Adaptive geostatistical design and analysis for prevalence surveys.

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

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

Background

- Can we improve the efficiency of sampling to capture **fine scale malaria heterogeneity** by using **adaptive sampling**?
- Most malaria surveys provide average prevalence estimates at national and regional level. These do not take into account the widely varying level of transmission at local and sub-district level.
- Random sampling methods provide disease prevalence estimate with a level of precision around that estimate.
- Malaria shows small scale variation - Fig. **1**, Chikwawa study site; describing such heterogeneity can guide targeted intervention strategies.

Non-adaptive Geostatistical Designs

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 ***semi-inhibitory*** design - Figures 2 and 3

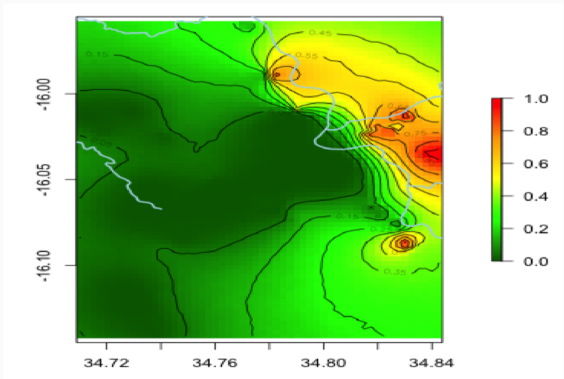

Fig. 1: Prevalence hotspots, source: Giorgi *et. al.*, 2014.

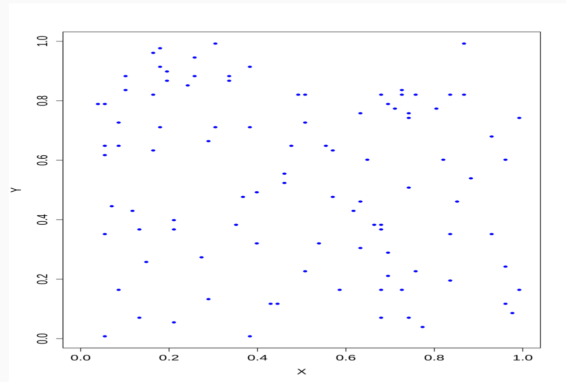

Fig. 2: Random design, $\delta = 0$; $n = 100$.

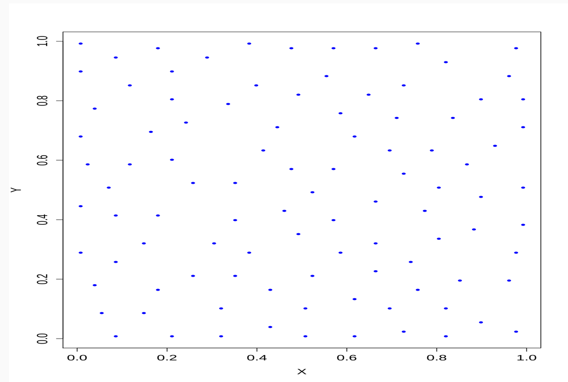

Fig. 3: Inhibitory design, $\delta = 0.08$; $n = 100$.

Adaptive Geostatistical Designs (AGD)

- New locations are added to the sample if they meet defined criteria, e.g. locations x^* at which predicted values of $S(x)$ have high prediction variance.
- We performed simulation studies to compare the efficiency of specific adaptive and non-adaptive designs in terms of predictive efficiency.
- *Singleton adaptive sampling*: locations are chosen sequentially, allowing x_{k+1} to depend on data obtained at locations $x_1, \ldots, x_k$; whereas *Batch adaptive sampling*: locations are chosen in batches (clusters) of size $b > 1$, allowing a new cluster, $\{x_{kb+1}, \ldots x_{(k+1)b}\}$, to depend on data obtained at locations $x_1, \ldots, x_{kb}$.
- Using **Minimum Distance Batch Adaptive Sampling**, we allow locations in a new batch to be at least a prescribed distance δ from each other and from all existing $x_1, \ldots, x_{kb}$ locations.
- This design ensures wide coverage of the study region's spatial extent, which brings benefits in terms of high efficiency (low variance) of spatial predictions.

Materials and methods

- We use data from the initial wave of sampling from large-scale malaria transmission reduction study currently being implemented in Majete wildlife reserve in Malawi to demonstrate how we are applying AGDs.
- We fit a standard geostatistical model for prevalence data:

$$\log[p(x_i)/\{1 - p(x_i)\}] = d(x_i)'\beta + S(x_i) \quad (1)$$

Main results and application

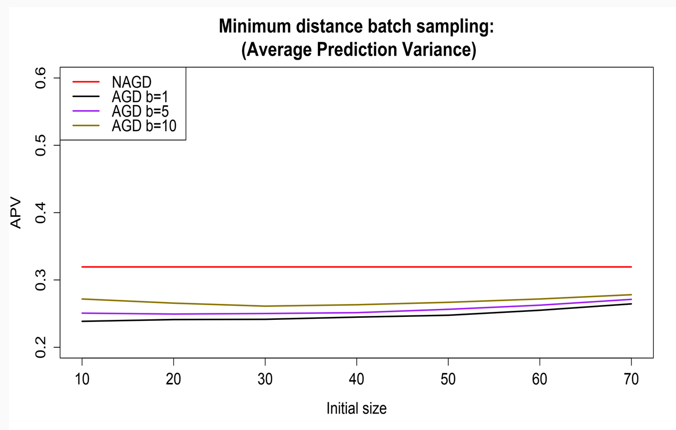

Fig. 4: Minimum distance batch adaptive design with 3 different batch sizes, in comparison with NAGD, $n = 100$ in each case.

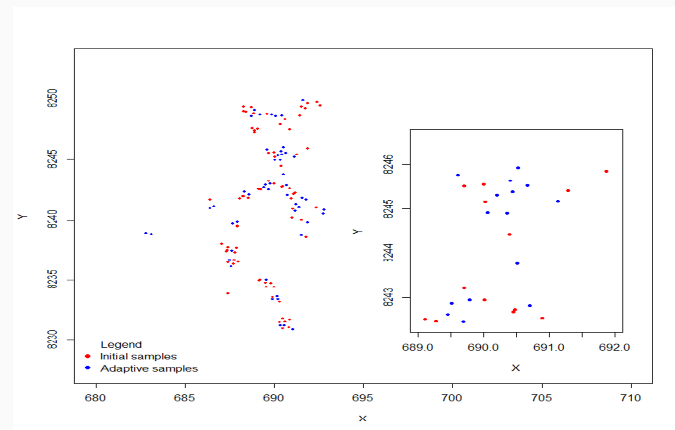

Fig. 5: Initial inhibitory sampling design locations (red dots) and adaptive sampling design locations (blue dots). Inset shows a subset of locations.

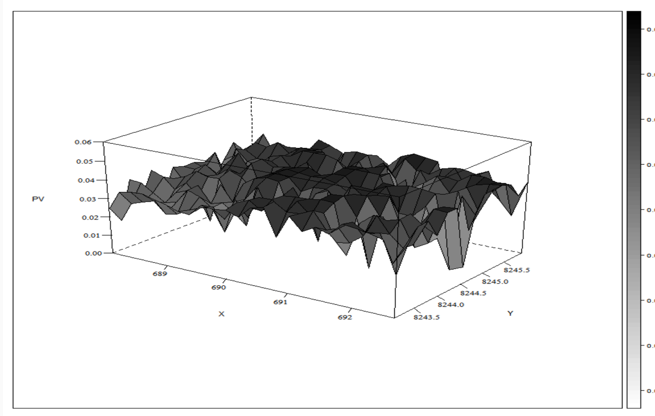

Fig. 6: Prediction variance surface for the inset sub-region from Figure 5

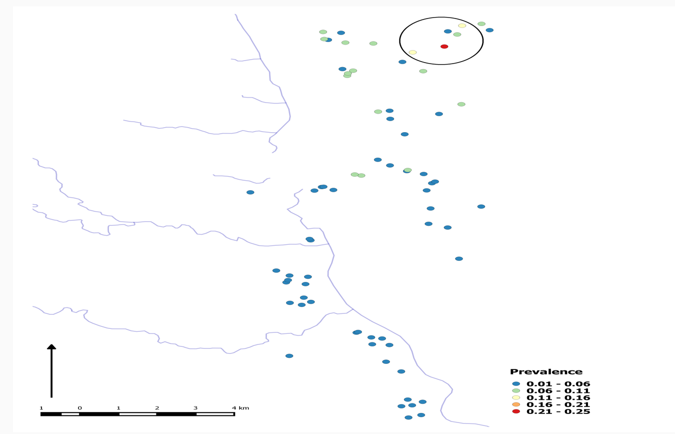

Fig. 7: Prevalence estimates at observed locations (NDVI, Elevation and Interaction).

We apply AGD sampling to a rolling Malaria Indicator Survey (rMIS) around Majete Wildlife Reserve perimeter

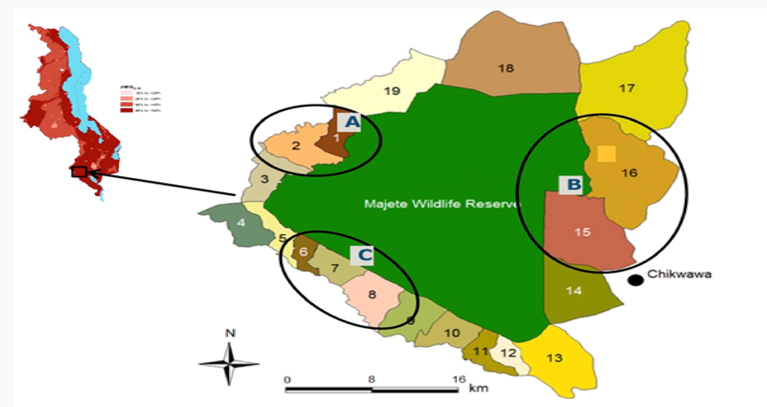

Fig. 8: Focal areas: **A, B, and C** around the Majete Wildlife Reserve perimeter where rMIS is being implemented.

Discussion and Conclusion

- Adaptive sampling is more efficient than non-adaptive sampling.
- Increasing the batch size is associated with a small loss of efficiency in predictive performance.
- Adaptive sampling allows effective detection and subsequent evaluation of hotspots as it results in progressive concentration of sampling into areas of high disease prevalence.
- Minimum distance batch adaptive sampling results in more efficient mapping of malaria disease prevalence.

References

Diggle, P. J. and Ribeiro, P. J. *Model-based geostatistics.* Springer, 2007.