SIMULATION EXPERIMENTS TO EVALUATE THE ROBUSTNESS OF THE OPTIMAL CONSTRUCTION OF MONITORING NETWORKS

M. GAÁL[1*] – SCHMIDTKE, J.[2] – RASCH, D.[4] – SCHMIDT, K.[2] – NEEMANN, G.[3] – KARWASZ, M.[3]

**e-mail: Marta.Gaal@uni-corvinus.hu*

[1]*Department of Mathematics and Informatics, Faculty of Horticultural Sciences, Corvinus University of Budapest H-1118 Budapest, Villányi út 29–33, Hungary (phone: +36-1-372-6261; fax: +36-1-466-9273)*

[2]*BioMath GmbH, Schnickmannstr. 4., 18055 Rostock, Germany (phone: +49-381-496-5810; fax: +49-381-496-5813)*

[3]*BLaU, Göttingen, Germany Wiesenstr. 8, 37073 Göttingen, Germany (phone: +49-551- 703435; fax: +49-551- 703536)*

[4]*Alpen-Adria Universität Klagenfurt, Institut für Mathematik Universitätsstr. 65-67, 9020 Klagenfurt (phone: +43-463-2700-3717; fax: +43-463-2700-3099)*

(Received 5th May 2004; accepted 2nd December 2004)

Abstract. OptiNet is a PC program for optimal network selection. The aim of this study is to test the program by simulation experiments and to investigate the robustness of the optimal selection. One of the most important results is that the area should not be represented by an equidistant grid to calculate the maximum Kriging prediction variance, it is sufficient to investigate the boundary points only. Effects of the parameters of the covariance function, the number of selected points and of a possible factor were also investigated. All simulations are based on the Gauss-Krüger coordinates with 1 m raster in the area of Brandenburg.
Keywords: *spatial statistics*, *optimal design, simulation, kriging, covariance function*

Introduction

In agriculture there is often need for using monitoring networks to detect changes within an agro-ecosystem. Monitoring is usually an expensive and time consuming work. Therefore monitoring networks should be planned and organized in an optimum way. To do this OptiNet, a program for optimal network selection was developed by BioMath GmbH in Rostock [6]. This is a PC program for:

- Constructing an optimal network in a given area.
- Extending an existing network in a given area.
- Reducing an existing network in a given area.
- Graphical representation of the candidate (starting) points and the selected points.

http://www.ecology.kee.hu · ISSN 1589 1623

The selection is based on the optimality criterion of the maximal Kriging prediction variance in the area, which is based on an exponential covariance function. A network is optimal if it minimizes this criterion within a given area (Brandenburg in our case) over the set of all possible networks. To find the maximum prediction variance, the area should be represented by a regular (equidistant) grid with a given distance *D* between neighbouring points in a row and a column.

Several problems, which up to now have not been solved theoretically, were investigated by simulation experiments. During the simulations, components of OptiNet were used and tested.

Our aim was to answer the following questions:

- How large can *D* be chosen so that a wrong network would be selected with a small probability?
- How can we interpret the parameters of the covariance function for practical purposes?
- What is the influence of incorrectly selected parameters of the covariance function?
- Is there any optimal number of selected points for a network?

The area examined

The simulation was based on the Gauss-Krüger coordinates with 1 m grid in the area of Brandenburg. This means that locations can be identified as different only if at least one of their coordinates differs by more than 1 m.

The sites were selected from an existing monitoring network for ground water level in Brandenburg. Most of the simulations were done with two sets of starting points, one with 20 sites, the other one with 24 sites. *Figure 1.* shows the 20 selected sites with circles and the squares indicate the additional points. Later in some parts of the simulation further points were added up to 32 or 40.

http://www.ecology.kee.hu · ISSN 1589 1623

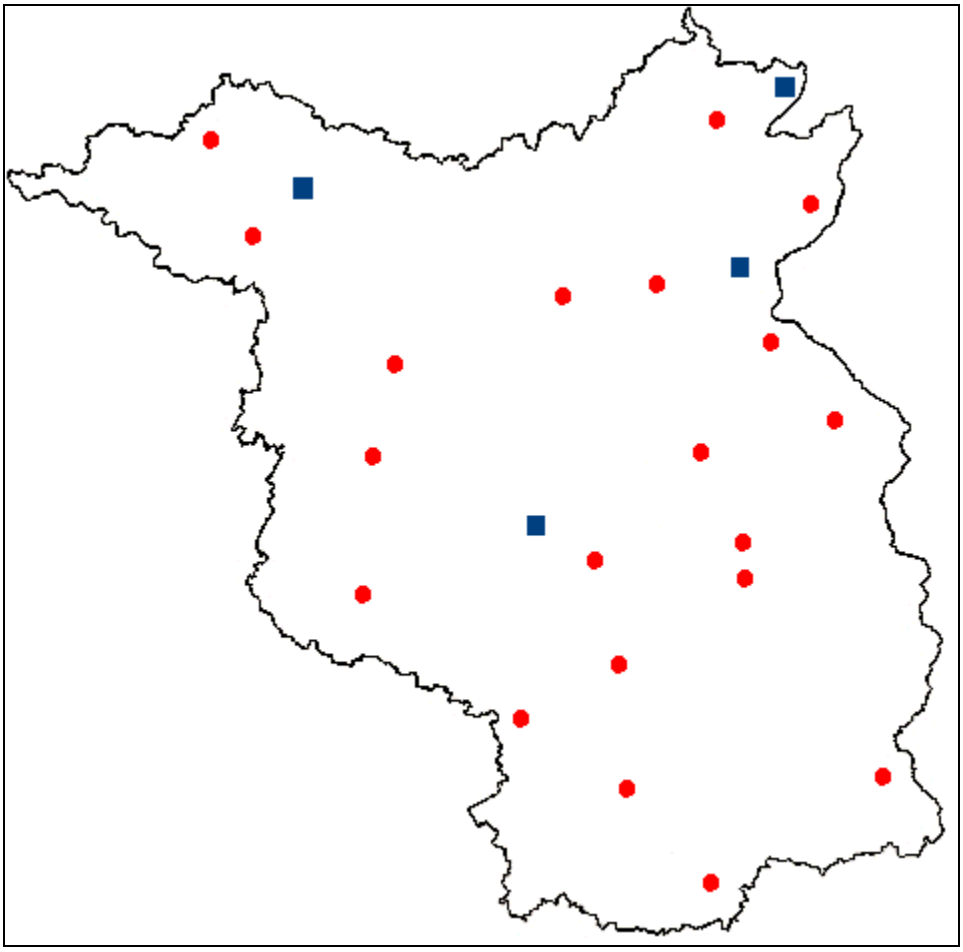


Figure 1. *The selected sites in the region Brandenburg (the circles indicate the 20 points, the squares the additional points).*

The covariance function

To calculate the Kriging prediction variance, the exponential covariance function was chosen. It is given by:

$$\mathrm{cov}(h_{ij}) = \alpha + \beta \cdot e^{\gamma h_{ij}}; \quad \alpha, \beta > 0, \quad \gamma < 0, \; i \neq j; i, j = 1, \ldots, N \qquad \text{(Eq.1)}$$

where h_{ij} represent the Euclidean distances between the pairs of candidate points P_i and P_j.

The exponential function in (Eq.1) is an intrinsically non-linear function [7] with linear parameters α and β and non-linear parameter γ.

Figure 2. shows the graphs of three exponential covariance functions with different γ values. The value of γ depends on the dimension in which h is measured. As the scale of x-axis (representing the distances) is changed from kilometres to meters, the value of γ must be divided by 1000 as it can be seen from (1) because the product γh must be invariant due to rescaling. During the calculations the coordinates, and by this the distances were given in meters. But in this reports we use the km-scale.

http://www.ecology.kee.hu · ISSN 1589 1623

$$f(x) = \alpha + \beta \cdot e^{(\gamma \cdot x)}$$

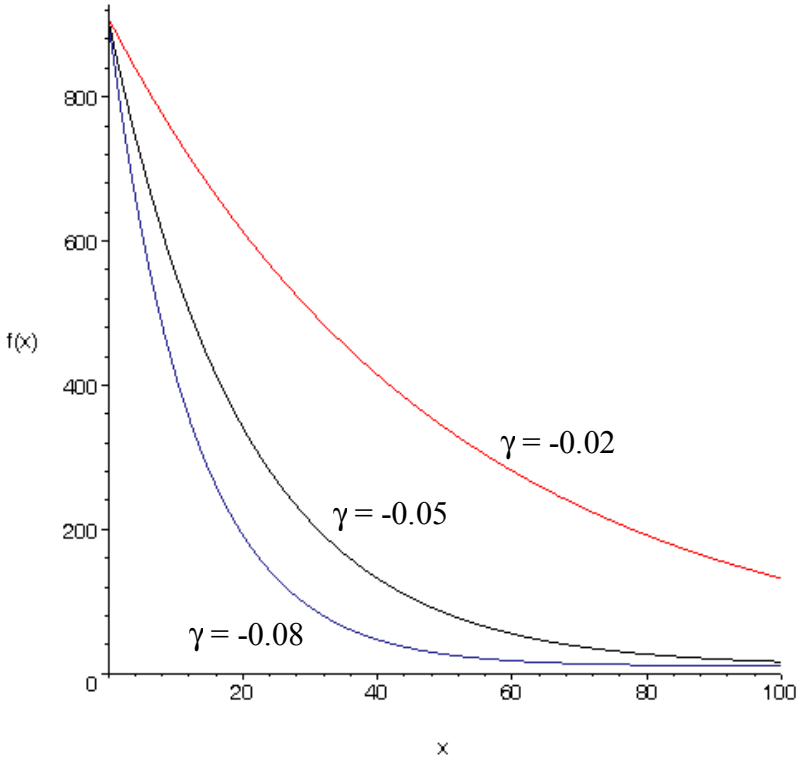


Figure 2. *The graph of the exponential covariance function with α = 10, β = 900 and different γ values.*

The value of $\alpha+\beta$ represents the variance (covariance for the distance 0), α is the asymptote for x to infinity, and γ describes the spatial correlation. In the examples above, we can consider the middle one (γ= -0.05) as a general case, while the other two represent extreme situations. For practical purposes we can say that in case of γ= -0.05 the spatial correlation has effect within about 60 km distance, in case of γ= -0.02 also in more than 100 km, while in case of γ= -0.08 only within 30-40 km.

Simulations and results

Grid distances

Based on the former study [6], the area should be represented by an equidistant grid with a given distance D to find the maximum prediction variance. The grid-sizes were examined with D-values of 5, 10, 15, 20, 25 and 30 km. A special case, labelled with 00, with only four grid-points out of the region was applied additionally. The parameters of the covariance function were α = 10, β = 900 and γ_1 = -0.05, γ_2 = -0.02 and γ_3 = -0.08. The number of the candidate points were N = 20 or 24, and in both cases we selected n = 12 points for the network. The main results were the followings:

http://www.ecology.kee.hu · ISSN 1589 1623

- The time needed for the calculation depends on N and n (according to the $\binom{N}{n}$ possible selections) and on the grid-size, too. For example in case of $N = 24$, $n = 12$ the total number of possible selections is almost 3 million (2704156). In our computer the running times were for $D = 10$ km 4 days, in case of $D = 5$ km more than 17 days.
- The values of the prediction variance usually increase as the grid-sizes decrease, but it is not a monotonous change.
- If γ has lower (more negative) value, the standard deviation of the criterion values is much lower, however the criterion values are higher.
- Comparing the different grid-sizes, the results of 5, 10 and 25 km grids proved to be similar.
- The grid-points, where we can find the lower criterion values, are located in almost every case on the boundary of the region or near to it, if there is no point on the boundary (like in case of the grid with 30 km).

Calculations based on the boundary points

Since in former simulations the maximum criterion values were always found on the boundary of the region, in our calculations we were searching the maximum prediction variance for evaluating different subsets on the boundary only instead of the looking at the whole grid.

This in this time heuristical approach is in the meantime justified by a paper of Haberl [3]. There it is shown that even in the case of a non-convex region (like Brandenburg) the maximum of the prediction variance is always on the (outer) boundary of the region and a gap inside the region (Berlin in Brandenburg) plays no role.

The distances between the points on the boundary were 5, 10, 15, 20, 25 and 30 km. These calculations take significantly less time than in investigating the whole grid.
We can also observe that:

- Results obtained with 5 or 10 km distances can be considered as equal.
- As in case of the grids, if γ has lower value, the standard deviation of the criterion values is much lower and the criterion values are higher.
- We can say that γ has an effect on the selected sets.

Effects of α and β

As we know that γ has an effect on the selected sets, it is interesting whether the two other parameters, α and β have also some influence or not. The covariance function was used with $\gamma = -0.05$ and with the following parameters:

a) $\alpha = 10$, $\beta = 900$
b) $\alpha = 5$, $\beta = 450$
c) $\alpha = 5$, $\beta = 200$
d) $\alpha = 5$, $\beta = 900$
e) $\alpha = 20$, $\beta = 900$

Regarding these parameters we got the following results:

http://www.ecology.kee.hu · ISSN 1589 1623

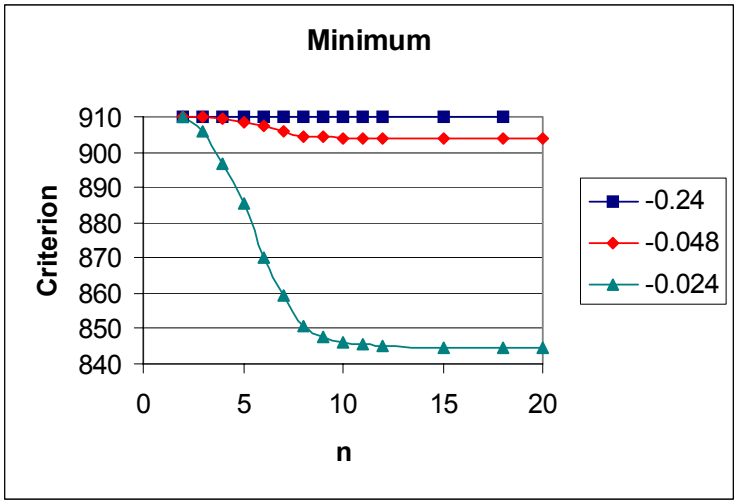
- The values of α and β do have no effect on the selected sets, we get always the same results. This was expected because these parameters are linear parameters of the intrinsically non-linear exponential function.
- The criterion values are determined only by β (in this case more or less equal to the value of β).
- As β is increasing, the difference between the minimum and maximum values of the criterion is also increasing.
- The results of D = 5 and D = 10 km can be considered as equal, also the minimum and maximum criterion values are the same, but in case of D = 5 the standard deviations are a little bit higher.
- With the same β, the criterion values were a little bit higher in case of N = 20 than in case of N = 24.

Effect of the selected sets

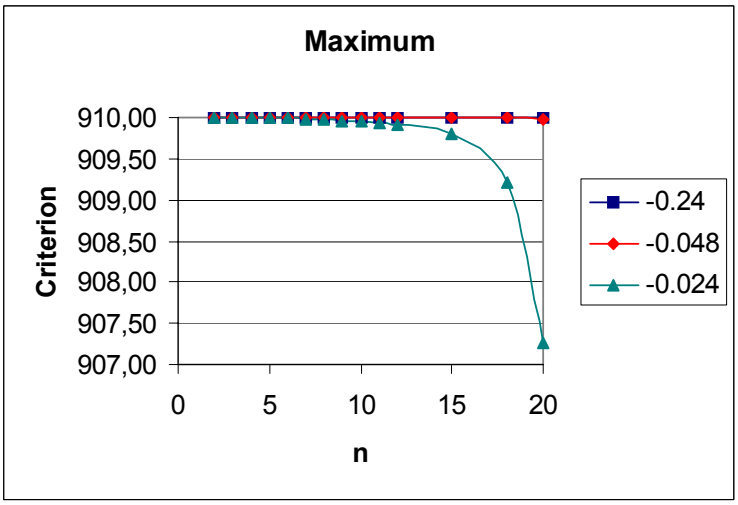
One of the most important questions is, whether there are an optimal number of selected points for a network. To answer this, a series of selections were made from N = 24 points, with the parameters α = 10, β = 900, γ_1 = -0.24, γ_2 = -0.048 and γ_3 = -0.024 (the different γ values indicate that the spatial correlation may be significant within 10, 50 or 100 km). The distances between the points on the boundary were 10 km, as we could see now, that it gives the same results as D = 5 km, but the calculation is quicker. Simulations showed that:

- When the number of the selected points (n) increases, the criterion values decrease.
- The change is more rapid in case of the minimum of the criterion values, and not significant in case of the maximum criterion values until $n \leq N/2$. It means also that if the number of the selected points increases, the differences between the minimum and maximum values of the criterion (which shows the mistake in case of a completely wrong selection) are also increasing in the first part, but then they decrease a little bit, when $n \approx N$.
- When we suppose that the spatial correlation exists only within a very short distance (in this example γ_1), the criterion values are the same for all selections and we cannot find an optimal subset of points.
- In case of larger distances (the γ value is near to zero) the differences are larger, so the optimal selection is really important.
- When we select 12 or more points (from 24), the change in the minimum value of the criterion is not important (*Figure 3/a*). This could mean that there is no reason to select more points for an optimal network, or we can select them arbitrarily. However, in the *Figure 3/c* we can see, that the difference between the minimum and maximum values of the criterion is still high (especially in case of γ_3), does not decrease with n, and the selection of additional points is also important.

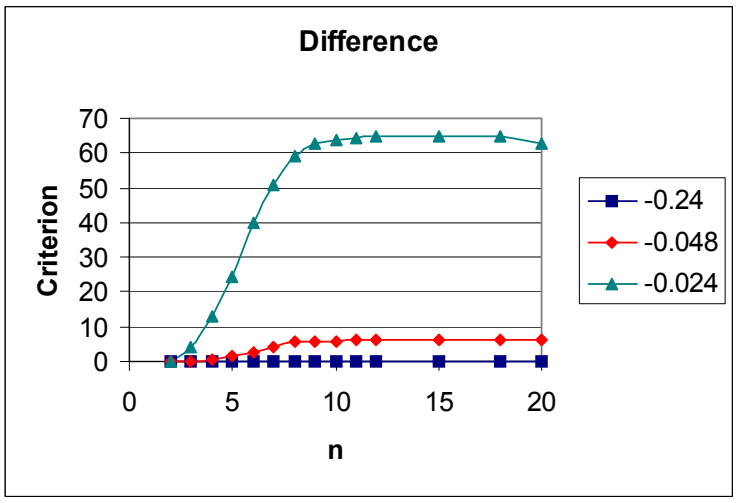
http://www.ecology.kee.hu · ISSN 1589 1623



a)



b)



c)

Figure 3. *The change of the criterion values by the selected sets (n), in case of different γ values (N = 24, D = 10 km, α = 10, β = 900).*

http://www.ecology.kee.hu · ISSN 1589 1623

Table 1. shows a series of selected sets. It can be seen that if we want to select a small set, the optimal selection is very important, as the selected points are changing. Than it seems that the formerly selected points remain, only we have additional selections too.

Table 1. *Selection series in the case N = 24 (D = 10 km, α = 10, β = 900, γ_3 = -0.024).*

	Candidate points																							
N	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
5		X										X		X						X			X	
6	X				X							X			X					X			X	
7	X		X					X		X		X						X	X					
8	X					X					X					X			X	X			X	
9	X	X				X						X				X			X	X			X	
10	X	X				X						X				X			X	X	X		X	
11	X	X				X			X			X			X		X		X	X	X		X	
12	X	X		X		X			X			X			X		X		X	X	X		X	X
15	X	X	X	X		X			X			X			X		X	X	X	X	X		X	X
18	X	X	X	X		X	X	X	X	X		X			X		X	X	X	X	X		X	X
20	X	X	X	X	X	X	X	X	X	X		X	X		X		X	X	X	X	X		X	X

From this simulation we cannot conclude that the change of the criterion values is not significant selecting more than 12 points (half of the starting points). Therefore we repeated the simulation with more starting points, like N = 32 and N = 40, using only γ= -0.024. It seems that when we select from many points, the minimum of the criterion values changes more rapidly, but the maximum changes slower, so the differences are higher.

As the running time does not allow calculating all the possible selections (in case of N = 40 and n = 8 the running time is more than 300 hours, in case of N = 32 and n = 9 more than 8 days), we suggest fitting a non-linear curve to estimate the missing values. The tangent hyperbolicus function proved to be good for these fittings (*Figure 4.*).

Based on the fitted curves we can conclude that the change of the minimum values of the criterion is not important when $n \geq N/2$.

Effect of a possible factor

All the previous simulations supposed that the area examined is more or less homogeneous. Since in most cases we have at least one environmental factor (like climate, soil type) defining different conditions at the candidate points, we also carried out a simulation supposing that we have a factor with four levels. The candidate points were divided into four groups (the whole area was divided into four sub-regions with equal number of sites in each). The sites for the network must be selected so that we have points from all sub-regions. (When it is possible with the same number of points from each sub-region.)

http://www.ecology.kee.hu · ISSN 1589 1623

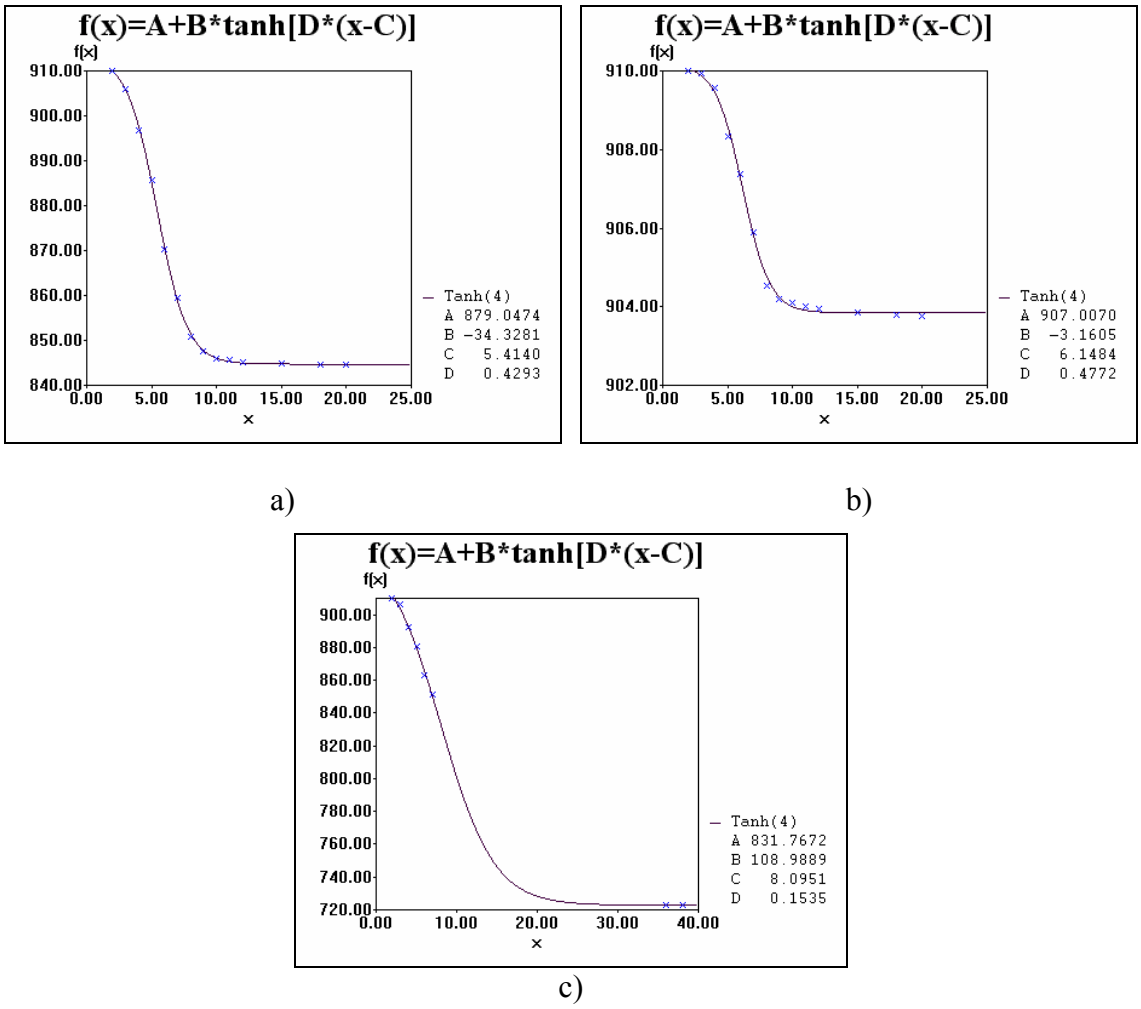


Figure 4. *Change of the criterion values by the selected sets (n) after fitting. a) N = 24, γ = -0.024, b) N = 24, γ = -0.048 c) N = 40, γ = -0.024*

As formerly, a series of selections were made from $N = 24$, $N = 32$ and $N = 40$ points, with the parameters $\alpha = 10$, $\beta = 900$ and $\gamma = -0.024$. We found that:

- The minimum values of the criterion are in almost every case the same as without sub-regions, but the maximum values and also the differences decrease more rapidly.
- Using sub-regions or not, the selected sites are in almost every case the same (therefore are the minimum values of the criterion the same).
- The calculation time is shorter, because we have a restricted number of possibilities due to blocking in sub-regions.

An example for utilising the results of simulation experiments

A monitoring character for Genetically Modified Plants

Genetically modified plants (GMP's) - once they have been placed on the market - have to be observed for suspected adverse environmental effects in accordance to

http://www.ecology.kee.hu · ISSN 1589 1623

Annex VII of the EU Directive 2001/18/EC and the Monitoring Guidance Note 2002/811/EC.

The aims are to monitor dissemination tendencies of the GMP's and of outcrossing events concerning the transfer of recombinant DNA to wild relatives (formation of recombinant hybrids). An example for a crop species possibly tending to disseminate and, additionally, having a number of wild relatives living in natural habitats could be rape (*Brassica napus* L.) [1]. Since genetically modified crop species - as the non-modified lines - will be cultivated in an agricultural environment the supposed GM crop depending environmental effects have to be monitored in ecosystems typical for agricultural landscapes.

Dissemination tendencies of GM crops and of hybrids of GM crops and wild relatives could be expected if the recombinant traits offer some ecological advantages to the recombinant plants. The new traits may change the behaviour of the species in a way the recombinant descendants becoming feral. As a consequence, recombinant crop species may get the opportunity to emerge not only in agro-ecosystems but also in special natural habitats. Mainly, more or less frequently disturbed habitats (so-called "ruderal" sites) will be appropriate for the growth of feral GM crops and of GM hybrids out of cultivated fields [5].

Consequently, to observe the behaviour of recombinant feral crops and hybrids special target (cultivated fields) and non-target ecosystems (natural and semi-natural habitats) typical for agricultural landscapes have to be investigated in a GMO monitoring. Accordingly, a monitoring network has to consider the landscape structure of agricultural landscapes concerning cultivated fields and natural habitats.

Neemann et al. [4] found the proportion of annual weeds of the plant communities, especially of the non-agricultural ecosystems, to be a suitable indicator for the differentiation of habitats of greater or minor importance for the dissemination of GM crops. Higher proportions of annual weeds are typical for ecosystems being frequently disturbed. Additionally, the more frequent natural ecosystems in agricultural landscapes are disturbed, the more often feral crops or wild relatives are found, too. To differentiate disturbed habitats of higher or minor importance for the dissemination of feral GM crops and wild hybrids a ranking system was developed. However, some agro-ecosystems (e.g. fields cultivated with summer annual crops) may also be of importance for the reproduction of feral GM crops.

As a consequence, for the implementation of a GMP monitoring in agricultural landscapes a character is needed offering information about the quality of the agricultural and non-agricultural ecosystems for dissemination. Additionally, the character must give information about the quantitative importance of disturbed (ruderal) sites and of fields cultivated with summer annual crops within specific landscapes. Considering the ranking system the character summarizes the area-based proportion of disturbed habitats of great importance for dissemination and of the area-based proportion of fields cultivated with summer annual crops. Accordingly, the character may be described as "dissemination character".

An optimal monitoring network for Genetically Modified Plants

The construction of an optimal network for agricultural landscapes considering the dissemination character firstly means to analyse landscape sectors quantitatively being important for GM crops and GM hybrids dissemination. This work has to be done by environmental specialists. Secondly, to reduce the monitoring expenditure within a

http://www.ecology.kee.hu · ISSN 1589 1623

landscape the optimal numbers of the selected sectors and the optimal places for them have to be found.

For the optimization example calculated below the composition of agro- and non-agro-ecosystems were investigated in six agricultural landscapes of the German regional state of Brandenburg. 22 landscape sectors representing the habitat composition of the six landscapes were analysed in detail concerning their habitat and field characteristics. Each of the sectors had an extension of 3 km x 3 km.

For the agricultural landscapes considered the optimal places of monitoring sectors have to be found by the optimization procedure. The distribution of the exponential covariance function was determined from measured values representing the "dissemination character" from the 22 habitat sectors. The model parameter ß was estimated from the range of the values to be 38.23 %. The parameter γ describing the magnitude of the spatial dependency was – due to the results of the simulation study – fixed to -0.05. With this model it is possible to construct an optimal network for monitoring the dissemination within habitats, i.e. to find the optimal number and location of habitat sectors of 3x3 km from possible candidate points. For the region Brandenburg we took the 22 candidate points from which 6, 12 or all points should be selected for a future monitoring network – considering the factor "landscape". In *Table 2*. the X characters represent the best selection for a given number of selected sets and the criterion values can be seen, too.

Table 2. *Results of the selection series.*

LOCATION				**Selected points**		
Name	**Coord. x**	**Coord. y**	**Landscape**	**6**	**12**	**22**
Dahnsdorf-Ost	5341244	5778382	Fläming			X
Dahnsdorf-West	5338225	5777500	Fläming		X	X
Illmersdorf	5381900	5752046	Fläming			X
Hohenseefeld	5385900	5752200	Fläming	X	X	X
Schönefeld	5387244	5773400	Nuthe-Urstromtal		X	X
Stülpe	5384300	5771300	Nuthe-Urstromtal	X	X	X
Bliesdorf	5443525	5842237	Oderbruch	X	X	X
Altlangsow-Nord	5458844	5828555	Oderbruch		X	X
Altlangsow-Süd	5461855	5826730	Oderbruch			X
Mallnow	5464644	5817600	Oderbruch			X
Rädikow	5432400	5842800	Barnim-Lebuser-Platte		X	X
Frankenfelde	5435600	5838400	Barnim-Lebuser-Platte			X
Seelow	5455844	5822900	Barnim-Lebuser-Platte			X
Schönfließ	5464244	5809400	Barnim-Lebuser-Platte	X	X	X
Postlin	5283231	5899400	Prignitz	X		X
Blüthen	5284131	5896000	Prignitz		X	X
Pirow	5293200	5904600	Prignitz			X
Burow	5296400	5903200	Prignitz		X	X
Falkenhagen	5416531	5916200	Uckermark		X	X
Ellingen	5421831	5914500	Uckermark			X
Göritz-Malchow	5427531	5922281	Uckermark	X	X	X
Dauer	5428417	5917900	Uckermark			X
criteria*				**2,716**	**3,285**	**3,335**

* criteria = (estimation of variance - optimality criterion) * 10^3

http://www.ecology.kee.hu · ISSN 1589 1623

The optimality criterion is the maximal prediction error for the dissemination character in the Brandenburg region. With increasing the number of selection points the optimality criterion decreases, i.e. the prediction error for sites without measurement will become smaller. By selecting 12 points (two per landscape) the criteria enhances by 20.95 % compared to the selection of only 6 points. Selecting all (=22) points the criteria enhances only by 1.52 % (with reference to 12 points).

The example shows that with the knowledge about the features of the covariance function and the distribution of the characteristic under examination it is easy to select different numbers of optimal sites from networks. This is helpful for example in designing time and cost effective monitoring plans for example for GMO´s in balancing the number of sites to be maximised and the prediction error to be minimised.

Discussion

Construction of monitoring networks is often needed in agriculture and environmental sciences. The aim of this study was to test the OptiNet program developed by BioMath GmbH [6]. One of the most important results is that the area should not be represented by an equidistant grid to calculate the maximum Kriging prediction variance, it is sufficient to investigate the boundary points only, that makes the selection more rapid. In the meantime it was justified by a paper of Haberl [3], too.

The optimality criterion of the maximal Kriging prediction variance in the area, based on an exponential covariance function is effective especially in case we select less then half of the candidate points ($n \leq N/2$).

It is also very important that when we suppose that the spatial correlation exists only within a very short distance, the criterion values are the same for all selections and we cannot find an optimal subset of points.

A case study regarding the monitoring of GMP's was also demonstrated. It showed that the program is a helpful tool in designing time and cost effective monitoring plans.

Acknowledgements. The research work was supported by the Eötvös Scholarship of the Hungarian Scholarship Board.

REFERENCES

[1] Chèvre, A.M., Ammitzbøll, H., Breckling, B., Dietz-Pfeilstetter, A., Eber, F., Fargue, A., Gomez-Campo, C., Jenczewski, E., Jørgensen, R., Klein, E., Meier, M.S., den Nijs, H.C.A., Pascher, K., Seguin-Swartz, G., Sweet, J., Stewart, N., Warick, S. (2004): A review on interspecific gene flow from oilseed rape to wild relatives. - In: den Nijs, H.C.A., Bartsch, D., Sweet, J. (eds.): Introgression from Genetically Modified Plants into Wild Relatives. CABI Publishing, Wallingford, pp. 235–251.

[2] Guttorp, P. (2003): Environmental Statistics – A Personal View - International Statistical Review 71:2, pp. 169-179.

[3] Haberl, J. (2004): Verallgemeinerte Konvexität und Richtungsmonotonie bei der Maximierung von Funktionalen in topologischen Vektorräumen. - Habilarbeit am Institut für Mathematik der Alpen-Adria Universität Klagenfurt, eingereicht, Dez. 2003, verteidigt am 17.11. 2004.

http://www.ecology.kee.hu · ISSN 1589 1623

[4] Neemann, G., Karwasz, M., Weitemeier, M. (2004): Umweltbeobachtung der Ausbreitung von Raps (*Brassica napus* L., spp. *oleifera*) in ausgewählten Flächenstrukturen. - In: LUA Brandenburg (Hrsg.): Anbau gentechnisch veränderter Pflanzen. Koexistenz und Umweltbeobachtung im Agrarraum. Studien und Tagungsberichte des Landesumweltamtes, Band 48: 38–48.

[5] Pessel, F.D., Lecomte, J., Emeriau, V., Krouti, M., Messean, A., Gouyon, P.H. (2002): Persistence of oilseed rape (*Brassica napus* L.) outside of cultivated fields. - Theor. Appl. Genet. 102: 841–846.

[6] Pilz, J., Rasch, D., Schmidtke, J. (2005, in press): Construction of an Optimal Network for Monitoring GMOs in Brandenburg. – In: J. Pilz (Ed.): Central European Series on Applied Statistics in Business, Environment, Finance and Technology, Vol. I, Gustav Heyn, Klagenfurt

[7] Rasch, D. (1995): Mathematische Statistik. - Joh. Ambrosius Barth**,** Berlin, Heidelberg (851 p.)

[8] Rasch, D. Herrendörfer, G., Bock, J., Victor, N., Guiard, V. (1996): Verfahrensbibliothek Versuchsplanung und -auswertung, Vol. I, 2nd edition - Oldenbourg Verlag München, Wien (940 p.)

http://www.ecology.kee.hu · ISSN 1589 1623