Department for Applied Statistics
Johannes Kepler University Linz

IFAS Research Paper Series 2011-53

Beyond space-filling: an illustrative case

Werner G. Müller, Luc Pronzato[a] and Helmut Waldl

March 2011

[a]Université de Nice-Sophia Antipolis/CNRS

Abstract

When collecting spatial data, it has become a standard practice to position the measurement points spread out uniformly across the available space. These so-called space-filling designs are now ubiquitous in corresponding publications and conferences. The statistical folklore is that such designs have superior properties when it comes to prediction and estimation of response functions. In this presentation we want to review the circumstances under which this superiority holds, provide some new arguments and clarify the motives to go beyond space-filling. We will accompany these findings with a simple two-dimensional example with seven observations.

1 Introduction

The predominance of space-filling designs for collecting spatial data $Y(x)$ with coordinates $x \in \mathcal{X}$ is, from a model-based viewpoint, based on the notion that such designs are good for maintaining the maximal prediction (kriging) variance

$$\max_{x\in\mathcal{X}} \mathrm{Var}[\hat{Y}(x)]$$

at small values. This observation turns up frequently in the applied literature (*e.g.*, (1)), theoretically it was shown most rigorously (2) for a certain class of space-filling designs and increasingly weaker correlations (see also (3) for a motivation for the limit-kriging approach). For the relationship among the various types of space-filling designs see (4).

The model underlying these considerations and our investigations is the correlated random field, given by

$$Y(x) = \eta(x, \beta) + \varepsilon(x) \ . \tag{1}$$

Here, β is an unknown vector of parameters in R^p, $\eta(\cdot,\cdot)$ a known function and the random term $\varepsilon(x)$ has zero mean, (unknown) variance σ^2 and a parameterized spatial error correlation structure such that $\mathbb{E}[\varepsilon(x)\varepsilon(x')] = \sigma^2 c(x, x'; \rho)$. It is often assumed that the deterministic term has a linear structure, *i.e.*, $\eta(x,\beta) = f^\top(x)\beta$, and that the random field $\varepsilon(x)$ is Gaussian, allowing estimation of β, σ and ρ by Maximum Likelihood.

Note that setup (1) is used in such diverse areas of spatial data analysis (cf. (5)) as mining, hydrogeology, natural resource monitoring and environmental sciences, and has become the standard modeling paradigm in computer simulation experiments.

2 Beyond space-filling

It is conventional practice that all unknown parameters are estimated from the same data set, but clearly the classic kriging variance does not reflect the additional uncertainty resulting from the estimation of the covariance parameters. A first-order expansion of the kriging variance for $\hat{\rho}^n$ around its true value is used in (6) (see also (7) for more precise developments), which yields an explicit correction term to augment the (normalized) kriging variance. This naturally leads to a design criterion,

$$\max_{x \in \mathcal{X}} \left\{ \mathrm{Var}[\hat{Y}_t(x)] + \mathrm{tr}\left\{ M_\rho^{-1} \mathrm{Var}[\partial \hat{Y}_t(x)/\partial \rho] \right\} \right\}, \tag{2}$$

to be minimized, which is called EK(empirical kriging)-optimality in (8) (see also (9) for a similar criterion). Here M_ρ stands for the information matrix for the covariance parameters ρ and the objective is to take the dual effect of the design into account: obtaining accurate predictions at unsampled sites and improving the accuracy of the estimation of the covariance parameters (those two objectives being conflicting, see (10)) through the formulation of a single criterion.

It is quite evident now that space-filling designs are not very efficient with respect to the EK-criterion, since they lack short distances that are required for capturing the local correlation. Let us demonstrate this on our reference example. We will assume the Ornstein-Uhlenbeck process on $\mathcal{X} = [0,1]^2$, which is a special case of (1) with $\eta(x,\beta) = \beta$, *i.e.* $f(x) \equiv 1$, and $c(x,x';\rho) = \rho^{|x-x'|}$, setting $\sigma^2 \equiv 1$ to avoid identifiability problems (see, *e.g.*, (11)). For a similar one-dimensional case, some results have been reported in (12). Furthermore, we will here report only results for $n = 7$ observations and $\rho = \exp\{-7\} \sim 0.001$, although similar, albeit perhaps more trivial results were achieved for other choices of ρ. For growing n there is expectedly a tendency towards approaching space-filling, which however is counterbalanced by decreasing ρ. This relation requires more detailed investigations in the future, but the below given necessarily limited cases seem to encapture the general behaviour well.

Fig. 1 displays a scatter-plot of the EK-criterion values against the minimal distances among all 7 points for 1000 uniformly randomly generated designs on $[0,1]^2$, the latter criterion to be maximized being commonly referred to as maximin distance, see (2). From this display it is evident that designs that are good in the maximin (space-filling) sense (right part of the plot) do not tend to achieve acceptably low levels for the corrected kriging variance (2), the EK-criterion.

The EK-criterion as a function of the design points is very rough and far from being convex. It is thus very unpromising if we try to find the optimal design on the entire unit square and we confine the search to a uniform grid of 21×21 points where the 7 design points may be located.

The iterative search started with a random design, in each iteration one of the design points was exchanged following the Federov-exchange-algorithm combined with simulated annealing. That is, we exchanged one point of the actual design with 100 randomly selected non-design-points on the grid. For each of these 100 new designs we computed the EK-criterion. If the minimal EK-criterion of these 100 new designs was less than the EK-criterion of the actual design, the corresponding new design was accepted and the temperature-parameter *temp* of the simulated annealing process decreased. If the best new design was worse than the actual design it was accepted with probability $P = \exp(\frac{EK_{old} - EK_{new}}{temp})$ where EK_{new} and EK_{old} are the EK-criterion values of the best new and the actual design respectively. After 20 successive iterations without improvement of the EK-criterion the search was stopped with the best design found. The EK-optimal design found on our grid is shown in Fig. 2, the EK-value of this design is 1.1821.

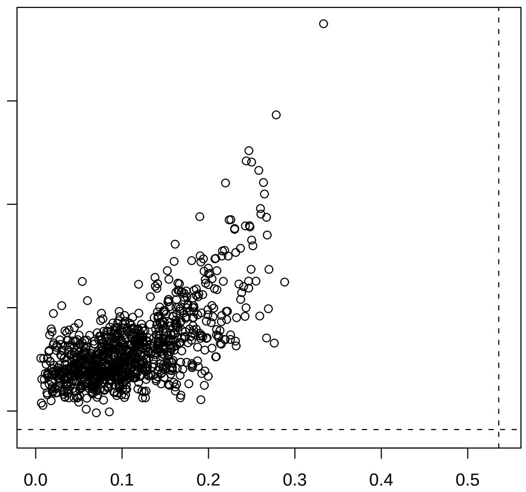

Figure 1: Scatterplot of minimal distances (horizontal) versus EK-criterion values (vertical) for 1000 uniformly randomly generated designs.

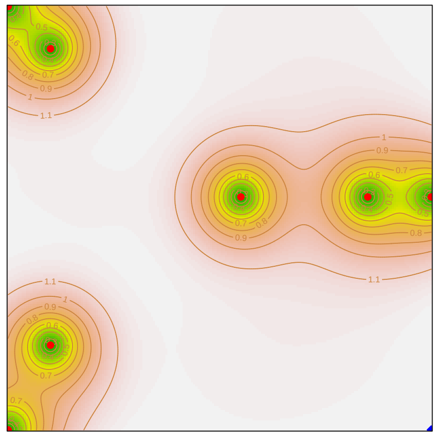

Figure 2: Level plot of the EK-criterion for the EK-optimal design (red dots) found on a grid of 21×21 points.

3 A substitute criterion

To avoid space-filling, we could thus now proceed by finding designs that optimize the EK-criterion. However, the EK-criterion is computationally complex, since each test of candidate design requires to evaluate (2), which in turn requires the evaluation of the target function for all points in the candidate set. It is evident that this is unfeasible for higher dimensions and that it would thus be useful to have an alternative criterion that can substitute EK and similarly reflects the total prediction uncertainty.

In the design of experiments literature, a connection between prediction-oriented and estimation-oriented criteria is well known and runs under the heading "equivalence theory". It goes back to the celebrated paper by Kiefer and Wolfowitz (13) who established the equivalence of optimal designs between two criteria of optimality for the regression with independent error case, one related to parameter estimation, the other related to prediction (*i.e.*, equivalence between D- and G-optimality). Since the EK-criterion is analogous to G-optimality for the correlated error case, this has motivated (14) to suggest to maximize a compound criterion with weighing factor α,

$$|M_\beta(\xi, \rho)|^\alpha \cdot |M_\rho(\xi, \rho)|^{(1-\alpha)}, \tag{3}$$

which consists of determinants of information matrices (D-optimality) corresponding to trend and covariance parameters respectively. For the detailed definition of these information matrices see, *e.g.*, (14) or (15) for computationally efficient implementations. We shall call criterion (3) CD_α-optimality (compound D-optimality) in the following.

Let us see how CD_α-optimality relates to EK-optimality by looking at the same randomly generated 1000 designs from above. Two scatterplots for $\alpha = 0.7$ and $\alpha = 0.8$ are displayed in Fig. 3 and 4 respectively. It is quite

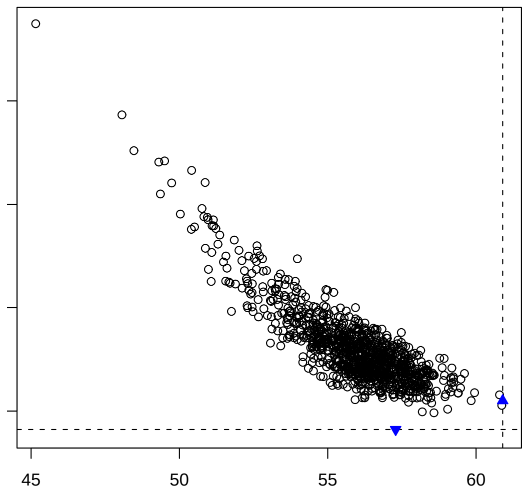

Figure 3: Scatterplot of $\mathrm{CD}_{0.7}$-criterion values (horizontal) versus EK-criterion values (vertical) for 1000 uniformly randomly generated designs.

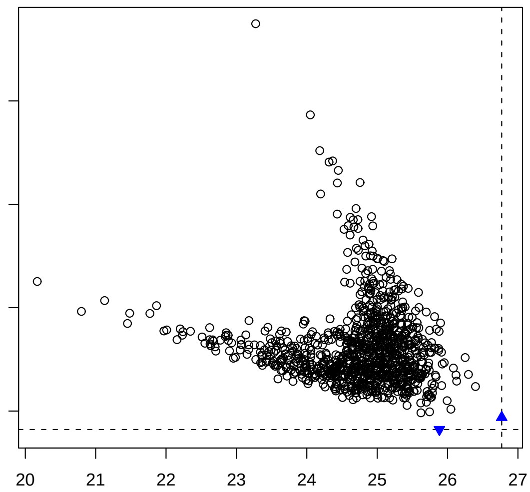

Figure 4: Scatterplot of $\mathrm{CD}_{0.8}$-criterion values (horizontal) versus EK-criterion values (vertical) for 1000 uniformly randomly generated designs.

evident from these pictures that the two criteria are in general good accordance (although the scatter branches out for higher values of α). However, of relevant importance is only the behaviour in the lower right corner, *i.e.* in the region of the desired extremes of the criteria. The achievable extremes are indicated by the dashed lines and the triangles indicate the values for the designs found optimal under the respectively other criterion. It turns out that especially the $\mathrm{CD}_{0.8}$-optimal design approaches the mininum EK-value of 1.182 quite closely by a value of 1.194.

The accordance of the two criteria breaks down if too small (below 0.5) or too high (above 0.85) values are chosen for α. It is clear that an efficient way of choosing an appropriate α is therefore needed but for this example it sufficed to try several values to come up with a reasonable choice. For proper α both criteria are seeking to find a compromise between space-filling behavior (*i.e.*, estimation the trend parameter in (1) and minimization of the kriging variance component in (2)) and clustering (*i.e.*, estimation of the covariance parameter ρ and minimization of the correcting term component in (2)). This can be observed by looking at Fig. 5 and 6 which display the EK-criterion contour lines for $\alpha = 0.7$ and $\alpha = 0.8$ CD_α-optimal designs respectively. The position x where the maximum is reached in the EK-criterion (2) is indicated by a diamond.

We may also learn something from looking at the worst designs. It is (unfortunately) much easier to maximize the EK-criterion (2) than to minimize it. The maximization of the EK-criterion yields a design as displayed in 7 with a criterion value of 2.90. This design strongly resembles the maximin design for seven points (cf. www.packomania.com), the EK-criterion contours are given in 8 with a maximum value of 2.68, just slightly better than the worst design! This again demonstrates the need to go beyond space-filling for

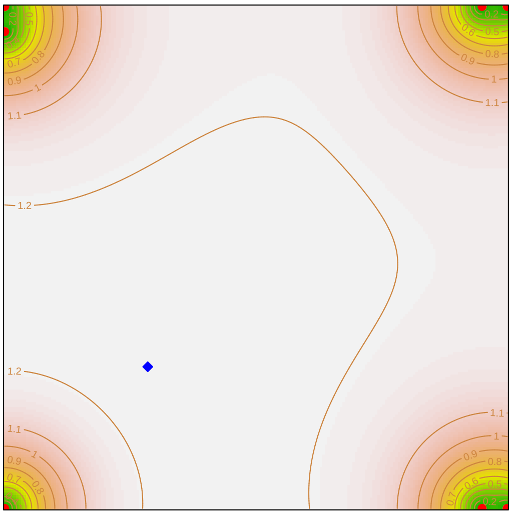

Figure 5: Level plot of the EK-criterion for the $CD_{0.7}$-optimal design (red dots) found on a grid of 21×21 points.

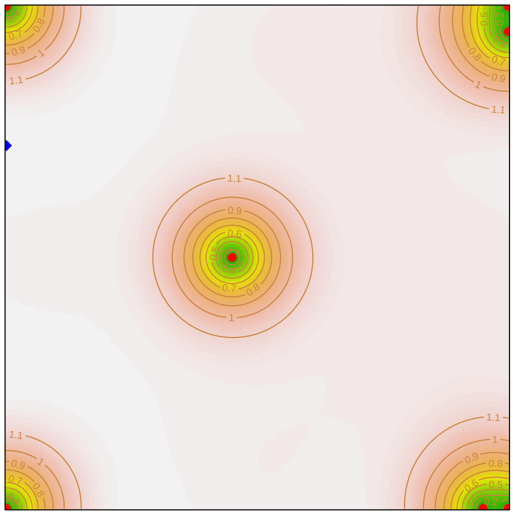

Figure 6: Level plot of the EK-criterion for the $CD_{0.8}$-optimal design (red dots) found on a grid of 21×21 points.

these type of problems.

4 Outlook

From these examples and similar one-dimensional considerations in (12) we are thus led to believe that good EK-efficiencies can be produced for CD_α-optimal designs in more complex setups and/or higher dimensions as well. This would be very advantageous since EK-optimal designs are much more difficult to generate than CD_α-optimal designs, since they require embedded optimizations over the candidate sets. Our quasi equivalence will allow to replace the very demanding optimization (2) by the much less intensive (3) without much loss in efficiency.

Acknowledgement

This work was partially supported by a PHC Amadeus/OEAD Amadée grant FR11/2010.

References

[1] van Groenigen J (2000) The influence of variogram parameters on optimal sampling schemes for mapping by kriging. Geoderma 97(3-4):223–236.

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

[3] Joseph V (2006) Limit kriging. Technometrics 48(4):458–466.

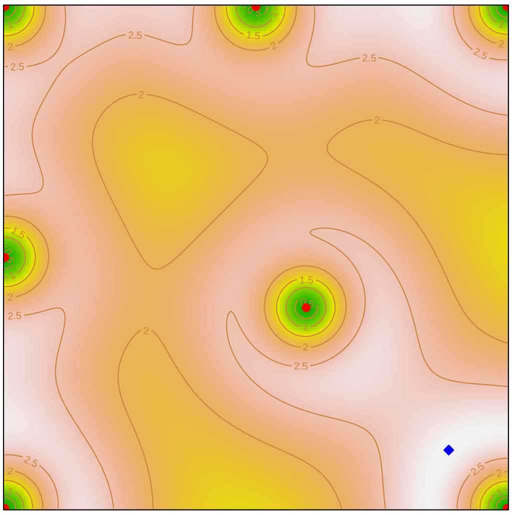

Figure 7: Level plot of the EK-criterion for the worst EK-optimal design (red dots) found on a grid of 21×21 points.

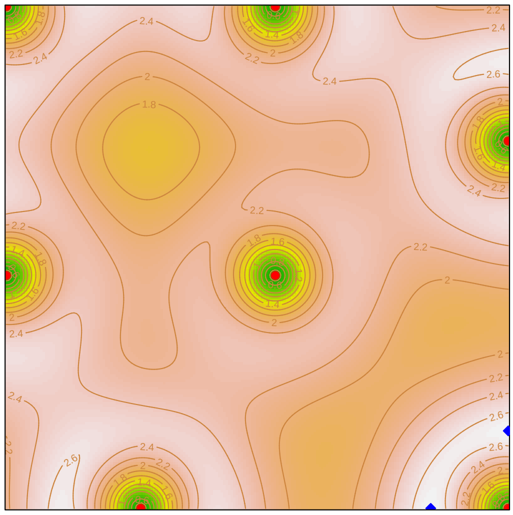

Figure 8: Level plot of the EK-criterion for the maximin design (red dots).

[4] Pronzato L, Müller W (2011) Design of computer experiments: space filling and beyond. In revision for Statistics and Computing.

[5] Cressie, N. A. C., 1993. Statistics for Spatial Data (Wiley Series in Probability and Statistics), rev. sub Edition. Wiley-Interscience.

[6] Harville, D. A., Jeske, D. R., 1992. Mean squared error of estimation or prediction under a general linear model. Journal of the American Statistical Association 87 (419), 724–731.

[7] Abt, M., 1999. Estimating the prediction mean squared error in gaussian stochastic processes with exponential correlation structure. Scandinavian Journal of Statistics 26 (4), 563–578.

[8] Zimmerman, D. L., 2006. Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. Environmetrics 17 (6), 635–652.

[9] Zhu, Z., Stein, M. L., 2006. Spatial sampling design for prediction with estimated parameters. Journal of Agricultural, Biological, and Environmental Statistics 11 (1), 24–44.

[10] Pronzato, L., 2008. Optimal experimental design and some related control problems. Automatica 44 (2), 303–325.

[11] Ying, Z., 1991. Asymptotic properties of a maximum likelihood estimator with data from a Gaussian process. Journal of Multivariate Analysis 36, 280–296.

[12] Müller, W. G., Pronzato, L., Waldl, H., 2011. Relations between designs for prediction and estimation in random fields: an illustrative case. In Porcu, Montero and Schlather (Eds.) Advances And Challenges In Space-time Modelling Of Natural Events, Lecture Notes in Statistics, Springer Verlag, Berlin, to appear .

[13] Kiefer, J., Wolfowitz, J., 1960. The equivalence of two extremum prob-

lems. Canadian Journal of Mathematics 12, 363–366.

[14] Müller, W. G., Stehlík, M., 2010. Compound optimal spatial designs. Environmetrics 21 (3-4), 354–364.

[15] Smirnov, O. A., 2005. Computation of the Information Matrix for Models with Spatial Interaction on a Lattice. Journal of Computational and Graphical Statistics, 14 (4), 910–927.