Model-based Geostatistics: geospatial statistical methods for public health applications

Peter J Diggle and Emanuele Giorgi

CHICAS, Lancaster University

Geostatistical design

Geostatistical design

- **Non-adaptive**

  1. **choose complete set of measurement locations,** $\mathcal{X} = \{x_i \in A : i = 1, \ldots, n\}$
  2. **collect measurement data Y**

- **Adaptive**

  1. **choose initial set of measurement locations,** $\mathcal{X}_1 = \{x_i \in A : i = 1, \ldots, b\}$
  2. **collect initial measurement data Y_1**
  3. **analyse initial data and use results to inform choice of** $\mathcal{X}_2 = \{x_i \in A : i = b + 1, \ldots, 2b\}$
  4. **continue until design is complete**

In adaptive design, b is called the batch size

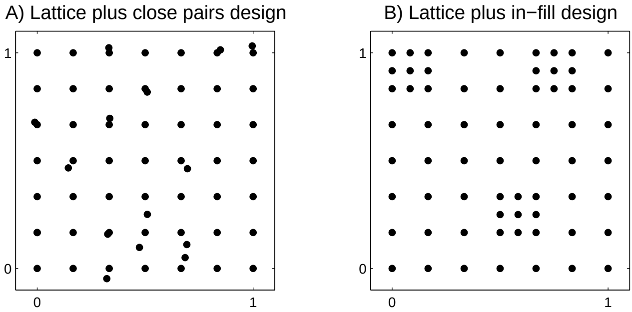
Naive design folklore

- **Spatial correlation decreases with increasing distance.**
- **Therefore, close pairs of points are wasteful.**
- **Therefore, spatially regular designs are a good thing.**

Less naive design folklore

- **Spatial correlation decreases with increasing distance.**
- **Therefore, close pairs of points are wasteful if you know the correct model.**
- **But in practice you need to estimate unknown model parameters.**
- **And to estimate model parameters, you need your design to include a range of small inter-point distances.**
- **Therefore, spatially regular designs should be tempered by the inclusion of some close pairs of points.**

Examples of lattice-based designs

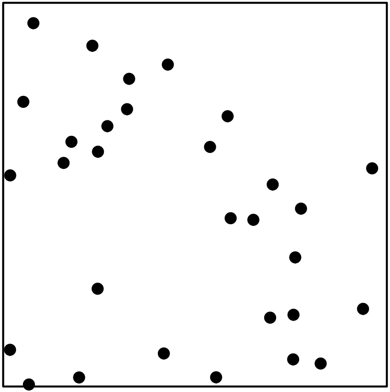


Limitations of lattice-based designs

- in some applications, available sampling locations are pre-specified, eg farms or villages
- absence of a probability sampling framework

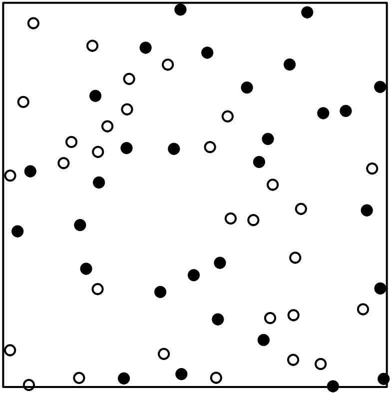
Lattice-free spatially regular sampling designs

Sample at random subject to a minimum distance constraint



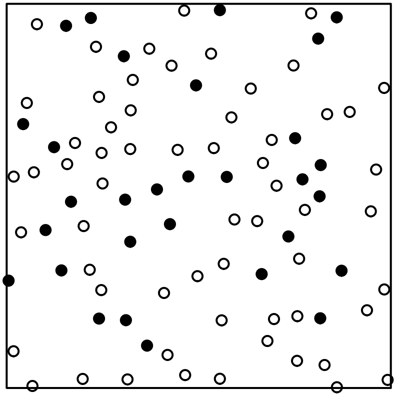
Lattice-free spatially regular sampling designs

Sample at random subject to a minimum distance constraint



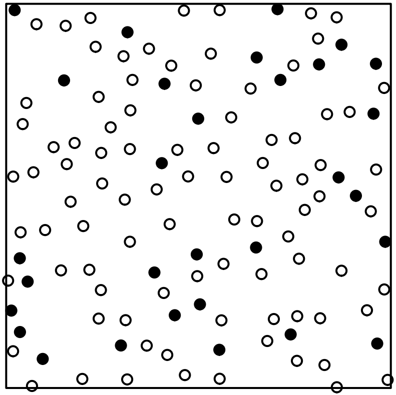
Lattice-free spatially regular sampling designs

Sample at random subject to a minimum distance constraint



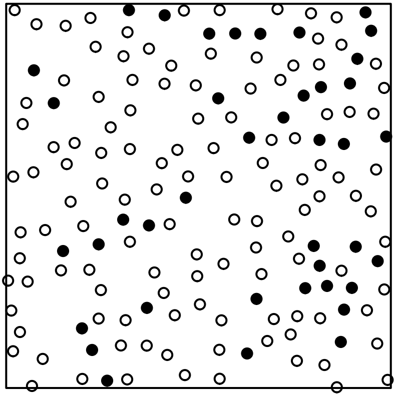
Lattice-free spatially regular sampling designs

Sample at random subject to a minimum distance constraint



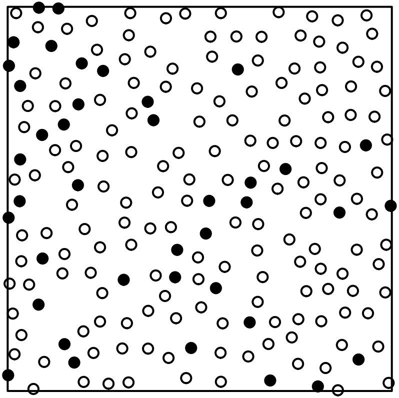
Lattice-free spatially regular sampling designs

Sample at random subject to a minimum distance constraint



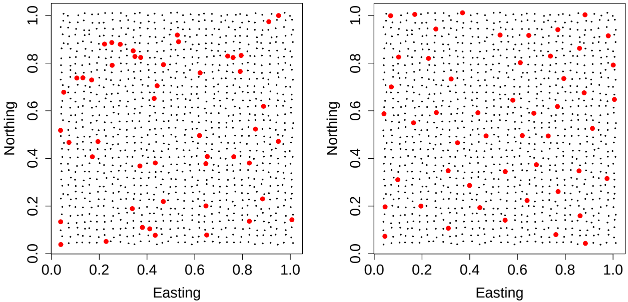
Lattice-free spatially regular sampling designs

Sample at random subject to a minimum distance constraint



Spatially regular sampling from a pre-specified set of locations

- **again achieve spatial regularity by sampling at random subject to a minimum distance constraint**



- **Tempering spatial regularity with close pairs still a good idea**

R code for regular sampling designs

```
discrete.sample<-function(xy.all,n,delta,k=0) {
#
# Arguments
# xy.all:  set of potential sample locations
# n:  size of required sample
# delta:  minimum distance between any two locations in
preliminary sample
# k:  number of close pairs (must be between 0 and n/2)
# #
# Result
# array of dimension n by 2 containing the final sampled
locations
#
...
}
```

Example: sampling in a regular plus close pairs design

```
set.seed(16713); N<-500; par(pty="s",mfrow=c(1,2))
xy<-cbind(runif(N),runif(N))
#
# generate spatially random sample of 50 locations
#
plot(xy[,1],xy[,2],pch=19,cex=0.25,xlab="Easting",
ylab="Northing",cex.lab=1,cex.axis=1,cex.main=1)
xy.sample<-xy[sample(1:dim(xy)[1],50,replace=FALSE),]
points(xy.sample[,1],xy.sample[,2],pch=19,col="red")
#
# generate spatially regular sample with 10 close pairs
#
plot(xy[,1],xy[,2],pch=19,cex=0.25,xlab="Easting",
ylab="Northing",cex.lab=1,cex.axis=1,cex.main=1)
xy.sample<-discrete.sample(xy,50,0.1,10)
points(xy.sample[,1],xy.sample[,2],pch=19,col="red")
```

A class of adaptive designs

Singleton adaptive

- Suppose we have measurements $\mathbf{Y_0}$ at a set of locations $\mathcal{X}_0$ and we are allowed to choose one additional location x
- Calculate the predictive variance of S(x),

$$\text{PV}(x) = \text{Var}(S(x)|\mathbf{Y_0}),$$

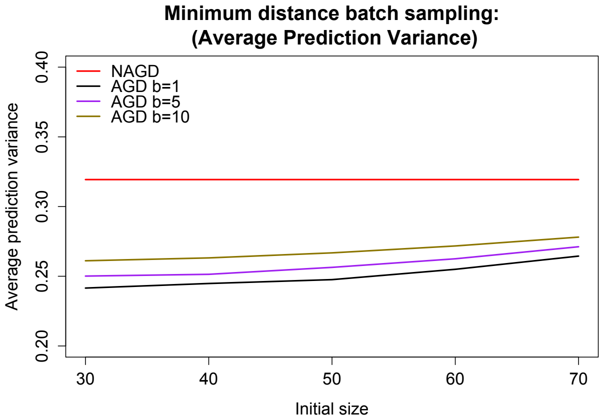
add the available location x with the biggest value of PV(x)

Batch adaptive

- Now suppose we are allowed to choose b additional locations x
- Add b available locations with biggest values of PV(x)?
- A bad idea...typically gives tight cluster of additional locations
- To avoid this, we again impose a minimum distance constraint

Non-adaptive (NAGD) vs minimum distance batch adaptive (AGD) sampling

- $n_0 = 30$ initial locations, $n = 100$ total locations
- minimum distance $\delta = 0.03$, batch sizes $b = 1, 5, 10$



Further remarks on geostatistical design

1. **Conceptually more complex problems include:**
   1.1 **design when some sub-areas are more interesting than others;**
   1.2 **design for best prediction of non-linear functionals of $S(\cdot)$;**
   1.3 **spatio-temporal: predicting a moving target**
2. **Theoretically optimal designs may not be practicable**
3. **A more realistic goal is to suggest constructions for good, general-purpose designs.**