Lab 8: Prevalence mapping

Peter Diggle & Emanuele Giorgi

Lancaster Medical School, Lancaster University, Lancaster, UK

Model-based geostatistics: geospatial statistical methods for public health applications, 5-9 October 2015

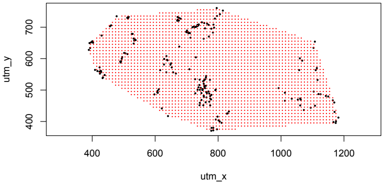
Overview

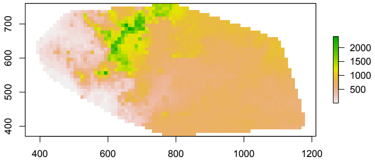
- Prevalence mapping using linear models.
- Prevalence mapping using binomial models.
- Computation of multivariate properties of a prevalence surface.

R packages: `PrevMap`, `raster`, `rgdal`, `splancs`.

Prediction grid and predictors extraction

```
> library(PrevMap)
> load("Loaloa_fits.RData")
> str(lm.geo.fit)
List of 9
 $ estimate  : ...
 $ covariance: ...
 $ log.lik   : ...
 $ y         : ...
 $ D         : ...
 $ coords    :  ...
 $ method    : ...
 $ kappa     : ...
 $ call      :  ...
>
> poly <- lm.geo.fit$coords[
+              chull(lm.geo.fit$coords),]
> library(splancs)
> grid.pred <- gridpts(poly,xs=10,ys=10)
>
> par(mfrow=c(2,1),mar=c(4,4,2,2))
> plot(lm.geo.fit$coords,asp=1,cex=0.5,pch=20)
> points(grid.pred,col=2,pch=20,cex=0.2)
>
> elev <- raster("CMR_NGA_alt.tif")
> elev <- projectRaster(elev,crs="+init=epsg:32632")
>
> predictors <- data.frame(ELEVATION=extract(elev,grid.pred*1000))
> plot(rasterFromXYZ(cbind(grid.pred,predictors$ELEVATION)))
```

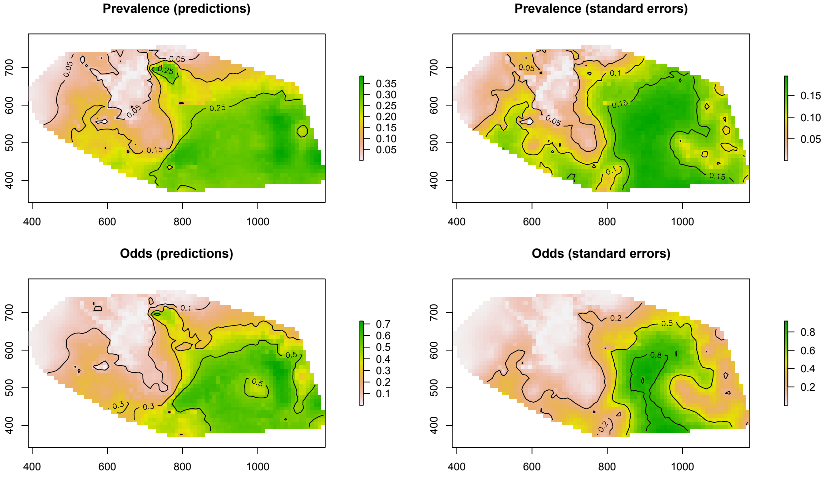




Prevalence mapping using linear models (1)

```
> pred.linear <- spatial.pred.linear.MLE(
+                lm.geo.fit, grid.pred=grid.pred,
+                predictors=predictors, type="joint",
+                scale.predictions=c("prevalence","odds"),
+                standard.errors=TRUE,
+                thresholds=0.2,
+                scale.thresholds="prevalence")
Type of prevalence predictions: joint (this step might be demanding)
Spatial predictions: prevalence
Spatial predictions: odds
>
> par(mfrow=c(2,2),mar=c(2,4,4,6))
> plot(pred.linear,type="prevalence",summary="predictions",
+      main="Prevalence (predictions)")
> contour(pred.linear,type="prevalence",summary="predictions",
+         add=TRUE,levels=c(0.05,0.15,0.25))
>
> plot(pred.linear,type="prevalence",summary="standard.errors",
+      main="Prevalence (standard errors)")
> contour(pred.linear,type="prevalence",summary="standard.errors",
+         add=TRUE,levels=c(0.05,0.1,0.15))
>
> plot(pred.linear,type="odds",summary="predictions",
+      main="Odds (predictions)")
> contour(pred.linear,type="odds",summary="predictions",
+         add=TRUE,levels=c(0.1,0.3,0.5))
>
> plot(pred.linear,type="odds",summary="standard.errors",
+      main="Odds (standard errors)")
> contour(pred.linear,type="odds",summary="standard.errors",
+         add=TRUE,levels=c(0.2,0.5,0.8))
```

Prevalence mapping using linear models (2)



Prevalence (predictions)

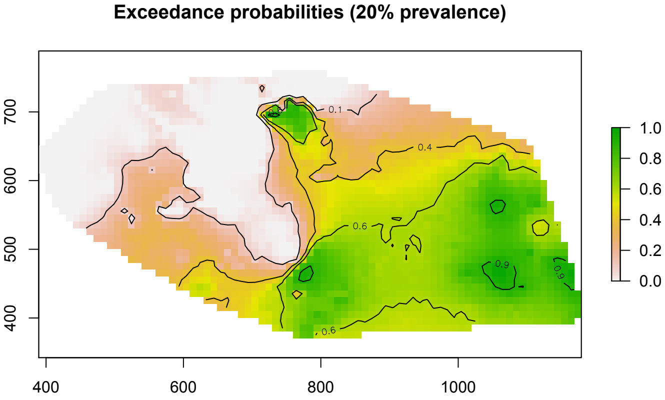
Prevalence (standard errors)

Odds (predictions)

Odds (standard errors)

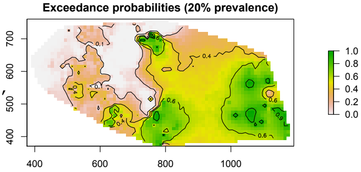
Prevalence mapping using linear models (3)

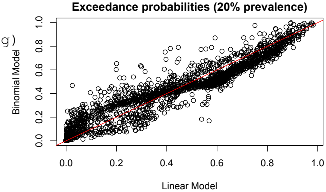
```
> plot(pred.linear,summary="exceedance.prob",
+      main="Exceedance probabilities (20% prevalence)",
+      zlim=c(0,1))
> contour(pred.linear,summary="exceedance.prob",
+         add=TRUE,levels=c(0.1,0.4,0.6,0.9))
```



Prevalence mapping using binomial models

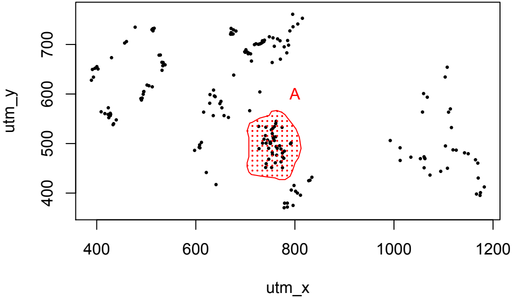
```
> c.mcmc <- control.mcmc.MCML(n.sim=10000,
+           burnin=2000,thin=8,
+           h=1.65/(nrow(loaloa)^(1/6)))
> pred.bin.MCML <- spatial.pred.binomial.MCML(bin.geo.fit,
+                  grid.pred=grid.pred,predictors=predictors,
+                  control.mcmc=c.mcmc,type="joint",
+                  scale.predictions="prevalence",
+                  standard.errors=TRUE,
+                  thresholds=0.2,
+                  scale.thresholds="prevalence")
Conditional simulation (burnin=2000, thin=8):
Iteration 10000 out of 10000
Type of predictions:  joint  (this step might be demanding)
Spatial predictions: prevalence
>
> par(mfrow=c(2,1),mar=c(4,4,2,4))
> plot(pred.bin.MCML,summary="exceedance.prob",
+      main="Exceedance probabilities (20% prevalence)",
+      zlim=c(0,1))
> contour(pred.bin.MCML,summary="exceedance.prob",
+         add=TRUE,levels=c(0.1,0.4,0.6,0.9))
> plot(pred.linear$exceedance.prob,
+      pred.bin.MCML$exceedance.prob,
+      xlab="Linear Model", ylab="Binomial Model",
+      main="Exceedance probabilities (20% prevalence)")
> abline(0,1,col=2)
```





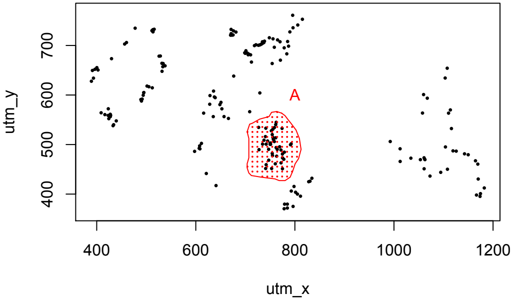
Computing multivariate properties (1)

- $M(A) = \max_{x \in A} P(x)$.



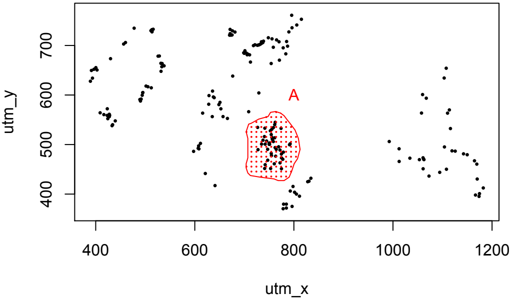
Computing multivariate properties (1)

- $M(A) = \max_{x \in A} P(x)$.
- $M(A) \approx \max_{g \in A} P(g)$



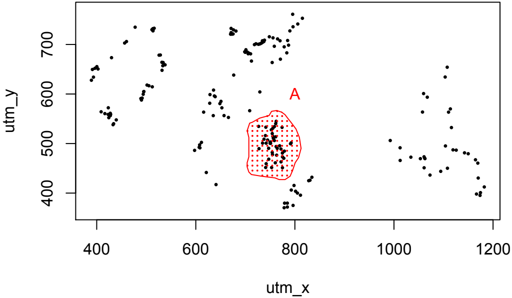
Computing multivariate properties (1)

- $M(A) = \max_{x \in A} P(x)$.
- $M(A) \approx \max_{g \in A} P(g)$
- $\hat{M}(A) = \max_{g \in A} \hat{P}(g)$



Computing multivariate properties (1)

- $M(A) = \max_{x \in A} P(x).$
- $M(A) \approx \max_{g \in A} P(g)$
- $\hat{M}(A) = \max_{g \in A} \hat{P}(g)$

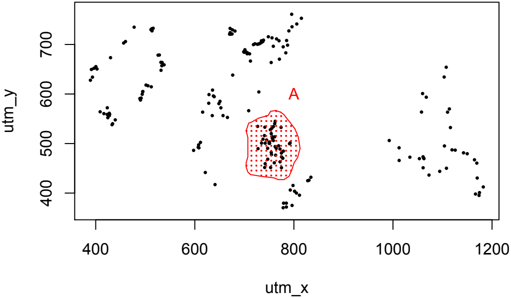


The mean prevalence

- $m(A) = \frac{1}{|A|} \int_A P(x)\, dx$

Computing multivariate properties (1)

- $M(A) = \max_{x \in A} P(x)$.
- $M(A) \approx \max_{g \in A} P(g)$
- $\hat{M}(A) = \max_{g \in A} \hat{P}(g)$

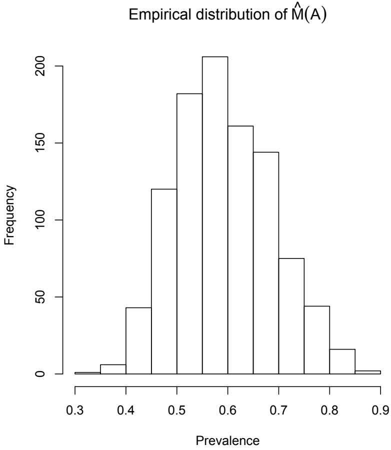


The mean prevalence

- $m(A) = \frac{1}{|A|} \int_A P(x)\, dx$
- $m(A) \approx \frac{1}{n_g} \sum_{g \in A} P(g)$
- $\hat{m}(A) = \frac{1}{n_g} \sum_{g \in A} \hat{P}(g)$

Computing multivariate properties (2)

```
> load("areaA.RData")
> plot(lm.geo.fit$coords,pch=20,cex=0.5,asp=1)
> lines(A,type="l",col=2)
> text(800,600,"A",col=2)
> library(splancs)
> indA <- which(inout(grid.pred,A))
>
> points(grid.pred[indA,],pch=20,col=2,cex=0.2)
>
> prev.samples <- exp(pred.bin.MCML$samples)/
+                 (1+exp(pred.bin.MCML$samples))
>
> dim(grid.pred)
[1] 2264    2
> dim(prev.samples)
[1] 2264 1000
> n.samples <- dim(prev.samples)[2]
> max.prev.A.samples <- rep(NA,n.samples)
> for(i in 1:n.samples) {
+  max.prev.A.samples[i] <- max(prev.samples[indA,i])
+ }
> hist(max.prev.A.samples,xlab="Prevalence",
+      main=
+      expression("Empirical distribution of"~hat(M)(A)))
> mean(max.prev.A.samples)
[1] 0.5927877
> sd(max.prev.A.samples)
[1] 0.09420895
> quantile(max.prev.A.samples,c(0.025,0.975))
     2.5%     97.5%
0.4323474 0.7945100
```



Exercise

River-blindness prevalence mapping in Liberia

Load all the data by running the following code.

```
load("LiberiaRemo_fits.RData")
str(lm.geo.fit) # Linear geostatistical model fit
str(fit.MCML) # Binomial geostatistical model fit
load("Liberia_grid_pred.RData")
str(grid.pred)  # Prediction grid (10x10 km)
load("Liberia_elevation.RData")
str(ELEVATION.pred) # Elevation at prediction locations
load("GB_mask.RData")
str(ind.GrandBassa) # Logical vector indicating whether each prediction location
                    # is in (TRUE) or out (FALSE) Grand Bassa county in Liberia
```

- Predict prevalence at the locations in `grid.pred` using both the linear a binomial models fits. Compare the the prevalence estimates and exceedance probabilities from the two models (use a prevalence threshold of 20$\%$).
- Compute the mean and the 95% confidence interval of the predictive distribution of the mean prevalence in Grand Bassa county.