Adaptive Geostatistical Designs

Michael G Chipeta[1,2], Dianne J. Terlouw[3] and Peter J. Diggle[4]

[1]University of Malawi, College of Medicine, Blantyre, Malawi; [2]University of Liverpool, Liverpool, UK; [3]Liverpool School of Tropical Medicine, Liverpool, UK; [4]Lancaster Medical School, Lancaster University, Lancaster, UK;

Background

Objective:
To predict a surface $S(x)$ using noisy measurements of $S(x_i)$ at n locations x_i as efficiently as possible for fixed n; where x_i are a subset of N potential sampling locations.

The design problem:
At which n sites in $\mathcal{X} = \{x_1, \ldots, x_N\}$ in a region of interest $\mathcal{D}$ should we collect the data $Y = \{y(x_i), \ldots, y(x_n)\}$?

Non-adaptive geostatistical designs (NAGD).

In *non-adaptive* geostatistical designs, sampling locations x_i are fixed in advance of any data-collection and cannot be changed.

Random sampling is efficient for parameter estimation, whilst *Regular* sampling is efficient for spatial prediction when model parameters are known. [1]

A good compromise is a *semi-regular* design. We define this as a set of sample points chosen at random subject to the constraint that no two sampled points are less than a prescribed distance δ apart

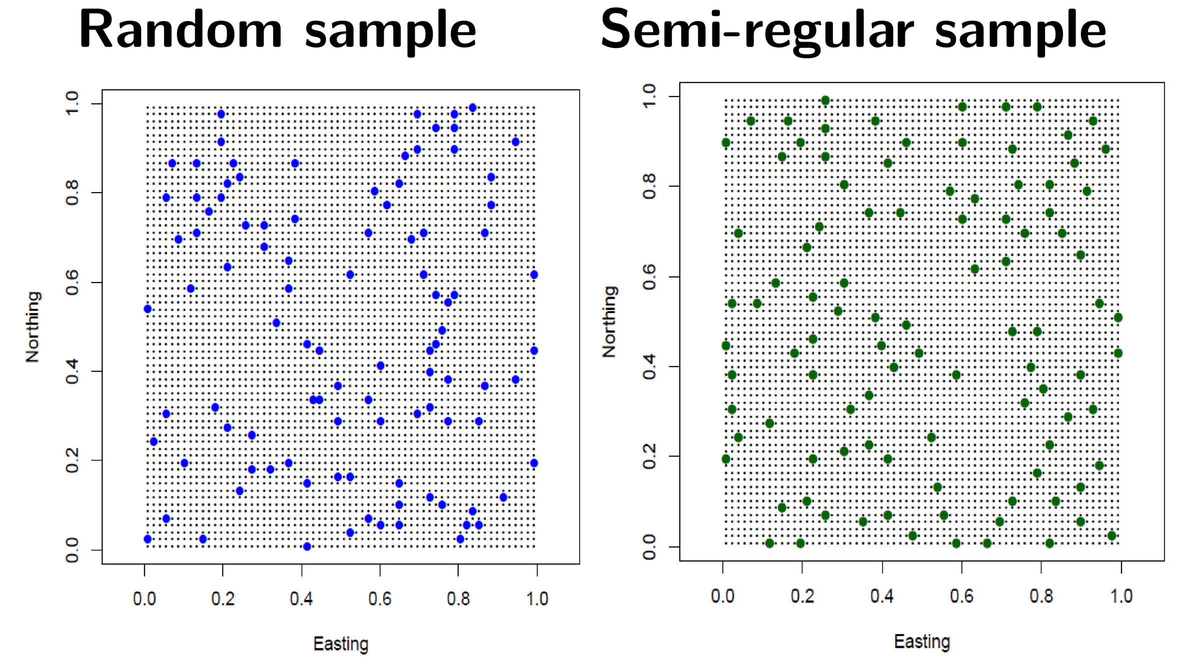


Figure 1: Completely random design, $\delta = 0$ (left panel) and Semi-regular design, $\delta = 0.05$ (right panel). δ is the minimum distance between any two locations, $n = 100$ in each case.

Adaptive geostatistical designs (AGD)

Definitions
Singleton adaptive sampling: locations are chosen sequentially, allowing x_{k+1} to depend on data obtained at locations $x_1, \ldots, x_k$.

Batch adaptive sampling: locations are chosen in batches of size $b > 1$, allowing a new batch, $\{x_{kb+1}, \ldots x_{(k+1)b}\}$, to depend on data obtained at locations $x_1, \ldots, x_{kb}$.

Batch adaptive sampling cannot be more efficient than singleton adaptive sampling, but is almost always more realistic in practice.

New locations are added to the sample when they meet pre-defined criteria, e.g. locations x at which predicted values of $S(x)$ have high prediction variance.

Batch AGD Sampling

Minimum Distance Batch Sampling:
All locations in new batch, $\{x_{kb+1}, \ldots x_{(k+1)b}\}$, should be at least a prescribed distance δ from each other and from all existing $x_1, \ldots, x_{kb}$ locations.

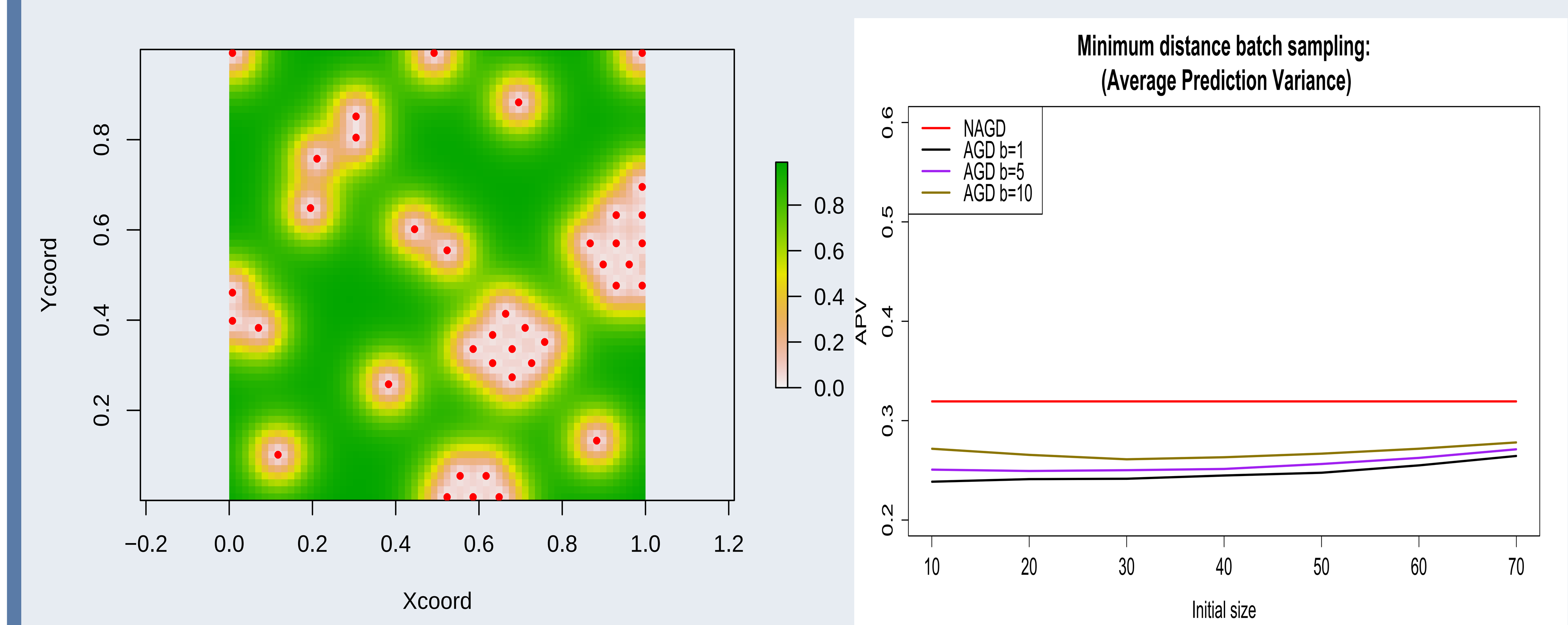
This design results in wide coverage of the study region's spatial extent, which brings benefits in terms of high efficiency (low variance) of spatial predictions.

Fig. 2a shows a sample of size $n = 40$ based on batches of size $b = 10$ with $\delta = 0.03$. Fig. 2b compares average prediction variance for samples of size $n = 100$ using non-adaptive sampling and minimum distance batch sampling with $\delta = 0.03$ and three different batch sizes, $b = 1$, 5 and 10.

Considerations for AGDs

- When the underlying process is not known (which is always the case in practice), adaptive sampling strategies are more efficient than non-adaptive strategies;
- Adaptive sampling strategies gain information by *learning* from the information obtained from previously sampled locations before choosing the next set of locations.

Adaptive and Non-adaptive Design Comparisons



(a) Adaptive sampling (see presentation).

(b) Batch sampling with 3 different batch sizes, $n = 100$.

Figure 2: Minimum distance batch adaptive design in comparison with NAGD

- The practical objective can and should inform the design strategy:
  - minimum distance batch sampling is appropriate for efficient mapping of the surface $S(x)$
  - detection and subsequent evaluation of hotspots would require proegressive concentration of sampling into areas fo high prevalence.

Application: Majete Malaria Project

- Sampling design for a rolling Malaria Indicator Survey (rMIS)
- Identification of hotspots to guide more targeted disease control interventions
- Estimation of effects of environmental, epidemiological and other risk-factors by contrasting areas of high and low prevalence.

References

[1] P. J. Diggle and Ribeiro P. J.
Model-based geostatistics.
Springer, 2007.

Acknowledgements

This work is supported by the Economic and Social Research Council [ES/J500094/1] under the Malawi - Liverpool joint PhD programme.

Contact Information

- Email: `Michael.Chipeta@liverpool.ac.uk`
- Phone: +265 999 412 605

University of Malawi · University of Liverpool · Lancaster University · LSTM · NorthWest Doctoral Training Centre · ESRC Economic & Social Research Council