Department for Applied Statistics
Johannes Kepler University Linz

IFAS Research Paper Series
2008-37

Compound optimal spatial designs

Werner G. Müller and Milan Stehlík

October 2008

Abstract

A single purpose design may be quite inefficient for handling a real-life problem. Therefore, we often need to incorporate more than one design criterion and a common approach is simply to construct a weighted average, which may depend upon different information matrices. Designs based upon this method have been termed compound designs. The need to satisfy more than one design criterion is particularly relevant in the context of random fields. It is evident that for precise universal kriging it is important not only to efficiently estimate the spatial trend parameters, but also the parameters of the variogram or covariance function. Both tasks could for instance be comprised by applying corresponding design criteria and constructing a compound design from there. Modern techniques for such first and second order characteristics will be suggested and reviewed in the presentation. A new hybrid stochastic exchange type optimization algorithm is proposed and an illustrating example of the design of a water-quality monitoring network is provided.
Key words: Optimum design; Monitoring network; Efficiency; Equidistant design; Parameterized covariance functions.

1 Introduction and Setup

Probably the main reason why optimum design principles are not frequently used in spatial data analysis is that the observations are correlated. Consequently, the corresponding optimal design questions must cope with the existence and detection of an error correlation structure, problems largely unaccounted for by traditional optimal design theory. In all of these situations there arise a number of issues, which require special techniques - for a recent discussion see Müller W.G. and Stehlík M. (2008). A statistical model we consider in the paper is the so called random field, given by

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

with design points (coordinates of monitoring sites) $\xi = \{x_1 \neq x_2 \neq \cdots \neq x_n\}$ forming a so-called replication-free design taken on a compact design space $\mathcal{X}$. The parameters β are unknown and the variance-covariance structure of the errors depends on parameters θ. In the paper we consider the following cases:

Case 1 We are interested only in the trend parameters β and consider θ as known or a nuisance.

Case 2 We are interested only in the covariance parameters θ.

Case 3 We are interested in both sets of parameters.

Furthermore, in the article we consider for the sake of simplicity the linear(ized) model $\eta(x, \beta) = f^T(x)\beta$. The paper can be seen as an extension of Müller W.G. (2005), which is concentrated on case 1. The example from this paper, the design of the water-quality network in the Südliche Tullnerfeld in Austria, will be employed here as well.

We assume that the errors $\varepsilon(x)$ are correlated and the correlation between two measurements depends on the distance $d = ||x - x'||$ between pairs of particular design points x and x' (isotropic fields). Perhaps the most widely

used parametric model that satisfies the above conditions is the so-called spherical covariance function

$$\mathrm{cov}(x, x'; \theta) = \begin{cases} \theta_1 + \theta_2, & d = 0 \\ \theta_2 \left\{ 1 - \frac{3}{2}(\frac{d}{\theta_3}) + \frac{1}{2}(\frac{d}{\theta_3})^3 \right\}, & 0 < d \leq \theta_3 \\ 0, & d > \theta_3 \end{cases} .$$

This function decreases monotonically from a 'nugget effect of θ_1 near the origin to a 'sill value' of $\theta_1 + \theta_2$, which is attained at the 'range' θ_3.

Recently, the so called Matérn model family is more and more frequently employed for its flexibility due to a smoothness parameter θ_2. It is given by

$$\mathrm{cov}(x, x + d) = \sigma^2 \cdot \left\{ \frac{1}{2^{\theta_2 - 1}\Gamma(\theta_2)} \left(\frac{d}{\theta_1} \right)^{\theta_2} K_{\theta_2} \left(\frac{d}{\theta_1} \right) \right\}, \tag{2}$$

where K_{θ_2} denotes a modified Bessel function of order θ_2. It encompasses a number of widely used models, e.g. the so called exponential (by $\theta_2 = 1/2$)

$$\mathrm{cov}(x, x + d) = \sigma^2 e^{-\frac{d}{\theta_1}}, \tag{3}$$

and a recent review of its history and properties can be found in Guttorp, P. and Gneiting, T. (2006) . If we further assume that the errors are Gaussian, we obtain a special case of Ornstein-Uhlenbeck process when the covariance is exponential (3).

The main purpose of statistical analysis is estimation of parameters β, θ and/or prediction of Y based upon the parametric models. Having estimated the spatial trend by a linear response a universal kriging estimator is typically used (see Müller W.G. , 2007). Universal kriging can be viewed as a two stage procedure involving GLS estimation of the trend surface and best linear prediction, as pointed out by Fedorov V.V. (1989). For a reverse two-stage interpretation see Cressie N.A.C. (1993).

The related optimum design question, which will be the core problem of the present paper is then how to select the inputs such, that we gain the maximum available information from the experiment. In the following we will define optimality of a design always strictly in the tradition of Kiefer (see e.g. Kiefer, J. (1959)), where the inputs are selected such, that a prespecified design criterion (e.g. the determinant of the trend parameters variance-covariance matrix, so-called D-optimality) is optimized. The classic Fisher information $M(\beta) = E\left[(\frac{\partial \ln f(x,\beta)}{\partial \beta})^2 \right]$, which is the basis for the D-optimality assumes differentiability with respect to the parameter (see e.g. Rao CR. 1965). This opens a problem of interpretation of the formally defined (e.g. by one-sided limits) information matrix for such values (see Stehlík M. (2004)). Still, the classic Fisher information can be well defined over some open set. Thus the D-optimality criterion has the form $\Phi(M) = \det(M)$, for details in this context see Stehlík M. and Müller W.G. (2008). Note that there exists a well developed theory for standard i.i.d. regression based on Kiefers (1959) concept of design measures, cf. Atkinson, A.C., Donev, A.N., and Tobias, R. (1992, 2007) for a recent textbook.

2 General Approaches for Multipurpose Designs

At many points of the previous section it became clear that a single purpose design may be quite inefficient for handling a real-life problem. Therefore we often need to incorporate more than one design criterion and a common approach is simply to construct a weighted average

$$\bar{\Phi}[\xi|\alpha] = \alpha\Phi[M(\xi)] + (1-\alpha)\Phi'[M'(\xi)], \tag{4}$$

(cf. Läuter, E. (1976)), which may depend upon different information matrices M and M'. The weighting parameter $0 \leq \alpha \leq 1$ has to be selected by the user and it is not very clear (due to the generally different scaling of $\Phi[\cdot]$ and $\Phi'[\cdot]$), which choice of α corresponds to an intended weighting. Related ideas for combining the purposes of parameter estimation and model discrimination already appear in Fedorov, V.V. (1972). Designs based upon (4) and its straightforward generalization to more than two terms have been termed compound designs; they also prove useful for the situation of multiple (e.g. pollution or seismic) sources influencing the response (see e.g. Steinberg, D.M., Rabinowitz, N., Shimshoni, Y., and Mizrachi, D. , 1995).

Another method to satisfy multiple design criteria is the one of constrained optimum designs (for a survey, see Cook, D. and Fedorov, V. , 1995). Instead of (4) one could for instance choose to find

$$\xi^* = \arg\max_{\xi\in\Xi} \Phi[M(\xi)] \quad \text{s.t. } \Phi'[M'(\xi)] > \kappa(\alpha),$$

i.e. an optimum design for criterion $\Phi[\cdot]$ which ensures sufficiently good estimation with respect to the second criterion $\Phi'[\cdot]$ (or vice versa). A disadvantage of constrained optimum designs is the asymmetric involvement of the considered goals. The relationship between constrained and compound designs is explored in detail in Cook, R.D. and Wong, W.K. (1994). Two different ways of combining criteria have been suggested recently by Mcgree et al. (2008), but contrary to their claim, also there a specific weighing is implicated. It is sometimes argued for "standardizing" the criteria Φ and Φ' (see e.g. Biedermann et al. (2007) or also Mcgree et al. (2008)) by employing

$$\tilde{\Phi}[\xi|\alpha] = \alpha\Phi[M(\xi)]/\Phi[M(\xi^*)] + (1-\alpha)\Phi'[M'(\xi)]/\Phi'[M'(\xi^*)]$$

instead of (4), which is not a major issue here as will be seen below.

2.1 $D-$optimal designs for estimating trend and covariance parameters

The need to satisfy more than one design criterion is particularly relevant in the context of collecting spatial data. Recalling the discussion in Section 1 it is evident that e.g. for precise universal kriging it is important not only to efficiently estimate the spatial trend parameters β, but also the parameters θ of the variogram or covariance function. Both tasks could for instance be comprised by applying corresponding design criteria $\Phi[M(\cdot,\beta)]$ and $\Phi'[M'(\cdot,\theta)]$ and constructing a compound design from $\bar{\Phi}[\cdot]$. Techniques, such as (4) and

its alternatives, for combining designs for first and second order characteristics have been reviewed in Müller W.G. (2007). For an indication of how many different criteria could be combined to a sensible overall criterion see a recent application by Park, S.Y., Choi, J.H., Wang, S., and Park, S.S. (2006).

Simultaneous optimization according to (4) can still be an unfeasible task, even with today's computer technology. A straightforward alternative is to firstly construct an optimum design for trend estimation with $n_0 < n$ sites by applying classic algorithms and then to use an augmentation rule for the $n - n_0$ remaining sites to cover variogram fitting. What then seems of decisive importance is a good choice of n_0, the number of sites to be allocated to the trend estimation design. If kriging is the ultimate aim of the analysis it might be helpful to determine which relative allocation (to variogram estimation and to prediction itself) gives the best overall performance of the kriging procedure (a first attempt to assess this performance by simulation and its application to monitoring network design can be found in Haas, T.C. , 1992). Zhu, Z. and Stein, M. (2006) e.g. find that only 3 to 10% of the observations need to be assigned to estimation rather than prediction.

Another similar approach to multipurpose design was recently devised by Lesch, S.M. (2005), who suggests to combine space-filling and information based procedures in the spirit of constrained optimum designs. He provides simulation results for a variety of settings motivated by a salinity survey application.

However, one can, if one is willing to make distributional assumptions, employ ML-estimators. For the full parameter set $\{\beta, \theta\}$ the information matrix then exhibits the block diagonal form

$$E \left\{ \begin{array}{cc} -\frac{\partial lnL(\beta,\theta)}{\partial\beta\partial\beta^T} & -\frac{\partial lnL(\beta,\theta)}{\partial\beta\partial\theta^T} \\ -\frac{\partial lnL(\beta,\theta)}{\partial\theta\partial\beta^T} & -\frac{\partial lnL(\beta,\theta)}{\partial\theta\partial\theta^T} \end{array} \right\} = \left(\begin{array}{cc} M(\xi;\theta,\beta) & 0 \\ 0 & M'(\xi;\theta) \end{array} \right).$$

Here

$$M(\xi) = \frac{1}{n} \sum_{x \in \xi} \sum_{x' \in \xi} f(x) [C^{-1}(\xi)]_{x,x'} f^T(x'), \tag{5}$$

and

$$\{M'(x,\xi)\}_{ii'} = \{C^{-1}(\xi)\}_{x,\cdot} \frac{\partial C(\mathcal{X})}{\partial \theta_i} C^{-1}(\xi) \frac{\partial C(\mathcal{X})}{\partial \theta_{i'}} \{C^{-1}(\xi)\}_{\cdot,x}.$$

We also use now the notation $[C(\xi)]_{ii'} = c(x_i, x_{i'}; \theta)$ to emphasize the dependence of C on the design and we assume knowledge of θ.

Again all the entries can be condensed to a design criterion Φ. If we now want to determine a D-optimal design for the whole parameter set, we can obviously - due to the orthogonality of the first and second order parameters - simply use the product of the respective determinants as an optimum design criterion. More generally a weighted version would yield

$$\bar{\Phi}'[\xi|\alpha] = |M(\xi)|^{\alpha} \cdot |M'(\xi)|^{(1-\alpha)}. \tag{6}$$

It is obvious that (6) conforms with (4) after logarithmic transformation. Also unless we are not simultaneously optimizing for α, standardizing the criteria will not effect the resulting ξ^*.

Most importantly (6) encompasses the three interesting cases identified in Section 1, namely $\alpha = 1$ corresponds to Case 1, $\alpha = 0$ corresponds to Case 2

and any other α to a specific instance of Case 3. Thereby we are allowed to tackle all cases within a common framework.

Note that Xia G., Miranda M.L. and Gelfand A.E. (2006) suggest to use the traces rather than the determinants, but eventually continue with considering only the first block. Also note that for such designs the impact of the amount of correlation on the design criterion can be quite paradoxical. Müller W.G. and Stehlík M. (2008) illustrate this point by showing that depending on the amount of local correlation (governed by θ), the criterion values increase and decrease respectively.

2.2 Design algorithms (with a new proposal)

For solving the respective optimization problem the techniques described in Sections 3.1 and 3.2 of Müller W.G. (2005) can be adapted. It is therefore natural to suggest in a particular situation to construct designs by adding points

$$x^+ = \arg \max_{x \in \mathcal{X} \setminus \xi} \phi(x, \xi)$$

maximizing and deleting points

$$x^- = \arg \min_{x \in \xi} \phi(x, \xi)$$

minimizing the so-called sensitivity function, the directional derivative of $\bar{\Phi}$. Starting with an arbitrary ξ and doing this iteratively leads to a standard exchange type algorithm as proposed by Fedorov, V.V. (1972), which eventually leaves us with an improved design corresponding to the prescribed finite number of observations (exact design). These algorithms can be quite demanding, especially in the spatial case with large designs and candidate sets (see Royle (2002) for a comprehensive review), thus in this paper we have used the following

hybrid stochastic version:

- Make the best exchange between a point from ξ and a randomly (uniformly weighted) chosen point $x* \in \mathcal{X} \setminus \xi$.
- If there is no improvement multiply weights by the distances $\|x_i - x*\|$, normalize weights and draw anew. Repeat this step if necessary.
- The above could be complemented by using a stochastic acceptance operator (decreasing temperature, similar to what was e.g. proposed in Zhou (2008)) to improve performance.

Alternatively Pázman, A. and Laca, J. (2008) employ the techniques based upon approximate information matrices. Specifically, they suggest to replace their diagonal entries by $\epsilon \left(\xi(x) - \kappa\right)^2 / \xi(x)$ and supplement the directional derivative $\phi(x, \xi) = g(x, \xi)/\xi$ by

$$g'(\xi, x) = -\mathrm{tr}\{\nabla' \Phi' \cdot M'(x, \xi)\},$$

where $\nabla' \Phi'$ is the respective submatrix of the gradient. This allows them to eventually propose a much simplified exchange algorithm with an allocation rule (compare with (9) in Müller W.G. (2005))

$$x^{(s)} = \arg \max_{x \in \bar{\mathcal{X}}_s} \epsilon [g(x, \xi_{(s)}) + g'(x, \xi_{(s)})] \left[1 - \left(\frac{\kappa}{\xi(x)}\right)^2\right] [1 - \xi(x)], \qquad (7)$$

and some heuristics on how to operate the tuning parameter ϵ. They still do not, however, report an extensive set of simulations or examples, that allow a judgement about the usefulness of this procedure.

A very different 'direct' approach can again be taken if the complete probabilistic structure of the random field model is known. An information theoretic approach based upon the Shannon entropy, which was firstly formalized by Caselton, W.F. and Zidek, J.V. (1984), is then possible. Their basic idea was that the uncertainty about some aspect (say parameter estimation, prediction, etc.) must be somehow reduced by the data and that a design that produces such data should minimize the overall entropy

$$\mathcal{E}[y] = E[-\log p(y) \mid \tilde{p}(y)].$$

Here, y is the collection of potential measurements and $p(\cdot)$ its probability density function, $\tilde{p}(\cdot)$ represents the state of complete ignorance and is usually chosen as an improper density $\tilde{p}(\cdot) \propto 1$. A comprehensive overview over this technique and some discussion contrasting it to those presented so far is given in Section 11 of Le, N.D. and Zidek, J.V. (2006).

2.3 A digression into prediction

Rather than estimating underlying parameters it is frequently the aim to predict the random field at given unsampled sites or over a continuous region as precise as possible. Usually then minimizing the maximum kriging variance at these sites or over a region $\mathcal{X}$, interpreted as the unconditional mean squared prediction error for the best linear unbiased predictor, is undertaken, i.e.

$$\min_{\xi} \max_{x \in \mathcal{X}} E[(\hat{y}(x|\xi) - y(x))^2]. \tag{8}$$

If the covariance parameters are estimated from the same dataset, the resulting additional uncertainty needs to enter the criterion. This uncertainty is frequently approximated by the trace $\operatorname{tr}\left\{M_\theta^{-1}\operatorname{Var}[\partial\hat{y}(x_0)/\partial\theta]\right\}$, cf. Harville and Jeske (1992) and Zimmerman and Cressie (1992). Consequently, Zimmerman (2006) regards

$$\min_{\xi} \max_{x \in \mathcal{X}} \left\{\operatorname{Var}[\hat{y}(x)] + \operatorname{tr}\left\{M_\theta^{-1}\operatorname{Var}[\partial\hat{y}(x)/\partial\theta]\right\}\right\} \tag{9}$$

as the design problem, which he terms EK-(empirical kriging-)optimality. These designs are much more demanding to achieve than parameter estimation designs, since they require embedded optimizations over the candidate sets.

Note that in the D-optimal uncorrelated case we have $\phi(x, \xi) = \operatorname{Var}[\hat{y}(x)]$ and thus by the celebrated equivalence theorem by Kiefer and Wolfowitz (1960) this special case will yield the same optimal designs for (6) and (9). We conjecture that in the general case one can find a specific α, such that one can - if not achieve - but come very close to such an equivalence as well. That is we do believe that optimal designs for estimation can be reasonably useful for predictions purposes, which would decisively reduce the computational burden.

3 Examples

3.1 1 dimensional example

Let us first motivate the development by the simpler one-dimensional setup.

Case 1 In Kiseľák J. and Stehlík M. (2008) it is shown, that the equidistant design is D-optimal in the case of Ornstein-Uhlenbeck process with constant trend. Their **Theorem 4** is also providing the form of the FIM for parameter β when the process is OU with constant trend, i.e.

$$M_\beta(n) = 1 + \sum_{i=1}^{n-1} \frac{e^{rd_i} - 1}{e^{rd_i} + 1},$$

where $r = 1/\theta$. This theorem is some extension of the **Theorem 3.6** in Dette H., Kunert J. and Pepelyshev A. (2007). Therein is proved, that for $r \to 0$ the exact n-point D-optimal design in the linear regression model with exponential covariance converges to the equally spaced design.

Case 2 The form of the FIM for parameter r when the process is OU is given by

$$M_r(n) = \sum_{i=1}^{n-1} \frac{d_i^2(e^{2rd_i} + 1)}{(e^{2rd_i} - 1)^2},$$

where $r = 1/\theta$. This was also independently observed by Zagoraiou and Baldi-Antognini (poster, Moda 8, 2007) and personally communicated.

For a proof first notice that an Ornstein-Uhlenbeck process possesses a Markovian property and thus the structure of the inverse covariance matrix is tridiagonal (see for details Kiseľák J. and Stehlík M., 2008). From this fact and the structure of M_θ we conjecture $M_r = \sum_{i=1}^{n-1} g(d_i, r)$, where $g(d_i, r)$ is a differentiable function of both variables. We have $g(d_i, r) = \frac{d_i^2(e^{2rd_i}+1)}{(e^{2rd_i}-1)^2}$ (see Stehlík M. , 2005). To complete this proof, we can now use either the fundamental theorem of calculus and the fact, that

$$\frac{\partial M_r}{\partial d_i}|_{d_i=x} = -\frac{2\,x\,(-e^{(4\,r\,x)} + 1 + x\,r\,e^{(4\,r\,x)} + 3\,x\,r\,e^{(2\,r\,x)})}{(e^{(2\,r\,x)} - 1)^3}$$

or complete induction.

Note that typically, when no nugget effect is present, the D-optimal design will be collapsing unto a single point (see Müller W.G. and Stehlík M. , 2008 and Stehlík M., Rodríguez-Díaz J.M., Müller, W.G. and López-Fidalgo, J. , 2008).

Case 3 If we are interested in estimating of both parameters simultaneously, typically, the correlation parameter has a higher impact on the design. This can also lead to the collapsing of the design or some counterintuitive behavior.

3.2 Spatial case

Here we will employ the same example as given in Müller W.G. (2005): The Südliche Tullnerfeld is a part of the Danube river basin in central Lower Austria and due to its homogeneous aquifer well suited for a model-oriented geostatistical analysis. It consists of 36 official water quality measurement

stations, which are irregularly spread over the region. A graphical representation of the sites (after rescaling to a unit of approximately 31 kms so that $\mathcal{X} = [-1, 1]^2$) on a 21 times 95 point grid is given in Figure 1. Note that due to the irregularity of the region only a fraction of the grid is actually used. The data set used in some calculations contains daily averages of chloride (Cl) concentrations in mg/l over the period 1992-1997 on all time points, for which at least one measurement was taken. As a spatial trend we take again a simple

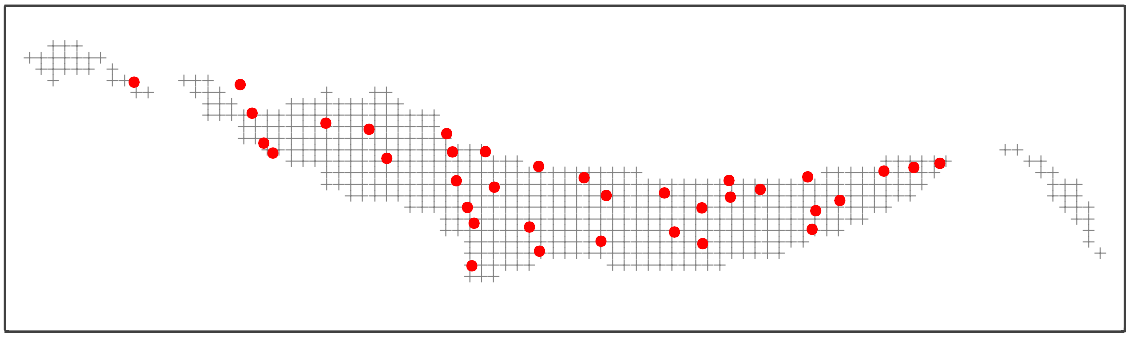

Figure 1: Chloride monitoring network in the Südliche Tullnerfeld (Austria).

plane, i.e. $f(x) = \binom{1}{x}$ and for the covariance function we employ an isotropic spherical covariance model with local parameter guesses $\hat{\theta} = \{4.89, 1.86, 0.81\}$ as given in Müller W.G. (2005).

Case 1 $\alpha = 1$: Resulting designs have already been presented therein (e.g. Figure 2), essentially the same emerge from using the present algorithm with efficiency improvements of around 30%.

Note that for the OU processes Stehlík and Kiseľák have observed that a triangular design will be close to the 3-point D-optimal designs. Moreover, a number of authors have investigated the problem of spatial sampling design assuming the correlation parameters to be known. For instance Yfantis, E.A., Flatman, G.T. and Behar, J.V. (1987), who provided empirical evidence, that when using kriging as prediction method and the average of maximum kriging variance as criterion, the equilateral triangular grid is apparently nearly optimal.

Case 2 $\alpha = 0$: For the construction of a local guess for M' one requires the covariance function and its derivatives for the preestimated spherical variogram. The latter are given by

$$\frac{\partial c(.)}{\partial \theta_1} = \begin{cases} 1, & h = 0 \\ 0, & \text{else} \end{cases}, \quad \frac{\partial c(.)}{\partial \theta_2} = \begin{cases} 1, & h = 0 \\ 1 - \frac{3h}{2\theta_3} + \frac{h^3}{2\theta_3^3} & 0 < h < \theta_3 \\ 0, & \text{else} \end{cases},$$

$$\text{and } \frac{\partial c(.)}{\partial \theta_3} = \begin{cases} \frac{3\theta_2 h}{2\theta_3^2}\left(1 - \frac{h^2}{\theta_3^2}\right) & 0 < h < \theta_3 \\ 0, & \text{else} \end{cases}.$$

By again applying our hybrid exchange type algorithm from Section 2.2 and employing $\Phi[M'] = \det M'(\xi)$ a reasonably good design is found after only

30 exchanges, yielding a value of $\Phi[M'(\xi^*)] = 226704 \times 10^5$, which is a vast improvement over the $\Phi[\xi'(A)] = 3954 \times 10^5$ of the existing design. The resulting ξ^* is displayed in Figure 2 and shows the in these instances usual separation of large and small distances (cf. Müller and Zimmerman (1999)).
Case 3 $\alpha = 0.5$

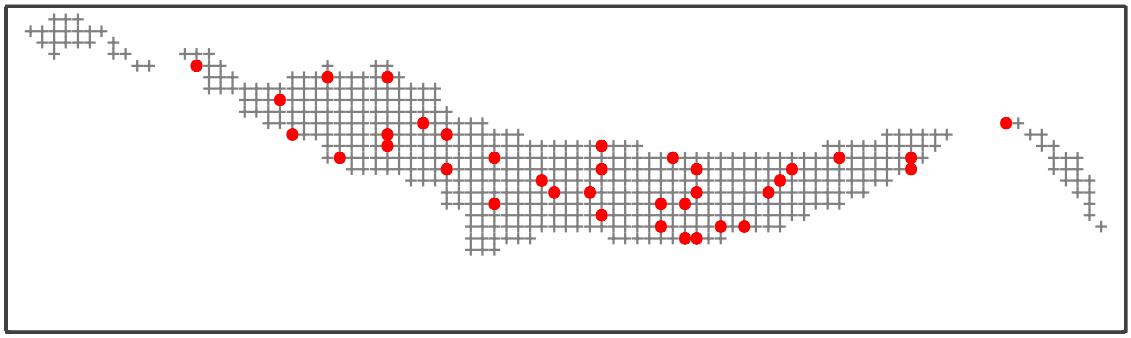

Figure 2: D-optimal network for estimating covariance parameters only.

Figure (3) displays the network after optimizing criterion (6) and clearly shows the balance between the two concurring criteria. It results in an about 7-fold improvement in the compound criterion. However, a fairer efficiency comparison, not distorted by the scaling effect, is found when we reduce the numbers of design points, such that we can still achieve the same criterion value as with the initial network. This leads to what is displayed in figure (4): an equivalent network with 21 sites (15 less than what we started with).

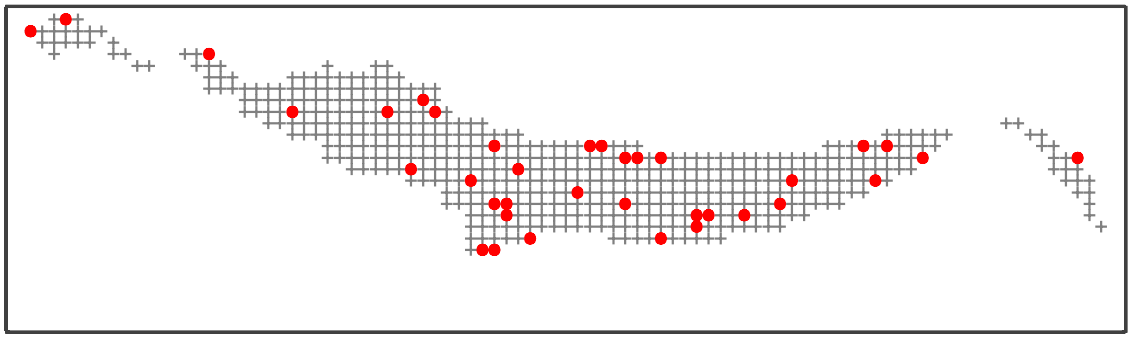

Figure 3: D-optimal compound design network, 36 sites, $\alpha = 0.5$.

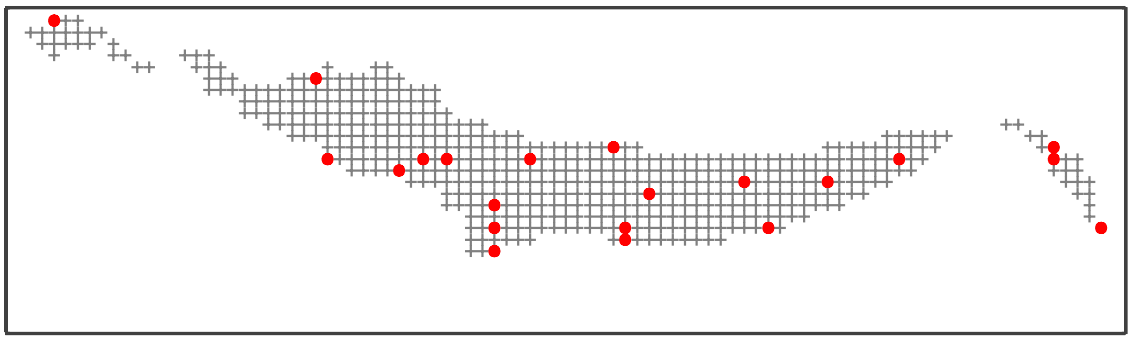

Figure 4: D-optimal compound design network, 21 sites, $\alpha = 0.5$.

Acknowledgments

The authors research was partially supported by Acciones Integradas ES 18/2008.

References

Atkinson, A.C., Donev, A.N., and Tobias, R. (1992, 2007). *Optimum Experimental Designs, with SAS*. Oxford Statistical Science Series, Oxford University Press.

Biedermann, S., Dette, H., and Zhu, W. (2007). Compound optimal designs for percentile estimation in doseresponse models with restricted design intervals. *Journal of Statistical Planning and Inference*, 137(12):3838–3847.

Caselton, W.F. and Zidek, J.V. (1984). Optimal monitoring networks. *Statistics & Probability Letters*, 2:223–227.

Cook, D. and Fedorov, V. (1995). Constrained optimization of experimental design (with discussion). *Statistics*, 26:129–178.

Cook, R.D. and Wong, W.K. (1994). On the equivalence of constrained and compound optimal designs. *Journal of the American Statistical Association*, 89:687–692.

Cressie N.A.C. (1993). Statistics for Spatial Data, 2nd ed., New York: Wiley.

Dette H., Kunert J. and Pepelyshev A. (2007). Exact optimal designs for weighted least squares analysis with correlated errors. *Statistica Sinica* 2008; 18:135–154.

Fedorov, V.V. (1972). *Theory of Optimal Experiments*. Academic Press, New York.

Fedorov V.V. (1989). Kriging and other estimators of spatial field characteristics. *Atmospheric Environment*, 23(1):175-184.

Guttorp, P. and Gneiting, T. (2006). Miscellanea: Studies in the history of probability and statistics XLIX. On the Matérn correlation family. *Biometrika*, 93(4):989–995.

Haas, T.C. (1992). Redesigning continental-scale monitoring networks. *Atmospheric Environment*, 26A(18):3323–3353.

Harville DA, Jeske DR. Mean squared error of estimation or prediction under a general linear model. *Journal of the American Statistical Association* 1992; **87**:724-731.

Kiefer, J. (1959). Optimal experimental designs (with discussion). *Journal of the Royal Statistical Society, Series B*, 272–319.

Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems. *Canadian Journal of Mathematics*, 12:363–366.

Kiseľák J. and Stehlík M.(2008). Equidistant D-optimal designs for parameters of Ornstein-Uhlenbeck process, *Statistics and Probability Letters* 78, 1388–1396.

Läuter, E. (1976). Optimal multipurpose designs for regression models. *Mathematische Operationsforschung und Statistik, Series Statistics*, 7:51–68.

Le, N.D. and Zidek, J.V. (2006). *Statistical Analysis of Environmental Space-Time Processes.* Springer, New York.

Lesch, S.M. (2005). Sensor-directed response surface sampling designs for characterizing spatial variation in soil properties. *Computers and Electronics in Agriculture*, 46:153–179.

Mcgree, M., J., Eccleston, A., J., Duffull, and B., S. (2008). Compound optimal design criteria for nonlinear models. *Journal of Biopharmaceutical Statistics*, 18(4):646–661.

Müller W.G. (2005). A comparison of spatial design methods for correlated observations, *Environmetrics*, 16, 495-505.

Müller W.G. (2007). *Collecting Spatial Data*, 3rd edition, Springer Verlag.

Müller W.G. and Stehlík M. (2008). Issues in the Optimal Design of Computer Simulation Experiments, *Applied Stochastic Models in Business and Industry*, DOI: 10.1002/asmb.740.

Müller, W. G. and Zimmerman, D. L. (1999). Optimal designs for variogram estimation. *Environmetrics*, 10(1):23–37.

Park, S.Y., Choi, J.H., Wang, S., and Park, S.S. (2006). Design of a water quality monitoring network in a large river system using the genetic algorithm. *Ecological Modelling*, 199:289–297.

Pázman, A. and Laca, J. (2008). The method of virtual noise for optimal design in experiments with estimated covariance function. Proceedings of Probastat2006, *Tatra Mt. Math. Publ.* 39 (2008), 125–133.

Rao CR. (1965) *Linear Statistical Inference and Its Applications*. New York: Wiley

Royle, J. A. (2002). Exchange algorithms for constructing large spatial designs. *Journal of Statistical Planning and Inference*, 100(2):121–134.

Stehlík M. (2004). Further aspects on an example of D-optimal designs in the case of correlated errors, Report #1 of the Research Report Series of the Department of Statistics and Mathematics, Wirtschaftsuniversität Wien, Austria, 2004.

Stehlík M. (2005). Covariance related properties of D-optimal correlated designs, Proceedings of the 5th St. Petersburg Workshop on Simulation, St. Petersburg, Russia, June 26-July 2, Edited by S.M. Ermakov, V.B. Melas and A.N. Pepelyshev, 645-652.

Stehlík M., Rodríguez-Díaz J.M., Müller, W.G. and López-Fidalgo, J. (2008). Optimal allocation of bioassays in the case of parametrized covariance functions: an application in lung's retention of radioactive particles, *Test* 17: 56–68.

Stehlík M. and Müller W.G. (2008). Fisher information in the design of computer simulation experiments, accepted for *Journal of Physics: Conference Series*, 2008.

Steinberg, D.M., Rabinowitz, N., Shimshoni, Y., and Mizrachi, D. (1995). Configuring a seismographic network for optimal monitoring of fault lines and multiple sources. *Bulletin of the Seismological Society of America*, 85(6):1847–1857.

Xia G., Miranda M.L. and Gelfand A.E. (2006). Approximately optimal spatial design approaches for environmental health data, *Environmetrics*, 17: 363–385.

Yfantis, E.A., Flatman, G.T. and Behar, J.V. (1987). Efficiency of kriging estimation for square, triangular, and hexagonal grids. *Mathematical Geology*, 19:183–205.

Zhou, J. (2008). D-optimal minimax regression designs on discrete design space. *Journal of Statistical Planning and Inference*, 138(12):4081–4092.

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

Zimmerman DL. Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. *Environmetrics* 2006; **17**:635–652.

Zimmerman DL, Cressie N. Mean squared prediction error in the spatial linear model with estimated covariance parameters. *Annals of the Institute of Statistical Mathematics* 1992; **44**(1):27–43.