Setting-up kriging-based adaptive sampling in metrology

D. Romano[1], R. Ascione[2]
[1]*University of Cagliari, Italy*
[2]*ENEA, Italy*

daniele.romano@dimcm.unica.it

Abstract

Statistical sampling is a fundamental tool in science, and metrology is no exception. The merit of a sample is its efficiency, i.e. a good trade-off between the information collected and the sample size. Although the sample sites are ordinarily decided prior to the measurements, a different option would be to select them one at a time. This strategy is potentially more informative as the next site can be decided based also on the measurements taken up to that time. The core of the method is to drive the next-site selection by a non-parametric model known as *kriging*, namely a stationary Gaussian stochastic process with a given autocorrelation structure [1,2]. The main feature of this model is the ability to promptly reconfigure itself, changing the pattern of the predictions and their uncertainty each time a new measurement comes in. Since the model is re-estimated after each added point the sampling procedure is an adaptive one. The next sampling site can be selected via a number of model-based criteria, inspired by the principles of reducing prediction uncertainty or optimizing an objective function, or a combination of the two.

The methodology has been applied by the authors [3,4] to design inspection plans for measuring geometric errors using touch-probe Coordinate Measuring Machines (CMM). Results showed that both the non adaptive statistical plans (Random, Latin Hypercube sampling, uniform sampling) and two adaptive deterministic plans from the literature were largely outperformed by the proposed plans both in terms of accuracy and cost.

Here we further investigate on a number of important questions related to adaptive kriging: which is the best trade-off between the number of adaptive and non-adaptive points (the latter chosen according to uniform coverage), which next-site selection criteria are more suitable to capturing extreme values of the signal in order to provide a good estimate of the geometric errors.

1 Kriging modeling

Response $y(x)$ is considered a realization of the Gaussian Process (GP) defined as

$$Y(x) = \mu + Z(x), \tag{1}$$

where μ is a constant and $Z(x)$ is a GP with zero mean and stationary covariance:

$$\mathrm{E}(Y(x)) = \mu, \qquad Cov(Y(x), Y(x+h)) = \sigma_Z^2 R(h;\theta). \tag{2}$$

In (2) σ_Z^2 is the process variance, R the correlation function depending only on the displacement vector h between any pair of points in the domain and on a parameters set θ. The model defined by (1) and (2) is known as simple kriging. A flexible choice for the GP correlation structure is the power exponential function:

$$R(h;\theta) = \prod_{i=1}^{d} \exp\left\{-\theta_i^{-1}|h_i|^{p_i}\right\}, \quad 0 < \theta_i < \infty, \ 0 < p_i \le 2, \ d = 1,2 \tag{3}$$

where θ=($\theta_1,\ldots, \theta_d, p_1,\ldots,p_d$), is a vector of unknown scale parameters ($\theta_1,\ldots, \theta_d$) and smoothing parameters ($p_1,\ldots,p_d$) respectively. Parameter θ_i describes how rapidly correlation decays in direction i with increasing distance $|h_i|$. Parameter p_i describes the shape of the correlation decay (see Figure 1).

2 Criteria for next-point selection

We use three kinds of criteria: objective-specific, informative, and a combination of the two. The objective-specific criterion (MaxF) point to maximize an objective function, e.g. the signal itself or the geometric error. The two informative criteria select next-point to inspect where the uncertainty of predictions by the current kriging model is maximum (MaxPVar), and where prediction uncertainty weighted by the distance from the nearest point already inspected is maximum, thus promoting uniform coverage (MaxWPVar). Finally, composite criteria (switch rules) are defined which select the next point as the one producing the maximum increase of the objective function wrt the previous step; if no increase is possible the rule switches to one of the two informative criteria (switch rule 1 (SR1): MaxF or MaxPVar; switch rule 2 (SR2): MaxF or MaxWPVar). Prediction uncertainty is evaluated empirically by using the Jackknife variance operator as it proved to convey much more information than the so-called kriging variance. The latter, in

fact, holds only when the parameters of the correlation function are known, which is hardly the case in practice.

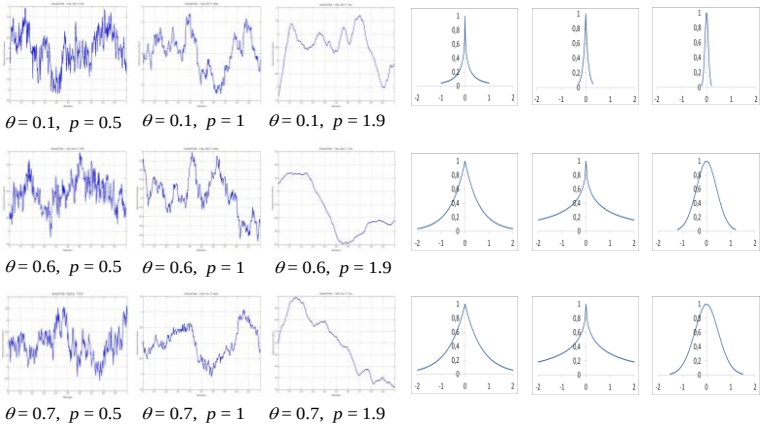

Figure 1: Nine one-dimensional signals (left) synthesized by as many GP models via their power exponential correlation function (right)

3 Scope of the analysis

We look for the most effective settings of the adaptive procedure in terms of the size of the initial non-adaptive sample, chosen as a Latin Hypercube one, and the criterion for next-point selection, in view of maximizing the objective function, i.e. the measured error after 40 inspected points (err_40). For this purpose we set up a planned experiment whose factors (levels) are: range and shape parameters θ (0.1, 0.6, 0.7) and p (0.5, 1.0, 1.9), the size of the initial LHS sample, n (4 to 38, step 2), and the criteria for next-point selection (five criteria, see section 2). Nine one-dimensional signals (Figure 1), spanning a sizable interval of information content, are generated by a random walk from each GP model obtained by crossing the three levels of parameters θ and p. Ten replicates of the adaptive procedure are run.

4 Results

Main effects and two-factor interactions for the response err_40 are shown in Figure 2. The measured error after 40 inspected points is generally accurate, ranging from 90% to 100% of the true error. The shape parameter, p, of the correlation function

has a significant effect on the procedure's ability to getting close to the true error while the range parameter θ is much less active. Low values of p, corresponding to very noisy signals, make the error's estimate less accurate. The most interesting result is that a 50%-50% allocation of the initial LHS points and the successive adaptive points proves to be a good and robust (small variation of the response for n in the interval from 8 to 34) choice for all the signals. The mean performance of the five adaptive criteria seems similar and always superior to Random sampling and LHS.

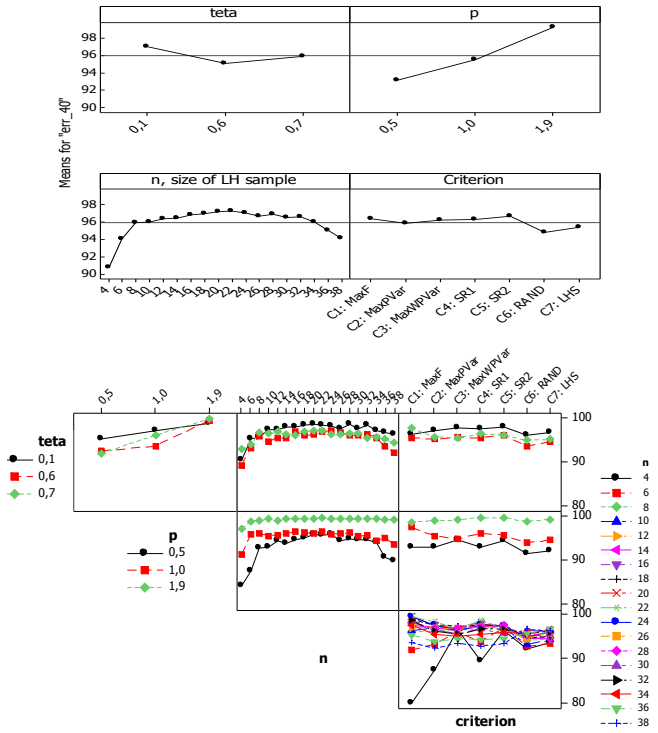

Figure 2: Plots of main effects (top) and two-factor interactions (bottom) for "err_40"

Valuable information is conveyed by the interaction effects. Noisy signals are better captured by allocating more points to the initial LHS plan before starting the adaptive procedure, while a few LHS points are enough for smooth signals; however, starting with four points only, (the minimum number for allowing the

estimation of kriging parameters μ, θ, p and σ_Z^2), generally lowers the performance especially when the criteria devoted to maximize the objective function (MaxF, SR1) are adopted, see the *n-Criterion* interaction in Figure 2 and Figure 3 (left). The same interaction also shows that the MaxF criterion is the most effective in estimating the true error with *n* ranging in a quite wide interval, say from 20 to 34, see Figure 3 (right).

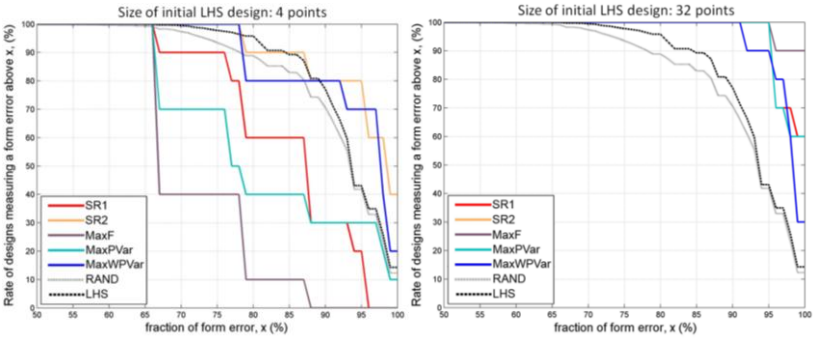

Figure 3: geometric error (%) estimated by all criteria for signal 1, starting with 4 (left) and 32 (right) initial LHS points.

References:

[1] Krige DG. A statistical approach to some basic mine valuation problems on the Witwatersrand. Journal of the Chemical, Metallurgical and Mining Society of South Africa, 1951;52(6):119–39.

[2] Cressie NAC. Statistics for spatial data. 1st Edition New York: Wiley Interscience; 1993.

[3] Pedone P., Vicario G., Romano D. Kriging-based sequential inspection plans for coordinate measuring machines. Applied Stochastic Models in Business and Industry 2009;25(2):133–49.

[4] Ascione R., Moroni G., Petrò S., Romano D. Adaptive inspection in coordinate metrology based on kriging models, Precision Engineering, 2013;37(2013): 44–60.