Gears Design with Shape Uncertainties using Monte Carlo Simulations and Kriging

Rodolphe Le Riche[1], Victor Picheny[2] and David Ginsbourger[3]
Ecole Nationale Supérieure des Mines de Saint-Etienne, France

André Meyer[4]
INSA de Rouen, France

and

Nam-Ho Kim[5]
Mechanical and Aerospace Eng. Dept., University of Florida, Gainesville, USA

This article presents an approach to the optimization of helical involute gears for geometrical feasibility, contact ratio, teeth sliding velocity, stresses and static transmission error (STE). The teeth shape is subject to random perturbations due to wear (a randomized Archard's wear). The consequences of shape inaccuracies are statistically expressed as a 90% percentile of the STE variation, which is optimized. However, estimating a 90% STE percentile by a Monte Carlo method is computationally too demanding to be included in the optimization iterations. A method is proposed where the Monte Carlo simulations are replaced by a kriging metamodel during the optimization. An originality of the method is that the noise in the empirical percentile, which is inherent to any statistical estimation, is taken into account in the kriging metamodel through an adequately sized nugget effect. The kriging approach is compared to a second method where the STE variation for an average wear profile replaces the percentile estimation.

I. Introduction

Gears are a fundamental mechanical system involved when torques must be transmitted with high efficiencies, which is the case of transmissions in cars, windmills, and other special machines. Although gears have been designed for a long time, controlling gears performance under teeth shape uncertainties is a recent, difficult and important design objective : teeth shape variations result from manufacturing (e.g., heat treatment) and from wear and induce, in particular, noise.

The applicative objective of this article is to design helical gears so as to account for teeth shape inaccuracies. Computational challenges that are familiar in Reliability Based Design Optimization (RBDO) accompany this problem: the teeth in gears are moving solids in contact with each other whose detailed simulation by finite elements is computationally intensive. With the added computational cost of accounting for shape uncertainties (through Monte Carlo simulations, reliability index calculation or polynomial chaos expansion), the robust design of these systems requires careful methodological developments.

Gears design is traditionally approached by parameterizing the shape of the teeth as helical and involute[1]. Most gears in use today are chosen in standardized tables and are non-optimal. In Ref. 2, helical involute gears are designed for geometrical feasibility and large contact ratio by semi-analytical approaches. In Ref. 3, the size of

---

[1] CNRS Permanent research associate, Ecole Nationale Supérieure des Mines de Saint-Etienne (ENSM-SE), Saint-Etienne, France.
[2] PhD student, ENSM-SE and Dept. of Mechanical and Aerospace Eng., University of Florida, Gainesville.
[3] Post-doctorate student, IM & CHYN, Neufchatel university, Switzerland.
[4] Permanent Teaching Associate, Dept. of Mechanical Eng., INSA de Rouen, 76801 Saint-Etienne du Rouvray cedex, France.
[5] Assistant Professor, Dept. of Mechanical and Aerospace Eng., University of Florida, Gainesville FL-32611, USA.

helical gear sets is minimized where the gears stress analysis is based on standardized AGMA formula. A finite element analysis of the gears was implemented in Ref. 4 and served to optimize the pressure angle for minimum von Mises stresses. Modified teeth shapes were proposed in Ref. 4 to make the gears less sensitive to shape and positions perturbations. The gears performance criteria were bearing contact, transmission error and stress state.

In the present article, gears are designed for geometrical feasibility, contact ratio, teeth sliding velocity, stresses and Static Transmission Error (STE). Efficient stress and STE analyses are performed using Hertz contact formula and beam theory. Teeth wear is described by superposing Archard's wear model[6] and random material removal. Gears design is decomposed into i) the optimization of the 11 involute parameters for deterministic responses ii) the optimization of the crowning to control teeth shape random perturbations, i.e., to minimize the 90th percentile of the STE variation, ΔSTE. ΔSTE is one of the primary origins of the noise made by gears[7].

From a methodological standpoint, it is not computationally affordable to calculate the ΔSTE percentile by simple Monte Carlo simulations at each optimization step. A kriging metamodel of the percentile is built before the optimization. The optimization uses the kriging prediction of ΔSTE instead of the computationally intensive Monte Carlo simulations, but other optimization criteria, which are not subject to uncertainties, are calculated with the original gears model.

The contributions of this work are twofold: firstly, a complete formulation of the gears optimum design problem, involving the 90th percentile of the STE variation, is solved; secondly, a methodology for robust design based on kriging as a substitute for the Monte Carlo simulations is proposed. The noise in the statistical estimator of the performance is taken into account in the kriging metamodel through its "nugget" intensity. The kriging approach is compared to an alternative method where the STE variation for the average wear is minimized. The comparison is made in terms of the achieved 90% percentile of the STE variation.

II. Problem formulation

A. Gears analysis

Geometrical gears analysis is based on classical involute helical representation[1,8]. 12 design variables are considered, some of which are illustrated in Figure 1: Z_p, the number of teeth of pinion 1 ; x_p and x_r the addendum modification coefficients of the two pinions ; ha_{0p} and ha_{0r} the cutting tool addendum of the two pinions, hf_{0p} and hf_{0r} the cutting tool dedendum of the pinions ; ρ_{ap} and ρ_{ar} the tip fillet radiuses, ρ_{fp} and ρ_{fr} the root fillet radiuses ; β_0 the helix angle.

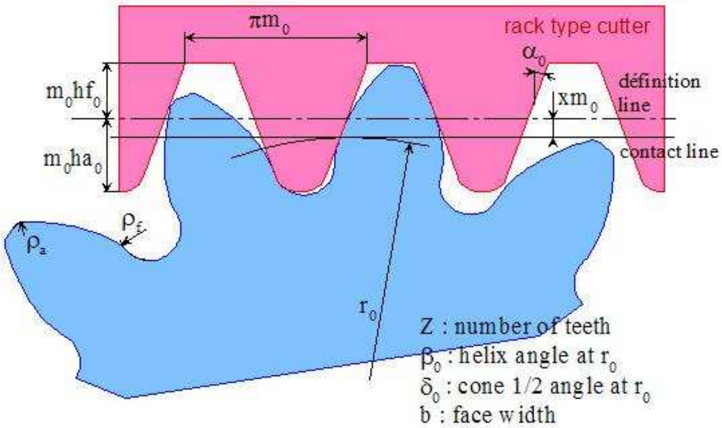

Figure 1: Gears parameters definitions.

The optimization criteria are:

(1) the contact ratio, r_c, i.e., the average number of teeth in contact, which is related to gears silence;
(2) the specific slip ratio, g_s, which is a measure of the tangential velocity of a tooth with respects to the tooth in contact and which is related to wear;
(3) the maximum contact pressure between teeth, F (using Hertz's model for contact between cylinders) ;
(4) σ_{vm}, the maximum von Mises stresses (at the teeth root);
(5) ΔSTE, the difference between the maximum and the minimum of the Static Transmission Error.

The *STE* describes the gears position errors due to teeth deformations under loads and due to teeth shape errors. It is estimated by a fast model based on Hertz contact and beam theories. All these aspects of gears analysis were implemented in the *Filengrene* software[8].

Wear profiles are perturbed Archard's profiles. They are generated by multiplying Archard's profile (proportional to *F*V*, where *F* is the contact pressure and *V* the relative sliding velocity) by a Gaussian random processes[13], **Θ**, as described in Figure 2. The Gaussian processes have mean 1, a range of 0.2 *rad* and a variance of 0.01.

The randomness in the wear profiles represents uncertainties in the load and number of cycles each gear will endure, hence inducing deviations (in amplitude and shape) from the nominal Archard's profile. Shape perturbations due to wear mainly affect the STE amplitude, ΔSTE, which is therefore a random function of the design variables, **x**, and the wear profiles, ΔSTE(**x**, **Θ**). All other gears design criteria (contact and specific slip ratios, stresses, etc.) are deterministic functions of **x** since they are marginally affected by shape variations of the order of 1 μm. A deterministic optimization criterion is obtained from ΔSTE by taking its 90th percentile, $P^{90}_{\Delta STE}(\mathbf{x})$. Such a criterion is natural when optimizing a random response because it ensures that 90% of the actual systems (gears here) will achieve the declared $P^{90}_{\Delta STE}$ performance. Computing this percentile involves Monte-Carlo simulations (MCS), which will be detailed in section 3.

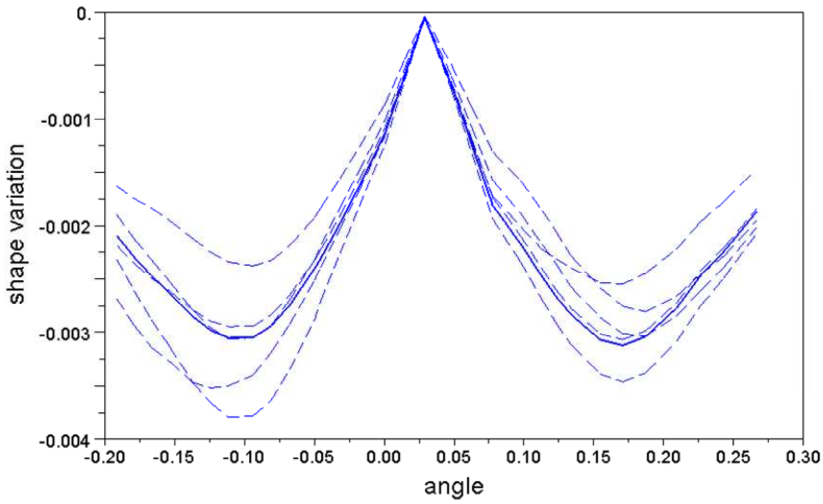

Figure 2: Profile wear (mm) vs. angle (rad). *The central continuous line corresponds to -0.1*F*V, i.e. wear according to Archard's model. F is the contact pressure and V the relative sliding velocity. Wears (dashed lines) are Gaussian random processes multiplied by Archard's model : Θ*(-0.1*F*V). Wear is then proportional to F*V curves with a maximum of 5 μm.*

B. Gears optimization formulation

Eventhough the Filengrene gears simulator is rapid (about 2 CPU seconds per gears analysis) it is not possible to directly include the Monte Carlo simulations inside the optimization loop because the computational costs are multiplied by each other : for 100 Monte Carlo simulations and 1000 optimization steps (which is an underestimation of the needed cost for globally optimizing in 11 dimensions), a single run would take 100*1000 = 56 CPU hours. To tackle the computational cost barrier, designing gears is first formulated as a deterministic problem in 11 variables, and then as a robust optimization problem in 6 variables. The initial deterministic optimization permits to fix 11-6=5 variables and provides an entry-level design to compare to.

Deterministic sub-problem:

The gears teeth involute shape is described by the 12 variables introduced earlier : Z_p, x_p, x_r , ha_{0p} , ha_{0r} , hf_{0p} , hf_{0r} , ρ_{ap} , ρ_{ar}, ρ_{fp} , ρ_{fr} and β_0. *Zp*, the number of teeth of the pinion, is the only integer variable. It is taken out of the formulation by setting its value a priori : with only one integer variable, optimizing the number of teeth boils down to repeating the procedure described in this article for various values of Z_p.

Let therefore **x** be the 11 design variables of the deterministic problem,

$$\mathrm{x} = \left[x_p, x_r,\ ha_{0p},\ ha_{0r},\ hf_{0p},\ hf_{0r}, \rho_{ap}, \rho_{ar}, \rho_{fp}, \rho_{fr}, \beta_0 \right] \tag{1}$$

The deterministic gears design problem is,

$$\text{Min } \Delta\text{STE}(\text{x}) \tag{2}$$

such that there is no teeth interference and

$$\begin{aligned} & r_c > 1.25, \quad g_s < 2, \quad \sigma_{vm} < \text{Re}, \quad F < P_{hertz}^{\max} \\ & \text{-}1 \le x_p, x_r \le 2, \quad 0.8 \le ha_{0p}, ha_{0r}, \ hf_{0p}, hf_{0r} \le 1.7, \\ & 0 \le \rho_{ap}, \rho_{ar} \le m_0, \quad 0.05\, m_0 \le \rho_{fp}, \rho_{fr} \le m_0 \quad and \quad 0 \le b_0 \le 30. \end{aligned} \tag{3}$$

The number of teeth of the second pinion and the module are not part of the variables vector because they are solved by satisfying two equality constraints. The gears ratio (0.9) yields

$$Z_r = \text{ integer}\left(Z_p / 0.9 \right) \tag{4}$$

m_0, the module (a scale factor) is calculated by solving a non linear equation stating that the distance between the gears centers is equal to *110 mm*. The details of this equation are beyond the scope of this article. Other working data are : pressure angle $\alpha_0=18$ *deg*, the torque transmitted by the pinion is 200Nm, the material is a 20 NC 6 cemented steel ($E=200$ *GPa*, $\nu=0.29$, $Re=980$ *MPa*, $P_{hertz}^{max}=1550$ *MPa*) and the gears thickness is *10 mm*.

The deterministic sub-problem is handled with the Covariance Matrix Adaptation Evolution Strategy, CMA-ES[9], which is the state-of-the-art evolutionary optimizer. Constraints are treated by a static linear penalization of the objective function, i.e., by minimizing

$$f_p(x) \quad = \quad \frac{\Delta EST}{\Delta_{ref}} + p \times \left[\left(1 - \frac{r_c}{1.25}\right)^+ + \left(\frac{g_s}{2} - 1\right)^+ + \left(\frac{\sigma_{vm}}{R_e} - 1\right)^+ + \left(\frac{F}{P_{hertz}^{\max}} - 1\right)^+ \right] \tag{5}$$

where $(...)^+ \equiv \max(0,...)$ and $p = 10$. CMA-ES is a stochastic optimizer, it is therefore not sensitive to the slope discontinuity at 0 introduced by our penalization scheme. This penalization scheme has, on the other hand, the advantage when compared to quadratic penalty functions of allowing convergence to feasible solutions at a finite, reasonably small values of *p*[14].

Gears simulations are not possible for every choice of the design variables: for example, there are configurations when the involute equation cannot be solved. These cases are handled by setting the objective function equal to a large number (100 here), which makes the final penalized objective function non-continuous. This is another reason for using a zero order optimizer such as CMA-ES. Note that other evolutionary optimizers (CMA-ES only handles continuous variables) are also often applied to gears design because the number of teeth is an integer variable[10,11].

Bounds on the variables are handled by repeating the probabilistic CMA-ES point generations until an in-bound point is proposed. Of course, out-of-bounds points are rejected without further analysis, so the numerical cost of this strategy is negligible when compared to gears analyses.

Robust optimization sub-problem

The robust optimization sub-problem is much more computationally intensive than the deterministic sub-problem because it involves estimating at each optimization step the 90% percentile of ΔSTE. Two strategies are proposed to decrease the computational cost : reducing the design space dimension and approximating percentiles with kriging (see Section 3). The reduction in dimension is achieved by considering the 6 design variables $\mathbf{x} = [x_p, x_r, ha_{0p}, ha_{0r}, hf_{0p}, hf_{0r}]$, while the rest of the variables, ρ_{ap} , ρ_{ar}, ρ_{fp} , ρ_{fr} and β_0 , are set equal to their deterministic optimal values. The reasons for this reduction are that *i)* the fillet radiuses have a more local influence than the addendum and dedendum and ii) the effect of the helix angle is almost completely decoupled from the effects of the other parameters (it acts in another dimension) : β_0 could eventually be tuned a posteriori.

The effect of teeth shape fluctuations due to wear is controlled by solving:

$$\text{Min } P^{90}{}_{\Delta STE}(\text{x}) \tag{6}$$

by changing $\text{x} = \left[x_p, x_r, \ ha_{0p}, \ ha_{0r}, \ hf_{0p}, \ hf_{0r} \right]$

such that there is no teeth interference and

$$
\begin{aligned}
&r_c > 1.25, \quad g_s < 2, \quad \sigma_{vm} < \mathrm{Re}, \quad F < P_{hertz}^{\max} \\
&\text{-}1 \le x_p, x_r \le 2, \quad 0.8 \ \le \ ha_{0p}, ha_{0r}, \ hf_{0p}, hf_{0r} \ \le \ 1.7 .
\end{aligned}
\tag{7}
$$

The next section describes how we proceed to calculate the 90th percentile.

III. Percentile estimation through Monte-Carlo simulations and kriging

A. Computing percentiles of ΔSTE

In section II, we described wear profiles as random processes modifying Archard's profiles. Hence, ΔSTE is a random variable, whose distribution can be estimated by Monte Carlo simulations (MCS). To do so, we first generate a large number (*k*) of wear profiles, $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ..., \boldsymbol{\theta}_k$ (again, details of the random wear profiles generation were given in Figure 2). For each of these profiles, we compute the corresponding $\Delta\mathrm{STE}(\mathbf{x}, \boldsymbol{\theta}_i)$. Then, the 90th percentile $P^{90}_{\Delta\mathrm{STE}}(\mathbf{x})$ is estimated from the sample $\{\Delta\mathrm{STE}(\mathbf{x}, \boldsymbol{\theta}_1), ..., \Delta\mathrm{STE}(\mathbf{x}, \boldsymbol{\theta}_k)\}$. We call $\hat{P}^{90}(\mathbf{x}^i; \ \ \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ..., \boldsymbol{\theta}_k)$ such an estimate.

A preliminary study is performed on four designs to choose *k*, the number of MCS (see Table 1). For each design, 500 MCS are performed. Using a Lilliefors test[15], we find that all the samples $\{\Delta\mathrm{STE}(\mathbf{x}_i, \boldsymbol{\theta}_1), ..., \Delta\mathrm{STE}(\mathbf{x}_i, \boldsymbol{\theta}_k)\}, i = 1, \ldots, 4$, follow normal distributions. Hence, we assume that ΔSTE is normally distributed for any design **x**.

Table 1: preliminary statistical analysis of four designs.

Design #	$\mathbf{x_p}$	$\mathbf{x_r}$	$\mathbf{ha_{0p}}$	$\mathbf{ha_{0r}}$	$\mathbf{hf_{0p}}$	$\mathbf{hf_{0r}}$	Mean(ΔSTE)	SD(ΔSTE)	$\hat{P}^{90}$
1	0.49	0.49	1.20	1.11	1.00	1.09	5.82	0.41	6.34
2	0.00	0.20	1.02	1.00	0.90	1.00	6.22	0.66	7.06
3	1.10	1.00	1.30	1.40	0.90	0.90	7.30	0.44	7.87
4	0.40	-0.40	1.70	1.00	0.90	1.60	6.66	0.45	7.24

The normality of the samples allows us to increase the accuracy of the estimated percentile. Indeed, an unbiased estimate $\hat{P}^{90}$ of $P^{90}_{\Delta\mathrm{STE}}(\mathbf{x})$ is then

$$\hat{P}^{90} = m + 1.28 s \tag{8}$$

where

$$
\begin{aligned}
m &= \frac{1}{k} \sum_{i=1}^{k} \Delta\mathrm{STE}(\mathbf{x}, \boldsymbol{\theta}_i) \quad , \text{ and} \\
s &= \sqrt{\frac{1}{k-1} \sum_i \left[\Delta\mathrm{STE}(\mathbf{x}, \boldsymbol{\theta}_i) - m \right]^2} \quad .
\end{aligned}
\tag{9}
$$

The variance of $\hat{P}^{90}$ can now be expressed analytically. Since *m* and *s* are independent random variables,

$$\mathrm{var}\left(\hat{P}^{90}\right) = \mathrm{var}(m) + 1.28^2 \,\mathrm{var}(s) . \tag{10}$$

Let $\sigma^2_{\Delta STE}$ denote the actual variance of $\Delta\mathrm{STE}(\mathbf{x}, \boldsymbol{\theta})$. We have[15]:

$$\mathrm{var}(m) = \frac{\sigma^2_{\Delta STE}}{k} \tag{11}$$

$$\mathrm{var}(s^2) = \frac{2\sigma^4_{\Delta STE}}{k-1} . \tag{12}$$

It can be shown using a first order Taylor expansion that the variance of *s* is approximately equal to

$$\mathrm{var}(s) = \frac{\sigma^2_{\Delta STE}}{2} \left(1 - \sqrt{\frac{k-3}{k-1}} \right) . \tag{13}$$

Finally,

$$\mathrm{var}\left(\hat{P}^{90}\right)=\sigma_{\Delta STE}^{2}\left[\frac{1}{k}+\frac{1.28^{2}}{2}\left(1-\sqrt{\frac{k-3}{k-1}}\right)\right] \tag{14}$$

Estimating a percentile requires a large number of calls to the gears simulation. To reduce this computational cost, we approximate the percentile by a simple model, often called *metamodel* or *surrogate model*. We chose a kriging metamodel (described in section III-B) because, contrarily to deterministic metamodels (neural networks, response surfaces, support vector machines, …), it allows to account for the noise in the estimated percentile.

B. A kriging metamodel for noisy observations

We make the assumption that the $\hat{P}^{90}$ estimate is equal to the true function $P^{90}_{\Delta\mathrm{STE}}(\mathbf{x})$ plus a random noise (due to the Monte Carlo sampling process):

$$\hat{P}^{90}\left(\mathbf{x}^{i};\quad \boldsymbol{\theta}_{1},\boldsymbol{\theta}_{2},...,\boldsymbol{\theta}_{k}\right)=P_{\Delta STE}^{90}\left(\mathbf{x}^{i}\right)+\varepsilon^{i} \tag{15}$$

The variance of ε^i is defined here as the variance of the estimated percentile (14).

$P^{90}_{\Delta\mathrm{STE}}$ is observed at n distinct locations $\mathbf{X}$:

$$\begin{aligned}\mathbf{X}&=\left\{\mathbf{x}^{1},...,\mathbf{x}^{n}\right\}\\ \hat{\mathbf{P}}^{\mathbf{90}}&=\left\{\hat{P}^{90}\left(\mathbf{x}^{1}\right),...,\hat{P}^{90}\left(\mathbf{x}^{n}\right)\right\}\end{aligned} \tag{16}$$

$\mathbf{X}$ is called the *design of experiments* (DoE) and $\hat{\mathbf{P}}^{\mathbf{90}}$ the *observations vector*.

$P^{90}_{\Delta\mathrm{STE}}$ is approximated by a simple model M, called metamodel, based on hypotheses on the nature of $P^{90}_{\Delta\mathrm{STE}}$ and on its observations $\hat{\mathbf{P}}^{\mathbf{90}}$ at the points of the DoE. In this article, we consider the ordinary kriging (OK) metamodel, which assumes that the function to approximate is one realization of a Gaussian process[13] Y of the form

$$Y\left(\mathbf{x}\right)=\mu+Z\left(\mathbf{x}\right) \tag{17}$$

where μ is an unknown constant mean and $Z(\mathbf{x})$ a stationary gaussian process of known covariance. With these assumptions, the probability density function of $Y(\mathbf{x})$ knowing the observations $\hat{\mathbf{P}}^{\mathbf{90}}$ is Gaussian and analytically known

$$\left[Y\left(\mathbf{x}\right)\mid\hat{\mathbf{P}}^{\mathbf{90}}\right]\sim N\left(m_{OK}\left(\mathbf{x}\right),s_{OK}^{2}\left(\mathbf{x}\right)\right)\ . \tag{18}$$

The OK prediction mean and variance at $\mathbf{x}$ are

$$m_{OK}\left(\mathbf{x}\right)=\mu+\mathbf{k}\left(\mathbf{x}\right)^{T}\mathbf{K}_{\Delta}^{-1}\left(\hat{\mathbf{P}}^{\mathbf{90}}-\mu\mathbf{1}\right) \tag{19}$$

$$s_{OK}{}^{2}(\mathbf{x})=\sigma^{2}-\mathbf{k}^{T}\left(\mathbf{x}\right)\mathbf{K}_{\Delta}{}^{-1}\mathbf{k}\left(\mathbf{x}\right)+\frac{\left(1-\mathbf{k}\left(\mathbf{x}\right)\mathbf{K}_{\Delta}{}^{-1}\mathbf{1}\right)^{2}}{\mathbf{1}^{T}\mathbf{K}_{\Delta}{}^{-1}\mathbf{1}} \tag{20}$$

where

- σ^2 is the process variance,
- $\mathbf{1}$ is a $m\times 1$ vector of ones,
- $\mathbf{k}^{T}\left(\mathbf{x}\right)=\left[k\left(\mathbf{x},\mathbf{x}_{1}\right)\quad k\left(\mathbf{x},\mathbf{x}_{2}\right)\quad\dots\quad k\left(\mathbf{x},\mathbf{x}_{m}\right)\right]$, an $m\times 1$ vector,
- $\mathbf{K}_{\Delta}=\mathbf{K}+\boldsymbol{\Delta}$,
- $\mathbf{K}=\left(k\left(\mathbf{x}_{i},\mathbf{x}_{j}\right)\right)_{1\le i,j\le n}$, an $m\times m$ matrix,
- $\boldsymbol{\Delta}=diag\left(\left\{\mathrm{var}\left(\varepsilon_{1}\right)\quad\mathrm{var}\left(\varepsilon_{2}\right)\quad...\quad\mathrm{var}\left(\varepsilon_{m}\right)\right\}\right)$, is the so-called *nugget* effect accounting for the noise in the observations
- and $k\left(\mathbf{u},\mathbf{v}\right)$ a covariance function.

For more complete proofs, see for instance Ref. 12 or Ref. 13. Figure 3 shows an example of OK model. Note that since the observations are noisy, the kriging mean does not interpolate the data.

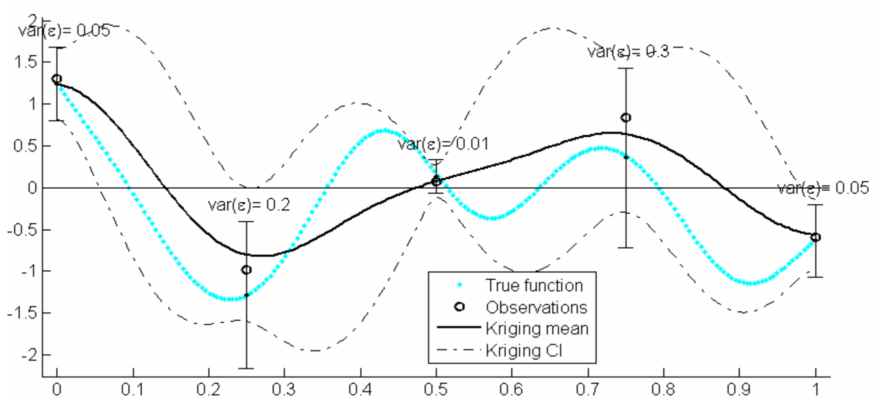

Figure 3: Example of Ordinary Kriging model with noisy observations. *The bars represent the noise standard deviation. Because of the noisy observations, the kriging mean does not interpolate the data and the kriging variance is non-zero at the observation points.*

C. Design of Experiments

We have seen that each observation $\hat{P}^{90}$ requires k calls to the gears simulator. Hence, a design of experiments (DoE) of size n requires $k \times n$ calls. Due to computational limitations, the number of calls cannot exceed 90,000. Both numbers must be chosen to ensure best trade-off between:

- Space-filling
- Reasonable variance of each estimate
- Affordable computational time

A larger n ensures a better filling of the design space, while a larger k reduces the variance of each observation. Empirical studies [16] show that for the kriging model, it is more accurate to have large variance and large DoEs. Hence, we choose k as small as possible. However, the hypothesis of normality of the error do not stand for very small samples (k <25), so we choose k = 30, and n = 3,000.
The preliminary analysis already discussed in Table 1 provides us with large ΔSTE samples at four design points, which can be used to validate our choice of k. Table 2 shows the variability of the percentile estimates, $\mathrm{var}\left(\hat{P}^{90}\right)$, from (14) assuming only 30 MCS are performed (the true variance $\sigma^2_{\Delta STE}$ is replaced by its accurate estimate based on 500 MCS).

Table 2: Variability of percentile estimates based on k = 30 Monte Carlo simulations.

Design #	$\mathrm{var}\left(\hat{P}^{90}\right)$ from (21)	$SD\left(\hat{P}^{90}\right)$	$\hat{P}^{90}$
1	0.010	0.10	6.34
2	0.027	0.16	7.06
3	0.012	0.11	7.87
4	0.013	0.11	7.24

For all four designs, we see that the standard deviations are very small compared to the mean value, so k = 30 is sufficient.

To ensure good space-filling properties, the training point locations are chosen from a Latin Hypercube Sampling (LHS) optimized for a maximum minimum distance (maximin) criterion. To each training point corresponds three response values: the ΔSTE with no wear, the ΔSTE with wear taken from the nominal Archard's profile (written ΔSTE_A), and the 90th percentile of ΔSTE with random wear (estimated by $\hat{\mathbf{P}}^{\mathbf{90}}_{\mathbf{\Delta STE}}$). Out of the 3,000 points, 825 are found as infeasible designs by the *Filengrene* software, so the final DoE consists of 2,175 observations. Table 3 summarizes the values taken by the observations.

Table 3: Observation values of the DoE.

ΔSTE (no wear)		ΔSTE_A (Archard's profile)		$\hat{P}^{90}_{\Delta STE}$	
Mean	Range	Mean	Range	Mean	Range
4.83	[0.49 ; 42.70]	7.31	[1.47 ; 37.75]	8.20	[2.12 ; 39.67]

D. Data analysis : is a robust optimization approach really necessary ?

First, to validate our approach, we look at the correlations between the responses. Indeed, there is no evidence *a priori* that the optimal design for the deterministic problem (2) is different from the robust optimal design (6).

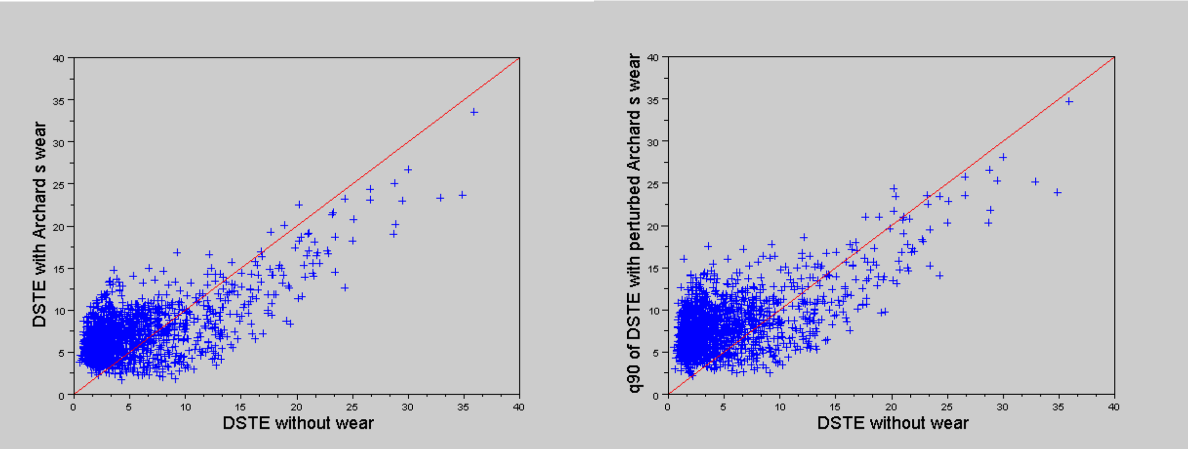

Figure 4: ΔSTE without wear vs. ΔSTE with Archard's wear profile (ΔSTE_A, left), and vs. the 90% percentile of ΔSTE with randomly perturbed Archard's wear (right).

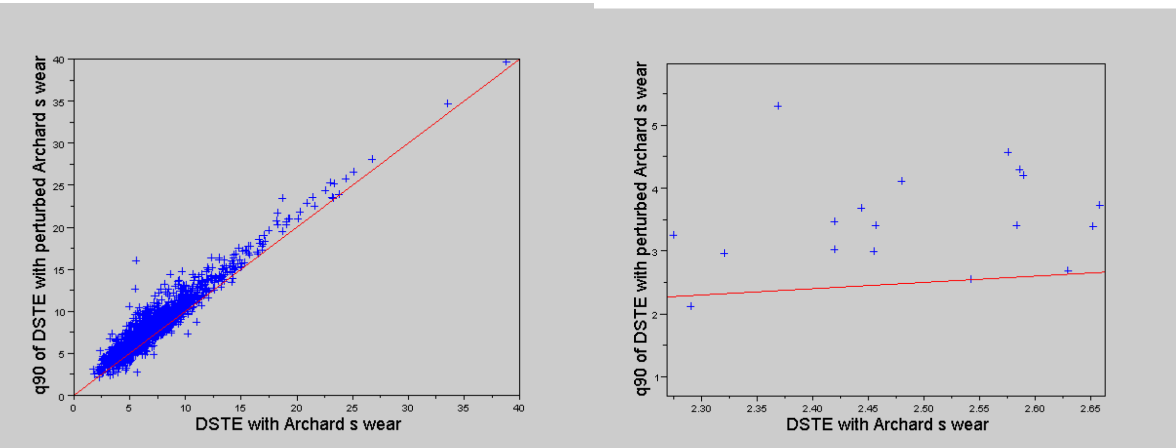

Figure 5: ΔSTE_A (with Archard's wear profile) vs. $\hat{P}^{90}_{\Delta STE}$ (30 Monte Carlo simulations). *Left: full scale, right: zoom.*

From Figure 4, we see that the ΔSTE with no wear differ in shape and amplitude from the ΔSTE with wear. Hence, the deterministic and robust optimizations are likely to have different solutions. On the contrary, ΔSTE from the nominal Archard's profile and the percentile from the stochastic profile are strongly correlated, the percentile being shifted by a margin. So it seems that for large scale optimization, the two problems are equivalent. However, for small values of ΔSTE (Figure 5, left), the correlation is weaker. It is difficult to determine if the difference is due to the noise in $\hat{P}^{90}$ or not.

To refine the analysis, we pick six points out of the 2,175 observations that have similar values of ΔSTE with Archard's wear (ΔSTE_A between 2.4 and 2.5) but correspond to different designs. At each design, we run 500 MCS to have an accurate estimate of the percentile (so the noise is negligible). The values of the accurate percentile estimates and the ΔSTE with deterministic wear are reported in Table 4.

Table 4: Comparison of six designs with similar **ΔSTE_A**.

Design #	**x_p**	**x_r**	**ha_{0p}**	**ha_{0r}**	**hf_{0p}**	**hf_{0r}**	**ΔSTE (Archard)**	**P^{90} ***
1	0.25	1.37	1.37	0.85	1.31	0.80	2.42	3.03
2	0.18	0.69	1.54	1.01	1.08	1.05	2.44	3.07
3	-0.21	1.11	0.90	1.23	1.61	0.90	2.46	3.34
4	-0.02	1.10	1.03	0.81	1.41	1.00	2.42	3.21
5	0.15	0.71	1.44	1.07	1.63	1.16	2.48	2.88
6	-0.12	-0.08	1.25	1.19	1.63	1.61	2.45	2.84

** Computed using 500 MCS*

From Table 4, we see that P^{90} and ΔSTE do not always behave similarly. For instance, designs #1 and #4 have the same best ΔSTE_A out of the 6 selected designs, but they have the third and fifth P^{90} . Design #5 is the worst design in terms of ΔSTE_A, but the second best in terms of P^{90} . This shows that optimizing ΔSTE with nominal (Archard) and stochastic wear is likely to lead to different solutions.

IV. Optimization results

A. Optimization without wear

The penalized objective function $f_p(x)$ (Formula *(5)*) has been minimized with the Covariance Matrix Adaptation Evolution strategy (CMA-ES[9]). The penalty parameter *p* was set to 10 and $\Delta_{ref} = 0.5$. The 11 design variables were scaled by their range, the initial CMA-ES "step size" was 0.3, there were 10 parents and 5 child designs at each CMA-ES iteration, and the runs were 12000 analyses long with a restart (taking the best-so-far design as new starting point and re-setting the step size to 0.3) at 6000 analyses. The optimization was repeated from three initial points randomly chosen inside the bounds. One of these initial (therefore non optimal) design point is presented in Figure 6. The best performing gear out of the three optimizations, which we consider as the optimum of the deterministic gears design formulation, is given in Figure 7. Comparison of the starting and optimal designs show that, at the deterministic optimum, the teeth mesh has no play and the helix angle is at its upper bound. Large helix angles induce forces in the gears axes direction, but the induced axis force is not a criterion of the current design formulation *(2)* and *(3)*. The optimum STE profile has four modes (i.e., four maxima) for one mesh period, versus one for the starting design shown in Figure 6. Each mode corresponds to a teeth pair entering or leaving contact. It is likely that larger teeth numbers would be optimal for minimizing the static transmission error variation by enabling more modes in a mesh period.

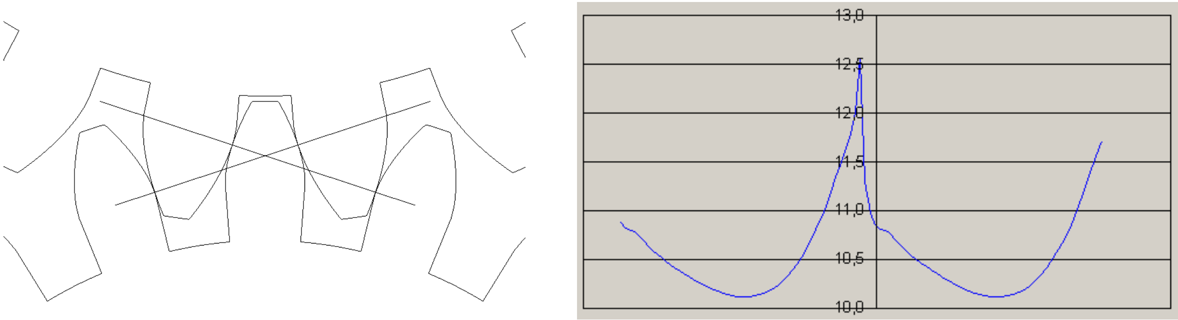

Figure 6 : Example of an optimization run starting design. Gears cross-section (left) and static transmission error (STE) in µm (right). $x_p = -0.2$, $x_r = 0.2$, $ha_{0p} = 1.5$, $ha_{0r} = 1.4$, $hf_{0p} = 1.3$, $hf_{0r} = 1.17$, $\rho_{ap} = 0.2$, $\rho_{ar} = 0.1$, $\rho_{fp} = 0.05$, $\rho_{fr} = 0.1$, $\beta_0 = 10$. $r_c = 2$, $g_s = 592$, $\sigma_{vm} = 1556$, $F = 11683$, $\Delta STE = 2.45$. Although there are no teeth interferences, this design is infeasible w.r.t. the maximum slip ratio, the maximum von Mises stress and the contact pressure. All dimensions but the STE are in *mm*, stresses are in *MPa*. The STE is plotted for one meshing period.

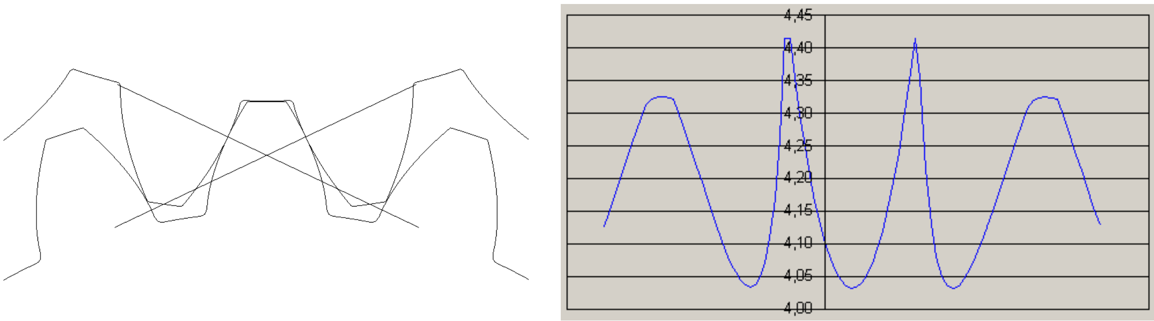

Figure 7 : Optimum design without wear. Gears cross-section (left) and STE in µm (right). $x_p = 0.49$, $x_r = 0.49$, $ha_{0p} = 1.20$, $ha_{0r} = 1.11$, $hf_{0p} = 1.00$, $hf_{0r} = 1.09$, $\rho_{ap} = 0.0003$, $\rho_{ar} = 0.19$, $\rho_{fp} = 0.64$, $\rho_{fr} = 0.23$, $\beta_0 = 30$. $r_c = 1.66$, $g_s = 1.47$, $\sigma_{vm} = 443$, $F = 938$, $\Delta STE = 0.38$. The design is feasible. All dimensions but the STE are in *mm*, stresses are in *MPa*. The STE is plotted for one meshing period.

B. Optimization with wear based on a kriging metamodel

Designing gears while accounting for probabilistic wear has been formulated in Equations *(6)* and *(7)* by replacing the variation in static transmission error, ΔSTE, by its 90th percentile, $P^{90}_{\Delta STE}$. Section III has explained how $P^{90}_{\Delta STE}(x)$ can be estimated by Monte Carlo simulations at a reasonable number (*n* = 2175) of design points *x*'s and then approximated by an ordinary kriging metamodel. The kriging metamodel is defined by its mean and variance at each *x*, $m_{OK}(\mathbf{x})$ and $s^2_{OK}(\mathbf{x})$, respectively. A basic idea of our approach to robust optimization with kriging is, first, to make the costly Monte Carlo simulations before the optimization iterations and estimate the percentiles at the sampled points, then learn them with a kriging metamodel which, finally, replaces the percentile estimation during the optimization. Although kriging provides $m_{OK}(\mathbf{x})$ and $s^2_{OK}(\mathbf{x})$, i.e., a complete Gaussian probability density function at each point x, we only use here the kriging mean, $m_{OK}(\mathbf{x})$, for two reasons. Firstly, contrarily to global optimization methods based on kriging such as EGO[17], we do not iteratively update the kriging model. We are therefore not interested in its uncertainty $s^2_{OK}(\mathbf{x})$. Secondly, the data points gathered in $\hat{\mathbf{P}}^{90}$ and learned by the kriging model are noisy since estimated from 30 Monte Carlo simulations. The kriging mean with nugget effect, $m_{OK}(\mathbf{x})$, acts as a filter of the observations noise.

All other optimization constraints, r_c, g_s, σ_{vm} and F, are calculated by a single call to the gears simulator because they are mainly insensitive to wear. To sum up, the robust optimization problem with kriging solved is,

$$\begin{gathered}\min_{x}\ \mathrm{m_{OK}}(\mathrm{x}) \\ \text{by changing } \mathrm{x} = \left[x_p, x_r,\ ha_{0p},\ ha_{0r},\ hf_{0p},\ hf_{0r}\right] \\ \text{such that there is no teeth interference and} \\ r_c > 1.25, \quad g_s < 2, \quad \sigma_{vm} < \mathrm{Re}, \quad F < P_{hertz}^{\max} \\ -1 \le x_p, x_r \le 2, \quad 0.8 \le ha_{0p}, ha_{0r},\ hf_{0p}, hf_{0r} \le 1.7.\end{gathered} \tag{22}$$

Like in the deterministic optimization problem, the constraints were handled by minimizing the penalized objective function *(5)* where ΔEST was replaced by $\mathrm{m_{OK}}$. The problem was solved using the CMA-ES optimization algorithm with the same settings as in the deterministic optimization of paragraph A, including the search length, restart and variables bounds handling. However, there are only six unknowns in the robust problem, the other variables being fixed at their deterministic optimal value, $\rho_{ap} = 0.0003, \rho_{ar} = 0.19, \rho_{fp} = 0.64, \rho_{fr} = 0.23, \beta_0 = 30$.

The best design found when solving *(22)* is described in Figure 8. We will henceforth refer to it as the "kriging design".

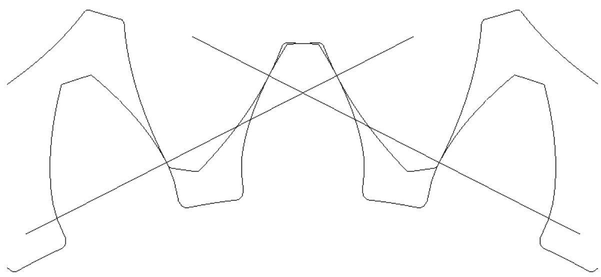

Figure 8 : Optimum robust design found using the kriging metamodel instead of the empirical estimator of the ΔSTE 90% percentile. $x_p = 0.20,\ x_r = 1.26,\ ha_{0p} = 1.44,\ ha_{0r} = 1.49,\ hf_{0p} = 1.27,\ hf_{0r} = 0.92$. $m_{ok} = 1.97$ (but $\hat{P}_{90} = 3.37$ based on 500 simulations), $r_c = 1.71$, $g_s = 2.00$, $\sigma_{vm} = 558$, $F = 1189$. The design is feasible. All dimensions but the STE are in *mm*, stresses are in *MPa*. The STE is plotted for one meshing period.

C. Optimization with wear based on a deterministic noise representative

An alternative formulation of the gears design problem with wear is to solve replace the percentile estimation by a single, deterministic, instance of wear. Here, this instance is simply chosen as the average of the wear profiles, i.e., a non-perturbed Archard's wear profile since the perturbation are centered on it. The design problem solved with Archard's wear is,

$$\begin{gathered}\min_{x}\ \Delta\mathrm{STE_A}(\mathrm{x}) \\ \text{by changing } \mathrm{x} = \left[x_p, x_r,\ ha_{0p},\ ha_{0r},\ hf_{0p},\ hf_{0r}\right]\end{gathered} \tag{23}$$

such that there is no teeth interference and

$$r_c > 1.25, \quad g_s < 2, \quad \sigma_{vm} < \mathrm{Re}, \quad F < P_{hertz}^{\max}$$
$$-1 \le x_p, x_r \le 2, \quad 0.8 \le ha_{0p}, ha_{0r}, hf_{0p}, hf_{0r} \le 1.7.$$

Note that the numerical cost of evaluating the objective function ΔSTE_A involves two calls to the gears simulator : a first call to evaluate the contact forces and sliding rates, and a second call to remove material from the teeth surfaces and recalculate all gears performance criteria. This numerical cost is twice that of the deterministic formulation but it is much lower than a complete Monte Carlo simulation. Problem *(23)* is solved with the CMA-ES optimizer is the same fashion as all other optimization problems discussed in this article. The optimum design for this problem, called "Archard's design", is described in Figure 9 .

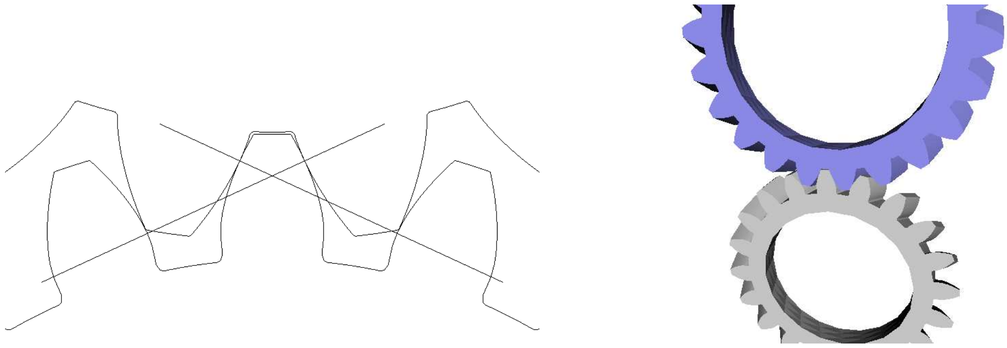

Figure 9 : Optimum design with Archard's wear. (2D cross-section, left, and 3D drawing, right). $x_p = 0.028, x_r = 0.86, \quad ha_{0p} = 1.26, \quad ha_{0r} = 1.32, \quad hf_{0p} = 1.20, hf_{0r} = 0.81$.

$\Delta STE_A = 1.49$ (but $\hat{P}_{90} = 2.64$ based on 500 simulations), $r_c = 1.62$, $g_s = 2.00, \sigma_{vm} = 631, F = 1115$.

D. Comparison of approaches

The four methods that have been seen up to now for designing gears are now compared :

1. Neglecting wear and solving Equations *(2)* and *(3)*. The solution to this deterministic problem was shown in Figure 7. The numerical cost of the method is one call to the gears simulator per optimization analysis (objective function and constraints).
2. Replacing Monte Carlo simulations with a kriging metamodel, as stated in Equation *(22)*. The obtained optimum design was described in Figure 8. The numerical cost of the procedure is one call to the gears simulator plus one call to the metamodel (negligible) per optimization analysis. In addition, there is an initial cost for calculating the design of experiments ($k \times n$ simulations) and inverting the $n \times n$ kriging covariance matrix $\mathbf{K}_\Delta$.
3. Optimizing the static transmission error for a deterministic nominal wear profile (Archard's profile). This formulation is stated in Equation *(23)* and the resulting design described in Figure 9. Each optimization analysis costs two calls to the gears simulator.
4. Designing the gears by selecting the best point of the initial maximin latin hypercube design of experiments (DoE). There are n points where k Monte Carlo simulations (MCS) are performed for a total cost of $k \times n$ simulations. Note that this DoE was used to build the kriging metamodel. The so-called "empirical MCS design" is described in Figure 10.

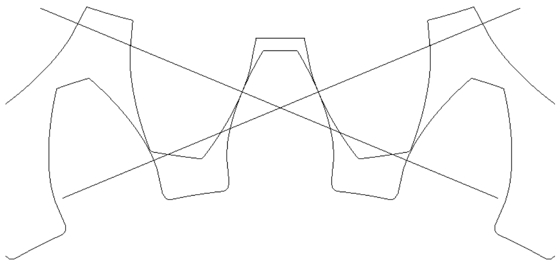

Figure 10 : Empirical MCS design, i.e., best design of the original latin hypercube set of points with 30 Monte Carlo simulations per point. $x_p = -0.004$, $x_r = 0.37$, $ha_{0p} = 1.37$, $ha_{0r} = 1.51$, $hf_{0p} = 1.27$, $hf_{0r} = 0.87$ **.** $\hat{P}_{90} = 2.69$ (but $\hat{P}_{90} = 3.29$ with 500 simulations), $r_c = 1.72$, $g_s = 1.96$, $\sigma_{vm} = 614$, $F = 1228$ **.**

Table 5 summarizes the designs associated to each approach. The last column (P^{90}) is the targeted design criterion, estimated with 500 Monte Carlo analyses. It is clear from Table 5 that neglecting wear from the ΔSTE calculation (row 1) leads to designs with poor ΔSTE performance in the presence of wear. Accounting for wear (compare rows 2 to 4 to row 1) decreases the pinion addendum modification coefficient x_p and increases its counterpart x_r . It also increases the teeth heights (ha's and hf's). Minimizing the kriging wear prediction is an improvement over neglecting wear. This method yielded a design that was not present in the initial DoE, but its P^{90} performance is similar to that of the best design in the DoE. This is because the initial 6-dimensional design space is quite large for fitting a metamodel with about 2000 data points. The kriging P^{90} prediction at the kriging optimum design is of poor quality (1.97 versus 3.37 µm). This is another piece of evidence that optimizing with a metamodel requires an iterative enrichment of the DoE around the optimum design area (other examples can be found in, e.g., Ref. 17). The overall best design was the Archard's design. For long optimization runs, it is also the most expensive method out of the ones considered here since each evaluation of the objective function needs two gears simulations. Nevertheless, for the 3*12000 analyses long optimization runs performed here, this extra-cost remains inferior to that of building the kriging database (3000*30 analyses). In robust optimization problems where *i)* the number of variables is of the order of 10 or less and *ii)* one knows a noise sample leading to a reliable design, the representative noise sample approach is to be preferred.

Table 5 : comparison of optimum gears for various formulations : without wear, with Archard's wear, with kriging percentile approximation, and with Monte Carlo simulations (best point of LHS). All these designs are feasible.

Design	**Numerical cost (in simulation call)**	x_p	x_r	ha_{0p}	ha_{0r}	hf_{0p}	hf_{0r}	**ΔSTE used criterion**	P^{90} *
1) without wear	1 per analysis	0.49	0.49	1.20	1.11	1.00	1.09	0.38	6.38
2) Archard's wear	2 per analysis	-0.028	0.86	1.26	1.32	1.20	0.81	1.49	2.64
3) kriging	1 per analysis + initial DoE	0.20	1.26	1.44	1.49	1.27	0.92	1.97	3.37
4) Monte Carlo (30 MCS / LHS point)	Initial DoE	-0.004	0.37	1.37	1.51	1.27	0.87	2.69	3.29

** Computed using 500 MCS*

V. Conclusions and perspectives

This article represents a first step in the optimization of gears while accounting for random teeth wear through the static transmission error (STE). Two robust optimization approaches have been proposed where the statistical estimation of the performance (here a 90% percentile of the STE variation) is replaced either by a kriging metamodel or by fixing the noise to an adequate value (here the average wear profile). This study confirms that the kriging approach is feasible because we observed that it leads to reasonable designs. However, it is seen that the kriging metamodel needs to be updated to allow a convergence to optimal designs. This is the methodological perspective of this article. The optimization with the wear profile fixed at its average value leads to the overall best design and is the currently advised method for solving the stated gears design problem. Regarding gears design, the results presented here should be completed by varying the number of teeth and considering teeth profile corrections.

Acknowledgments

This work was founded in part by the CETIM foundation through the "Optimisation multidisciplinaire d'engrenages" (gears multidisciplinary optimization) project.The authors also thank Emmanuel Rigaud and Joel Perret-Liaudet for their fruitful comments.

References

[1] J.R. Colbourne, *The geometry of involute gears*, Springer Verlag, New York, 1987.

[2] A. L. Kapelevich and R. E. Kleiss, *Direct gear design for spur and helical involute gears*, Gear Technology, Sept-Oct 2002, pp.29-35.

[3] X. Li, G.R. Symmons and G. Cockerham, *Optimal design of involute profile helical gears*, Mech. Mach. Theory, vol. 31, No. 6, Elsevier, 1996, pp. 717-728.

[4] C.-H. Li, H.-S. Chiou, C. Hung, Y.-Y. Chang and C.-C. Yen, *Integration of finite element analysis and optimum design on gear systems*, Finite elements in Analysis and Design, Elsevier, vol. 38, 2002, pp. 179-192.

[5] F. L. Litvin, I. Gonzales-Perez, A. Fuentes, K. Hayasaka and K. Yukishima, *Topology of modified surfaces of involute helical gears with line contact developed for improvement of bearing contact, reduction of transmission errors, and stress analysis*, Mathematical and Computer Modelling, vol. 42, 2005, pp. 1063-1078.

[6] A. Kahraman, P. Bajpai and N. E. Anderson, *Influence of tooth profile deviations on helical gear wear*, transactions of the ASME, vol. 127, July 2005, pp. 656-663.

[7] N. Driot and J. Perret-Liaudet, *Variability of modal behaviour in terms of critical speeds of a gear pair due to manufacturing errors and shaft misalignments*, Journal of Sound and Vibration, vol. 292, n$^{o.}$ 3-5, 2006, pp. 824-843.

[8] A. Meyer, *Filengrene : a software for gears analysis and manufacturing*, distributed by TEO+ comp., Rouen, France, 2008.

[9] A. Auger and N. Hansen, *A Restart CMA Evolution Strategy With Increasing Population Size*, in IEEE Congress on Evolutionary Computation, CEC 2005, Proceedings, 2005, pp. 1769-1776.

[10] S. Missoum, P. Hernandez and Zafer Gürdal, *Optimization of a gearbox using mathematical programming and genetic algorithms*, in Proceedings of the 3rd World Congress of Structural and Multidisciplinary Optimization, Buffalo, New-York, USA, May 1999, pp. 487-490.

[11] K. Deb and S. Jain, *Multi-speed gearbox design using multi-objective evolutionary algorithms*, KanGAL technical report No. 2002001, Indian Institute of Technology, Kanpur, India, 2002.

[12] Matheron, G., Le Krigeage Universel, Cahiers du Centre de Morphologie Mathematique, No. 1 Fontainebleau, France.

[13] Cressie N., Statistics for spatial data, Wiley Series in Probability and Mathematical Statistics, 1993

[14] Le Riche, R. and Guyon, F., *Dual Evolutionary Optimization*, Artificial Evolution, Lecture Notes in Computer Science, No. 2310, selected papers of the 5th International Conference, Evolution Artificielle, Collet, P., Lutton, E., Schoenauer, M., Fonlupt, C. et Hao, J.-K., eds, Springer Verlag publ., 2002, pp. 281-294.

[15] Kenney, J. F. and Keeping, E. S., Mathematics of Statistics, 2nd ed., Princeton, NJ: Van Nostrand, 1951.

[16] Ginsbourger D., Picheny V., Roustant O., Richet Y., "*A new look at Kriging for the Approximation of Noisy Simulators with Tunable Fidelity*", 8th European Network for Business and Industrial Statistics conference (ENBIS8), Cd-Rom, Athens, September 2008.

[17] D.R. Jones, M. Schonlau, and W. Welch., Effcient *global optimization of expensive black-box functions*, Journal of Global Optimization, vol.13, 1998, pp. 455–492.