



### **Science Arts & Métiers (SAM)**

is an open access repository that collects the work of Arts et Métiers Institute of Technology researchers and makes it freely available over the web where possible.

This is an author-deposited version published in: <https://sam.ensam.eu>  
Handle ID: <http://hdl.handle.net/10985/9177>

#### **To cite this version :**

Benjamin LAMOUREUX, Jean-Rémi MASSÉ, Nazih MECHBAL - Diagnostics of an Aircraft Engine Pumping Unit Using a Hybrid Approach based-on Surrogate Modeling - In: IEEE Conference on Prognostics and Health Management, United States, 2013 - IEEE PHM 2013 - 2013

Any correspondence concerning this service should be sent to the repository

Administrator : [scienceouverte@ensam.eu](mailto:scienceouverte@ensam.eu)



# Diagnostics of an Aircraft Engine Pumping Unit Using a Hybrid Approach based-on Surrogate Modeling

Benjamin Lamoureux, Jean-Rémi Massé  
R&D department – Regulation Systems Division  
Safran Snecma  
Moissy-Cramayel, France  
benjamin.lamoureux@sneema.fr

Nazih Mechbal  
DYSCO – Dynamic of Structures, Systems and Control  
Arts et Métiers ParisTech - PIMM UMR 8006  
Paris, France  
nazih.mechbal@ensam.eu

**Abstract**—This document introduces a hybrid approach for fault detection and identification of an aircraft engine pumping unit. It is based on the complementarity between a model-based approach accounting for uncertainties aimed at quantifying the degradation modes signatures and a data-driven approach aimed at recalibrating the healthy syndrome from measures. Because of the computational time costs of uncertainties propagation into the physics based model, a surrogate modeling technic called Kriging associated to Latin hypercube sampling is utilized. The hybrid approach is tested on a pumping unit of an aircraft engine and shows good results for computing the degradation modes signatures and performing their detection and identification.

**Keywords**-Prognostics and Health Management, hybrid approach, detection, identification, degradations, surrogate modeling, kriging, Latin hypercube sampling

## I. INTRODUCTION

In recent years, availability was raised to the status of strategic challenge for many industries mainly because a great amount of the average costs is attributable to Maintenance, Repair and Overhaul (MRO) and Delays and Cancellations (D&C). For example, in aeronautics, they represent a quarter of the total expenses for airliners. In order to optimize availability, novel MRO strategies are under development, based on failure anticipation and real-time optimization of maintenance plan. Most of them are issued from Prognostics and Health Management (PHM), and the most proven one is Conditioned Based Maintenance (CBM) [1].

Typically, PHM is either model-based [2] or data-driven [3] but unfortunately, both of these methods have their limitations. On the one hand, instead of its good precision in degradation modeling and sensitivity analysis, model-based approaches have some difficulties adapting to environmental variations and unclassifiable, but significant, perturbations. On the other hand, data-driven approaches are suitable ways to take into account uncertainties but they show their limit when confronted to the problem of lack of available measured data. Actually, this problem is dual: First, measured data are needed in the upstream development stages in order to perform the learning step of data-driven algorithms but in practice they are

not available before the entry into service, which is too late if the system suffers from prohibitive controller retrofit costs. Then, measured data of degraded states are needed for creating a degradation signature database but if the system has a good reliability, these data are not numerous enough to manage the stochastic nature of PHM.

As both approaches have their weaknesses, it is interesting to combine them in a hybrid approach. This kind of approach is quite uncommon in the scientific community, but it has already been applied. For example, Garga et al. [4] have described a hybrid reasoning approach that integrates domain knowledge with test and operational data from an industrial gearbox and Kumar et al. [5] have proposed a hybrid prognostics framework aimed at performing the PHM of electronic products. In this paper, a hybrid approach is also used. It combines the advantages of following approaches: the model-based one to quantify the degradation modes signatures and their related damage model in upstream development stages and the data-driven one to recalibrate the healthy syndrome in downstream stage.

The remainder of the document is organized as following: In the next section, the hybrid approach is introduced. The second section addresses surrogate modeling and particularly Kriging and section IV and V are dedicated to an application on a pumping unit of an aircraft engine.

## II. HYBRID APPROACH

In this section, the monitoring of a system  $\mathcal{S}$  using a hybrid approach is addressed. Its deterministic numerical model is  $\mathcal{f}$  so that:

$$(\eta_1, \dots, \eta_h) = \mathcal{f}(\mathbf{U}, \beta_1, \dots, \beta_p, \gamma_1, \dots, \gamma_d) \quad (1)$$

where  $\mathbf{U} \in \mathbb{R}^{n \times k}$  is the matrix of the model inputs (variables evolving with time during a run),  $k$  is the number of samples,  $\eta_1, \dots, \eta_h$  are the Health Indicators (HI) values,  $\beta_1, \dots, \beta_p$  are the context parameters (constant during a run) and  $\gamma_1, \dots, \gamma_d$  are the degradation values. A degradation is defined by its mode  $j$  (index of the degradation parameter) and its magnitude

$\omega$  (value of the parameter). The nominal magnitude, or healthy magnitude of a degradation mode  $j$  is written  $\omega_0^j$ .

### A. Model-Based Framework

#### 1) Uncertainties Management

Taking into account uncertainties consists in replacing the deterministic parameters  $\beta_1, \dots, \beta_p$  by the random variables  $\beta_1, \dots, \beta_p$ . They can be characterized by their probability density functions (**pdf**). A pdf is completely defined by its type (uniform, normal, exponential...) and its parameter vector  $\Theta$ . These pdf are often determined through experience feedback or expertise.

Thanks to uncertainties, it is possible to compute stochastic HI distributions from a deterministic model by randomly sampling model parameters values according to their pdf. This operation is called uncertainties propagation [6]. Many tools are available for uncertainties propagation but the most famous and proven one is the Monte-Carlo algorithm. In this case, Equ.1 can be written in its stochastic form:

$$(\eta_1, \dots, \eta_h) = \mathcal{f}(\mathbf{U}, \beta_1, \dots, \beta_p, \gamma_1, \dots, \gamma_d) \quad (2)$$

where  $\eta_1, \dots, \eta_h$  are random variable corresponding to the HI distributions.

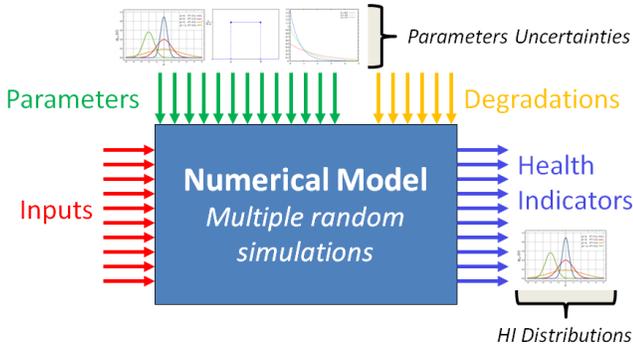


Figure 1. Model-based framework scheme

#### 2) Simulated Health Indicators

The purpose of the model-based approach is to create simulated data to overcome the lack of measured data in upstream development stages. Simulated pdf Parameters (**SP**) are defined for each triplet (HI  $i$ , mode  $j$ , magnitude  $\omega$ ) as following:

$$\mathbf{SP}(i, j, \omega) = \left( {}^i_j\theta_1(\omega), \dots, {}^i_j\theta_r(\omega) \right)^T \quad (3)$$

where  $r$  is the number of parameters for the chosen type of pdf and  ${}^i_j\theta_k(\omega), k = 1, \dots, r$  is the value of the pdf parameters of  $\eta_i$  obtained by Maximum Likelihood Estimation (**MLE**) for  $\mathcal{f}(\mathbf{U}, \beta_1, \dots, \beta_p, \omega\delta_{1j}, \dots, \omega\delta_{jj}, \dots, \omega\delta_{dj})$  with  $\delta_{kl}$  Kronecker delta. It can be noticed that if  $\omega$  is equal to 0, then the SP corresponds to the parameter vector of a healthy distribution.

#### 3) Damage Models

Fixing the HI index  $i$  and the degradation mode  $j$ , it is supposed that the SP has been computed for  $k$  values

of magnitude  $\omega_1, \dots, \omega_k$ . From these values, it is possible to construct regression models for  ${}^i_j\theta_k(\omega), k = 1, \dots, r$  via Least Square Estimation (**LSE**) as shown in Equ.4.

$${}^i_j\theta_k(\omega) = \mathcal{g}_k^T(\omega) {}^i_j\lambda_k \quad (4)$$

where  $\mathcal{g}_k$  are the regression functions and  ${}^i_j\lambda_k$  the regression coefficients. The Damage Model of HI  $i$  and degradation mode  $j$  is defined as the following matrix:

$${}^i_j\mathbf{D} = \left[ {}^i_j\lambda_1 \quad \dots \quad {}^i_j\lambda_r \right] \quad (5)$$

Thus,  $\mathcal{g}^T(\omega) {}^i_j\mathbf{D}$ ,  $\mathcal{g} = (\mathcal{g}_1, \dots, \mathcal{g}_r)$  is the LSE of  $\mathbf{SP}(i, j, \omega)$ .

#### 4) Degradation Mode Simulated Signatures

Considering a given degradation mode  $j$  and its magnitude  $\omega$ , the Simulated Syndrome (**SSy**) is a matrix defined as following:

$$\mathbf{SSy}(j, \omega) = \left[ \mathcal{g}^T(\omega) {}^1_j\mathbf{D} \quad \dots \quad \mathcal{g}^T(\omega) {}^h_j\mathbf{D} \right]^T \quad (6)$$

where  $h$  is the number of HI. Then, the simulated healthy syndrome (**ShSy**) is defined whatever the degradation mode  $j$  as following:

$$\mathbf{ShSy} = \mathbf{SSy}(j, \omega_0^j) \quad (7)$$

The Maximal Admissible Magnitude (**MAM**) is defined as the magnitude for which the system reaches a failure state and is written  $\omega_{MAM}^j$  for degradation mode  $j$ . The Signature (**Sg**) of a degradation mode  $j$  is the difference between simulated MAM and simulated healthy syndromes. It is defined as following:

$$\mathbf{Sg}(j) = \mathbf{SSy}(j, \omega_{MAM}^j) - \mathbf{ShSy} \quad (8)$$

### B. Data-Driven Part

The data-driven process consists first in an extraction of real HI. Then, supposing that the data correspond to a healthy state of the system, which is normally the case just after the entry into service, a MLE of the measured HI distributions pdf parameters is performed to compute the measured healthy syndrome (**MhSy**) which is defined as following:

$$\mathbf{MhSy} = \left[ \mathbf{MhSy}_1 \quad \dots \quad \mathbf{MhSy}_h \right]^T \quad (9)$$

where  $\mathbf{MhSy}_k = (\theta_1^k, \dots, \theta_r^k)^T$ ,  $r$  is the number of parameters for the chosen type of pdf and  $\theta_k^i, k = 1, \dots, r$  are the MLE values of the pdf parameters of the measured distribution of HI  $i$ . Provided both MhSy and damage model are available, it is possible to calculate the Recalibrated Syndromes (**RSy**) defined as following:

$$\mathbf{RSy}(j, \omega) = \mathbf{SSy}(j, \omega) - \mathbf{ShSy} + \mathbf{MhSy} \quad (10)$$

The recalibrated failure syndrome (**RfSy**) is defined for a given degradation mode  $j$  as the syndrome corresponding to the MAM:

$$\mathbf{RfSy}(j) = \mathbf{RSy}(j, \omega_{MAM}^j) \quad (11)$$

The RfSy is particularly useful for the detection process because it is derived from the true healthy values of HI.

### C. Detection and Identification Performance

In order to quantify the performance of the detection and identification processes, some performance metrics called Key Performance Indicators (**KPI**) need to be defined.

For detection, the KPI are usually the False Positive ratio (**FP**) and the True Positive ratio (**TP**) [7] computed for the best threshold. FP is equivalent to the rate of type I error or false alarms and TP is equivalent to the rate of true alarms. The best threshold is defined as the threshold for which the couple (FP; TP) is the further from the no-discrimination line and it can be determined graphically from the analysis of Receiver Operating Characteristic (**ROC**) curves. An example of a ROC curve from the Matlab© dataset is given in figure 2. FP and TP are computed for each degradation mode  $j$  between **ShSy** and **RfSy(j)**.

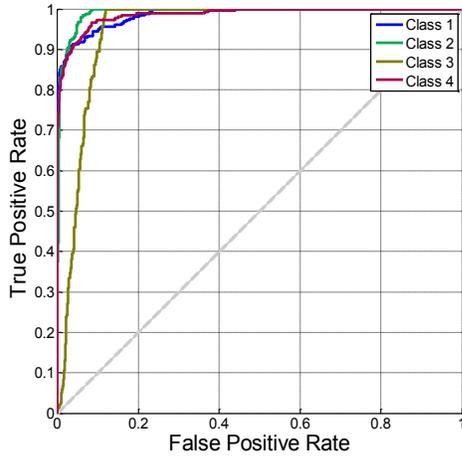


Figure 2. Example of ROC curve from matlab. The grey dashed line is the no-discrimination line.

For identification, the KPI is called the Distinguishability vector, written  $\mathbf{D} = (D_1, \dots, D_r)^T$ , with coefficient  $D_k$  defined for two signatures as following:

$$D_k(i, j) = \cos^{-1} \left( \frac{\mathbf{Sg}_k(i)^T \mathbf{Sg}_k(j)}{\|\mathbf{Sg}_k(i)\| \|\mathbf{Sg}_k(j)\|} \right) \quad (12)$$

where  $\mathbf{Sg}_k(i)$  and  $\mathbf{Sg}_k(j)$   $k = 1, \dots, r$  design the  $k^{\text{th}}$  column of the signature matrices. Typically, specifications require  $TP > 80\%$  and  $FP < 5\%$ . The vector  $\mathbf{D}$  is a way to quantify the angle between two signatures so in fine the potential for discrimination between degradation modes.

Finally, it is obvious that in order to compute the damage models, a huge amount of measures is needed (c.f. Equ.4.). However, if the numerical physics based model of the system is expensive in terms of computation time, it is not possible to run all the needed simulations. In the next section, a solution is introduced through surrogate modeling whose aim is basically to construct a “low cost” model of model.

## III. SURROGATE MODELING AND KRIGING

### A. Surrogate Modeling Basics

When physics based models are expensive in terms of computation time, it is interesting to use surrogate modeling. For instance, it is used to optimize aerospace design [8]. A surrogate model is a mathematic function:

- Of negligible computation cost
- Approximating the physics based model responses

A surrogate modeling process is composed of the following steps, also presented in figure 3:

- Determination of the variation range of influential parameters and their pdf  $\rightarrow$  uncertainty management
- Choice of the surrogate model type (ex: Kriging)
- Design sample construction (Design of Experiment)
- Scatter plot of the learning sample (Visual Checking)
- Surrogate model construction

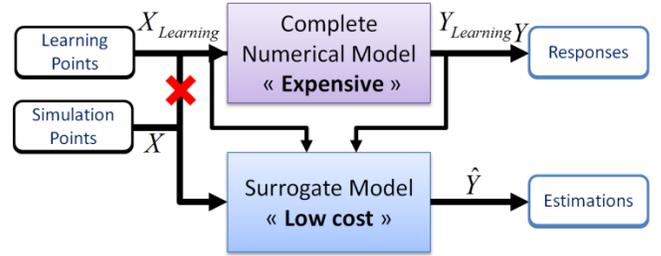


Figure 3. Principle of surrogate modeling. The “low cost” model is constructed from  $X_{Learning}$  and  $Y_{Learning}$

### B. Design of Experiment

As shown in figure 3, some learning points, also called design sites, are requested in order to build a surrogate model. In order to optimize the sites selection, Design of Experiment (**DOE**) are constructed. Even if the type of DOE to be used depends on the surrogate model, the choice is typically made among low discrepancy DOE in order to browse in an exhaustive way the variation range of model parameters space.

For instance, Latin Hypercube Sampling (**LHS**) is widely used with Kriging in order to create design sites from multidimensional pdf of  $p$  variables. LHS consists in the following steps:

1. Discretization of the  $p$  pdf into  $n$  intervals with equal probability. Intervals are noted:  $\mathcal{J}_1^p, \dots, \mathcal{J}_n^p$ .
2. Creation of a permutation matrix  $\mathbf{A} \in \mathbb{N}^{n \times p}$ :

$$\mathbf{A} = \begin{bmatrix} \sigma^1(1) & \dots & \sigma^p(1) \\ \vdots & \ddots & \vdots \\ \sigma^1(n) & \dots & \sigma^p(n) \end{bmatrix} \quad (13)$$

where  $\sigma^1, \dots, \sigma^p$  are permutations of  $\llbracket 1; n \rrbracket$

3. Random sampling according to the different pdf in order to construct the DOE.

$$DOE = \begin{bmatrix} rand(\mathcal{J}_{\sigma^1(1)}^1) & \cdots & rand(\mathcal{J}_{\sigma^p(1)}^p) \\ \vdots & \ddots & \vdots \\ rand(\mathcal{J}_{\sigma^1(n)}^1) & \cdots & rand(\mathcal{J}_{\sigma^p(n)}^p) \end{bmatrix} \quad (14)$$

where *rand* is a function drawing randomly a value according to an interval pdf.

The main advantage of LHS is that it can produce low discrepancy DOE not only in the global space but also in each single dimension. However, the quality of LHS DOE is not homogeneous and it can be interesting to use an optimization criterion such as in [9].

### C. Kriging

Originally, Kriging was developed by the mining engineer Daniel Krige for interpolation in geostatistics before being applied to numerical modeling. See [10] for a recent survey. A Kriging model can be written as following:

$$Y(\mathbf{x}) = \mathbf{f}^T(\mathbf{x})\mathbf{b} + Z(\mathbf{x}) \quad (15)$$

where  $\mathbf{x}$  is a point in a  $d$ -dimensional input space,  $\mathbf{f}^T(\mathbf{x})\mathbf{b}$  is a regression model and  $Z$  is a Gaussian process of mean zero and covariance  $\sigma^2\mathcal{R}(\boldsymbol{\theta}, \mathbf{x}_i, \mathbf{x}_j)$  with  $\mathcal{R}$  an assumed correlation function between outputs and inputs so that:

$$\mathcal{R}(\boldsymbol{\theta}, \mathbf{x}_i, \mathbf{x}_j) = \prod_{k=1}^d \mathcal{R}_k(\theta_k, x_{ik}, x_{jk}) \quad (16)$$

The Kriging model parameters  $\boldsymbol{\theta}$ ,  $\mathbf{b}$  and  $\sigma^2$  are generally computed by MLE. Some examples of correlation functions are given in table I. These functions imply that  $Y(\mathbf{x}_i)$  and  $Y(\mathbf{x}_j)$  are more correlated as their input locations  $\mathbf{x}_i$  and  $\mathbf{x}_j$  are closer. The choice of the correlation function is of paramount importance because it determines the quality of the Kriging model estimations. It depends on the characteristics of the model, for example, a Gaussian correlation suits generally well to linear models whereas exponential correlation is more adapted to non-linear models.

TABLE I. TYPES OF CORRELATION FUNCTIONS

Correlation Type	$\mathcal{R}_k(\boldsymbol{\theta}_k, x_{ik}, x_{jk})$
Exponential	$\exp(-\theta_k  x_{jk} - x_{ik} )$
Gaussian	$\exp(-\theta_j  x_{jk} - x_{ik} ^2)$
Exponential - Gaussian	$\exp(-\theta_j  x_{jk} - x_{ik} ^{\theta_{n+1}}), \quad 0 < \theta_{n+1} \leq 2$

From  $n$  observations  $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)^T$  corresponding to design sites  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$ , Kriging uses Best Linear Unbiased Predictor (BLUP) criterion minimizing the Mean Squared Error (MSE) of the predictor. For a point  $\mathbf{x}_{n+1}$ , the Kriging predictor is:

$$\hat{Y}(\mathbf{x}_{n+1}) = \mathbf{f}^T(\mathbf{x}_{n+1})\mathbf{b} + \mathbf{r}(\mathbf{x}_{n+1})^T \mathbf{R}^{-1}(\mathbf{Y} - \mathbf{F}\mathbf{b}) \quad (17)$$

where  $\mathbf{b}$  is the matrix of the regression coefficients,  $\mathbf{R}$  is the correlation matrix,  $\mathbf{F} = (\mathbf{f}(\mathbf{x}_1), \dots, \mathbf{f}(\mathbf{x}_n))^T$  and  $\mathbf{r}$  is the correlation function between  $\mathbf{x}_{n+1}$  and design sites so that:

$$\mathbf{r}(\mathbf{x}_{n+1}) = [\mathcal{R}(\boldsymbol{\theta}, \mathbf{x}_1, \mathbf{x}_{n+1}) \cdots \mathcal{R}(\boldsymbol{\theta}, \mathbf{x}_n, \mathbf{x}_{n+1})]^T \quad (18)$$

It can be proven that if  $\mathbf{x}_{n+1}$  coincides with a design site, the predictor equals the observation. Thus, the Kriging predictor is an exact interpolator. Eventually, it is possible to calculate the variance of the prediction  $\Sigma^2$  at any point  $\mathbf{x}$ :

$$\Sigma^2(\mathbf{x}) = \sigma^2 \left( 1 - \mathbf{r}^T \mathbf{R}^{-1} \mathbf{r} + \frac{(1 - \mathbf{1}^T \mathbf{R}^{-1} \mathbf{r})^2}{\mathbf{1}^T \mathbf{R}^{-1} \mathbf{1}} \right) \quad (19)$$

Finally, the Kriging predictor has three main advantages: it is a BLUP, it is an exact interpolator on design sites and it is capable of estimating its own prediction variance. A Kriging toolbox named Design and Analysis of Computer Experiments (DACE) is available for Matlab© [11]. It includes an efficient algorithm for the estimation of Kriging hyperparameters  $\boldsymbol{\theta}$ ,  $\mathbf{b}$  and  $\sigma^2$ .

### D. Kriging Model Validation

The main drawback of Kriging is that if the construction step is not rigorous enough, the model can rapidly provide aberrant results. That is why some validation criteria must be used. For example, a criterion based on cross validation can be used [12]. The cross validation is based on the computation of the surrogate model prediction error  $e_{-i}$  with:

$$e_{-i} = f_i - \tilde{f}_{-i} \quad (20)$$

where  $f_i$  is the exact output value of design point  $i$  and  $\tilde{f}_{-i}$  is the prediction given by the Kriging metamodel learnt without design point  $i$ . The Cross validation Root Mean Squared Error  $RMSE_{CV}$  is defined as follows:

$$RMSE_{CV} = \sqrt{\frac{\sum_{i=1}^n (f_i - \tilde{f}_{-i})^2}{n}} \quad (21)$$

where  $n$  is the number of design points. The closer to zero the  $RMSE_{CV}$ , the better the Kriging model.

The cross validation curve is defined as the curve of  $\tilde{f}_{-i}$  in function of  $f_i$ . The closer to the curve  $\tilde{f}_{-i} = f_i$  the scatterplot, the more efficient the Kriging model.

## IV. APPLICATION: SYSTEM MODELING

### A. System Presentation

The studied system is a pumping unit of an aircraft engine fuel system [13]. This system is composed of a centrifugal low pressure pump and a gear high pressure pump. The location of these pumps in the fuel system is presented in figure 4.

On figure 4, it can be noticed that five different valves are presented. Three of these valves are interesting piloted by the pressure difference  $\Delta P = P_{HP} - P_{LP}$ :

- **BSV**: Burning Stage Valve to switch between 1 and 2 injector lines.
- **TBV**: Transient Bleed Valve to produce a discharge in the high-pressure compressor when needed.
- **HPSOV**: High Pressure Shut Off Valve to maintain the pressurization of the system and turn on or shut off the fuel injection.

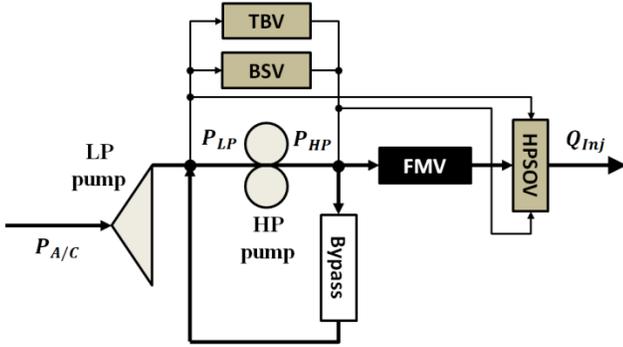


Figure 4. Aircraft Fuel System simplified Scheme with  $P_{A/C}$  aircraft pressure supply,  $P_{LP}$  and  $P_{HP}$  respectively low and high pressures of the system and  $Q_{inj}$  injection flow.

### B. System Modeling

Some related works on the subject have addressed the issue of modeling a gear pump and its degradation modes. For example, Casoli et al. [14] have proposed a method to model a gear pump with AMESim<sup>®</sup> and Fritz and Scott have developed a wear model [15]. In this paper, the model is built with the software AMESim<sup>®</sup>, based on the Bond Graph theory. The pump outlet flow is then expressed as:

$$Q = \alpha \cdot \eta_{vol} \cdot dis \cdot \omega \quad (22)$$

where  $Q$  is the pump outlet flow,  $dis$  the pump displacement,  $\eta_{vol}$  the volumetric efficiency,  $\omega$  the pump rotation speed and  $\alpha$  an empirical constant. The equation used to compute the volumetric efficiency is:

$$\eta_{vol} = 1 - \left\{ \left[ 1 - \left( \beta - \frac{\gamma \cdot \Delta P}{\omega} \right) \right] \cdot \left[ 1 + \delta \cdot \frac{T_{fuel}}{\omega} \right] \right\} \quad (23)$$

where  $\Delta P$  is the pressure drop between pump inlet and pump outlet,  $T_{fuel}$  is the fluid temperature at the pump inlet and  $\beta, \gamma, \delta$  are empirical constant values.

The other components of the system are modeled from the following standard AMESim<sup>©</sup> libraries: hydraulic, hydraulic component design, 1-D mechanical. Interface Blocks are also used for running cosimulation with Matlab-Simulink<sup>©</sup>.

### C. System Analysis

#### 1) Model Parameters

The modeling of the previously introduced system is based on 9 parameters divided into 3 context parameters and 6 degradation parameters. One can notice that this problem is a quite simple one because these types of models are generally composed of tens or hundreds of parameters.

#### 2) Uncertainties quantification

For each of the 3 context parameters, uncertainties must be quantified in order to be propagated into the model. In this application, without loss of generality, it is supposed that all the context parameters pdf follow a Gaussian law. The context parameters are the fuel temperature  $T_{fuel}$ , the inlet fuel pressure  $P_{A/C}$  and the injection pressure  $P_{inj}$ . The quantification is performed by analyzing experiment feedback of similar systems. The result is given in Equ.24.

$$\begin{cases} T_{fuel} \sim \mathcal{N}(\mu = 14, \sigma^2 = 144) \\ P_{A/C} \sim \mathcal{N}(\mu = 2, \sigma^2 = 1e^{-2}) \\ P_{inj} \sim \mathcal{N}(\mu = 1, \sigma^2 = 1e^{-4}) \end{cases} \quad (24)$$

#### 3) Degradation Modes

Thanks to experience feedback and experts advices, six degradation modes were identified. These degradation modes can make the system non-compliant with its functional requirements and lead to flight cancellations. Among this modes, two of them were declared critical through a failure analysis: pump internal leakage and pump external leakage. Indeed, these modes have very important occurrence rates compare to others.

To simulate the influence of all the potential faults, they are introduced into the AMESim<sup>®</sup> model. For example, internal leakage is modeled by a diaphragm with a variable section between pump inlet and pump outlet (Figure 5) and external leakage is modeled by a diaphragm between pump outlet and external tank at atmospheric pressure (Figure 5).

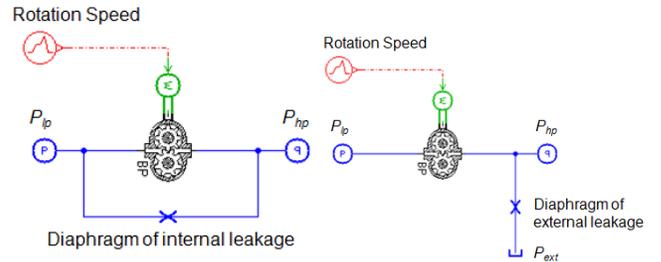


Figure 5. AMESim<sup>©</sup> modeling of degradation modes internal leakage (left) and external leakage (right)

The Maximal Admissible Magnitude (**MAM**) of a degradation mode is defined as the magnitude for which the system reaches a failure state. The following table introduces all the degradation modes and their associated MAM.

TABLE II. DEGRADATION MODES OF THE PUMPING UNIT

Degradation Mode	Degradation Direction	MAM*
1. HP pump Internal Leakage diameter	Diameter increase	1.5mm
2. HP pump External Leakage diameter	Diameter increase	1mm
3. TBV Stiction	Stiction Increase	15N
4. TBV Internal Leakage diameter	Diameter increase	0.05mm

Degradation Mode	Degradation Direction	MAM*
5.HPSOV Stiction	Stiction Increase	20N
6.HPSOV Internal Leakage diameter	Diameter increase	0.1mm

\* Maximal Admissible Magnitude

#### 4) Health Indicators

In order to monitor the system status, three different HI are defined. They correspond to the rotation speed of the pump at the opening of the valves i.e. the rotation speed for which hydraulic power is high enough to open respectively the BSV, the TBV and the HPSOV. They are named  $wBSV$ ,  $wTBV$  and  $wHPSOV$ . The simulation time of one call of the model is about  $t_{sim} = 120s$  so it is too expensive to run a Monte-Carlo algorithm for uncertainties propagation. Indeed, it would need  $n$  Monte-Carlo simulations for each of the  $m$  magnitudes of the  $d$  degradation modes so if  $(n, m, d) = (2000, 20, 6)$ , it would represent  $n * d * m * t_{sim} = 333$  day with a dual core, 2.93 GHz CPU. The next section presents how Kriging can be used to decrease these computational time costs.

#### D. Kriging Model

The Kriging surrogate model gives the estimation of the values of  $wBSV$ ,  $wTBV$  and  $wHPSOV$  in function of both the context parameters and the degradations so  $\mathbf{x}$  is replaced by  $(wBSV, wTBV, wHPSOV)^T$  in Equ.15. The design sites are constructed via a Latin hypercube sampling with  $n = 400$ . Then, the Kriging model is built using a first degree polynomial regression model and an exponential correlation. Eventually, Monte-Carlo algorithms of 10000 runs are run for each triplet (HI  $i$ , mode  $j$ , magnitude  $\omega$ ) with linearly growing magnitudes  $\omega$ .

In order to validate the Kriging model, the Kriging prediction error  $e_{-i}$  is computed for each design site  $i$ . The cross validation curve for  $wBSV$  is shown in figure 6. The curve is rather close to the  $\tilde{f}_{-i} = f_i$  curve, which indicates that the Kriging model can be validated.

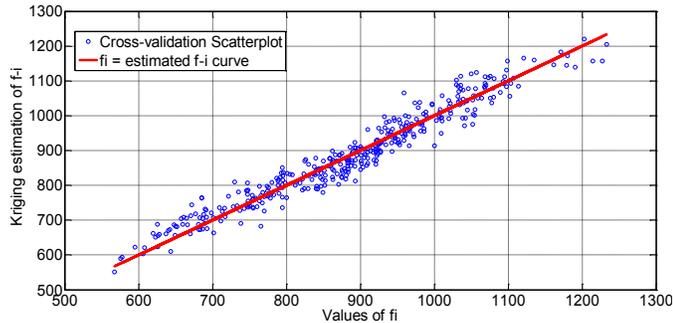


Figure 6. Cross-Validation curve for HI wBSV.

#### E. Damage Model

In this document, linear regression is used for damage models so  $\mathcal{G}_k(\omega) = (\omega, 1)^T$ . Without loss of generality, the Gaussian hypothesis is also made for HI so  $r = 2$ . On figure 7,

the distribution of  $wHPSOV$  (HI 3) for degradation mode 6 is shown for 10 linearly growing magnitudes.

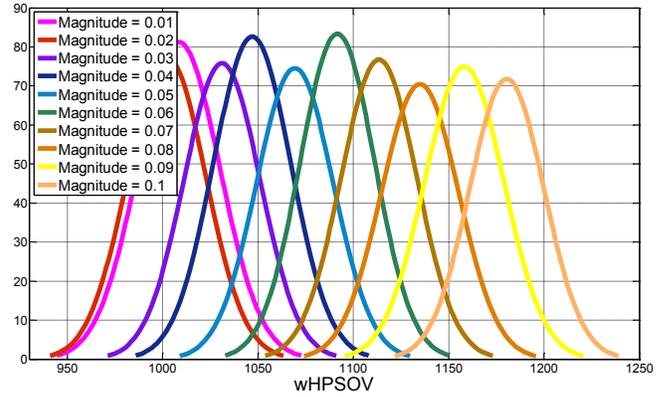


Figure 7. Progression of the  $wHPSOV$  simulated distribution for linearly growing magnitudes of degradation mode 6 (HPSOV internal leakage)

The associated damage model can be computed from the previous distributions by using Equ.4 and Equ.5 with  $r = 2$ :

$$\mathfrak{z}D = \begin{bmatrix} 2.1e^3 & -8.54 \\ 970 & 20.7 \end{bmatrix} \quad (25)$$

The following figure 8 shows how the damage model of Equ.25 is computed for parameter 1. This figure shows that parameter 1 is relevant to represent the system health status because it evolution with the degradation mode magnitude has a linear trend. However, parameter 2 does not shows any trend and consequently it is not representative of the system health status. As the system is composed of 6 degradation modes and 3 HI, 18 damage models are calculated.

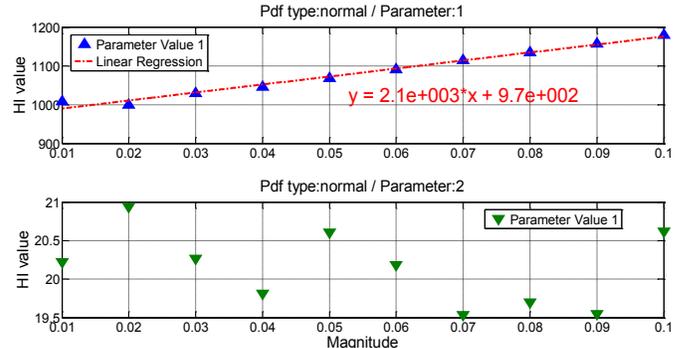


Figure 8. Example of Damage Model Computation via linear regression of  $wHPSOV$  versus Magnitude for parameter 1

## V. APPLICATION: RESULTS

### A. Degradation Modes Signatures

Thanks to the damage models, it is possible to compute the signatures of the degradation modes (c.f. Equ.8). The results for all the degradation modes are the following ones, with  $ShSy$  the simulated healthy syndrome:

$$ShSy = \begin{bmatrix} 714 & 61.7 \\ 903 & 20.9 \\ 1008 & 20.2 \end{bmatrix} \quad (26)$$

$$\begin{aligned}
\mathbf{Sg}(1) &= \begin{bmatrix} 181 & -3.0 \\ 154 & 0.6 \\ 257 & 2.7 \end{bmatrix}, \mathbf{Sg}(2) = \begin{bmatrix} 83 & -2.5 \\ 82 & -0.2 \\ 82 & 2.1 \end{bmatrix} \\
\mathbf{Sg}(3) &= \begin{bmatrix} 28 & -0.9 \\ 19 & 0.7 \\ 20 & -1.6 \end{bmatrix}, \mathbf{Sg}(4) = \begin{bmatrix} 7 & -5.4 \\ -28 & -1.3 \\ 192 & -1.6 \end{bmatrix} \\
\mathbf{Sg}(5) &= \begin{bmatrix} -30 & -2.8 \\ 12 & 4.4 \\ -22 & 0.6 \end{bmatrix}, \mathbf{Sg}(6) = \begin{bmatrix} -7 & -6.2 \\ 39 & -1.7 \\ 8 & -2.4 \end{bmatrix}
\end{aligned}$$

The bar plot of the first component of these syndromes (the mean) are presented in figure 9. The results given in this figure are obtained through simulations only but so it is necessary to replace them in a realistic framework. Thus, experts were asked to validate or invalidate these results by a physical approach. Their conclusions are the following:

- Degradation modes 1 and 2 generate an increase of all the HI, which is physically explained by the fact that pump leakages create a loss of outlet fuel flow so a reduction of hydraulic power.
- Degradation mode 4 generate a huge variation of HI 3 which can be explained by the fact that because of its internal leakage, the TBV utilizes more flow and so the remaining flow is not sufficient to move the HPSOV.
- The other variations are almost negligible and experts have not seen any inconsistency.

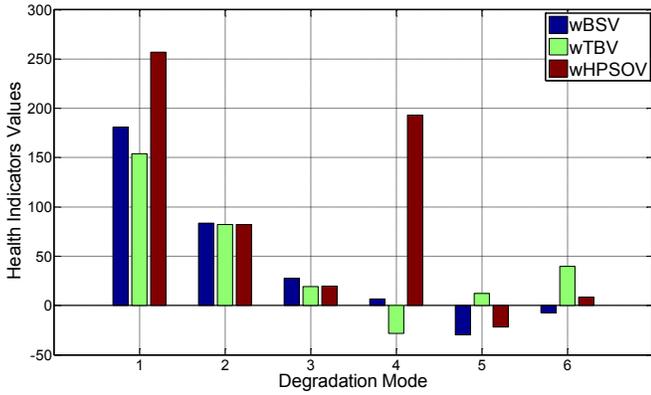


Figure 9. Degradation mode signatures for the 3 HI.

### B. Recalibrated Healthy Syndrome

For this application, 195 datasets of HI measured on the real operational system are available. As they have been computed at the beginning of its operational life, it is supposed that they correspond to healthy behaviors. Figure 10 compares the measured and simulated healthy distribution of wBSV with their respective Gaussian fitting obtained by MLE. It can be noticed that despite their similarity, both the simulated mean and standard deviation are a little bit underestimated. The purpose of the syndrome recalibration is to fix this approximation.

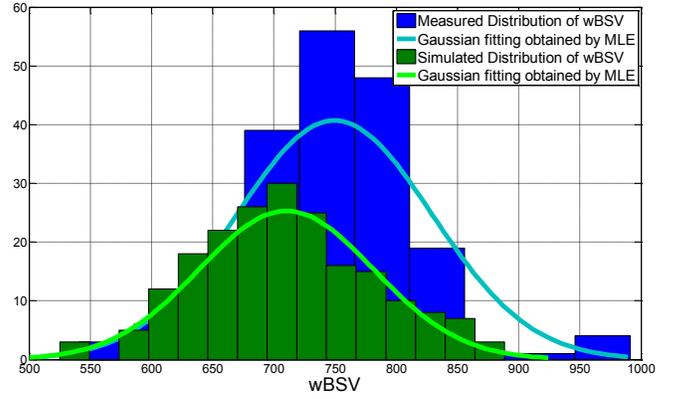


Figure 10. Comparison between measured and simulated healthy distributions of HI wBSV with their gaussian fitting

The syndrome recalibration is done by using Equ.10 and knowing the  $MhSy$  and the maximal admissible degradation, it is possible to compute the  $RfSy$ :

$$\begin{aligned}
\mathbf{MhSy} &= \begin{bmatrix} 750 & 79.9 \\ 945 & 15.2 \\ 1060 & 23.5 \end{bmatrix} \\
\mathbf{RfSy}(1) &= \begin{bmatrix} 931 & 76.9 \\ 1099 & 15.8 \\ 1317 & 26.2 \end{bmatrix}, \mathbf{RfSy}(2) = \begin{bmatrix} 833 & 77.4 \\ 1027 & 15.0 \\ 1142 & 25.6 \end{bmatrix} \\
\mathbf{RfSy}(3) &= \begin{bmatrix} 778 & 79 \\ 964 & 15.9 \\ 1080 & 21.9 \end{bmatrix}, \mathbf{RfSy}(4) = \begin{bmatrix} 757 & 74.5 \\ 917 & 13.9 \\ 1252 & 21.9 \end{bmatrix} \\
\mathbf{RfSy}(5) &= \begin{bmatrix} 720 & 77.1 \\ 957 & 19.6 \\ 1038 & 24.1 \end{bmatrix}, \mathbf{RfSy}(6) = \begin{bmatrix} 743 & 73.7 \\ 984 & 13.5 \\ 1068 & 21.1 \end{bmatrix}
\end{aligned} \quad (27)$$

### C. Detection and Identification Performance

In this section, it will be supposed that the specifications are the following ones:

$$\left\{ \begin{array}{l} TP > 80\% \\ FP < 5\% \\ \forall (i,j) \in \{1,2,3\}, \exists k \in \{1,2,3\} : r_{ij} / D_k(i,j) > 0.8 \text{ rad} \end{array} \right. \quad (28)$$

#### 1) Detection KPI

TP and FP are computed from  $MhSy$  and  $RfSy$  and their values are presented in table III for HI 1, 2 and 3. The results show that KPI of degradation modes 1, 2 and 6 are compliant with the specifications but HI are not sensitive enough to detect the others. However, as explained before, those degradation modes are not the critical ones because their occurrence rate is low. Eventually, the results traduce a good potential performances for detection.

TABLE III. DETECTION KPI – TP AND FP VALUES FOR BEST THRESHOLD

DM	Best Thresholds	True Positive Rates	False Positive Rates
1	(810,986,1108)	(93, 100, 100)	(6, 0, 0)

DM	Best Thresholds	True Positive Rates	False Positive Rates
2	(765,952,1031)	(76, <b>98, 98</b> )	(23, <b>2, 2</b> )
3	(701,922,978)	(56,50,67)	(43,27,30)
4	(664,931,990)	( <b>80, 85, 70</b> )	(79,16,47)
5	(735,920,998)	(56,67,73)	(39,30,33)
6	(708,895,1092)	(60,78, <b>100</b> )	(55,27, <b>0</b> )

## 2) Identification KPI

Distinguishability vectors are computed for all the couples of degradation modes and the values in *rad* are presented in table IV. The results show that all the distinguishability vectors have at least one coefficient above the specification except for the couple (1; 2). However, as degradation modes 1 and 2 are located on the same equipment, it does not affect the localization. Moreover, the main validation criterion is that the critical degradation modes are identifiable from the others in order not to pollute their detection. Finally, as this criterion is verified in our case, results can be considered satisfactory.

TABLE IV. IDENTIFICATION KPI – DISTINGUISHABILITY VALUES

Degradation modes	1	2	3	4	5	6
1	$\begin{vmatrix} 0 \\ 0 \end{vmatrix}$	$\begin{vmatrix} 0.22 \\ 0.21 \end{vmatrix}$	$\begin{vmatrix} 0.30 \\ 1.72 \end{vmatrix}$	$\begin{vmatrix} 0.82 \\ 1.08 \end{vmatrix}$	$\begin{vmatrix} 2.30 \\ 0.93 \end{vmatrix}$	$\begin{vmatrix} 1.07 \\ 1.16 \end{vmatrix}$
2	$\begin{vmatrix} 0.22 \\ 0.21 \end{vmatrix}$	$\begin{vmatrix} 0 \\ 0 \end{vmatrix}$	$\begin{vmatrix} 0.17 \\ 1.76 \end{vmatrix}$	$\begin{vmatrix} 1.04 \\ 0.99 \end{vmatrix}$	$\begin{vmatrix} 2.19 \\ 1.12 \end{vmatrix}$	$\begin{vmatrix} 0.97 \\ 1.07 \end{vmatrix}$
3	$\begin{vmatrix} 0.30 \\ 1.72 \end{vmatrix}$	$\begin{vmatrix} 0.17 \\ 1.76 \end{vmatrix}$	$\begin{vmatrix} 0 \\ 0 \end{vmatrix}$	$\begin{vmatrix} 1.09 \\ 0.96 \end{vmatrix}$	$\begin{vmatrix} 2.32 \\ 1.10 \end{vmatrix}$	$\begin{vmatrix} 1.11 \\ 0.91 \end{vmatrix}$
4	$\begin{vmatrix} 0.82 \\ 1.08 \end{vmatrix}$	$\begin{vmatrix} 1.04 \\ 0.99 \end{vmatrix}$	$\begin{vmatrix} 1.09 \\ 0.96 \end{vmatrix}$	$\begin{vmatrix} 0 \\ 0 \end{vmatrix}$	$\begin{vmatrix} 2.25 \\ 1.29 \end{vmatrix}$	$\begin{vmatrix} 1.52 \\ 0.08 \end{vmatrix}$
5	$\begin{vmatrix} 2.30 \\ 0.93 \end{vmatrix}$	$\begin{vmatrix} 2.19 \\ 1.12 \end{vmatrix}$	$\begin{vmatrix} 2.32 \\ 1.10 \end{vmatrix}$	$\begin{vmatrix} 2.25 \\ 1.29 \end{vmatrix}$	$\begin{vmatrix} 0 \\ 0 \end{vmatrix}$	$\begin{vmatrix} 1.25 \\ 1.33 \end{vmatrix}$
6	$\begin{vmatrix} 1.07 \\ 1.16 \end{vmatrix}$	$\begin{vmatrix} 0.97 \\ 1.07 \end{vmatrix}$	$\begin{vmatrix} 1.11 \\ 0.91 \end{vmatrix}$	$\begin{vmatrix} 1.52 \\ 0.08 \end{vmatrix}$	$\begin{vmatrix} 1.25 \\ 1.33 \end{vmatrix}$	$\begin{vmatrix} 0 \\ 0 \end{vmatrix}$

## CONCLUSION

In this document, a hybrid approach for diagnostics based on the complementarity between model-based and data-driven technics has been introduced through some definitions and framework. Then, the principle of Kriging and its utility in time computational costs reduction was addressed. Eventually, the approach was tested on an aircraft engine pumping unit and had showed good results not only in computing the degradation modes signatures but also in evaluating detection and identification performances. However, this example was a simplified one and for future research, the main purpose is to extend the approach to the entire fuel system. The difficulty will then be to manage the large number of parameters and it will certainly be necessary to use some sensitivity analysis technics. In parallel, a further literature review on Kriging optimization and validation methods needs to be done in order to increase the quality of the models. Finally, Kriging was just

one surrogate modeling method among others and it would be interesting to perform a comparative study of all this technics.

## REFERENCES

- [1] "OSA-CBM Website," [Online]. Available: <http://www.osacbm.org/>.
- [2] R. Isermann, «Model-Based fault-detection and diagnosis - status and applications,» *Annual Reviews in Control*, vol. 29, n° %11, pp. 71-85, 2005.
- [3] H. Wang, T. Y. Chai, J. L. Ding et M. Brown, «Data Driven Fault Diagnosis and Fault Tolerant Control: Some Advances and Possible New Directions,» *Acta Automatica Sinica*, vol. 35, n° %16, p. 739–747, 2009.
- [4] A. Garga, K. McClintic, R. Campbell, C. C. Yang, M. Lebold, T. Hay et C. Byington, «Hybrid Reasoning for Prognostic Learning in CBM Systems,» vol. 6, pp. 2957-2969, 2001.
- [5] S. Kumar, M. Torres, Y. C. Chan et M. Pecht, «A hybrid prognostics methodology for electronic products,» chez *Neural Networks, 2008. IJCNN 2008. (IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on*, 2008.
- [6] E. De Rocquigny, N. Devictor, S. Tarantola, F. Mangeant, C. Schwob, R. Bolado-Lavin, J. R. Massé, P. Limbourg, W. Kanning and P. Van Gelder, *Uncertainty in industrial practice: A guide to quantitative uncertainty management*, Wiley, 2007.
- [7] T. D. Wickens, *Elementary Signal Detection Theory*, Oxford University Press, 2002.
- [8] A. I. J. Forrester et A. J. Keane, «Recent advances in surrogate-based optimization,» *Progress in Aerospace Sciences*, vol. 45, n° %11-3, pp. 50-79, 2009.
- [9] F. A. Viana, G. Venter et V. Balabanov, «An algorithm for fast optimal Latin hypercube design of experiments,» *International journal for numerical methods in engineering*, vol. 82, n° %12, pp. 135-156, 2010.
- [10] J. P. C. Kleijnen, «Kriging metamodeling in simulation: a review,» *European Journal of Operational Research*, vol. 192, n° %13, pp. 707-716, 2009.
- [11] H. B. Nielsen, S. N. Lophaven et J. Sondergaard, *DACE - A Matlab Kriging Toolbox*, Lyngby - Denmark: Informatics and Mathematical Modelling, Technical University of Denmark, DTU, 2002.
- [12] R. H. Myers, D. C. Montgomery et C. M. Anderson-Cook, *Response surface methodology: process and product optimization using designed experiments*, Hoboken, New Jersey, USA, : Wiley, 2009.
- [13] B. Lamoureux, J. R. Massé et N. Mechbal, «An Approach to the Health Monitoring of the Fuel System of a Turbofan,» chez *IEEE Conference on Prognostics and Health Management (PHM)*, Denver (CO), 2012.
- [14] P. Casoli, V. Andrea and G. Franzoni, "A Numerical Model for the Simulation of External Gear Pumps," in *Proceedings of the 6th JFPS International Symposium on Fluid Power*, Tsukuba, 2005.
- [15] R. Frith and W. Scott, "Wear in external gear pumps : a simplified model," *Wear* 172, pp. 121-126, 1994.