Modelling a Fleet of Machines Using PCA

Farah Ankoud, Gilles Mourot, and José Ragot

Centre de Recherche en Automatique de Nancy, UMR 7039 - Nancy-Université,
CNRS, 2, avenue de la Forêt de Haye, 54516 Vandoeuvre-lès-Nancy Cedex France
{farah.ankoud, gilles.mourot, jose.ragot}@ensem.inpl-nancy.fr

Abstract. *In this paper, we propose a method to model a fleet of identical machines. These latest may work in the same or in different operating conditions. Firstly, the method consists in estimating a linear model, using PCA, based on the data collected on the machine itself. Secondly, the common parts of the models of the different machines are identified. New models for the machines are then generated taking into account the identified common parts. An academic example is finally presented to illustrate the results of the method.*

Keywords. *Modelling, fleet of machines, common parameters, optimization, constraints, PCA.*

1 Introduction

In many industrial sectors, a fleet of identical machines (nuclear power plants, wind farms, etc.) can be exploited by the same process; however those machines may work in different conditions. The problem consists in modelling each machine of the fleet using not only the knowledge issued from the machine itself, as it is done classically, but also all the data collected from the identical machines of the fleet. This problem is commonly known as multi-task learning problem.
Under linearity hypothesis of the models and considering that the environmental variables appear explicitly as explanatory variables in addition to the ones peculiar to the machines, we can suppose that the models are composed of two parts: a common one made up of the variables of the machines themselves (with the same structure and the same coefficients) and a distinct part related to the environmental variables. One can logically suppose that the explanatory power of the environmental variables is lesser than that of the other variables. The last part of the models can also be divided into two parts. The first one is formed by the variables shared by all the models but with different coefficients and the second one is made up of variables that affect the behavior of at least one machine. Identifying the common parts to the models may be of a benefit because it reduces the cost of the identification of the models of the machines, facilitates the construction of the model of a new machine of the fleet, and, on the long term, can reduce the cost of the maintenance system of the fleet of the machines.
Figure 1 gives an example of models of three machines. In this figure, a bar of the specified color of the machine number q ($q \in \{1, 2, 3\}$) at the position of the

variable x_i ($i \in \{1, \ldots, 7\}$) means that this variable constitutes an explanatory variable of the model of the q^{th} machine. The magnitude of the bar reflects the value of the coefficient of the corresponding variable.

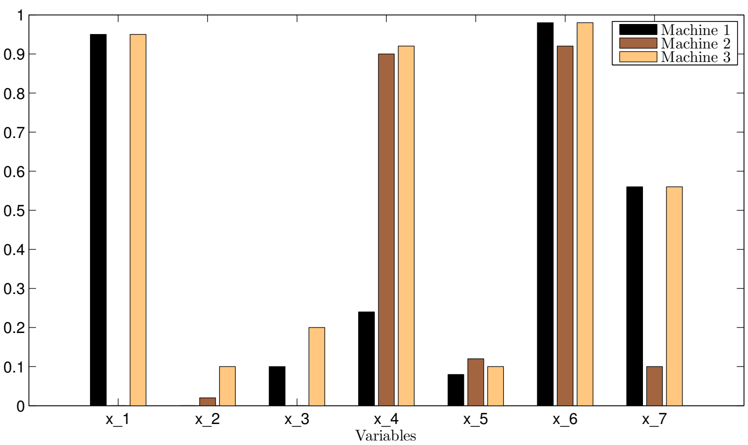

Fig. 1. Example on the fleet effect

From Fig. 1, we can conclude that: the model of the first machine is composed of variables x_1 and x_3 to x_7; the model of the second machine is composed of x_2 and x_4 to x_7 and the model of the third machine is composed of all the variables. The coefficient of the variable x_6 is common to all three models. However, although x_4 appears in the three models, its coefficient is common to models 2 and 3 but not 1. The coefficient of x_5 is common to the three models, however it is too small and one can wonder if omitting variable x_5 from the models does not give simpler models with the same results. The coefficients of x_1 and x_7 are common to the models of the machines 1 and 3.

In [3, 8], the authors present a method to identify simultaneously the structure and the coefficients of linear models based on data collected over all the machines via a generalization of the LASSO [10] regularization technique. However, the proposed method supposes that all the models have the same structure and it does not identify explicitly the common part of the models. The estimation of the parameters of two linear models having the same structure and identical coefficients is done in [6]. In [7], a method for estimating coefficients of models sharing some a priori known common part is proposed. In [9], authors present a method addressing the same problem in addition to the problem of partitioning models into different classes depending on the a priori known common part they share. When the common parts of the models are not a priori known, as it is commonly supposed, methods for identifying the models of different machines and their common parts were presented in previous works [1, 2].

However, all existing works deal with multiple-input/single-output models and suppose that each machine can be described by only one equation. Hence, two hypotheses are considered:

- the number of linear models, describing the normal behavior of the machines, is limited to one and their structures are known,
- only the output variables are measured with errors: the explanatory variables are error free.

In this paper, we make no previous hypothesis neither on the number of models of the machines nor their structures. In Sect. 2, an approach to identify models with no distinction between input and output variables is presented. These error-in-variables models take benefit of the presence of identical machines for modelling the behavior of each machine. Section 3 presents the results of the application of the proposed method on a simulation example. To conclude, Sect. 4 summarizes the results of this paper and the future perspectives.

2 PCA Modelling and Fleet Effect

Suppose there are Q identical machines in a fleet. For each machine q, a model describing its normal behavior is built. Suppose that n_q measurements for m variables, denoted x_i^q $(i = 1, \ldots, m)$, are available for each machine and set X^q the matrix containing these measurements ($X^q \in \mathbb{R}^{n_q \times m}$). The work presented in this paper enables to establish models for Q machines taking into account the fleet effect: these models share some common variables with identical coefficients. The proposed method consists of the following steps:

1. identify the models (and their number) describing the normal behavior of each machine independently from the others,
2. interpret these models to identify their structures and their common parts,
3. identify new models considering their structures their common parts,
4. validate the new models by a quantitative analysis of the residuals.

2.1 Identifying the Model of Each Machine Independently

As stated earlier, the first step of the method consists in identifying the models, describing the normal behavior of each machine, and their number. Consider the case of the machine number q and suppose that all data contained in X^q corresponds to a normal behavior of the machine. Principal Component Analysis (PCA) determines an optimal linear transformation of the data matrix X^q in terms of capturing the variation in the data [5]:

$$T^q = X^q P^q \quad \text{and} \quad X^q = T^q {P^q}^T \tag{1}$$

where $T^q \in \mathbb{R}^{n_q \times m}$ is the principal component matrix and the matrix $P^q \in \mathbb{R}^{m \times m}$ contains the principal vectors which are the eigenvectors associated to the eigenvalues λ_i^q of the covariance matrix Σ^q of X^q:

$$\Sigma^q = P^q \Lambda^q {P^q}^T \quad \text{with} \quad P^q {P^q}^T = {P^q}^T P^q = I_m \tag{2}$$

where $\Lambda^q = \text{diag}(\lambda_1^q \dots \lambda_m^q)$ is a diagonal matrix with diagonal elements in decreasing magnitude order and I_m is the identity matrix of appropriate dimension. The eigenvectors associated to null or quasi-null eigenvalues depict the linear or quasi-linear relationships between the variables $x_i^q, \forall \quad i$. Different approaches described in the literature permit to identify these eigenvectors. A simple way consists in fixing a threshold on the amplitude of the eigenvalues in order to identify those that can be considered quasi-null. We set $\tilde{P}^q$ the matrix containing the corresponding r_q eigenvectors.

2.2 Comparaison of the PCA Models

Suppose that, for each machine q, matrix $\tilde{P}^q$ is known. The column vectors of all the matrices $\tilde{P}^q$ define, after appropriate transformations, the same subspace [5]. Hence, in order to compare the elements of those different matrices, we propose to transform them such that "0" and "1" appear at specific places. More explicitly, we can easily find a transformation function g^q:

$$\begin{aligned} g^q : \mathbb{R}^{m \times r_q} &\longmapsto \mathbb{R}^{m \times r_q} \\ \tilde{P}^q &\longmapsto \overline{P}^q \end{aligned} \tag{3}$$

such that:

$$\tilde{P}^q = \begin{pmatrix} \tilde{p}_{1,1}^q & \tilde{p}_{1,2}^q & \cdots & \tilde{p}_{1,r_q}^q \\ \tilde{p}_{2,1}^q & \tilde{p}_{2,2}^q & \cdots & \tilde{p}_{2,r_q}^q \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{p}_{r_q,1}^q & \tilde{p}_{r_q,2}^q & \cdots & \tilde{p}_{r_q,r_q}^q \\ \tilde{p}_{r_q+1,1}^q & \tilde{p}_{r_q+1,2}^q & \cdots & \tilde{p}_{r_q+1,r_q}^q \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{p}_{m,1}^q & \tilde{p}_{m,2}^q & \cdots & \tilde{p}_{m,r_q}^q \end{pmatrix} \longmapsto \overline{P}^q = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \\ \overline{p}_{r_q+1,1}^q & \overline{p}_{r_q+1,2}^q & \cdots & \overline{p}_{r_q+1,r_q}^q \\ \vdots & \vdots & \ddots & \vdots \\ \overline{p}_{m,1}^q & \overline{p}_{m,2}^q & \cdots & \overline{p}_{m,r_q}^q \end{pmatrix} . \tag{4}$$

Remark 1. The choice of the variables over which we normalize the eigenvectors is not unique. In order to avoid the numerical problems, we can chose to normalize these vectors considering the variables having the biggest coefficients.

Remark 2. Denote $r = \min(r_1, \dots, r_Q)$. For a matrix $\tilde{P}^q$ for which $r_q < r$, normalizing its columns using a transformation matrix of rank r_q may not be the best transformation to realize. We may try searching for all the possible transformation matrices of rank r between the columns of $\tilde{P}^q$ and then apply each transformation to the set of vectors concerned with this transformation. However, instead of having r_q new vectors, we will have $C_{r_q}^r$ vectors to form the new matrix $\overline{P}^q$ ($C_r^{r_q}$ is the number of r-combinations over r_q).

Denote $\overline{P}_{s_q}^q$ the s_q^{th} column of $\overline{P}^q$. In order to identify all the common parts to models of the machines number q and ℓ, a comparison of $\overline{P}_{s_q}^q$ should not only be made with $\overline{P}_{s_\ell}^\ell$ but also with $-\overline{P}_{s_\ell}^\ell$ ($s_q = 1, \dots, r_q$ and $s_\ell = 1, \dots, r_\ell$).

Example 1. Assume that $\overline{P}^q_{s_q} = (0 \quad 1 \quad -0.40 \quad -1.00 \quad -1.53)^T$ and $\overline{P}^\ell_{s_\ell} = (0 \quad 1 \quad 0 \quad 1.02 \quad 1.50)^T$. By comparing these vectors, we can conclude that only the coefficient of the second variable is common to models q and ℓ. However, $-\overline{P}^\ell_{s_\ell}$ has the coefficients of the fourth and fifth variables similar to those of $\overline{P}^q_{s_q}$. A larger common part is hence identified for the models of the machines q and ℓ.

Furthermore, while examining the elements of $\overline{P}^q_{s_q}$, $\forall q$, we can notice that some of them are significantly smaller than the others. This can lead to the conclusion that the associated variables to those coefficients may be omitted from the models (by setting their coefficients to zero).
In a general way, the identification of the common coefficients of the r_q models of the machines q with the r_ℓ models of the machine ℓ can easily be done:

1. calculate the difference between the coefficients $\overline{p}^q_{s_q,i}$ and $\overline{p}^\ell_{s_\ell,i}$ $(-\overline{p}^\ell_{s_\ell,i})$, $i = 1, \ldots, m$, and use a binary coding to reflect the proximity between the coefficients: "1" if the calculated difference is smaller than a predefined threshold δ and "0" if not. Hence, a vector of distances, denoted $d^{q,\ell}_{s_q,s_\ell}$, is obtained,
2. construct an occurrence vector $I^{q,\ell}_{s_q,s_\ell}$ for each couple of vectors $\overline{P}^q_{s_q}$ and $\overline{P}^\ell_{s_\ell}$: at the position i, put "1" if both $\overline{p}^q_{s_q,i}$ and $\overline{p}^\ell_{s_\ell,i}$ are larger than ε (a prefixed threshold under which the coefficients are considered null) and "0" if not,
3. calculate $c^{q,\ell}_{s_q,s_\ell} = d^{q,\ell}_{s_q,s_\ell} \otimes I^{q,\ell}_{s_q,s_\ell}$ ($\otimes$ stands for the binary product). This vector indicates the common coefficients to $\overline{P}^q_{s_q}$ and $\overline{P}^\ell_{s_\ell}$. In this work, we are only interested in vectors having at least two non zero elements.

2.3 Identification of New Models

All the common coefficients are now known, the ones small enough to be considered null are also identified, our aim is to identify the models of the different machines taking into account the common coefficients to the shared variables and the nullity of some coefficients. In order to identify the new models, we propose to solve a total least squares problem under equality constraints:

$$\begin{aligned} &\min_\theta \frac{1}{2} \|X\theta\|_2^2 \\ &\text{s.t.} \quad H\theta = 0 \end{aligned} \tag{5}$$

where $X = \begin{pmatrix} X^1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & X^Q \end{pmatrix} \in \mathbb{R}^{N \times M}$, $\theta = \begin{pmatrix} \theta^1 \\ \theta^2 \\ \vdots \\ \theta^Q \end{pmatrix} \in \mathbb{R}^M$ with $N = n_1 + \cdots + n_Q$ is the total number of observations and $M = mQ$ is the total number of measured variables. The selection matrix $H \in \mathbb{R}^{C \times M}$ (C is the total number of constraints) is constructed in a way to select the coefficients of the models to be set identical or those to be set to zero. It is only composed of 1, 0 and -1.
In order to solve the problem in (5), we can use a constraint elimination technique composed of three steps [4]:

1. transform the optimization problem under constraints into another one of lower dimension by eliminating the constraints,
2. solve the new problem,
3. deduce the solution of the original problem.

Transform the Original Problem into Another of Lower Dimension. Matrix H is of full row rank C. We can easily find a permutation matrix $\Pi \in \mathbb{R}^{M \times M}$ ($\Pi \Pi^T = I_M$) that transforms H into $H^* = H\Pi$ with:

$$H^* = [H_1^* \quad H_2^*] \quad \in \mathbb{R}^{C \times M} \tag{6}$$

where $H_1^* \in \mathbb{R}^{C \times C}$ is a regular matrix and $H_2^* \in \mathbb{R}^{C \times (M-C)}$.
Set $\theta^* = \Pi^T \theta$ and partition it into $\theta_1^* \in \mathbb{R}^C$ and $\theta_2^* \in \mathbb{R}^{M-C}$. The constraints $H\theta = 0$ lead to:

$$\theta_1^* = -{H_1^*}^{-1} H_2^* \theta_2^* \ . \tag{7}$$

Hence, the vector θ^* can be written as:

$$\theta^* = \mathcal{T} \theta_2^* \tag{8}$$

where $\mathcal{T} = \begin{pmatrix} -{H_1^*}^{-1} H_2^* \\ I_{M-C} \end{pmatrix}$ and I_{M-C} is the identity matrix of appropriate dimension $(M-C) \times (M-C)$. This way, the original problem (5) is transformed into:

$$\min_{\theta_2^*} \frac{1}{2} \| X \Pi \mathcal{T} \theta_2^* \|_2^2 \ . \tag{9}$$

Solution of the New Problem. Denote $Z = X\Pi\mathcal{T}$, the final problem to be solved is then:

$$\min_{\theta_2^*} \frac{1}{2} \| Z \theta_2^* \|_2^2 \ . \tag{10}$$

We impose a normality constraint on the vector θ_2^*:

$$\|\theta_2^*\|_2^2 = 1 \ . \tag{11}$$

The problem to solve is nothing else than a total least squares problem for which the solution $\hat{\theta}_2^*$ is obtained as the eigenvector associated to the smallest eigenvalue of the matrix $Z^T Z$.

Solution of the Original Problem. The solution of the initial problem (5) can easily be obtained:

- using (10) and (11), an estimate $\hat{\theta}_2^*$ is found,
- an estimate $\hat{\theta}^*$ of θ^* is obtained using (8),
- an estimate of the original vector of coefficients is $\hat{\theta} = \Pi \hat{\theta}^*$.

The new models are finally validated by examining the residuals and evaluating the possible loss of information compared to the models identified on each database independently from the others as illustrated in the following example.

3 Application

Suppose that four databases are available and that each database contains the measurements of 7 different variables. The coefficients of the models identified applying the PCA on each database independently from the others, after determining their structures, are given in Table 1. We denote $\theta^q_{s_q} = \left(\theta^q_{s_q,1}, \ldots, \theta^q_{s_q,7}\right)^T$ the vector of the coefficients associated to $\overline{P}^q_{s_q}$ $(q = 1, \ldots, 4)$.

Table 1. Estimated coefficients for each database independently from the others

	i						
	1	2	3	4	5	6	7
$\overline{P}^1_1$	1	0	0.45	-1.03	0	1.04	0
$\overline{P}^1_2$	0	1	-1.18	-1.51	0	1.08	0
$\overline{P}^2_1$	1	0	0	-0.51	-0.48	0	0
$\overline{P}^2_2$	0	1	0	0.49	1.50	0.98	0
$\overline{P}^2_3$	0	0	1	-1.03	2.05	0	1.04
$\overline{P}^3_1$	1	0	0	1.52	0	0	1.05
$\overline{P}^3_2$	0	1	0	0	1.48	-2.05	1.01
$\overline{P}^3_3$	0	0	1	-1.52	1.97	-1.02	0
$\overline{P}^4_1$	1	0	0	1.50	0	-0.96	-0.99
$\overline{P}^4_2$	0	1	0	0	1.59	-1.92	1.43
$\overline{P}^4_3$	0	0	1	-0.89	1.92	0	1.18

The analysis of the contents of Table 1 considering the proximity of the coefficients and the occurrence of the variables in the models enabled us to deduce the following constraints:

$$\begin{array}{lllll}
\theta^3_{1,1} = \theta^4_{1,1} & \theta^1_{2,2} = \theta^2_{2,2} & \theta^2_{2,2} = \theta^3_{2,2} & \theta^2_{2,2} = \theta^4_{2,2} & \theta^1_{2,4} = \theta^4_{1,4} \\
\theta^2_{3,3} = \theta^3_{3,3} & \theta^2_{3,3} = \theta^4_{3,3} & \theta^3_{1,4} = \theta^4_{1,4} & \theta^2_{3,4} = \theta^4_{3,4} & \theta^1_{2,6} = \theta^4_{1,6} \\
\theta^2_{2,5} = \theta^3_{2,5} & \theta^2_{2,5} = \theta^4_{2,5} & \theta^2_{3,5} = \theta^3_{3,5} & \theta^2_{3,5} = \theta^4_{3,5} & \theta^3_{2,6} = \theta^4_{2,6} \\
\theta^1_{2,6} = \theta^2_{2,6} & \theta^3_{2,7} = \theta^4_{2,7} & \theta^2_{3,7} = \theta^4_{3,7} & &
\end{array}$$

In addition to these constraints, constraints of type $\theta^q_{s_q,i} = 0$ are also considered and the total number of constraints is 59. The new estimated coefficients taking into account all the constraints are given in Table 2. In this table, we used the notation $\hat{\hat{P}}^q_{s_q}$ to denote a second sub-model issued from $\overline{P}^q_{s_q}$ after taking into account other constraints than those used to obtain $\hat{P}^q_{s_q}$.
In order to validate the new models, we examine the variation of the residual criteria, here considered as the means of the quadratic error, compared to the ones obtained by the applying the PCA on each database independently from the others. Table 3 gives the values of the residual criteria where $\overline{J}^q_{s_q}$ (resp. $\hat{J}^q_{s_q}$

Table 2. Estimated coefficients considering the constraints

	i						
	1	2	3	4	5	6	7
$\hat{P}_1^1$	1	0	0.45	-1.02	0	1.04	0
$\hat{P}_2^1$	0	1	-1.29	-1.51	0	1.05	0
$\hat{\hat{P}}_2^1$	0	1	-1.12	-1.52	0	1.01	0
$\hat{P}_1^2$	1	0	0	-0.50	-0.48	0	0
$\hat{P}_2^2$	0	1	0	0.49	1.63	1.05	0
$\hat{P}_3^2$	0	0	1	-0.93	2.018	0	0.97
$\hat{P}_1^3$	1	0	0	1.52	0	0	1.04
$\hat{P}_2^3$	0	1	0	0	1.63	-2.00	1.06
$\hat{P}_3^3$	0	0	1	-1.54	2.01	-1.08	0
$\hat{P}_1^4$	1	0	0	1.52	0	-0.98	-1.01
$\hat{\hat{P}}_1^4$	1	0	0	1.52	0	-1.01	-1.05
$\hat{P}_2^4$	0	1	0	0	1.63	-2.00	1.06
$\hat{P}_3^4$	0	0	1	-0.93	2.01	0	0.97

and $\hat{\hat{J}}_{s_q}^q$) corresponds to the residual criterion associated to the relation $\overline{P}_{s_q}^q$ (resp. $\hat{P}_{s_q}^q$ and $\hat{\hat{P}}_{s_q}^q$), $q = 1, \ldots, 4$ and $s_q = 1, \ldots, r_q$.

Table 3. Residual criteria

q	$\overline{J}_1^q$	$\hat{J}_1^q$	$\hat{\hat{J}}_1^q$	$\overline{J}_2^q$	$\hat{J}_2^q$	$\hat{\hat{J}}_2^q$	$\overline{J}_3^q$	$\hat{J}_3^q$
1	0.053	0.053		0.048	0.052	0.049		
2	0.017	0.017		0.060	0.076		0.074	0.081
3	0.053	0.053		0.083	0.096		0.073	0.077
4	0.071	0.072	0.074	0.172	0.180		0.101	0.110

Table 3 shows no major deterioration in the residual criteria. In order to verify that the identified common parts correspond to variables contributing the most in the models, we give at Table 4 the variations in the residual criteria resulting from omitting each variable at a time from the corresponding model.
From Table 4, we can notice that the largest loss of information is due, most of the time, to the variables having a shared coefficient by at least two models. This confirms the hypothesis made in the introduction of this paper that the fleet effect appears on the variables having the most explanatory power in the models. Figure 2 shows the evolution of x_1^4 and its estimates $\overline{x}_{1,1}^4$ and $\hat{x}_{1,1}^4$, obtained respectively by the old and the new models.
Figure 2 shows that $\hat{x}_{1,1}^4$ estimates x_1^4 as good as $\overline{x}_{1,1}^4$. The same phenomenon

Table 4. Variations in the residual criteria

Model	x_3	x_4	x_5	x_6	x_7
$\overline{P}_1^1$	0.02	0.33		0.16	
$\overline{P}_2^1$	0.08	0.52		0.17	
$\overline{P}_1^2$		0.24	0.09		
$\overline{P}_2^2$		0.19	0.26	0.22	
$\overline{P}_3^2$		0.24	0.37		0.18
$\overline{P}_1^3$		0.48			0.37
$\overline{P}_2^3$			0.26	0.34	0.38
$\overline{P}_3^3$		0.40	0.32	0.14	
$\overline{P}_1^4$		0.54		0.22	0.08
$\overline{P}_2^4$			0.03	0.26	0.16
$\overline{P}_3^4$		0.96	0.52		0.41

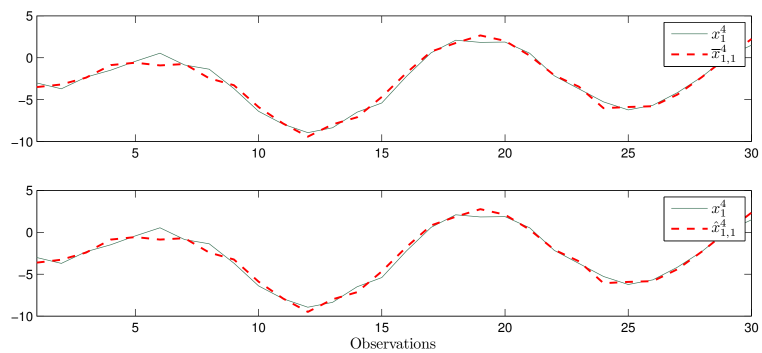

Fig. 2. Evolution of signals x_1^4, $\overline{x}_{1,1}^4$ and $\hat{x}_{1,1}^4$

is observed for the other available variables. Hence, considering the fleet effect does not affect the quality of estimation of the models of the four machines.

4 Conclusions

In this paper, we presented a method in order to identify linear models describing the normal behavior of identical machines of a fleet such that the models may share some common parts. The method supposes no previous knowledge neither on the number of models of each machine nor their structures. At first, the method proceeds by identifying the models describing the normal behavior of each machine independently from the others. Secondly, an analysis of the models is realized in order to determine their structures and their common parts. New models are then identified taking into account the identified common parts. These models are finally validated using quantitative analysis of the residuals.

The results of applying the proposed method on a simulation example are satisfying.
In perspective, a study on the best transformation matrices to apply to the principal component matrices will be conducted. The influence of omitting some of the variables from the models on the total explained variance by the models will be studied. On the long term, the contribution of this type of modelling to the diagnosis of a fleet of identical machines will be explored and the extension of the method to the dynamical models may be conceivable.

References

1. Ankoud, F., Mourot, G., Chevalier, R., Paul, N. and Ragot, J.: Estimation of a generic model for a fleet of machines. In *International Conference on Communications, Computing and Control Applications, IEEE CCCA'11*, March 3-5 (2011), Hammamat, Tunisia.
2. Ankoud, F., Mourot, G., Ragot, J., Chevalier, R. and Paul, N.: Modeling and fleet effect for the diagnosis of a system behavior. In *European Safety and Reliability Conference, ESREL 2011*, September 12-18 (2011), Troyes, France.
3. Argyriou, A., Evgeniou, T., and Pontil, M.: Multi-task feature learning. In *the 20th Annual Conference on Advances in Neural Information Processing Systems*, December 3-6 (2007), Vancouver, Canada.
4. Björck, A.: Numerical methods for least squares problems. SIAM (1996), Philadelphia.
5. Jolliffe, I. T.: Principal component analysis. Springer Series in Statistics. Springer-Verlag (2002), New York, second edition.
6. Kubokawa, T.: Double shrinkage estimation of common coefficients in two regression equations with heteroscedasticity. *Journal of Multivariate Analysis*, **67** (1998) 169–198.
7. Liu, A.: Estimation of the parameters in two linear models with only some of the parameter vectors identical. *Statistics & Probability Letters*, **29**(4) (1996) 369 – 375.
8. Lounici, K., Pontil, M., Tsybakov, A. B. and Van de Geer,S.: Taking advantage of sparsity in multi-task learning. In *Conference on Computational Learning Theory*, June 18-21 (2009), Montreal, Canada.
9. Porreca, R. and Ferrari-Trecate, G.: Partitioning datasets based on equalities among parameters. *Automatica*, **46**(2) (2010) 460–465.
10. Tibshirani, R.: Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society. Series B*, **58**(1) (1996) 267–288.