Probabilistic Principal Components and Mixtures, How This Works

Anna M. Bartkowiak, Radoslaw Zimroz

Probabilistic principal components and mixtures, how this works

Anna M. Bartkowiak[1] and Radoslaw Zimroz[2]

[1] Wroclaw University, Inst. of Computer Science, 50-383 Wroclaw PL
and Wroclaw School of Information Technology, 54-239 Wroclaw PL,
[2] Wroclaw University of Technology, Diagnostics and Vibro-Acoustics
Science Laboratory, 50-421 Wroclaw PL

Abstract. Classical Principal Components Analysis (PCA) is widely recognized as a method for dimensionality reduction and data visualization. This is a purely algebraic method, it considers just some optimization problem which fits exactly to the gathered data vectors with their particularities. No statistical significance tests are possible. An alternative is to use probabilistic principal component analysis (PPCA), which is formulated on a probabilistic ground. Obviously, to do it one has to know the probability distribution of the analyzed data. Usually the Multi-Variate Gaussian (MVG) distribution is assumed. But what, if the analyzed data are decidedly not MVG? We have met such problem when elaborating multivariate gearbox data derived from a heavy duty machine. We show here how we have dealt with the problem.
In our analysis, we assumed that the considered data are a mixture of two groups being MVG, specifically: each of the sub-group follows a probabilistic principal component (PPC) distribution with a MVG error function. Then, by applying Bayesian inference, we were able to calculate for each data vector x its a posteriori probability of belonging to data generated by the assumed model. After estimation of the parameters of the assumed model we got means - based on a sound statistical basis - for constructing confidence boundaries of the data and finding outliers.

Keywords: probabilistic principal components, multi-variate normal distribution, mixture models, un-mixing multivariate data, condition monitoring, gearbox diagnostics, healthy state, probabilities a posteriori, outliers.

1 Introduction

Classical Principal Components Analysis (PCA) is widely recognized as a method for dimensionality reduction and data visualization. However, PCA is a purely algebraic method, it considers just some optimization problem which fits exactly to the gathered data vectors with their particularities.

Yet, without a proper probability model it is impossible to formulate statistically significant statements.

On the opposite, Probabilistic Principal Components Analysis (PPCA) permits to tackle the data in a smoothed holistic way. It is easy to introduce into its models (formulated in d-dimensional data space) some q dimensional sub-models with q lower than d. Additionally, probabilistic principal components may be combined into a mixture model, which permits to model the non-Gaussian data as a mixture of several sub-groups, each of them having its own Gaussian distribution. We will show below how such a model (embedding PPCA into mixtures) may be useful in analysis of real data.

We will consider data obtained from vibration signals of a heavy-duty machine being in good state. Say, the data are contained in a real data matrix $\mathbf{B}$ of size $n \times d$, that is with n rows (time segments) and d columns (variables characterizing the segments). It is common to imagine the data vectors of such a matrix as d-dimensional points located in the d-dimensional data space. During operation, the condition of the machine may deteriorate. The very important question is: **how to determine, whether the condition of the machine is good (healthy), or - whether it starts to be** (or is already) **faulty.**

Methods of multivariate data analysis permit to answer the above question, provided that it is formulated in strict mathematical language. For instance, one may be concerned with the following questions:

- Is the machine in good or bad condition? How to carry out the monitoring of the state the machine? To answer these questions, one needs also data sample of a 'bad' machine. The bad data sample should be provided as another data matrix with d columns containing values of the same variables as those measured for the 'good' matrix B. A survey of methods and papers dealing with this question may be found in [1, 5, 6, 13, 8, 10, 21], and many others.
- For the problem: How to detect the fault possibly early? see, e.g. [11].
- For very common and widely elaborated problems falling under the topics Feature selection and/or Dimensionality reduction see, e.g. [20, 3, 18], and references therein.
- Say, we have data only for a machine in good condition. For its monitoring, we might specifically ask for the boundary in the data space delimiting the 'normal', that is 'healthy' data. This problem is usually solved using methods like one class classification, novelty or anomaly detection, and outlier identification, see, for example, [2, 8, 12, 14].

In the following we will be concerned only with the last item. We will consider only one machine being in good condition. Our novel contributions are related to a modelling of multidimensional diagnostic data using probabilistic approach. Our proposal is to combine *three* statistical models *into one common model*, which yields so called *probabilities a posteriori* (*posteriors*). Under way, we are able to reduce dimensionality of the considered data. The posteriors obtained from the common model permit to perform - according to one's wish - condition monitoring, anomaly or novelty detection, identification of outliers (if any), and dimensionality reduction.

In this paper we show generally how the common model may be formulated and how its parameters may be estimated – this is illustrated using the mentioned set B of the gearbox data. We show also - for the analyzed data set B - that the mentioned posteriors may be calculated directly and how they look like. The posteriors are the basis for further statistical inference - like anomaly detection, confidence boundaries construction, etc., however, for lack of space, this is not elaborated in the paper.

The paper is scheduled as follows. Actually, we are in Section 1, Introduction. Next Section 2 introduces the three basically used by us statistical methods, namely the Mixture model, Bayesian inference and the Probabilistic Principal Components – to construct a common model for the data. Some issues of dimensionality reduction are also considered. Section 3 contains a short description of essential features of the data serving as the basis for our analysis, also some details on constructing the learning and testing sample. In Section 4 we formulate the principles of our experiment and the goals to be achieved. We show, how the assumed common model works with our data. We show also, how the posteriors – calculated for our data look like – and what exactly they do mean. Section 5 contains some discussion and closing remarks.

2 Methodology of un-mixing multivariate data by using mixture model with probabilistic principal components

2.1 The mixture model

Suppose, we have M different groups of multivariate data, each of the groups containing data vectors $\mathbf{x}$ with d elements corresponding to d observed variables. The observed vector $\mathbf{x}$ belonging to group j ($j = 1, \ldots, M$), has its specific probability distribution denoted as $p(\mathbf{x} \mid j)$. The basic equation describing the overall probability distribution of all the data may be modelled as mixture composed from these M groups [15]:

$$p(\mathbf{x}) = \sum_{j=1}^{M} P(j) p(\mathbf{x} \mid j), \tag{1}$$

where the parameters $P(j)$ are called *mixing coefficients*. They have the properties:

$$\sum_{j=1}^{M} P(j) = 1, \qquad \text{and} \qquad 0 \leq P(j) \leq 1, \quad j = 1, \ldots, M.$$

The *overall* probability distribution function $p(\mathbf{x})$ defined above 1 is a proper pdf (probability distribution function) describing the probabilities of all the data mixed together into one common group. The derived pdf (1) is called the *total* pdf. The mixing coefficients $P(j)$ are called *priors* or *probabilities á priori.*

Using Bayes' theorem, it is common to define *posterior probabilities* (*posteriors*) as:

$$P(j \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid j) P(j)}{p(\mathbf{x})} \tag{2}$$

In the following we will consider mixtures models composed only of two components, that is $M = 2$. The group-conditioned pdf's will be MVG with spherical covariance matrix ($\boldsymbol{\Sigma}_j = \sigma_j^2 \boldsymbol{I}$):

$$p(\mathbf{x} \mid j) = \frac{1}{(2\pi\sigma_j^2)^{d/2}}\, exp\{-\frac{||\mathbf{x} - \boldsymbol{\mu}_j||^2}{2\sigma_j^2}\} \tag{3}$$

2.2 Probabilistic principal components and reduction of the variables space

The probabilistic principal components methodology is based on the assumption that the observed data vector $\mathbf{x}$ may be modelled as a linear combination of some latent variables defined in an - unobservable directly - latent variables space of dimension $q <= d$. The assumed model reads:

$$\mathbf{x} = \mathbf{W}\mathbf{z} + \boldsymbol{\mu} + \mathbf{e}. \tag{4}$$

Meaning of symbols appearing in the assumed model:

$\mathbf{x}$ - the observed d-dimensional data vector, called also data instance,
$\mathbf{z}$ - q-dimensional latent factor variable, with $\mathbf{0}$ mean and unit isotropic variance; $\mathbf{z}$ is distributed as $N_q(\mathbf{0}, \mathbf{I})$,
$\mathbf{W}$ - so called matrix of loadings, consists of constant real numbers playing the role of parameters of the model; it may be estimated e.g. by the Maximum Likelihood (ML) method,
$\boldsymbol{\mu}$ - some constants playing the role of shift parameters; have to be estimated; the ML method yields here the data means as estimates,
$\mathbf{e}$ - independent noise process distributed as $N_d(\mathbf{0}, \sigma^{\mathbf{2}}\mathbf{I})$.

Taking eq. (4) into account, the probability density model for the probabilistic principal component analysis (PPCA) reads:

$$p(\mathbf{x} \mid \mathbf{z}) = \frac{1}{(2\pi\sigma^2)^{d/2}}\, exp\{-\frac{||\mathbf{x} - \mathbf{W}\mathbf{z} - \boldsymbol{\mu}||^2}{2\sigma^2}\} \tag{5}$$

Tipping and Bishop [17] have shown how to obtain estimates of the unknown parameters appearing in eq. (5). By integrating out the latent variables $\mathbf{z}$ they got that the distribution of the observed variables $\mathbf{x}$ is

$$\mathbf{x} \sim N(\boldsymbol{\mu}, \mathbf{C}), \quad \text{where} \;\; \mathbf{C} = \mathbf{W}\mathbf{W}^{\mathbf{T}} + \sigma^{\mathbf{2}}\mathbf{I}. \tag{6}$$

Tipping and Bishop [17] have shown also that the ML methods yields the following estimates for the parameters $\mathbf{W}$ and σ^2 appearing in the probability model for PPCA shown in eq. (5):

$$\mathbf{W}_{ML} = \mathbf{U}_q(\Lambda_q - \sigma^2\mathbf{I})^{1/2}\mathbf{R}, \quad \text{where} \;\; \sigma^2_{ML} = \frac{1}{d-q}\sum_{j=q+1}^{d} \lambda_j, \tag{7}$$

and $\Lambda_q, \mathbf{U}_q$ denote, up to a rotation matrix $\mathbf{R}$, the first q largest eigenvalues and the connected with them eigenvectors of the covariance matrix $\mathbf{C}$.

The dimension q is kept constant in the above reasoning; it is declared by the user. The variance σ^2_{ML} is interpreted as the variance lost in the projection from the data space (dimension d) to the latent space (dimension q).

After estimation of all the parameters appearing in the general mixture models (1) and its components, the posteriors defined in eq. (2) will be the most important. They will play an essential role in our analysis of real gearbox data, which are described in next section.

3 The analyzed data sets: learning sample B500 and test sample Bres

In the following we will show an analysis conducted using true data from machines working in field conditions. The data were recorded by Bartelmus and Zimroz [1] from two gearboxes, one being faulty, i.e. in bad condition, the other being healthy, i.e. in good condition.

Taking as a new feature the sum of all the 15 variables, Bartelmus and Zimroz [1] were able to classify – on the base of the proposed feature – about 80 % of all data vectors. To classify the remainder, they needed an external variable, called ZWE, indicating for the actual load of the working machine.

The data were more thoroughly investigated in [19, 2]. It appeared that the distribution of the variables is not Gaussian, the data contain a considerable number of outliers, moreover, the covariance structure in the two groups (faulty and healthy) is markedly different.

In the following we will consider only the healthy data containing $n = 951$ data vectors. The entire healthy data set **B** was subdivided into a learning sample called **B500** and a testing sample called **Bres**. The learning sample **B500** was obtained from randomly chosen 500 rows of the original data set **B**. The remainder of the data containing 451 rows from **B**, was designated as **Bres** for testing the built model.

Apart from this, we got also for each data instance (i.e. data vector $\mathbf{x}$) the value of another variable, called ZWE. The ZWE variable represents value of averaged speed for short (1s) observation period called segment (of signal), from which one 15D feature vector $\mathbf{x}$ was derived. Value of ZWE may belong to speed range: 940-1000 rpm (rotations per minute). Typically, values ZWE$<=$ 990 denote a heavy load (HL); for ZWE$>$ 990 the load is considered to be small or none (NL).

The number of heavy and small/none loads in the investigated $B500$-sample happened to be: 439 instances HL, 61 instances NL.

In Fig. 1 we show the distribution of the variable ZWE in the entire data set **B** (top graph), in the derived learning sample B500 (middle graph), and in the test sample Bres (bottom graphs). One may notice that all the three displays are very similar with respect of their ZWE distributions.

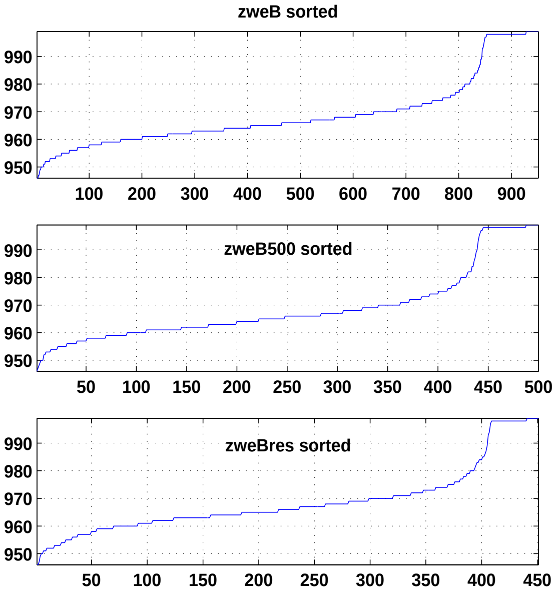

Fig. 1. Ordered values of ZWE in analyzed data **B** and its sub-samples B500 and Bres. *Top*: original set **B**. *Middle*: learning sample B500. *Bottom*: test sample Bres. Take notice thar the distributions of ZWE visible in the three displays look similar.

For easiness of identifying the further results, the data instances (i.e. the data vectors) from both samples were sorted according the their ZWE values (each sample was sorted separately). After sorting, the heavy load data instances (HL) appear first, and the no-load instances (NL) last.

Our further analysis will consist of:
(i) building a two-group mixture model with embedded probabilistic principal components of dimension $q = 2$,
(ii) calculating the posteriors (see eq. 2) allowing for statistical inference on fitness of the assumed model and on the normality or abnormality of consecutive data vectors (abnormality means here outliers or atypical observations, which are not concordant with the assumed population model).

4 Application of mixture model with embedded PPCs to real data; how this works

4.1 Preliminary settings

In this section we report our analysis when using the B500 and Bres samples of size 500 × 15 and size 451 × 15 appropriately. The rows $\mathbf{x} = [x_1, x_2, \ldots, x_{15}]$ of both samples are ordered according to increasing values of ZWE corresponding to their respective $\mathbf{x}$ vectors.

The B500 set is supposed to be the learning sample and the Bres set the test sample for the constructed probabilistic model.

Our main goal is to obtain for the B500 sample a decomposition into two Gaussian sub-samples numerated as $j = 1$ and $j = 2$. A second goal is to assert the connection of the derived sub-samples with the load variable ZWE. A third goal is to obtain an affirmation that the obtained decomposition (un-mixing of the original data set B into two component-sets from which it is composed) fits adequately to the gathered data.

We will show in next two subsections how these goals were realized for the B500 data set. Here we add only that we carried out the analysis using a special type neural network `gmm` from the Netlab library [15]. The network worked in an unsupervised way, i.e. it knew only that it has to divide the B500 sample into two sub-groups, however it did not know that the sub-groups are expected to be associated with the status of the variable ZWE, which was out of reach for the network during its work at this stage.

4.2 Modelling data from the B500 sample

The basic mixture model from eq. (1) with M=2 was applied. It says that we will consider the B500 sample as a mixture composed from two sub-groups, each of them having its own probability density function (pdf) $p(\mathbf{x}|j)$, $j = 1, 2$. Each of these pdf's is assumed be MVG with probabilistic principal components embedded into the expected values of the assumed MVG's – accordingly to eq. (3) and (4). There is a lot of parameters to estimate. The neural network `gmm` packs them into a structure called here `mixB500`. The structure contains in its subsequent fields values of the parameters needed for an analysis of the supplied data B500. The fields of `mixB500` and their contents are shown in Table 1. After initialization of the structure, Tthe fields are filled sequentially with advancing of the analysis.

Table 1. The structure mixB500 containing parameters used in our mixture model, before and after applying the EM estimation procedure

```
          type: 'gmm'
           nin: 15
      ncentres: 2
    covar_type: 'ppca'
      ppca_dim: 2
        priors: [0.1285 0.8715]
       centres: [2x15 double]
        covars: [9.3489e-004 0.0090]
             U: [15x2x2 double]
        lambda: [2x2 double]
          nwts: 98
```

```
          type: 'gmm'
           nin: 15
      ncentres: 2
    covar_type: 'ppca'
      ppca_dim: 2
        priors: [0.1318 0.8682]
       centres: [2x15 double]
        covars: [8.5744e-004 0.0093]
             U: [15x2x2 double]
        lambda: [2x2 double]
          nwts: 98
```

The fields of the structure `mixB500` are:
type - a kind of signature of the structure,
nin - number of the variables (columns) in the data matrix B500,
ncentres - how many sub-groups (components of the mixture) are desired,
covar_type - indicates how the covariance matrices have to be calculated; 'ppca' means the option, that the covariances should be calculated according to eq. (6),
ppca_dim - how many principal components (latent variables according eq. (4)) we wish to include into the model. We declared that we want to retain only 2 principal components,
priors - cardinalities of the two sub-groups of the mixture, centres - means of the two initialized sub-groups (left structure) after run of k-means, and re-adjusted after run of the EM algorithm (right structure),
covars - covariance matrices of the sub-model. In case of 'ppca' option the spherical covariance matrices are assumed by default. We have two sub-groups, each needs one real value as its variance,
U - eigenvectors from the matrix $\mathbf{C}$ given in eq. (6), for each sub-group separately.
lambda - the eigen-values associated with the eigen-vectors in $\mathbf{U}$,
nwts - the number of values memorized in the structure `mixB500`. In our case, the structure contains 98 constants, which are necessary when considering particular problems connected with the constructed mixture model. The parameters/weights are optimized by the Maximum likelihood method using the EM algorithm.

After finishing the estimation process, the structure `mixB500` is filled with data and estimates of parameters necessary for further calculations. In particular, we may find there the parameters necessary for evaluation of the two sub-groups into which the entire data set B was split. The un-mix of the mixture appearing in set B is done.

Next steps of calculations are optional. We will be concerned with the content of subgroups established by the `gmm` network, also how this content is connected with the load variable ZWE. This is considered in next subsection.

4.3 The content of subgroups obtained from the mixture model memorized in mixB500

The `gmm` network feeded with the B500 sample data has split the obtained data into two subgroups. Parameters useful for further calculations are stored in the structure `mixB500` (see Table 1).

We are mainly concerned, what is the content of these subgroups. To obtain answer to this question, we inspect the group probability densities (likelihoods) $L(\mathbf{x}|j)$ and their posteriors $P(j|\mathbf{x})$. They are shown in Fig 2.

The top graph in Fig. 2 shows likelihoods, obtained as values of the probability density function $p(\mathbf{x} \mid j)$ with parameters evaluated by the ML method. In our case we have in the mixture 2 groups of data. Each group has its Gaussian

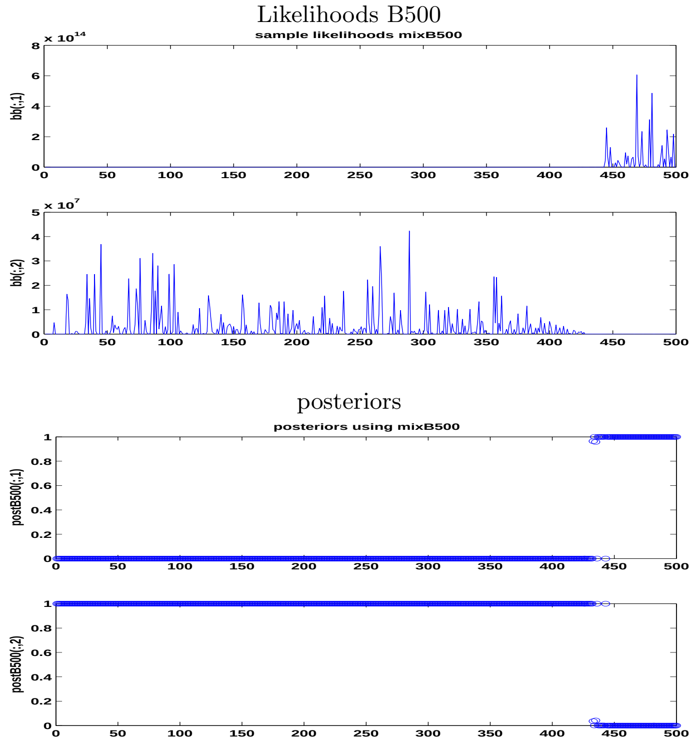

Fig. 2. Learning sample B500. Likelihoods and posterior probabilities of appearing data vectors $\mathbf{x}_i,\ i = 1, \ldots, 500$ in the mixture formed from two sub-groups. Counting from top to bottom: *First panel* : Likelihood of appearing in sub-group numbered $j = 1$. *Second panel* : Likelihood of appearing in sub-group numbered $j = 2$. *Third panel* : posterior of belonging to sub-group numbered $j = 1$. *Fourth panel* : posterior of belonging to sub-group numbered $j = 2$.

pdf with estimated parameters stored in the structure **mixB500**.Thus we are able to evaluate the value of the respective pdf (in other words, the likelihood) for every data vector **x**.

Taking the pdf of the first derived sub-group numbered as *j=1*, we substitute into this pdf in turn all data vectors **x** contained in the set B500; this yields the set of likelihood values displayed in the first panel of Fig. 2. The displayed likelihoods are numbered $1, 2, \ldots, 500$, that is similarly as the data vectors **x** serving to evaluation of the displayed likelihoods. Looking at the graph may notice that the pronounced values of the likelihoods appear only for the (about)last 50 data instances of B500. However, the sample B500 is sorted according its increasing ZWE values. Thus we may state: the subgroup *j=1* contains data instances with highest ZWE values, which means NL category of the load.

Taking the pdf of the second derived sub-group numbered as *j=2*, and repeating the actions as above, we obtain the series of likelihoods evaluated for subsequent values **x** of the data B500, however now the likelihoods are evaluated from the pdf characterizing the subgroup numbered *j=2*. The likelihoods evaluated in such a way are shown in the second panel of Fig. 2. One may notice here, that pronounced values of the likelihoods appear only for the (about) first 450 data instances. It happens that just these 450 data instances are HL (i.e. heavy loaded). Thus the subgroup numbered *j=2* contains data instances which are heavy loaded.

Analogous reasoning may be conducted when considering the probabilities a posteriori shown in the 3rd and 4th panel of Fig. 2. Here we see a clear group membership assignment. Moreover, the assignment is amazingly sharp. All data instances are allocated with a high probability. There are no doubtful assignments.

The final allocation of the 500 data instances is 66 + 434 (to sub-group *j=1* and *j=2* appropriately).

4.4 Analysis of the data set Bres

The Bres data set, counting 451 data instances, is composed from the remnants of the entire data set B after removing from it the sample B500. It constitutes test data for the mixture model `mixB500` built previously in subsection 4.3 from the B500 data. Now the Bres data could be considered using two possibilities :

(i) Looking at the behavior of the testing vectors $x \in$Bres by evaluating their likelihoods and posteriors on the basis of the mixture models whose parameters were kept memorized in the structure mixB500 obtained from an alien data set (B500).

(ii) Constructing a new, own data structure `mixBres`, and taking this new structure as basis for calculating the likelihoods and the posteriors for the Bres sample.

We have performed the analysis according both (i) and (ii). The results, displayed in a similar way as those in Fig. 2, are amazingly similar; for lack of space they could not be shown here. Performing a similar analysis as for the B500 data set we got very similar results. For lack of space we show here only the final allocations of the data vectors $\mathbf{x}$ from Bres:

When making allocation using the alien `mixB500` structure: $51 + 400$.

When building own mixture model and own structure `mixBres`: $50 + 401$.

5 Discussion and Concluding Remarks

We have considered so far only the simplest probabilistic principal component mixture models assuming Gaussian rank-2 sub-models with a spherical covariance matrix.

To our surprise, such a very simple model works amazingly well both for the learning sample B500 and the test sample Bres of the healthy data B. Indeed, we got an un-mixing of the entire data set B into two sub-models, one of them corresponding to the heavy_load and the other to the light/none_load state of the instances belonging to set B. Moreover, this was achieved using only sub-models of dimension $q = 2$ (the original data are 15-dimensional).

The main results are: The data for the healthy gearbox can be modelled as a mixture of two separate sub-groups, each of them having its own multi-variate Gaussian distribution. The subgroups are associated with an external variable ZWE, namely one subgroup has ZWE of category HL (heavy load), the other subgroup has ZWE of category NL (no or light load). The outliers stated in [2] have disappeared.

However this simple model is not valid for data coming from a faulty gearbox. Faulty data are essentially different (see [19]) and have to be modelled separately using a more complex model.

All the calculations were done using raw data without any standardization. It is known that neural networks (its optimization procedures) are favoring standardized data. Also the results in [21] were obtained using standardized data. It would be interesting to repeat the analysis using standardized data. Also, we feel it worthy to look for a similar model for the data from a faulty gearbox, which seems to be for the gearbox data from [1] a much more difficult task.

References

1. Bartelmus, W., Zimroz, R.: A new feature for monitoring the condition of gearboxes in nonstationary operating systems, Mechanical Systems and Signal Processing 23 (5), 1528-1534 (2009).
2. Bartkowiak, A., Zimroz, R.: Outliers analysis and one class classification approach for planetary gearbox diagnosis. Journal of Physics: Conference Series 305 (1), art. no. 012031 (2011).
3. Bartkowiak, A., Zimroz, R.: Data dimension reduction and visualization with application to multidimensional gearbox diagnostics data: Comparison of several methods. Diffusion and Defect Data Pt.B: Solid State Phenomena 180, 177-184 (2012).

4. C.M. Bishop, Neural Networks for Pattern Recognition. Oxford University Press (1995).
5. J. Chen et al.: Customized lifting multiwavelet packet information entropy for equipment condition identification. Smart Matter Struct 22 095022 (14pp), IOP Publishing.(2013).
6. J. Chen et al.: Planetary gearbox condition monitoring of ship-based satellite communication antennas using ensemble multivawelet analysis methods. Mech. Syst. Signal Processing (2014).
7. M. Cocconcelli, R. Zimroz, R. Rubini, W. Bartelmus: Kurtosis over energy distribution approach for STFT enhancement in ball bearing diagnostics. Condition Monitoring of Machinery in Non-Stationary Operations 2012, part I, 51-59 (2012).
8. T. Heyns, P.S. Heyns, J.P. deVilliers: Combining synchronous averaging with a Gaussian mixture model novelty detection scheme for vibration-based condition monitoring of a gearbox. Mech.Syst. Signal Process, 32 200215 (2012).
9. T. Heyns, P.S. Heyns, R. Zimroz: Combining discrepancy analysis with sensorless signal resampling for condition monitoring of rotating machines under fluctuating operations. Int. J. of Condition Monitoring 2 iss. 2, 52-58 (2012).
10. Jardine, A.K.S., Lin, D., Banjevic, D.: A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mech. Syst. Signal Process. 20, 1483-1510 (2006).
11. L. Jedlinski, J. Jonak: Early fault detection in gearboxes based on support vector machines and multilayer perceptron with a continuous wavelet transport. Applied Soft Computing Journal, in print (2015).
12. S.S. Khan and M.G. Madden: One-class classification: taxonomy of study. The knowledge Engineering Review, Cambridge Univ. Press (2014).
13. Y. Lei, J. Lin, M.J. Zuo, Z. He: Condition monitoring and fault detection of planetary gearboxes: A review. Measurement 48, 292-306 (2014).
14. L. Montechiesi, M. Cocconcelli, R. Rubini: Artificial immune system via Euclidean Distance Minimization for anomaly detection in bearings. Mech. Syst. Signal Processing. (2015), http://dx.doi.org/10,1016/j.ymssp.2015.04.017.
15. Ian T. Nabney, NETLAB, Algorithms for Pattern Recognition. Springer, London, Heidelberg (2002).
16. M.A.F. Pimentel, D.A. Clifton, L. Clifton, L. Tarassenko: A review of novelty detection. Signal Processing 99 215-249 (2014).
17. M.E. Tipping, C.M. Bishop: Probabilistic principal component analysis, J. Roy. Statist. Soc. B 61, 611-622.
18. Hanwei Zheng et al. : Dimensionality reduction by supervised neighbor embedding using Lapacian search. Computational and Mathematical Methods in Medicine (Hindawi), (2014).
19. Zimroz, R., Bartkowiak, A.: Investigation on spectral structure of gearbox vibration signals by principal component analysis for condition monitoring purposes. Journal of Physics: Conference Series 305 (1), art. no. 012075 (2011).
20. R. Zimroz, R., Bartkowiak, A.: Multidimensional data analysis for condition monitoring: features selection and data classification. The Ninth International Conference on Condition Monitoring and Machinery Failure Prevention Technologies Technologies, CM2012—MFPT2012. BINDT, 11-14 June, London. Electronic Proceedings, art no. 402, pp. 1-12 (2012).
21. Zimroz, R., Bartkowiak, A.: Two simple multivariate procedures for monitoring planetary gearboxes in non-stationary operating conditions. Mech. Syst. Signal Process. 38(1), 237-247 (2013).