# Simulation-based optimal design for estimating weed density in agricultural fields

Bel Liliane, Parent Eric

AgroParisTech/ INRA MIA 518, Paris, France. Liliane.Bel@agroparistech.fr

Makowski David

INRA, Thiverval-Grignon, France.

**Abstract:** In order to manage herbicide treatment we present a method for optimizing the locations of weed density measurements. The practical problem is to estimate weed density in each one of the $n$ quadrats of a field, assuming that $m$ measurements were already collected and using $p$ additional measurements optimally located. The proposed method consists in three steps: 1) fit a statistical model to the $m$ available measurements taking into account the nature of the data, 2) define possible locations of the $p$ additional measurements using a simulated-annealing algorithm, 3) assess the designs using weed density values simulated using the fitted statistical model. This method is applied to several wheat fields and the results show that it improves weed density predictions. Sensitivity to several tuning parameters is discussed.

## 1   Introduction

Weeds can induce important yield losses in agricultural fields. In order to prevent huge losses weed management is frequently based on herbicide application. But extensive herbicide application leads to a risk of water pollution by chemicals. Sometimes, herbicide application is useless and the need of precise knowledge of weed density in the field is crucial. In order to provide a map of weed density in a field without counting all the plants, it is necessary to design a spatial statistical model fitted from a limited number of measurements. The purpose of this paper is to present a method for optimizing the locations of weed density measurements in agricultural fields in order to manage herbicide treatment. Consider an agricultural field divided into $n$ quadrats and assume that weed density measurements were already collected in $m$ out of the $n$ quadrats, $m < n$. Our practical problem is to estimate weed density in each one of $n$ quadrats by using

    i) the $m$ available weed density measurements,

    ii) $p$ additional measurements, $p < n - m$, collected in other quadrats located in the same field,

and by estimating the weed density in the unmeasured quadrats with a statistical technique. Potentially, the use of $p$ additional measurements can lead to improved weed density estimates, but the degree of improvement depends on the experimental design i.e. on the number of additional measurements $p$ and on their locations in the field.

This paper presents a method for defining, assessing, and selecting experimental designs in order to determine an appropriate number $p$ of additional measurements and optimize their locations in the field.

The proposed method consists in three main steps:

1. fit a spatial statistical model to the $m$ available measurements taking into account that the data are countings or presence-absence data,

2. assess the design of $m + p$ quadrats using weed countings values simulated using the fitted statistical model to define the criterion,

3. define possible locations of the $p$ additional measurements using a simulated-annealing algorithm according to the previously defined criterion.

This method was applied to several wheat fields and the results showed that it could improve weed density predictions. Sensitivity to several tuning parameters is discussed.

## 2 Materials and Methods

### 2.1 Statistical model for mapping weeds

Assuming $m$ measurements are available, a standard technique to produce a map of weed countings in a field of $n$ quadrats is ordinary kriging. Kriging performs well when the data distribution is Gaussian or not far from Gaussian. Weed countings are discrete data and the Gaussian distribution is not well adapted. Models for Poisson, zero Inflated Poisson and binary data are designed involving a continuous Gaussian latent variable accounting for the spatial dependence. The kriging is performed on the latent variable.

### 2.2 Conditional simulations

Conditional simulations are simulations of a spatial field according to a spatial model which are constrained to take observed values in a set of locations. Given a design of $m + p$ sites, its quality is assessed with the root mean square difference between conditional simulations and kriging estimates.

### 2.3 Simulated Annealing Algorithm

The search of an optimal design is achieved by a simulated annealing algorithm: $p$ quadrats are randomly selected and added to the $m$ initial quadrats. Slight perturbations on the previous configuration are iteratively proposed to improve the

conditional simulation criterion. Configurations that do not improve the criterion are accepted with a decreasing probability in order to favour the exploration of the configurations domain.

# 3 Results

## 3.1 Simulated data

The procedure is evaluated on simulated data, sharing the same characteristics as the weed data (size of the field, number of quadrats, countings data of same magnitude). It turns out that the procedure gives a better design to estimate a weed map when the data are significantly spatially correlated provided that the variogram of the latent variable is well estimated. Not surprisingly when the data are not or slightly correlated a random design does the job as well. As usual the simulated annealing algorithm is sensitive to the temperature parameter that has to be tuned accordingly to the magnitude and the kind of the data (countings or presence-absence). Several ways to modify the design configuration (all the $p$ points or only one are randomly changed, the modification is random on one or two directions), have been tested but they result in equivalent outcomes.

## 3.2 Case study

Weeds have been measured exhaustively in a field divided in 92 quadrats on a grid $4 \times 23$. $m = 20$ regularly arranged sites are selected in such a way to cover the entire domain. We look for $p = 10$ other points to improve the estimate of the weed map. The procedure is achieved with 1000 iterations for the SAA. Figure 1 shows a) the original data and the 20 points of the initial design, b) the estimated map with 10 sites randomly selected added to the initial design and c) the final design with the optimized 10 additional sites. The RMSE has improved by 15%, and it is worth noticing that the procedure locates the new sites in the area where the weeds are numerous.

# 4 Concluding remarks

Accuracy of predicted infestation levels depends on locations of weed density measurements. We showed that locations leading to accurate predictions can be found using a simulation-based approach with a simulated-annealing step. This approach can be used to map weed infestation in agricultural fields and allows farmers to apply herbicides in highly infested areas only. The performance of the proposed approach depends on the spatial correlation of weed densities and on tuning parameters of the simulated annealing algorithm. An algorithm based on particle filter could also be used within the same framework.
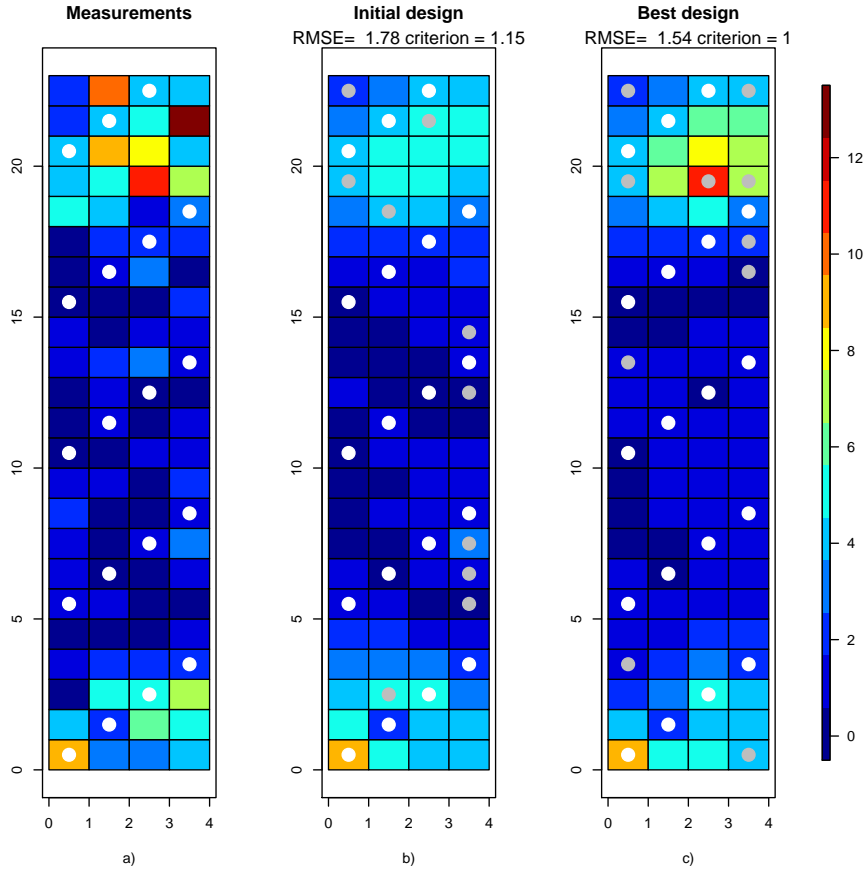
Figure 1: a) original data and initial design b) estimated map with 10 randomly additional sites c) estimated map with 10 optimized additional sites

# References

Amzal B., Bois F.Y., Parent E. (2006). Bayesian-optimal design via interacting particle filter. *Journal of the American Statistical Association* 101, 773-785.

Makowski D., Chauvel B., Munier-Jolain N. (2010). Improving weed population model using a sequential Monte Carlo method. *Weed Research*, vol. 50, no4, pp. 373-382.

Müller P. (1999). Simulation-based optimal design. *Bayesian statistics* 6, 459-474. J.M. Bernado, J.O. Berger, A.P. Dawid, A.F.M. Smith (Eds.). Oxford University Press.