

# MÉTHODOLOGIE D'ANALYSE STATISTIQUE DES MESURES DE PIÈCES FABRIQUÉES EN USINE POUR IDENTIFIER ET TOLÉRANCER LES COMPOSANTS GÉOMÉTRIQUES CRITIQUES

Thierry CEMBRZYNSKI (\*), Jacques DEMONSANT (\*)

(\*) RENAULT Technocentre 1 Avenue du Golf 78280 Guyancourt

## Introduction

La qualité d'une prestation Y délivrée par un système physique est la résultante :

- des spécifications établies par le bureau d'études, c'est-à-dire des valeurs des caractéristiques nominales et des tolérances des caractéristiques  $X_i$  des composants et des interfaces du système,
- du respect de ces spécifications par la fabrication.

La conception d'un système est dite robuste au regard d'une prestation Y délivrée si cette dernière n'est pas ou est peu sensible aux dispersions de fabrication ainsi qu'à la variabilité dans laquelle le système est amené à fonctionner.

Pour améliorer la qualité du système G. Taguchi préconise une démarche dénommée *Robust Engineering* [2] [3] [4] [5]. Cette démarche consiste à définir les valeurs nominales et les tolérances des différentes caractéristiques  $X_i$  de manière à ce que la prestation Y soit fournie en toute circonstance. Pour cela il préconise l'établissement d'une relation du type  $Y=f(X_1, X_2, \dots, X_p)$  à partir de données expérimentales. Pour estimer cette relation il recommande de réaliser un plan d'expériences comportant un nombre limité d'essais. Il arrive parfois qu'un plan d'expériences ne puisse pas être mis en œuvre facilement, notamment lorsque le nombre de facteurs potentiellement influents identifiés par les concepteurs est très élevé ou lorsque, fixer une valeur de réglage rigoureusement déterminée à chacun d'entre eux, conformément au plan d'expériences, est très difficile.

Il existe ainsi des situations où la réalisation d'un plan d'expériences n'est pas envisageable mais qu'en revanche, il soit possible d'observer différentes réalisations du système.

Nous nous proposons de montrer ici que dans de tel cas, il est possible d'arriver à des résultats concluants en s'appuyant sur une méthodologie d'analyse statistique à base de régression PLS.

Cette méthodologie comprend deux phases :

- identifier les caractéristiques critiques et établir leur influence sur la prestation Y,
- optimiser les spécifications du système.

## 1. Relier la prestation mesurée Y aux cotes géométriques $X_i$ du système

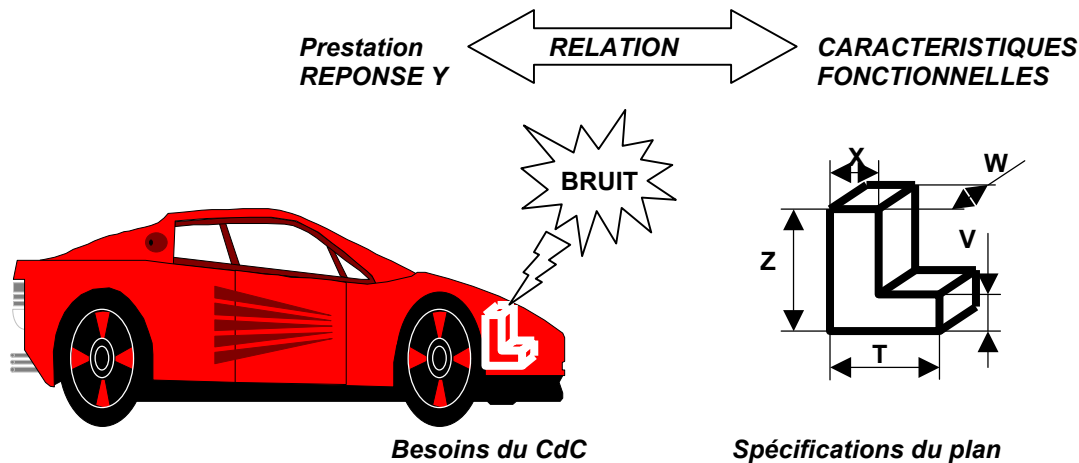
### 1.1. Construction d'un tableau de données Individus $\times$ Variables

La méthode requiert le recueil de n systèmes prélevés en usine comportant des « bons » et des « mauvais ». Chacun de ces systèmes est caractérisé par :

- une (ou q) variable(s) Y représentative(s) de la prestation à améliorer,
- p variables  $X_i$  explicatives candidates représentant certaines caractéristiques géométriques des composants du système.

La chronologie de fabrication et de prélèvement des n systèmes sera conservée dans le tableau de données.

### 1.2. La démarche d'analyse



**Figure 1** : recherche des caractéristiques fonctionnelles X, V, W, Z etc... influentes sur la prestation Y (bruit) délivrée par le système.

Elle a pour but d'établir une relation entre la prestation Y et les variables  $X_i$  (figure 1). Elle comporte les étapes suivantes :

- 1) détection et élimination des mesures aberrantes,
- 2) adimensionnement des variables  $X_i$  et Y,
- 3) ajout de variables transformées  $X_i^2$ ,  $X_i^3$ ,  $\exp(X_i)$ ,
- 4) modélisation à l'aide de la régression PLS,
- 5) sélection des variables  $X_i$  les plus influentes sur la prestation Y.

### 1.1.1. Détection et élimination des mesures aberrantes

Il convient, tout d'abord, d'éliminer les valeurs aberrantes susceptibles de biaiser la modélisation. Deux approches ont été envisagées :

1. La cotation ISO (cf. norme AFNOR E60-181) préconise en fonction de la nature des variables, un type de loi (loi de Gauss pour les cotes dimensionnelles, loi de Rayleigh ou loi dite de défaut de forme pour les défauts géométriques). Dans ce contexte, il est donc nécessaire d'estimer les paramètres de la loi retenue pour chaque caractéristique, puis d'éliminer les observations qui ont une probabilité inférieure à 0,1% d'être dépassée.
2. Construction d'intervalle de confiance à partir de l'estimation de la moyenne  $\bar{x}$  et de l'écart type s :

$$\text{Si } x_i \notin [\bar{x} - t_{99,9\%,n-1} \cdot s, \bar{x} + t_{99,9\%,n-1} \cdot s] \text{ alors } x_i \text{ est déclaré aberrant}$$

où  $t_{\alpha,n-1}$  représente le quantile d'ordre  $\alpha$  de la loi de Student à n-1 degrés de liberté

Par ailleurs, pour disposer d'une estimation robuste l'estimation  $s = \sqrt{\frac{1}{n-1} \sum_i (X_i - \bar{X})^2}$  étant sensible à la présence d'une valeur aberrante dans l'échantillon, on pourra remplacer cette estimation par :

$$s = \frac{1}{n-1} \sum_{i=2}^n \frac{w_i}{d_2(2)} \text{ où } w_i = |x_i - x_{i-1}|$$

où  $d_2(n)=1,128$  est l'espérance mathématique de la loi de l'étendue réduite pour un échantillon de n valeurs. On ne manquera pas de faire diagnostiquer les mesures déclarées aberrantes par le service métrologique.

C'est cette deuxième approche plus simple à mettre en œuvre que nous préconisons.

### 1.1.2. Adimensionnement des variables $X_i$ et de Y

Pour pouvoir comparer les effets des caractéristiques  $X_i$  sur la prestation Y, il est indispensable de travailler sur des variables sans dimension.

Diverses méthodes d'adimensionnement de variables sont possibles. La méthode la plus utilisée consiste à s'appuyer sur l'intervalle de tolérance [Lit, Lst], puis à se centrer sur le milieu de cet intervalle et à réduire en divisant par  $a/3$  où  $a$  représente la demi largeur de l'intervalle de tolérance<sup>1</sup>. Cette méthode présente l'inconvénient de ne pas être pertinente pour des caractéristiques décentrées.

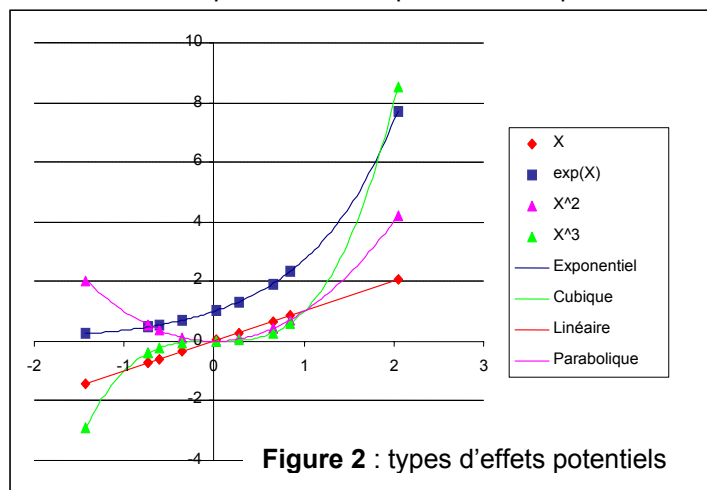
Nous préconisons donc l'adimensionnement par centrage et réduction à partir des moyennes et écarts types estimés sur les données après élimination des mesures aberrantes :

$$X_i^* \rightarrow \frac{X_i - \bar{X}_i}{s_i}$$

### 1.1.3. Ajout de variables transformées pour rechercher d'éventuels effets non linéaires.

Pour mettre en évidence d'éventuels effets non linéaires de type parabolique, cubique, voire exponentiel, il est nécessaire de créer des variables supplémentaires correspondant au carré  $X^2$ , au cube  $X^3$  ou à l'exponentielle  $e^X$ , des variables de base  $X$  comme l'illustre la figure 2.

Finalement on dispose donc de  $4p$  variables explicatives candidates.



### 1.1.4. Modélisation PLS

La méthode classique de régression (OLS) est la méthode la plus simple pour construire des modèles linéaires reliant des variables explicatives (numériques)  $X$  ou prédicteurs, à une ou plusieurs variables à expliquer  $Y$  ( $Y = \beta X + \varepsilon$ ) ; mais elle souffre de graves défauts :

- Incapacité à prendre en compte les données manquantes, ce qui conduit souvent le praticien à rejeter beaucoup d'observations incomplètes qui contiennent pourtant de l'information utile dans les champs renseignés.
- Grande sensibilité à la colinéarité entre les variables explicatives  $X$  : la colinéarité exacte rend la régression multiple impossible, et la colinéarité approchée la rend numériquement instable et fait perdre l'interprétation des coefficients  $\beta$ .
- Indétermination lorsque le nombre d'observations  $n$  est inférieur au nombre des variables explicatives  $p$ , situation pourtant fréquemment rencontrée en pratique dans l'industrie.

Pour faire face à ces difficultés, nous avons retenu la régression PLS (Partial Least Squares) [6] car cette méthode permet de pouvoir traiter beaucoup de variables (même corrélées entre elles) avec peu d'observations, voire avec des données manquantes comme c'est le cas après élimination des mesures aberrantes.

En revanche il n'existe pas pour la régression PLS, de méthode de sélection de variables automatique « pas à pas ». On obtient donc après ajout des transformations de variables ( $X^2$ ,  $X^3$ ,  $e^X$ ), un modèle linéaire comportant alors  $4p$  coefficients. Reste alors à déterminer les coefficients les plus importants afin de sélectionner les caractéristiques critiques, c'est-à-dire celles qui sont les plus influentes sur la prestation  $Y$ .

<sup>1</sup> On suppose conventionnellement que l'intervalle de tolérance correspond à  $6\sigma$ .

### 1.1.5. Sélection des variables $X_i$ les plus influentes sur la prestation $Y$

Cette sélection s'appuie sur deux types de résultats :

- l'analyse des corrélations linéaires entre les variables explicatives, leurs transformées et la variable de prestation  $Y$ ,
- l'analyse des coefficients du modèle linéaire PLS.

#### 1.1.5.1. Test de la significativité des corrélations $\rho(X_i, Y)$

A partir des  $n$  couples de mesures, on calcule les coefficients de corrélation linéaire

$$\rho_{X_i, Y} = \frac{\sum_{j=1}^n (X_{i,j} - \bar{X}_i)(Y_j - \bar{Y})}{\sqrt{\sum_{j=1}^n (X_{i,j} - \bar{X}_i)^2 \sum_{j=1}^n (Y_j - \bar{Y})^2}}$$

de chaque variable  $X_i$  avec la prestation  $Y$ .

La quantité  $\left( n - 2 \times \frac{\rho_{X_i, Y}^2}{1 - \rho_{X_i, Y}^2} \right) = \frac{CME_{reg.lin Y=aX_i+b}}{CME_{residuelle}} \approx F_{1, n-2}$  suit une loi de Fisher à 1 et  $(n-2)$  degrés de liberté.

On teste alors la significativité de la liaison linéaire  $Y=a.X_i + b$ . Pour ce test nous avons retenu un risque  $\alpha$  (fort) de 20% de rejet de l'hypothèse nulle d'absence de liaison ( $H_0 : \rho(X_i, Y)=0$ ) quand les deux variables  $X_i$  et  $Y$  ne sont réellement pas corrélées afin de limiter le risque  $\beta$  d'accepter à tort l'hypothèse nulle  $H_0$  d'absence de liaison quand les deux variables  $X_i$  et  $Y$  sont réellement corrélées.

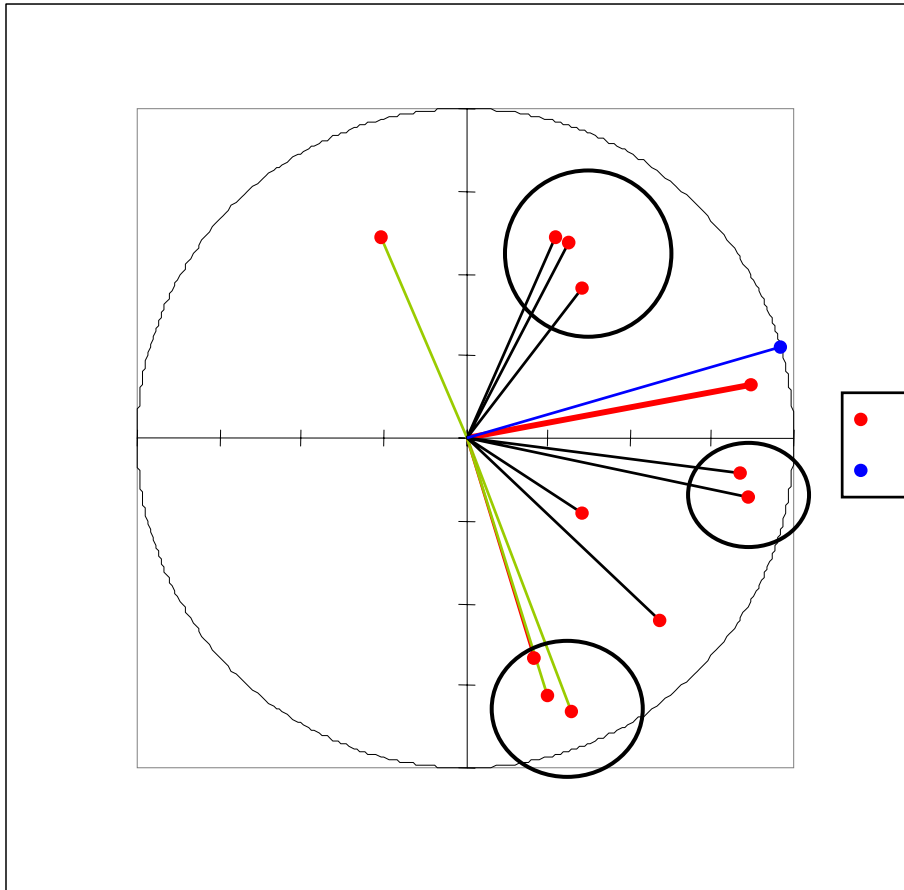
$$Si \quad n - 2 \times \frac{\rho_{X_i, Y}^2}{1 - \rho_{X_i, Y}^2} \leq F_{1, n-2}^{1-\alpha} \quad H_0 \text{ acceptée} \quad X_i \text{ éliminée}$$

#### 1.1.5.2. Analyse des corrélations $\rho(X_i, T)$ des variables $X_i$ , $Y$ avec les facteurs $(T_1, T_2)$ de la régression PLS

Les facteurs  $T_1, T_2$  orthogonaux et donc non corrélés de la PLS, sont les deux combinaisons linéaires des 4p variables explicatives initiales ( $T=X.w$ ) qui maximisent la corrélation  $\rho(T, Y)$ .

L'examen du plan principal  $T_1, T_2$  des corrélations des variables explicatives  $X_i$  (en rouge) avec la variable prestation  $Y$  (en bleu) renseigne alors beaucoup sur :

- les variables  $X_i$  les moins significatives c'est-à-dire celles qui apparaissent orthogonales à  $Y$  ; sur la figure 3 cela conduit, par exemple, à éliminer les variables  $X_1, X_1^2, e^{X_1}$  en jaune ;
- les variables  $X_i$  les plus influentes c'est-à-dire qui apparaissent colinéaires à  $Y$  ; sur la figure 3 on retient alors  $X_2$  ;
- les groupes de variables  $X_i$  redondantes ; la figure 3 montre ainsi trois groupes de variables fortement corrélées entre elles ; il suffit de retenir une variable dans chacun de ces groupes.



### 1.1.5.3. Sélection des variables $X_i$ critiques à partir des coefficients $\alpha_i$ du modèle PLS

Pour sélectionner les variables les plus influentes on retiendra évidemment les coefficients  $\alpha_i$  les plus grands en valeur absolue du modèle PLS, sachant que  $\alpha_i$  est l'effet engendré sur Y par une variation de  $X_i$  égale à  $1 \times \sigma$ .

Les variables sont alors classées en trois catégories en fonction de la valeur de  $k = \frac{\alpha_i - \bar{\alpha}}{s_\alpha}$  où  $\bar{\alpha}$  et

$s_\alpha$  désignent la moyenne et l'écart-type des coefficients  $\alpha_i$  des variables :

- si  $k > 1,65$ , la variable  $X_i$  est considérée comme très vraisemblablement influente,
- si  $1,04 < k < 1,65$ , la variable  $X_i$  est considérée comme vraisemblablement influente,
- si  $k < 1,04$ , la variable  $X_i$  est considérée comme non influente.

Les coefficients  $\alpha_i$  trouvés non influents confortent le diagnostic des analyses précédentes. Par ailleurs, la nature des variables sélectionnées (variable brute ou variable transformée) donne une indication sur la forme de l'effet (linéaire, quadratique, exponentiel,...) que produisent les variables  $X_i$  sur la prestation Y.

## 2. Optimiser les spécifications du système

Le cahier des charges du système exige que la prestation Y vérifie la condition suivante :

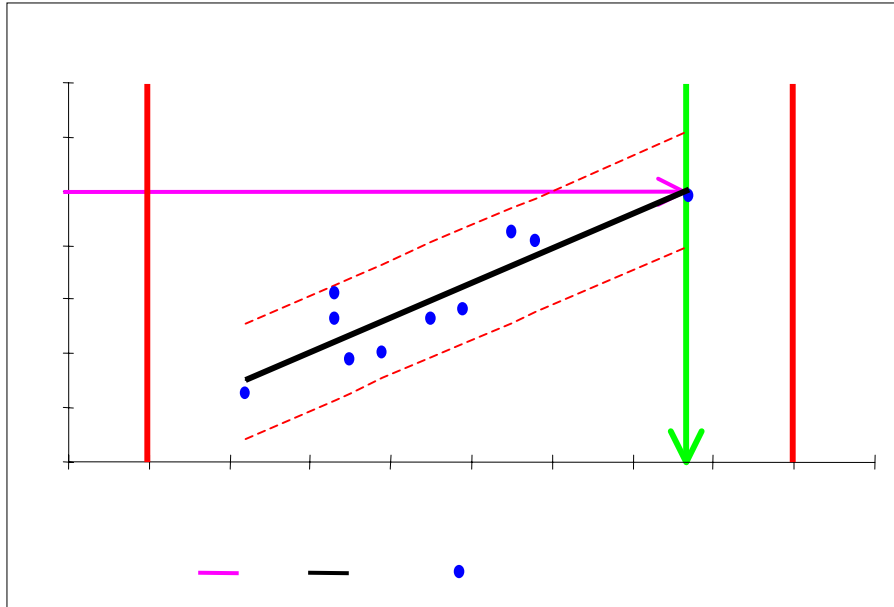
$$Y < Y_{\max}$$

On cherche alors quelle spécification appliquer aux variables  $X_i$  les plus critiques pour être assuré du respect de la condition ci-dessus.

Dans ce but, on utilise les résultats de la régression PLS pour les variables critiques retenues. Leur effet est modélisé par une relation linéaire, quadratique ou exponentielle selon l'effet produit.

L'exigence sur chaque variable  $X_i$  est alors définie à partir du modèle réciproque, sans tenir compte de l'incertitude de l'estimation, celle-ci étant supposée traduire l'effet des autres variables influentes. Sur

l'exemple de la figure 4, il est nécessaire de réduire l'intervalle de tolérance initial ( $10 \pm 0,4$ ) de 16% de la caractéristique X pour respecter le cahier des charges de la prestation Y.



### 3. Application à la maîtrise du bruit d'un groupe électro-pompe de direction assistée

La direction assistée est actionnée par un circuit hydraulique ; celui est mis en mouvement par un groupe électro-pompe ; ce groupe comprend des engrenages ; la discontinuité de l'engrènement est susceptible de produire un bruit bien connu des spécialistes de boîtes de vitesses.

Pour éviter de livrer des systèmes défectueux, un contrôle à 100% a été instauré en fin de ligne de fabrication chez le fournisseur. Ce contrôle porte sur un paramètre d'accélération  $\gamma$  ; la limite d'acceptation pour ce paramètre a été fixée à partir d'une campagne de mesures sur des véhicules de série. Compte tenu du coût de ce contrôle, nous avons engagé avec le fournisseur une démarche de recherche des caractéristiques influentes sur le bruit pour mieux spécifier le système (on agira sur la conception des plans) et pour mieux surveiller la fabrication (on optimisera le plan de surveillance).

Les experts techniques ont retenu 51 caractéristiques des pièces (dentures, paliers, carters...) sources potentielles de bruit.

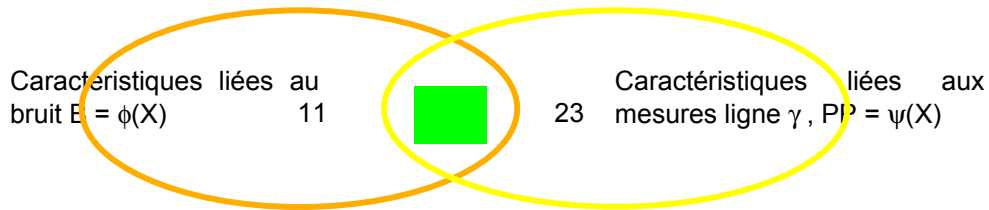
Le fournisseur a mesuré sur 19 systèmes de série :

- en ligne (100%), les Pressions Pulsées ( $N/m^2$ ) et les Accélérométries  $\gamma$  (g),
- les bruits aériens (dB) sur banc traduits en Energies Acoustiques ( $Pa^2$ ),
- les 51 caractéristiques fonctionnelles des pièces en métrologie, notamment des états de surface.

La régression PLS a permis de sélectionner les quatre caractéristiques X, les plus influentes (figure 5) à l'intersection des :

- 11 caractéristiques influentes sur le bruit (EA),
- 23 caractéristiques influentes sur les mesures lignes d'accélérométrie  $\gamma$  et de Pression Pulsée.

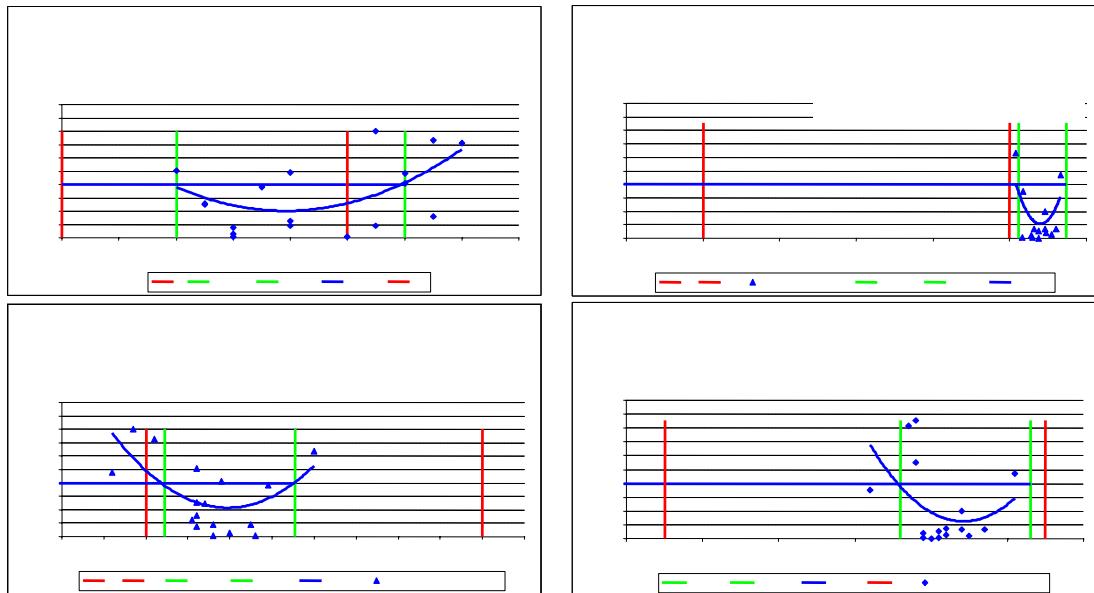
**Figure 5** : Sélection des caractéristiques les plus influentes



A partir du cahier des charges concernant le niveau de bruit maximum acceptable, pour les quatre caractéristiques critiques ( $X_1, \dots, X_4$ ) :

- on a redéfini la limite de contrôle sur les mesures ligne  $\gamma$  et Pressions Pulsées pour pouvoir déclencher une rétroaction en fabrication à partir des mesures ligne (optimisation du plan de contrôle) ;
- on a recentré les nominaux (Parameter Design) et proposé de réduire les intervalles de tolérance (Tolerance Design) pour affiner la conception (figure 6).

Pour des raisons de confidentialité, nous ne décrivons pas la nature de ces caractéristiques.



**Figure 6.** Proposition de nouveaux intervalles de tolérance

## Conclusion

Les plans d'expériences sont de plus en plus utilisés dans l'industrie pour optimiser la conception des produits et procédés de fabrication. Cependant, il n'est pas toujours possible de maîtriser expérimentalement les facteurs dont l'on souhaite mesurer l'influence. Dans ce cas, il est souvent possible d'effectuer des prélèvements en fabrication, d'effectuer des mesures sur les pièces prélevées. Pour analyser ces résultats qui comprennent souvent plus de mesures que d'observations, une méthodologie s'appuyant sur la régression PLS nous est apparue tout à fait appropriée. Nous l'avons traduit en procédure applicable par des techniciens tant dans notre bureau d'études que chez notre fournisseur.

## Bibliographie

- [1] Demonsant J. *Comprendre et mener les plans d'expériences*, AFNOR, Paris, 1996
- [2] Demonsant J., *L'ingénierie robuste*, in Encyclopédie de la qualité, Hermès, 2000
- [3] Fowlkes W., Creveling C., *L'ingénierie robuste*, Dunod, Paris, 1998
- [4] Taguchi G., *Introduction to Quality Engineering*, Asian Productivity Organisation, Tokyo, 1986
- [5] Taguchi G., Chowdhury S., Taguchi S., *Robust Engineering*, McGraw-Hill, New York, 1999
- [6] Tenenhaus M., *La Régression PLS, théorie et pratique*, Technip, Paris, 1998

