



**HAL**  
open science

## Two stage approaches for modeling pollutant emission of diesel engine based on Kriging model

El Hassane Brahmi, Lilianne Denis-Vidal, Zohra Cherfi, Nassim Boudaoud,  
Ghislaine Joly-Blanchard

► **To cite this version:**

El Hassane Brahmi, Lilianne Denis-Vidal, Zohra Cherfi, Nassim Boudaoud, Ghislaine Joly-Blanchard. Two stage approaches for modeling pollutant emission of diesel engine based on Kriging model. A.D.Rodic. Automation, Control: Theory and Practice, IN-TECH, pp.1-18, 2009, ISBN 978-953-307-039, 978-953-51-5858-5. 10.5772/7896 . hal-00565512

**HAL Id: hal-00565512**

**<https://hal.science/hal-00565512>**

Submitted on 13 Feb 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Chapter Number 18

## Two stage approaches for modeling pollutant emission of diesel engine based on Kriging model

El Hassane Brahmi, Lilianne Denis-Vidal, Zohra Cherfi, Nassim Boudaoud,  
Ghislaine Joly-Blanchard  
*University of technology of Compiègne  
France*

### 1. Introduction

The automotive industry faces the competing goals of producing better performing vehicles and keeping development time with low costs. It is crucial for the manufacturers to be able to produce fuel-economic vehicles, which respect pollutant emissions standards, and which meet the customers expectations. Accordingly, the complexity of the engines responses we have to optimize and the number of the parameters to control during the design stage, have increased rapidly, in the later years.

In order to deliver vehicles, which respond to these requirements, in a reasonable time scale. Companies use design of experiments (DOE) (Schimmerling et al., 1998) in one side, and modelling, in the other side. DOE is a power tool, but the cost of the experiments and their duration, particularly in the field of pollutant emissions, can be a limit to their use in automotive industry.

The engine developers use two main approaches to model engine behaviour. The first one is based on chemical and physical models, via differential system. This approach is not the subject of this article, because we have not such models. Furthermore, even when these models are available, generally, they are time-consuming, impractical for multi-objective optimisation routines, and fail to capture all the trends in the engine system described by measured data (like Zeldovich model). All this, is particularly true when the number of the control parameters is large and engine responses are complex.

Statistical modelling based on carefully chosen measured data of engine performance, according to an experimental design is an important alternative technique.

Strategies based on Lolimot (Castric et al., 2007) (Local Linear Model Tree) and Zeldovich mechanisms (Heywood, 1988) have been developed in order to predict emissions of NO<sub>x</sub>. In the first case, the corresponding model can lead to singular points, which reduces the precision of the results. In the second case, the results are not satisfactory enough.

The literature presents several methods based on statistical trainings such as neural networks. This method gives good results, even in the nonlinear case. However, it is not

adapted to our case, because it requires a great number of experiments to obtain a significant estimate of its parameters, and we are very limited by the small experiments number which the industrialist is able to realize. The techniques of the design of experiments (Cochran & Cox, 1957) were conceived to deal with this kind of problems. On the other hand, recent works (Sacks et al. 1989; Bates et al. 1996; Koehler & Owen, 1996) suggest that the polynomial models are not adapted to the numerical experiments. For example, a surface of response of order two is not enough flexible to model a surface admitting several extrema.

The aim of this paper is to present the result that we have obtained in the field of pollutants emissions prediction. These results were obtained without the increase of the number of the experiments that the industrialist can do. We call upon a sophisticated statistical model resulting from the field from geostatistic named Kriging.

We use this method, through two approaches, in order to improve the prediction of NO<sub>x</sub> (nitrogen oxide) emissions, and fuel consumption.

In the first stage, we estimate the response directly from the controllable factors like main injection timing, pressure in common rail. This can be assimilated to a black box modelling.

In the second stage we propose an innovative approach that allows us to predict the response from a functional data. More precisely we estimate the engine performance from signals like pressure and cylinder temperature, this signals are obtained from a model of combustion. The main advantage of the second approach is that it allows us to include a physical knowledge of combustion. This lack of knowledge is often criticized in the case of black boxes models.

The Kriging method is very well adapted for the second approach which predicts engine responses from the state variables (signals) obtained from a physical model of combustion. This means that this method can be recommended in cases where we have a lot of variables decision with a small number of experiences; this is because the method is based on the study of the similarity of the response of interest in different sites in relation to the distance separating them. We recall that, Software such as R and Matlab contain a toolbox to use this method, but unfortunately the latter being restricted to less than 3 dimensions. Adapting the method to higher dimensions has been considered.

To implement this second approach, a model reduction is needed; this reduction will be made in two steps:

- 1) Reducing the number of state variables from 10 to 2 by the study of correlations.
- 2) Reducing the number of points in each signal using the theory of Fourier.

Once the reduction is made, the Kriging can be applied to the model obtained.

This paper is organized as follows: In the second section, we describe the engine behaviour and recall the importance of controlling pollutant emissions. In the third section, the

ordinary Kriging techniques are recalled. In the fourth section, two different approaches for modelling our problem are proposed. An efficient reduction model strategy is considered in order to apply the Kriging method. Finally the Kriging method is applied to the reduced model. In the last section, numerical results are given followed by a short discussion.

## 2. Engine calibration

The engine calibration is a process which aims at defining the values of the engine control parameters. During the ten last years, the set of themes "engine calibration" took an important place in the process of the development of the internal combustion engines. Indeed, under the impulse of the standards more and more severe, the car manufacturers are brought to integrate more and more elaborated technologies in the power unit.

Under these conditions, the principles of control used in the past (cartographic approach, buckles open...) are not enough sufficient. Indeed, the output variables (quantity injected, advances in lighting or in injection) were primarily given starting from two variables of entry (speed and load). Today the use of new technologies in the conception of engines, in order to reduce the pollutant emissions, as for example EGR (exhaust Gas Recirculation) (Pierpont et al. 1995), multiply the number of parameters to control, as we can see it in Fig.1. This figure shows the exponential evolution of number of setting parameters due to the hardware complexity increase. This makes the cartographic approach impracticable.

Moreover, this kind of approach does not take into account the dynamic of system. The main drawback of this evolution is the increase of the difficulty to understand the engine behavior. To deals with all the parameters, we use a Kriging model which we define in the next section.

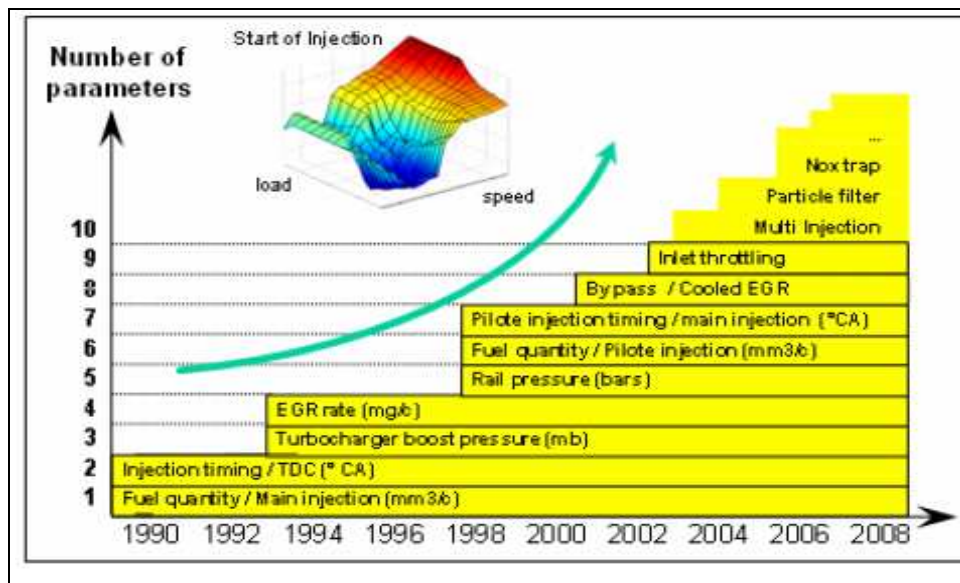


Fig.1. Parameters to tune a diesel engine function of technologies

### 3. Ordinary Kriging Techniques

Kriging methods are used frequently for spatial interpolation of soil properties (Krige, 1951; Matheron, 1963). Kriging is a linear least squares estimation algorithm. It is a tool for interpolation. The aim is to estimate the value of an unknown real function  $Z$  at point  $x_0^n$ , given the values of function  $Z$  at some other points  $x_i \in \mathbb{R}^n$  for each  $i = 1, \dots, n$ .

#### 3.1 Ordinary Kriging

The ordinary Kriging estimator  $\hat{Z}(x_0^n)$  is defined by:

$$\hat{Z}(x_0^n) = \sum_{i=1}^n \lambda_i Z(x_i) \quad (1)$$

Where  $n$  is the number of surrounding observations  $Z(x_i)$  and  $\lambda_i$  is the weight of  $Z(x_i)$ . The weights should sum to unity in order to make the estimator unbiased:

$$\sum_{i=1}^n \lambda_i = 1 \quad (2)$$

The weights are also determined such that the following Kriging variance is minimal under the constraint given by the equation 2:

$$\text{var}(\hat{Z}(x_0^n) - Z(x_0^n))$$

This leads to a classical optimization problem with equality constraint. The Lagrange multiplier theory is used in order to work out this problem. This gives a linear system to be solved (Davis, 1986).

#### 3.2 Variogram

The variogram is a function representing the spatial dependency. It is obtained from the stationarity definition. In fact, this stationarity hypothesis is an indispensable condition for the use of the Kriging method.

In the case of ordinary Kriging the expression of the variogram is obtained from the following definition of intrinsic stationarity:

- 1)  $E(Z(x_i + h) - Z(x_i)) = 0 \quad \forall i \in \{1, \dots, n\}$
- 2)  $\text{Var}(Z(x_i + h) - Z(x_i)) = 2\gamma(h) \quad \forall i \in \{1, \dots, n\} \text{ and } h$

More precisely the expression of theoretical variogram is deduced from the second condition of intrinsic stationarity. This condition means that the variation of a data set is only dependent on distance  $r$  between two locations, where the variables values are  $Z(x_i + h)$  and  $Z(x_i)$  with  $r = |h|$ .

Note that the variogram,  $\gamma(h)$ , is a function of the separation between points  $h$  and not a function of the specific location  $(x_i, x_i + h)$ . This mathematical definition is a useful abstraction, but not easy to apply to observed values.

Consider a set of  $n$  observed data:  $\{(x_1, z_1), (x_2, z_2), \dots, (x_n, z_n)\}$ , where  $x_i$  is the location of observation  $i$  and  $z_i$  is the associated observed value. There are  $\frac{n(n-1)}{2}$  unique pairs of observations. For each of these pairs we can calculate the associated separation distance:

$$r = |x_i - x_j|$$

To infer the variogram from observed data we will then use the common formula for the experimental variogram (Cressie, 1993).

$$\hat{\gamma}(r) = \frac{1}{2N(r)} \sum_{N(r)} [Z(x_i) - Z(x_j)]^2 \quad (3)$$

Where:

$$N(r) = \{(i, j) \text{ such that } |x_i - x_j| = r\}$$

$N(r)$  is the pair number of  $Z(x_i + h)$  and  $Z(x_j)$ ;  $\hat{\gamma}(r)$  is the experimental variogram.

### 3.3 Variogram Modeling

The experimental variogram presented in equation 3, estimates the theoretical variogram, for only a finite number of distances. Moreover, it does not necessarily form a valid variogram. This means, that maybe, it does not concern a negative conditionally function. Indeed, this condition is necessary to ensure the positivity of the variance of a sum of random variables (Christakos, 1984).

The experimental variogram is then modeled by a function of negative conditional type and is defined for all distances. This modeling makes the Kriging possible. A variogram model should be fitted to such variogram.

A model must be selected among the various forms of the variogram models, which exist in the literature and adjusted of experimental variogram (Arnaud & Emery, 2000). This means that the parameters of the model must be estimated. This adjustment can be done graphically, but it is usually done with an estimation method as the weighted least squares or maximum likelihood method.

Once the variographic model is chosen, and its parameters are estimated. We compute the weights  $\lambda_i$  which appear in (1) by solving the following system:

$AA = B$ , with

$$A = \begin{bmatrix} \gamma(r_{11}) & \gamma(r_{12}) & \dots & \gamma(r_{1n}) & 1 \\ \gamma(r_{21}) & \gamma(r_{22}) & \dots & \gamma(r_{2n}) & 1 \\ \dots & \dots & \dots & \dots & \dots \\ \gamma(r_{n1}) & \gamma(r_{n2}) & \dots & \gamma(r_{nn}) & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

$$A = [\lambda_1 \quad \lambda_2 \quad \dots \quad \lambda_n \quad \lambda]^T$$

And

$$B = [\gamma(r_{01}) \quad \gamma(r_{02}) \quad \dots \quad \gamma(r_{0n}) \quad 1]^T$$

Where  $\lambda$  is the Lagrange multiplier.

$\gamma(r_{ij})$  is the variogram model used for adjusting the experimental variogram.

The variance of the estimate  $s_i^2$  i.e. the square of the standard error at each point is obtained by the relationship:

$$s_i^2 = A^T \cdot B$$

If we assume that the estimation errors are normally distributed around the true value, then the probability that the true value will be in  $Z(x_i) \pm s_i$  is 68 %, while the probability that the true value will be in  $Z(x_i) \pm 2s_i$  is 95 %, (Davis, 1986).

### 3.4 Kriging Emulator Validation

The true test of the quality of the fitted emulator model is its ability to predict the response at untried factor values. In order to maximally exploit the data to aid model fitting, the emulators are validated using leave-one-out cross validation. This process involves taking the fitted model and re-fitting it to a subset of non used experimental data.

More precisely, for an experiment with  $d$  design factors  $y = y_1, \dots, y_d$ , the set of  $n$  experimental design points  $X = x_1, \dots, x_n$  and responses  $Z = z_1, \dots, z_n$ , contain the information used to build the Kriging model. A cross validation involves predicting at each design point in turn when that point is left out of the predictor equations. Let  $\hat{Z}(x_i)$  be the estimate of the  $Z(x_i)$  based on all the design points except  $x_i$ . The prediction error (the estimated root mean square error, RMSE) is then calculated as:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n [Z(x_i) - \hat{Z}(x_i)]^2}$$

An index of the accuracy of the emulator is made by expressing as a percentage of the range of the response  $Z$ ,

$$RMSE = 100 \times \frac{RMSE}{\max(Z) - \min(Z)}$$

## 4. Two approaches to model engine responses

In this section we present two stage approaches based on Kriging method for the prediction of NOx (nitrogen oxide) emissions, and fuel consumption.

-In the first stage, we estimate the response directly from the controllable factor like main injection timing, pressure in common rail (black box).

-In the second stage we propose an innovative approach that allows us to predict the engine response from signals like pressure and cylinder temperature (states variables of combustion chamber).

#### 4.1 First approach

We recall that the first approach consists to build a Kriging model from the controllable parameters. Thus the Kriging was trained on about 300 input/ output sets of examples generated by using D-optimal design method. The training examples cover engine speeds from 1000 rpm to 5000 rpm in 250 rpm intervals, and load vary from 1 to 23 bar. The data was generated to cover the cycle point of the engine map in order to construct a global Kriging emulator. For this reason our model takes into account the engine speed among the following parameters control:

- Prail : rail pressure,
- Mpil1: pilot1 injection quantity,
- Mpil2: pilot2 injection quantity,
- Ppil1: pilot1 injection timing,
- VEGR: EGR valve position,
- Main: Main injection quantity,
- Pmain: Main injection timing,
- Ppil2: pilot2 injection timing,
- VNT: turbine vane position,
- Volet: position component of admission,

#### 4.2 Second approach

##### 4.2.1 Modelling

In the second approach, we propose an innovative approach that allows us to predict the response from signals like pressure and cylinder temperature (states variables of combustion chamber).

More precisely, we decompose the problem of estimate of the engine responses, into two steps sub problems (Fig.2):

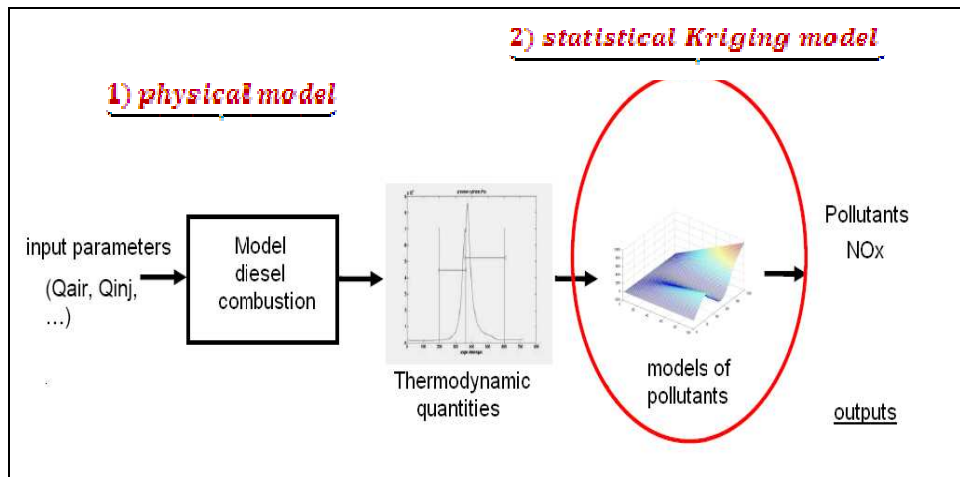




Fig.2. Coupling of the pollutants and consumption models with the combustion model.

- 1) The first step consists in simulating the various thermodynamic quantities from a physical model. In this work, we use the model developed by (Castric et al., 2007), which takes into account the input parameters. It leads to have a good representation of the experimental results. This model allows us to generate the following thermodynamic quantities:

- The cylinder low pressure (the alone quantity that we can measure)
- The temperature in the cylinder,
- The temperature of the fresh gas in the cylinder
- The temperature of the mixed gas in the cylinder,
- The temperature of the burned gas in the cylinder,
- The mass of the fresh gas in the cylinder,
- The mass of the entrained gas in the cylinder,
- The mass of the burned gas in the cylinder,
- The turbulence in the motor,
- The fuel vapor mass.

We precise, that each signal is represented by a vector of 1334 components.

- 2) The second step consists in building a statistical Kriging model, from the 11 thermodynamics quantities generated by the model of combustion.

It is true that the advantage of this procedure is that, it allows us to include a physical knowledge of combustion. But this approach requires a great time of computing. Indeed to build Kriging from 11 signals, can pose a serious problem in memory capacity and the computing time can be considerable. Thus, to be able to implement this procedure, a reduction of the model is essential.

#### 4.2.2 Model reduction

The data of the first model can be directly used for the Kriging. It is not the case for the second one. In the last case the data have to be reduced.

The reduction process begins by studying the different correlations between the state variables and their corresponding p-value. The chosen criterion consists in testing the p-value: if it is less than to 0.05, the correlation is significant. This analysis allows us to retain only two state variables: the cylinder pressure  $P$  and the mixed gas temperature in the cylinder,  $T_e$ .

In the second step, the number of components of the two remaining signals is reduced. This is accomplished by using the discrete Fourier transform. The function `fft` of Matlab returns the discrete Fourier transform (DFT) of a vector, computed with a fast Fourier transform (FFT) algorithm. After calculating the coefficients a minimum number of these are retained. This allows us to reproduce the initial signal, with a relative error approximately less than 0.02.

The reduction of the number of points of each signal is tantamount to minimizing the number of Fourier coefficients representing that signal. The two retained signals representing respectively the cylinder pressure and the temperature of the mixed gas in the cylinder, have been reduced to a 40 Fourier coefficients. Each signal has been reconstructed from the 40 kept coefficients, with an acceptable relative error. The following table.1 presents the relative error committed, for the reconstruction of the two signals from the 40 coefficients selected:

$$\text{Relative error} = \frac{\|S - S_{r_{FFT}}\|}{\|S\|}$$

S: is the experimental signal

$S_{r_{FFT}}$  is the reconstruction of the signal S using the fast Fourier transformation.

$\|\cdot\|$  is the Euclidian norm.

Type of signal	relative error
the cylinder low pressure	0.01
the temperature of the mixed gas in the cylinder	0.02

Table1. Relative error committed for the reconstruction of two signals.

Figure 3 shows the experimental signals, resulting from the combustion and their reconstruction by using the fft Matlab function.

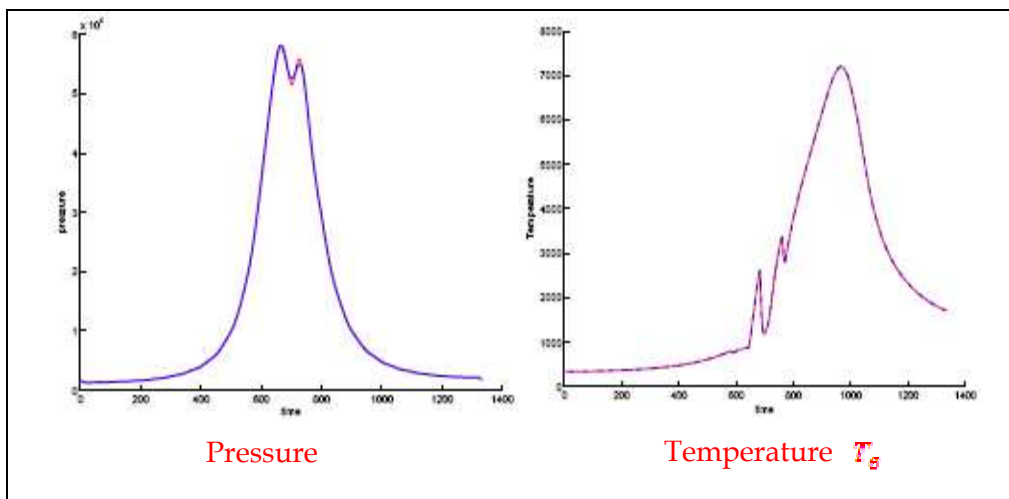


Fig.3 Rebuilding of the measured signals (red curve) by using the discrete Fourier transform (blue curve)

Such reduction makes the Kriging possible, The considered entries of the model are:

$$\begin{cases} P_{cyl,i}(t) \rightarrow [c_1(P), \dots, c_n(P)]_i \\ T_{cyl,i}(t) \rightarrow [c_1(T), \dots, c_n(T)]_i \end{cases}$$

Where:

-  $i$  is the index that correspond to the  $i$ th operating point of engine. An operation point of the engine is defined by engine speed and engine torque

-  $c_n(f)$  is the kept Fourier coefficient for the signal  $f$ , which correspond to the  $i$ th operating point of engine.

## 5. Application to the estimation of engine responses

In the previous section, we have presented and explained the two approaches used in this work, in order to model a behavior of diesel engine. Then, this section will be devoted to present the results respectively obtained by each approach, for the estimation of each response.

### 5.1 Numerical results using the first approach

We recall that the construction and the modeling of the experimental variogram is the most important step in the Kriging method. Thus, in this part, we will start by giving the chosen model.

#### *Variogram fitting:*

Variography modeling is a critical step and most difficult in the construction of a Kriging model. For this reason, several models were adjusted and then compared. It was difficult to select the better model graphically. The cross validation facilitates the work. It allows us to select the one, which minimizes the root mean square error.

For the NOx, the retained model is a Gaussian model which is expressed by the equation:

$$\gamma(r) = c_0 + c \left( 1 - \exp\left(-\frac{r^2}{a^2}\right) \right) \quad \text{as } r \geq 0$$

The value of the model parameters was founded using the least square method. So we obtain:

$$c_0=10.929, c=1.829, a=309.559.$$

For the Consumption, the model used is an exponential model, given by the equation:

$$\gamma(r) = c_0 + c \left( 1 - \exp\left(-\frac{r}{a}\right) \right) \quad \text{as } r \geq 0$$

This leads to:  $c_0=7.106, c=2.415, a=477.444$ .

Where:

- r is the distance
- $c_0$  is the Nugget effect
- $c_0 + c$  is the sill correspond to the variance of  $Z(x)$
- $\sqrt{3}a$  and  $3a$  are the range (the distance at which the variogram reaches the sill) for the Gaussian and exponential model respectively (Baillargeon et al., 2004).

Figures 4 shows the experimental variogram (red points), and Gaussian model (blue curve) corresponding to NOx response.

Figures 5 shows the experimental variogram (red points), and exponential model (blue curve) corresponding to consumption response.

The variogram is a tool that quantifies the spatial correlation of the response of interest. It measures the variability of NOx and consumption as a function of distance, we notice that, when the distance reached the range  $535 \approx \sqrt{3}a$  (Fig.4) and  $1300 \approx 3a$  (Fig.5) the variation becomes stationary. This explains that we can have a similar behavior of consumption and NOx on two different operating points, thus with a pattern of different control parameters

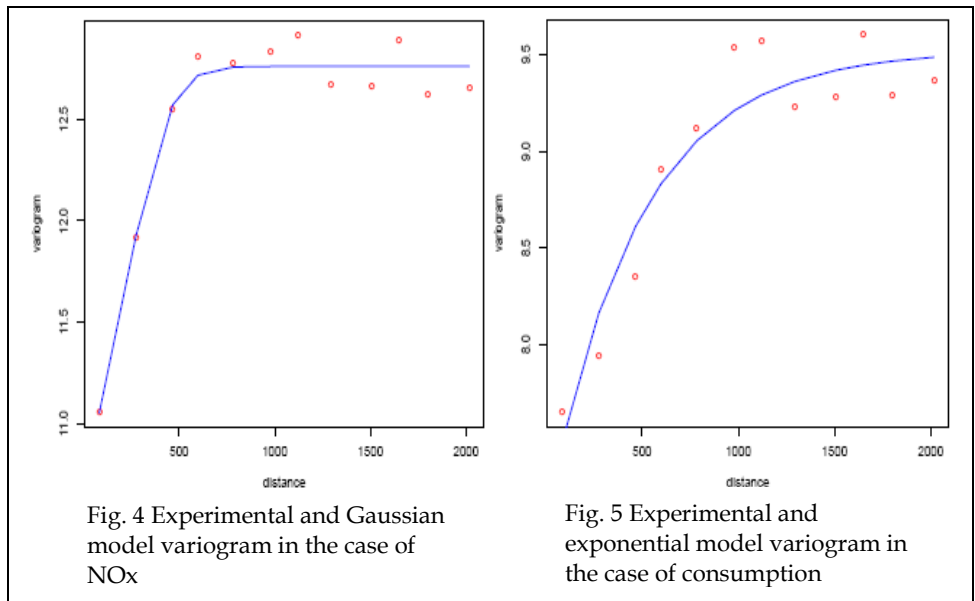


Fig. Experimental and model variogram

Figures 6 and 7 show the Cross-validation plots for the Kriging model, corresponding to the Gaussian and exponential variogram respectively. The plots contain the measured, the Kriging estimated value and a 10% errors bands.

The accuracy of predictions was similar for both validation data. Accuracy was good for both of the responses and still within 10% for the majority of operating conditions.

By against, graph 7 presents some observations which are poorly estimated. This is because they are far from the cloud of points used for the adjustment. On the other hand, this bad estimate is due to the experimental design used. The classical and optimal designs, in particular the D-optimal, are not suitable for Kriging, which is based on measuring similarity between sites. Indeed, the D-optimal design allows testing just a small number of levels for each variable and tend to generate points on the edges of the experimental field (Koehler & Owen, 1996) . This distribution of points, which is optimal to fit a polynomial model, cannot pick up any irregularities inside the experimental field and lead to some poorly estimated points. To address this problem, we recommend to use an appropriate designs for Kriging. Class 'space filling designs', such as Latin hypercubes, provide a good spatial distribution of points and is well adapted for modeling by Kriging (Stein, 1987), (McKay et al., 2000).

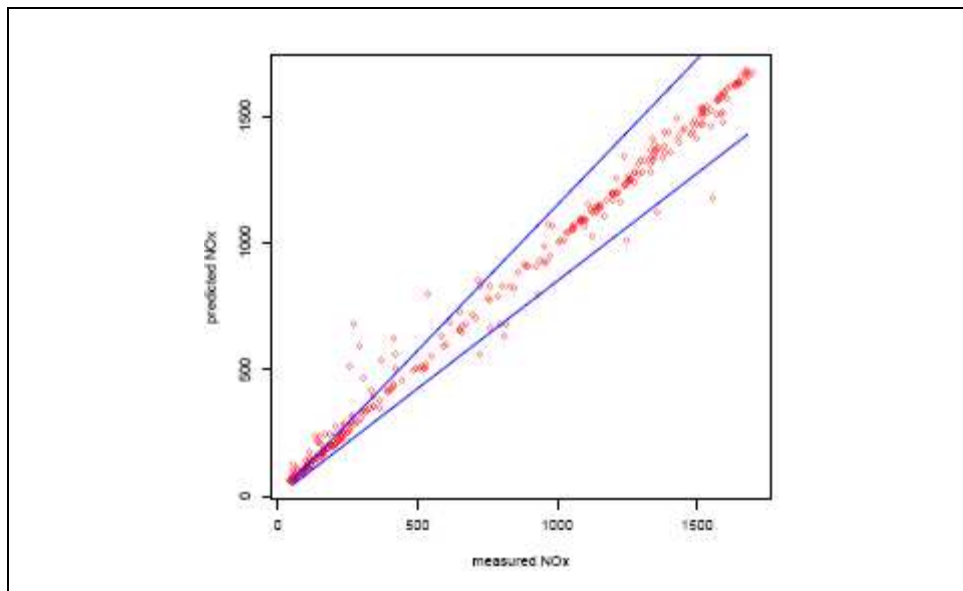


Fig.6. Measured and Kriging predicted NOx [ppm] with  $\pm 10\%$  error bands

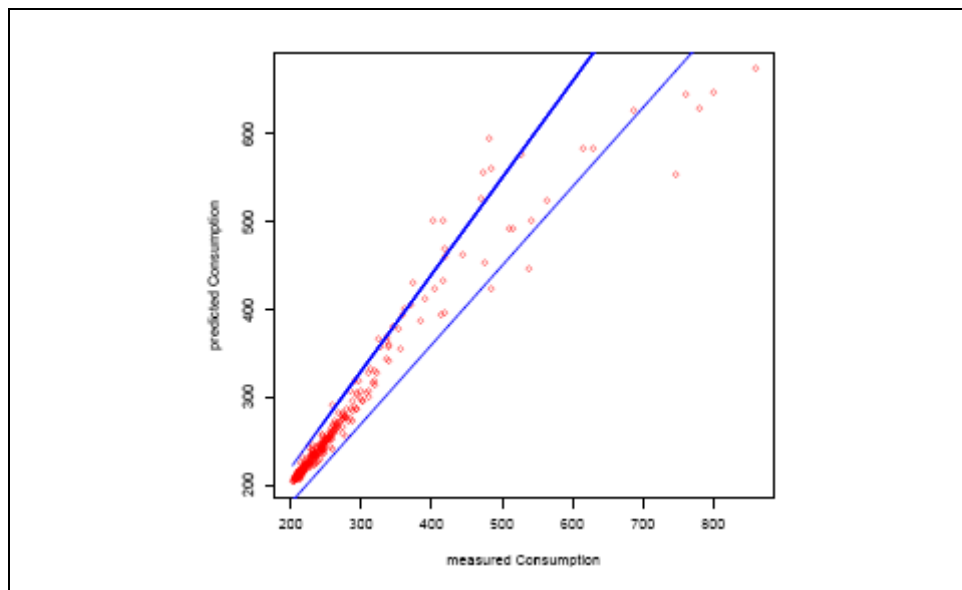


Fig.7. Measured and Kriging predicted consumption [g/kWh] with  $\pm 10\%$  error bands

The emulator model is fitted to each response in turn and the RMSE, percentage RMSE are recorded. These results are presented in Table2. The percentage RMSE results show that the model has a %RMSE less than 7% of the range of the response data. This indicates roughly, that if the emulator is used to predict the response at a new input setting, the error of prediction can be expected to be less than 7%, when compared with the true value.

	NOx	Consumption
RMSE	61.4	40.63
%RMSE	3.84	6.19

Table2. Kriging RMSE and %RMSE for each response: first approach case

### 5.2 Numerical results using the second approach

This subsection is devoted to the presentation of the numerical results obtained in the case of the second modeling more precisely we give the mathematical model used to adjust the experimental variogram.

#### *Variogram fitting:*

At the same way in the first approach, we build the experimental variogram and the model which adjusts it for each engine response.

For the NOx, the model used is a Gaussian model given by equation:

$$\gamma(r) = c_0 + cr^a \quad \text{as } r \geq 0 \text{ and } 0 \leq a < 2$$

The value of parameters model was founded using the least square method:

So,  $c_0=997.28$ ,  $c=0.00018$ ,  $a=1.52$ .

In this case the variogram does not show a sill, this means that the variance does not exist.

For the Consumption, the model used is an exponential model given by equation:

$$\gamma(r) = c_0 + c \left( 1 - \exp\left(-\frac{r}{a}\right) \right) \quad \text{as } r \geq 0$$

So  $c_0=5193$ ,  $c=0.0327 \cdot 10^5$ ,  $a=5.9536 \cdot 10^5$

Where:

- $r$  is the distance.
- $c_0$  is the Nugget effect.
- $c_0 + c$  is the sill correspond to the variance of  $Z(x)$ .
- $3a$  are the range (the distance at which the variogram reaches the sill) for the exponential model (Baillargeon et al., 2004).

Figures 8 shows the experimental variogram (red points), and power model (blue curve) corresponding to NOx response.

Figures 9 shows the experimental variogram (red points), and exponential model (blue curve) corresponding to consumption response.

We notice that when the distance reaches the range  $1.786110^6 \approx 3a$  (Fig. 9) the variation becomes stationary. In other term, it means that there is not a correlation beyond the distance  $3a$ . This explains that we have a similar behavior of consumption on two different operating points, thus with a pattern of different control parameters.

On the other hand, for the variogram of NOx, we notice that it is a power kind, which explains that for two different configurations engine, we will have a behavior completely different of the NOx emissions.

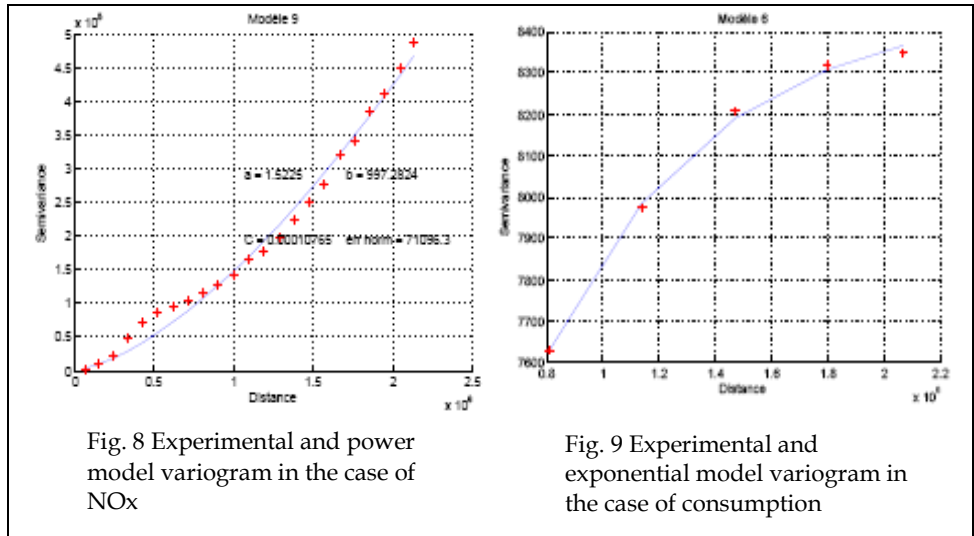


Fig. Experimental and model variogram

Figures 10 and 11 show the cross-validation plots for the Kriging model, corresponding to the power and exponential variogram respectively. The plots contain the measured, the Kriging estimated value and a 10% errors bands.

As we can see it, the accuracy of the predictions is similar for both response and still within 10% for the majority of operating conditions.

On the other hand, we notice that in the second approach, the accuracy of the predictions is improved for the two responses, compared to the first approach. This improvement is very clear for the consumption estimation.

We can explain this improvement, by the fact that in the second approach we include thermodynamic quantities such as the pressure for the prediction of the two responses. The inclusion of these quantities allows to bring back an additional knowledge for the prediction of the both responses. Indeed, this knowledge results from the fact, that these quantities represent the states variables of our system, and they characterize the behavior of combustion in the internal of the combustion chamber.



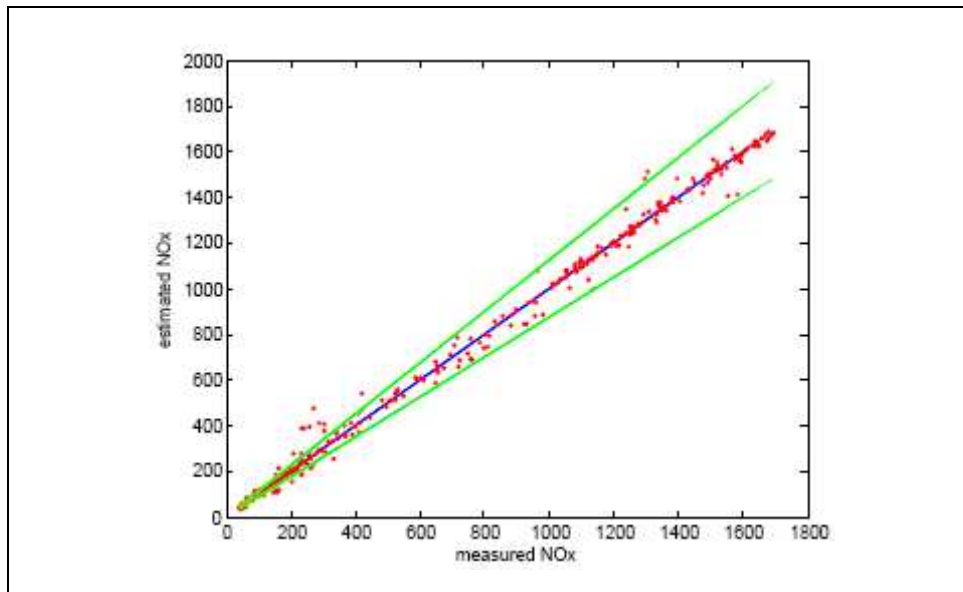


Fig.10. Measured and Kriging predicted NOx [ppm] with  $\pm 10\%$  error bands

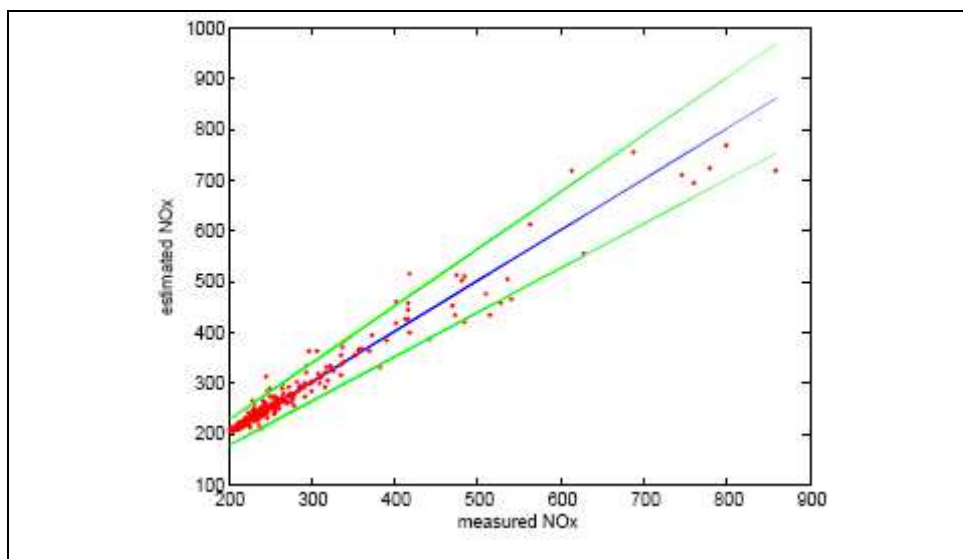


Fig.11. Measured and Kriging predicted consumption [g/kWh] with  $\pm 10\%$  error bands

The emulator model is fitted to each response in turn and the RMSE, percentage RMSE are recorded. These results are presented in Table3. The percentage RMSE results show that the model has a %RMSE less than 4% of the range of the response data. This indicates roughly, that if the emulator is used to predict the response at a new input setting, the error of prediction can be expected to be less than 4%, when compared with the true value.

	NOx	Consumption
RMSE	40.51	19.99
%RMSE	2.45	3.04

Table3. Kriging RMSE end %RMSE for each response: second approach case

## 6. Comparison and discussion

We recall that in the section 4, we presented two different approaches, based on the Kriging model. In this section we will try to make a comparison between these two approaches, and discuss the advantages and inconvenient of each approach.

### Case of NOx:

A legitimates question, that we can pose in the case of the estimate of NOx is the following one:

Why did we obtain a variogram Gaussian in the first approach, however in the second approaches the variogram is the power form, knowing that the pressure (observed) is obtained, from the same control parameters utilized in the first approach?

In fact, the power variogram obtained in the second approach represents more the true behavior of the emissions of NOx. Indeed, the interpretation of the power variogram suggests that the variability of the response (in this box the NOx) increases with the distance between the points. This interpretation joins the opinion of the experts, who say that for two various engine configurations, the quantity of the corresponding NOx emissions will be also different.

Obtaining a variogram Gaussian in the first approach, is explained by the fact that the speed parameter of the engine take a raised values compared to the other control parameters. For example, if we take the first and the second line of the table 5, which correspond to two speeds different engine. We notice that the behavior of NOx is similar, however the distance between these two points is very tall (caused by speed mode) what explains the sill on the variogram of the first approach.

Fortunately, that this change in the behavior of variogram does not have an influence on the prediction of NOx. But the interpretation of the variogram in the first approach can lead us to make false conclusions. Indeed, in the case of the first approach, the variogram makes us believe that the quantity of the NOx emissions remains invariant when we consider very different configurations of control parameters. What does not reflect reality. In the case, where we wish to use the variogram, to understand how a response varies. We advise to check the values of the data, or to standardize the factors of the model.

N	Prail	Main	Mpil1	Mpil2	Pmain	Ppil1	Ppil2	VNT	VEGR	Volet	NOx
1000	407,70	5,91	1,0	1,09	-4,43	-18,72	-11,22	79,99	36,01	75,95	67,06
2000	609,00	11,15	1,1	1,34	-5,99	-36,26	-15,22	67,49	34,51	75,93	64,15

Table5. Example of control parameters configuration

### Case of consumption:

To manage to highlight the contribution of the second approach in the improvement of the prediction of consumption we consider another representation of the results in figure 12.

We note that for the first approach, the Kriging method could estimate with a good accuracy all the points which are close to the cloud used for the adjustment. On the other hand the prediction of the points which are far from the cloud was bad (an explanation was given in section 5.1).

The use of the second approach brought back an improvement for the estimate of these points. This gives a force of extrapolation to the Kriging method.

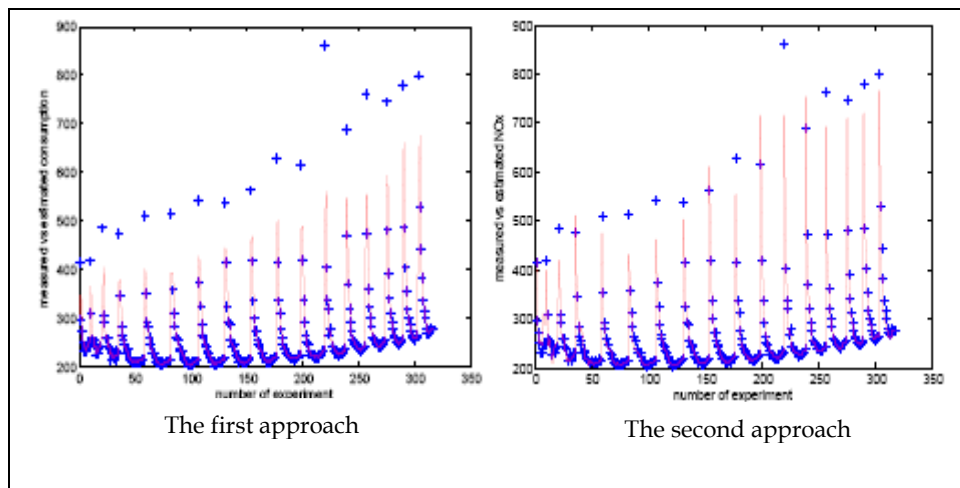


Fig.12. Comparison of consumption estimation for the two case approaches. (the + points are the experimental data and the red line is the model )

## 7. Conclusion

This paper deals with the problem of engine calibration, when the number of parameters of control is considerable. An efficient process of solving such problems falls into three successive steps: experimental design, statistical modeling and optimization. In this paper, we concentrate on the second step. And we demonstrated the important role of the experimental design on the quality of the prediction of the Kriging model in the case of consumption response. On the other hand the Kriging model was adapted to allow an estimate of the response in the case of higher dimensions. It was applied to predict the two engine responses NOx and consumption through two approaches. The first approach gave acceptable results. These results were clearly improved in the second approaches especially in the case of consumption. We demonstrated that the resulting model can be used to predict the different responses of engine. It is easy to generalize for various diesel engine configurations and is also suitable for real time simulations. In the future, this model will be coupled with the evolutionary algorithms for multi-objective constrained optimization of calibration.

## 8. References

- Arnaud, M.; Emery, X. (2000). Estimation et interpolation spatiale. *Hermès Science Publications*, Paris.
- Bates, R.A.; Buck, R.J.; Riccomagno, E.; Wynn, H.P. (1996). Experimental Design and Observation for large Systems. *J. R. Statist. Soc. B*, vol. 58, (1996) pp. 77-94.
- Baillargeon, S.; Pouliot, J.; Rivest, L.P.; Fortin, V.; Fitzback, J. interpolation statistique multivariable de données de précipitations dans un cadre de modélisation hydrologique, Colloque Géomatique 2004: un choix stratégique, Montréal (2004)
- Castric, S.; Talon, V.; Cherfi, Z.; Boudaoud, N.; Schimmerling, N. P. A, (2007) Diesel engine combustion model for tuning process and a calibration method. IMSM07 The Third International Conference on Advances in Vehicul Control and Safety AVCS'07, Buenos Aires, Argentine (2007).
- Castric, S. (2007) Readjusting methods for models and application for diesel emissions, PhD thesis, University of Technology of Compiègne, 2007.
- Christakos, G. (1984). On the problem of permissible covariance and variogram models. *Water Resources Research*, 20(2):251-265.
- Cochran, W. G.; Cox, G. M. (1957). *Experimental Designs*. Second edition. New York : Wiley. p 611.
- Cressie, N. A. C. (1993) *Statistics for spatial data*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Inc., New York. Revised reprint of the 1991 edition. A Wiley-Interscience Publication.

- Davis, J.C. *Statistics and Data Analysis in Geology*, second edition John Wiley and Sons. New York (1986).
- Edwards, S.P.; A.D.P.; Michon, S.; Fournier, G. The optimization of common rail FIE equipped engines through the use of statistical experimental design, mathematical modelling and genetic algorithms, *S.A.E paper*, vol. 106, n°3, (1997), pp. 505-523.
- Goers, A.; Mosher, L.; Higgins, B. (2003). Calibration of an aftermarket EFI conversion system for increased performance and fuel economy with reduced emissions, *S.A.E. paper*, vol. 112, n°3, March 2003, pp. 1390-1407, 2003-01-1051.
- Heywood, J. (1988) *Internal combustion engine fundamentals*, London : Mac Graw-Hill (1988)
- Koehler J.R.; Owen A.B. (1996) *Computer Experiments*. In Ghosh, S., Rao, C.R., (Eds.), *Handbook of Statistics, 13 : Designs and Analysis of Experiments*, North- Holland, Amsterdam, p.261-308. (1996)
- Krige, D.G. (1951) A statistical approach to some basic mine valuation problems on the Witwatersrand, *J. of Chem. Metal. and Mining Soc. of South Africa*. Vol. 52 pp 119-139 (1951).
- McKay M.D., Beckman R.J., Conover W.J. Comparison of three methods for selecting values input variables in the analysis of output from a computer code, *Technometrics*, Vol. 42, n°1, (February 2000) pp. 55 - 61, 239-245
- Matheron, G. (1963) Principles of Geostatistics, *Economic Geology*, v. 58, n° 8, (December 1963) pp. 1246-12688.
- Pierpont D. A.; Montgomery D. T.; Reitz R. D. Reducing particulate and NOx using multiple injection and EGR in a D.I. diesel, *S.A.E paper*, vol. 104, n°4 March(1995) , pp. 171-183 950217.
- Pilley, A.D.; A.J.B.; Robinson, D.; Mowll, D. (1994) Design of experiments for optimization of engines to meet future emissions target, *International Symposium on Advanced Transportation Applications* (1994).
- Sacks J., Schiller S.B., Welch W.J. (1989) Designs for Computer Experiments. *Technometrics*, vol. 31,41-47.
- Schimmerling, P.; J.C.S. ; Zaidi, A. (1998) *Use of design of experiments*. Lavoisier.
- Stein, M. Large sample properties of simulations using Latin hypercube sampling, *Technometrics*, vol. 29, n°2, (1987) pp. 143-151, 0040-1706.