Experimental designs and Kriging modelling: the use of strong orthogonal arrays

Nedka D. Nikiforova [1], Rossella Berni [1], Luciano Cantone [2]

[1]Department of Statistics Computer Science Applications "G. Parenti", University of Florence, Italy
[2]Department of Engineering for Enterprise "Mario Lucertini", University of Rome "Tor Vergata", Rome, Italy

ENBIS 2021 Spring Meeting
May 17-18 2021

Outline

1. Introduction
2. Main theory
    - LH designs based on strong orthogonal arrays
    - Outlined theory on Kriging
3. Emprical example
4. Conclusions

Introduction

Design and Analysis of Computer Experiments:

$\rightarrow$ physical experimentation for some complex engineering and technological problems appears often too costly or, in some cases, also impossible to be performed;
$\rightarrow$ a computer code or simulator is run to depict the physical system under study;
$\rightarrow$ the complexity of the simulator requires an approximation of it through a surrogate model that represent a valid approximation of the computer code, acting as a statistical interpolator of the simulated input-output data
$\rightarrow$ **KRIGING** (Krige, 1951; Sacks et al., 1989);
$\rightarrow$ a key-point for computer experiments: the planning of the experimental design;
$\rightarrow$ space-filling designs.

Cont'd

Space-filling designs:

→ spread the design points as uniformly as possible in order to observe the response in the entire experimental region;
→ several types of space-filling designs have been developed in the literature by also considering optimal designs (Pronzato and Müller, 2012);
→ Latin Hypercube (LH) designs, introduced by McKay et al. (1979), is one of the most commonly used class of space-filling designs;
→ a LH design achieves the maximum uniformity when projected in any one dimension. However, in practice, the design involves a large number of input variables, and thus the uniformity should be achieved in more than one dimension.

A few types of LH designs:

→ orthogonal LH designs (Ye, 1998; Bingham et al., 2009);
→ maximin LH designs (Johnson et al., 1990; Joseph et al., 2015);
→ LH designs based on orthogonal arrays (Tang, 1993);
→ LH designs based on strong orthogonal arrays (He and Tang, 2013).

Our proposal

A compelling approach for the design and analysis of computer experiments:

→ a suitable LH design for the computer experiment through a new class of orthogonal arrays, called strong orthogonal arrays (He and Tang, 2013);
→ main advantages of the proposed experimental design:

- the achievement of very good space-filling properties;
- a relatively low number of experimental runs.

→ analysis of the computer experiment through suitable Kriging models with anisotropic covariance functions;
→ an empirical example based on a real case-study which further demonstrate the validity of our proposal.

Main theory

Orthogonal Array (OA):

An OA of *strength* t is an $n \times d$ matrix where the j-column has s_j levels $(1, 2, ..., s_j;$ $j = 1, ..., d)$, and it is such that for any $n \times t$ submatrix, each possible level combination occurs with the same frequency (Tang, 1993; Hedayat et al., 1999). If $s_1 = ... = s_j = ... = s_d = s$, then the OA is symmetric and it is denoted by:

$$\mathrm{OA}(n, d, s, t) \tag{1}$$

→ Latin Hypercube design based on OA (OA-based-LH design) (Tang, 1993).

Strong Orthogonal Array (SOA):

A SOA of *strength* t is an $n \times d$ matrix with entries $(1, ..., s^t)$, such that any subarray of g columns, for any g with $1 \leq g \leq t$, can be collapsed into an OA $(n, g, s^{u_1} \times ... \times s^{u_g}, g)$ for any positive integer $u_1, ..., u_g$ with $u_1 + ... + u_g = t$ (He and Tang, 2013). We denote such an array as follows:

$$\mathrm{SOA}(n, d, s^t, t) \tag{2}$$

→ Latin Hypercube design based on SOA (SOA-based-LH design) (He and Tang, 2013).

Cont'd

Building a SOA-based-LH design:

Main steps (He and Tang, 2013):

1. define an OA according to the strength t; for example, if $t = 2$ then the OA should be an OA (n, d, s, t), while if $t = 3$ then the OA should be an OA $(n, d + 1, s, t)$;
2. build a Generalized Orthogonal Array (GOA) (n, d, s, t) from the corresponding OA defined at Step no.1;
3. from the GOA obtained in the previous step, construct the SOA with n rows, d columns, s^t levels and strength t. Once the SOA is obtained, the SOA-based-LH design could be generated. To this end, let λ be the index of the SOA defined by $\lambda = \frac{n}{s^t}$.
4. for each column of the SOA, replace the λ entries for the level c $(c = 1, ..., s^t)$ by any permutation of $(c-1)\lambda + 1, (c-1)\lambda + 2, ..., c\lambda$; lastly, the SOA-based-LH design is generated in the design space $[0, 1)^d$ through the usual methods (Lin and Tang, 2015).

Kriging: outlined theory

Universal Kriging:

Universal Kriging considers a non-constant trend function as follows (Krige, 1951; Sacks et al., 1989):

$$y_{\boldsymbol{x}} = \mu(\boldsymbol{x}) + Z(\boldsymbol{x}) \tag{3}$$

$$\mu(\boldsymbol{x}) = \boldsymbol{f}'(\boldsymbol{x})\beta \tag{4}$$

where $\boldsymbol{f}'(\boldsymbol{x}) = (f_1'(\boldsymbol{x}), ..., f_m'(\boldsymbol{x}))$ is a set of trend functions defined for each new point $\boldsymbol{x}$, $\boldsymbol{\beta}$ is the column vector $[m \times 1]$ of unknown parameters, and $Z(\boldsymbol{x})$ identifies a spatial stochastic process. By considering n selected points, $\boldsymbol{F}$ is defined as the model regression matrix of dimension $[n \times m]$ formed by the n independent functions $f(\boldsymbol{x})$: $\boldsymbol{F} = (f(\boldsymbol{x}_1), ..., f(\boldsymbol{x}_n))$.

$$E(Z(\boldsymbol{x})) = 0$$

$$Z(\boldsymbol{x}) = Cov(Z(\boldsymbol{x}), Z(\boldsymbol{x} + \boldsymbol{h})) = \sigma_y^2 R(\boldsymbol{h}; \boldsymbol{\omega}) \tag{5}$$

where σ_y^2 is the process variance, e.g. the variance of Y, and $\boldsymbol{\omega}$ is the vector of parameters defining the stationary stochastic process for the correlation function R

The Power Exponential covariance function and nugget issues:

The Power exponential covariance function and nugget issues:

- the Power exponential covariance function (Rasmussen and Williams, 2006):

$$R(h;\boldsymbol{\omega}) = \exp\Big(-\sum_{j=1}^{d} \frac{\mid h_j \mid^{p_j}}{\phi_j}\Big); j = 1, ..., d \qquad (6)$$

  In formula (6) $\boldsymbol{\omega} = (\phi, \boldsymbol{p})$ where ϕ is the vector of characteristic length scale parameters, while $\boldsymbol{p}$ is the vector related to the parameters of smoothness, with $0 < p_j \leq 2$;

- the inclusion of a nugget coefficient allows to avoid instabilities during the estimation (e.g. computational problems and jumps), and also account for potentially possible deviations of inaccurate assumptions on the stationarity of the process and on the chosen correlation function (Stein, 1999; Gramacy and Lee, 2012).

Empirical example based on a real case-study

Data and technical details:

- the empirical example is performed by using the data in Nikiforova et al. (2021);
- improve the payload distribution of freight trains in terms of in-train compression and tensile forces. High compression forces can lead to train derailment, while high in-train tensile forces may cause "train disruption";
- Nikiforova et al. (2021) considered a freight train unloaded in five different sections, where within each train section the payload distribution can be: uniform, triangular or trapezoidal;

Input variables:

- five inputs related to the shape of the payload distribution (one for each train section), e.g. $h = \{h_1, h_2, h_3, h_4, h_5\}$ (Arcidiacono et al., 2017; Nikiforova et al., 2021);
- five inputs related to the position of maximum load (one for each train section), e.g. $x = \{x_1, x_2, x_3, x_4, x_5\}$ (Arcidiacono et al., 2017; Nikiforova et al., 2021);

Cont'd

The SOA-based-LH design:

- starting from a symmetric OA $(64, 11, 2, 3)$, a GOA $(64, 10, 2, 3)$ is built, and the final SOA $(64, 10, 2^3, 3)$ is obtained (Nikiforova et al., 2021);
- the SOA-based-LH design space-filling properties: a stratification on a $2 \times 2 \times 2$ grid in any three-dimensional projection, and, in addition, on the finer grids of $2^2 \times 2$ and 2×2^2 in any two-dimensional projection.

Output variables:

Three in-train forces: i) compression forces computed at 2m, ii) compression forces computed at 10m, and iii) tensile forces computed at 2m. The true values of compression and tensile forces are calculated through the TrainDy software (Cantone, 2011), internationally certified for the computation of in-train forces of freight trains.

Kriging modelling:

Trend function, Power exponential covariance function and nugget issues:

- for each output variable, a Kriging model is estimated through the **R** package *DiceKriging* (Roustant et al., 2012);
- trend function: first order polynomial trend, plus the quadratic effects related to the position of maximum load, and the interaction terms between the position of maximum load and the shape of the payload distribution, strictly related to the same train section;
- the Kriging models are estimated by considering the anisotropic Power exponential covariance function, formula (6);
- a nugget term δ is also estimated;
- the validation of the Kriging models is performed through the leave-one-out cross validation method.

Kriging modelling results

Three goodness-of-fit measures for evaluating the estimated Kriging models:

-$Q2$: the $Q2$ predictivity coefficient;
-SE-LOO: the standard error of the leave-one-out residuals;
-RMSE-LOO: the root mean square error of the leave-one-out residuals.

Table 1: Goodness-of-fit measures for each estimated Kriging model

	2m Compression forces	10m Compression forces	2m Tensile forces
$Q2$	0.8561	0.8426	0.9226
SE-LOO	0.8017	0.7636	0.7642
RMSE-LOO	0.0197	0.0179	0.0009

Cont'd

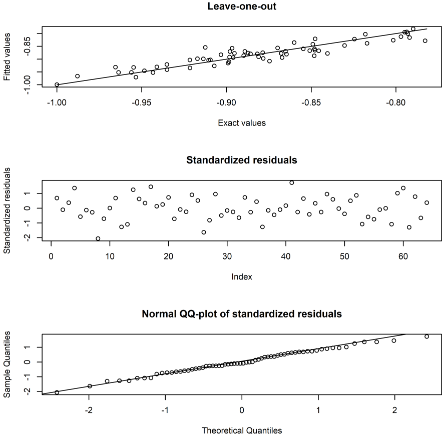

Figure 1: Compression Forces at 2m: Goodness of fit with leave-one-out method. The three plots presented are as follows: the residuals (top), the standardized variance of residuals (middle) and the Normal QQ plot of the residuals (bottom).

Cont'd

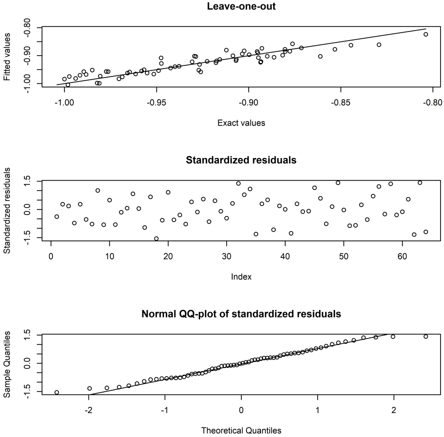

Figure 2: Compression Forces at 10m: Goodness of fit with leave-one-out method. The three plots presented are as follows: the residuals (top), the standardized variance of residuals (middle) and the Normal QQ plot of the residuals (bottom).

Cont'd

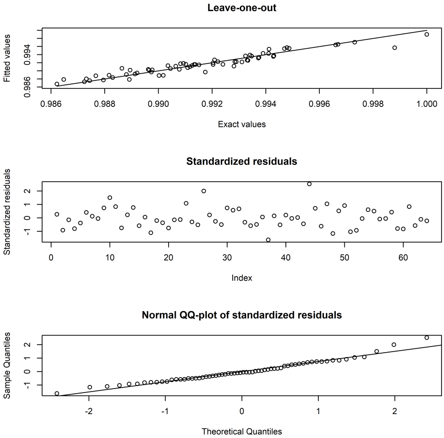

Figure 3: Tensile Forces at 2m: Goodness of fit with leave-one-out method. The three plots presented are as follows: the residuals (top), the standardized variance of residuals (middle) and the Normal QQ plot of the residuals (bottom).

Conclusions

- In this talk, we have presented a compelling approach for the design of computer experiments based on SOAs;
- the obtained SOA-based-LH design allows to achieve very good space-filling properties with a relatively low number of experimental runs and small computational efforts;
- the analysis of the computer experiments, performed through suitable Kriging models with anisotropic covariance functions, provides very satisfactory results confirming that the suggested approach could be successfully applied for solving similar technological problems;
- the choice of the covariance function is a key-point which should be further investigated in details.

THANK YOU FOR YOUR ATTENTION!

[1] Arcidiacono, G., Berni, R., Cantone, L. and Placidoli, P. (2017). Kriging models for payload-distribution optimization of freight trains. *International Journal of Production Research*, 55: 4878–4890.

[2] Bingham D., Sitter R.R. and Tang B. (2009). Orthogonal and nearly orthogonal designs for computer experiments. *Biometrika*, 96(1): 51-65.

[3] Cantone, L. (2011). TrainDy: the new Union Internationale des Chemins de Fer software for freight train interoperability. *Journal of Rail and Rapid Transit*, 225(1): 890–899.

[4] Gramacy R. and Lee H. (2012). Cases for the nugget in modelling computer experiments. *Statistics and Computing*, 22(3): 713–722.

[5] He Y. and Tang B. (2013). Strong orthogonal arrays and associated Latin hypercubes for computer experiments. *Biometrika*, 100(1): 254–260.

[6] Hedayat A.S., Sloane N.J.A. and Stufken J. (1999). Orthogonal Arrays: Theory And Applications. New York, NY: Springer.

[7] Johnson M.E., Moore L.M. and Ylvisaker D. (1990). Minimax and maximin distance designs. *Journal of Statistical Planning and Inference*, 26(2): 131-148.

[8] Joseph V.R., Gul E. and Ba S. (2015). Maximum projection designs for computer experiments. *Biometrika*, 102(2): 371–380.

[9] Krige D.G. (1951). A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, 52(9): 119–139.

[10] Lin C.D and Tang B. (2015). Latin hypercubes and space-filling designs. In Handbook of Design and Analysis of Experiments, Dean A, Morris M, Stufken J, Bingham D (eds). Chapman & Hall/ CRC Press: Boca Raton, USA, pp.: 593-625.

[11] McKay M., Beckman R.J. and Conover W.J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2): 239–245.

[12] Nikiforova, N.D., Berni, R., Arcidiacono, G., Cantone, L. and Placidoli, P. (2021). Latin Hypercube Designs based on Strong Orthogonal Arrays and Kriging Modelling to Improve the Payload Distribution of Trains. *Journal of Applied Statistics*, 48(3): 498-516, DOI: 10.1080/02664763.2020.1733943.

[13] Pronzato L. and Müller W. (2012). Design of computer experiments: space filling and beyond. *Statistics and Computing*, 22(3): 681-701.

[14] Rasmussen, C.E. and C.K.I. Williams, C.K.I. (2006). *Gaussian Processes for Machine Learning*, The MIT Press, Boston.

[15] Roustant O., Ginsbourger D. and Deville Y. (2012). DiceKriging, DiceOptim: two R packages for the analysis of computer experiments by Kriging-based metamodeling and optimization. *Journal of Statistical Software*, 51(1): 1–55.

[16] Sacks J., Welch W.J., Mitchell T.J. and Wynn H.P. (1989). Design and analysis of computer experiments. *Statistical Science*, 4(4): 409–423.

[17] Stein M.L. (1999). Interpolation of Spatial Data. New York, NY: Springer.

[18] Tang B. (1993). Orthogonal array-based Latin hypercubes. *Journal of the American Statistical Association*, 88(12): 1392–1397.

[19] Ye K.Q. (1998). Orthogonal column Latin hypercubes and their application in computer experiments. *Journal of the American Statistical Association*, 93(12): 1430-1439.

Appendix

Generalized Orthogonal Array - GOA (He and Tang, 2013):

Let us consider a matrix $\mathbf{B}$ of dimension $n \times dt$. The matrix $\mathbf{B}$ is such that the dt columns are arranged in d groups, i.e. $\mathbf{B} = \mathbf{B_1}, ..., \mathbf{B_d}$; each group $\mathbf{B_j}$, $j = 1, ..., d$, is in turn composed of t columns, i.e. $\mathbf{B_j} = \mathbf{b_{j1}}, ..., \mathbf{b_{jt}}$. The matrix $\mathbf{B}$ is called a GOA of size n, with d constraints, s levels and strength t if the sub-matrix $\mathbf{B}^*$ consisting of t columns $\mathbf{b_{i'j'}}$, where $i' = i'_1, ..., i'_g$ and $j' = 1, ..., u_{i'}$, is an OA with strength t for any $1 \leq g \leq t$, any $1 \leq i'_1 < ... < i'_g \leq d$, and any positive integer $u_1, ..., u_g$ with $u_1 + ... + u_g = t$. Such an array is denoted by:

$$\text{GOA}(n, d, s, t) \tag{7}$$