661(02)00096-5)
 - Warne RT, Larsen R (2014) Evaluating a proposed modification of the Guttman rule for determining the number of factors in an exploratory factor analysis. *Psychological Test and Assessment Modeling* 56(1): 104–123.
 - Wen Q, Sun L, Song X, Gao J, Wang X, Xu H (2020) Time series data augmentation for deep learning: A survey. *arXiv preprint arXiv:2002.12478*.
 - Wohlrab L, Fürnkranz J (2009) A comparison of strategies for handling missing values in rule learning. Technical Report TUD-KE-2009-03.
 - Xiao H, Huang D, Pan Y, Liu Y, Song K (2017) Fault diagnosis and prognosis of wastewater processes with incomplete data by the auto-associative neural networks and ARMA model. *Chemometrics and Intelligent Laboratory Systems* 161: 96–107. <https://doi.org/10.1016/j.chemolab.2016.12.009>
 - Xu C, Zhao S, Liu F (2017) Distributed plant-wide process monitoring based on PCA with minimal redundancy maximal relevance. *Chemometrics and Intelligent Laboratory Systems* 169: 53–63. <https://doi.org/10.1016/j.chemolab.2017.08.004>
 - Yang H-H, Huang M-L, Yang S-W (2015) Integrating auto-associative neural networks with Hotelling T² control charts for wind turbine fault detection. *Energies* 8(10): 12100–12115. <https://doi.org/10.3390/en81012100>
 - Zagoruyko NG (1999) Applied data and knowledge analysis methods. Institut matematiki Novosibirsk Publ., 260 pp. [in Russian]
 - Zavaljevski N, Gross KC (2000) Sensor fault detection in nuclear power plants using multivariate state estimation technique and support vector machines. In: *Proceedings Third International Conference of the Yugoslav Nuclear Society YUNSC 2000, Belgrade (Yugoslavia), October 2000*. Argonne National Lab., Argonne, IL.
 - Zhang Y, Hu Z (2011) On-line batch process monitoring using hierarchical kernel partial least squares. *Chemical Engineering Research and Design* 89(10): 2078–2084. <https://doi.org/10.1016/j.cherd.2011.01.002>
 - Zhang Y, Meratnia N, Havinga P (2009) Adaptive and online one-class support vector machine-based outlier detection techniques for wireless sensor networks. In: *23rd International Conference on Advanced Information Networking and Applications, AINA 2009, Workshops Proceedings, Bradford (UK), May 2009*. IEEE Computer Society, 990–995. <https://doi.org/10.1109/WAINA.2009.200>
 - Zhang Y, Qin SJ (2007) Fault Detection of nonlinear processes using multiway kernel independent component analysis. *Industrial & Engineering Chemistry Research* 46(23): 7780–7787. <https://doi.org/10.1021/ie070381q>
 - Zhang Y, Zhou H, Qin SJ, Chai T (2010) Decentralized fault diagnosis of large-scale processes using multiblock kernel partial least squares. *IEEE Transactions on Industrial Informatics* 6(1): 3–10. <https://doi.org/10.1109/TII.2009.2033181>
 - Zhao Y, Nasrullah Z, Hryniewicki MK, Li Z (2019b) LSCP: Locally selective combination in parallel outlier ensembles. In: Berger-Wolf T, Chawla N (Eds) *Proceedings of the 2019 SIAM International Conference on Data Mining, Calgary (Canada), May 2019*, 585–593. <https://doi.org/10.1137/1.9781611975673.66>
 - Zhao Y, Nasrullah Z, Li Z (2019a) PyOD: A Python toolbox for scalable outlier detection. *Journal of Machine Learning Research* 20(96): 1–7.
 - Zhu J, Ge Z, Song Z (2017) Non-Gaussian industrial process monitoring with probabilistic independent component analysis. *IEEE Transactions on Automation Science and Engineering* 14(2): 1309–1319. <https://doi.org/10.1109/TASE.2016.2537373>