

Ensayo de métodos de monitorización acústica en el Jardín Botánico de Maracaibo, municipio San Francisco, estado Zulia

JR Ferrer-Paris
Laboratorio de Ecología Espacial, Centro de Estudios Botánicos y
Agroforestales
Instituto Venezolano de Investigaciones Científicas

Borrador
Versión de 18 de febrero de 2015

A quien pueda interesar

Documento para uso exclusivo del personal científico y estudiantes del Laboratorio de Ecología Espacial, no difundir sin autorización expresa del autor. Se presenta un análisis estadístico basado en un muestreo de campo realizado entre el 9 y 10 de octubre de 2014 con la participación de Leinny González, Cecilia Lozano, Junior Larreal, Liset Zambrano, Thayma Vargas, Arlene Cardozo, José Soto, Ada Y. Sánchez Mercado y J.R. Ferrer Paris. Este documento es generado utilizando las funciones de **Sweave** desde una sesión de *R* (R Development Core Team, 2010), por tanto todas las tablas y figuras se generan y actualizan automáticamente a partir de los datos suministrados. Para acceso al código fuente en *R* y los archivos de datos contacte al autor. Dentro de *R* utilizamos los paquetes *unmarked* (?); *Rcpp* (?); *lattice* (?); *reshape* (?); *gdata* (?); *gstat* (Pebesma, 2004); *rgdal* (Keitt *et al.*, 2012); *plotrix* (Lemon, 2006); *RColorBrewer* (Neuwirth, 2011); *raster* (Hijmans & van Etten, 2012); *sp* (Pebesma & Bivand, 2005); *chron* (James & Hornik, 2011).

1. Puntos de muestreo

Los datos de las planillas están en un archivo en *Google Drive*, los bajamos en formato csv y los leemos en una sesión de *R*.

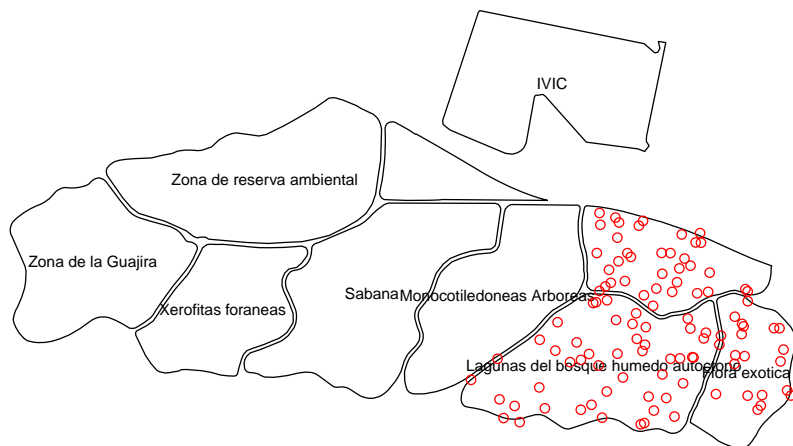
```
> Puntos <- read.csv("~/CEBA/data/JardinBotanicoMaracaibo/20150109_MA_Puntos.csv")
> Muestreo <- read.csv("~/CEBA/data/JardinBotanicoMaracaibo/20150109_MA_Muestreo.csv")
> Grabaciones <- read.csv("~/CEBA/data/JardinBotanicoMaracaibo/20150109_MA_Grabaciones.csv")
> Puntos$y <- ifelse(is.na(Puntos$Latitud.nueva), Puntos$Latitude, Puntos$Latitud.nueva)
> Puntos$x <- ifelse(is.na(Puntos$Longitud.nueva), Puntos$Longitude, Puntos$Longitud.nueva)
> dts <- merge(Grabaciones, Muestreo,
+             by.x=c("Fecha.muestreo", "Punto", "Micrófono"),
+             by.y=c("Fecha", "Punto", "Micrófono"))
> dts$Archivo <- with(dts, sprintf("%s_%s_%s", Punto, Fecha.muestreo, sub(":", "", Inicio)))
>
```

Vamos a empezar con la lista de puntos caracterizados. Debemos calcular la incertidumbre de la georeferencia, que en este caso es la suma de la distancia al punto (DAP), el error de posicionamiento o error de GPS (*Error.GPS*) y la extensión del punto del muestreo (*Extent*) que equivale a la distancia máxima de las fuentes de sonido que pueden ser detectadas con el micrófono, según Arlene esta es de aproximadamente 10 m. La incertidumbre (*Incertidumbre*) está expresada en metros.

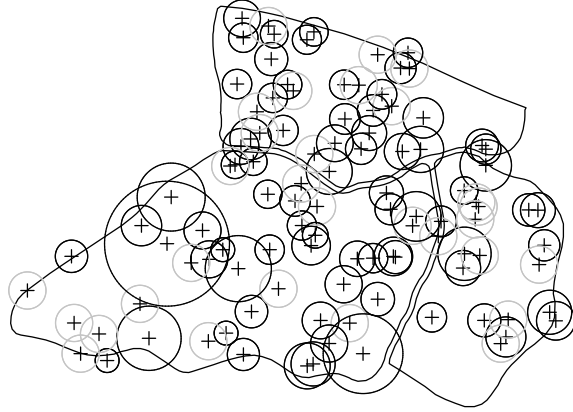
```
> Puntos$DAP <- ifelse(is.na(Puntos$Longitud.nueva),
+                     as.numeric(sub(" ", ".", Puntos$Distancia.al.punto)),
+                     as.character(Puntos$Distancia.al.punto)),
+                     0)
> Puntos$DAP <- ifelse(is.na(Puntos$DAP), mean(Puntos$DAP, na.rm=T), Puntos$DAP)
> Puntos$error.GPS <- Puntos$EPE
> Puntos$error.GPS <- ifelse(is.na(Puntos$error.GPS), mean(Puntos$error.GPS, na.rm=T), Puntos$error.GPS)
> Puntos$Extent <- 10
> Puntos$Incertidumbre <- Puntos$DAP + Puntos$error.GPS + Puntos$Extent
>
```

Para visualizar estos puntos en un mapa debemos definir sus atributos espaciales. Las funciones *coordinates* y *proj4string* del paquete *sp* permiten definir las coordenadas y proyección geográfica de los datos. Añadimos un mapa de la zonificación del jardín para observar su ubicación relativa en la zona oeste del JBM:

```
> JBM.ll <- shapefile("~/CEBA/data/JardinBotanicoMaracaibo/ZonificacionJardin.shp")
> plot(JBM.ll)
> text(JBM.ll, "nombre", cex=.75)
> points(Puntos, col=2)
```



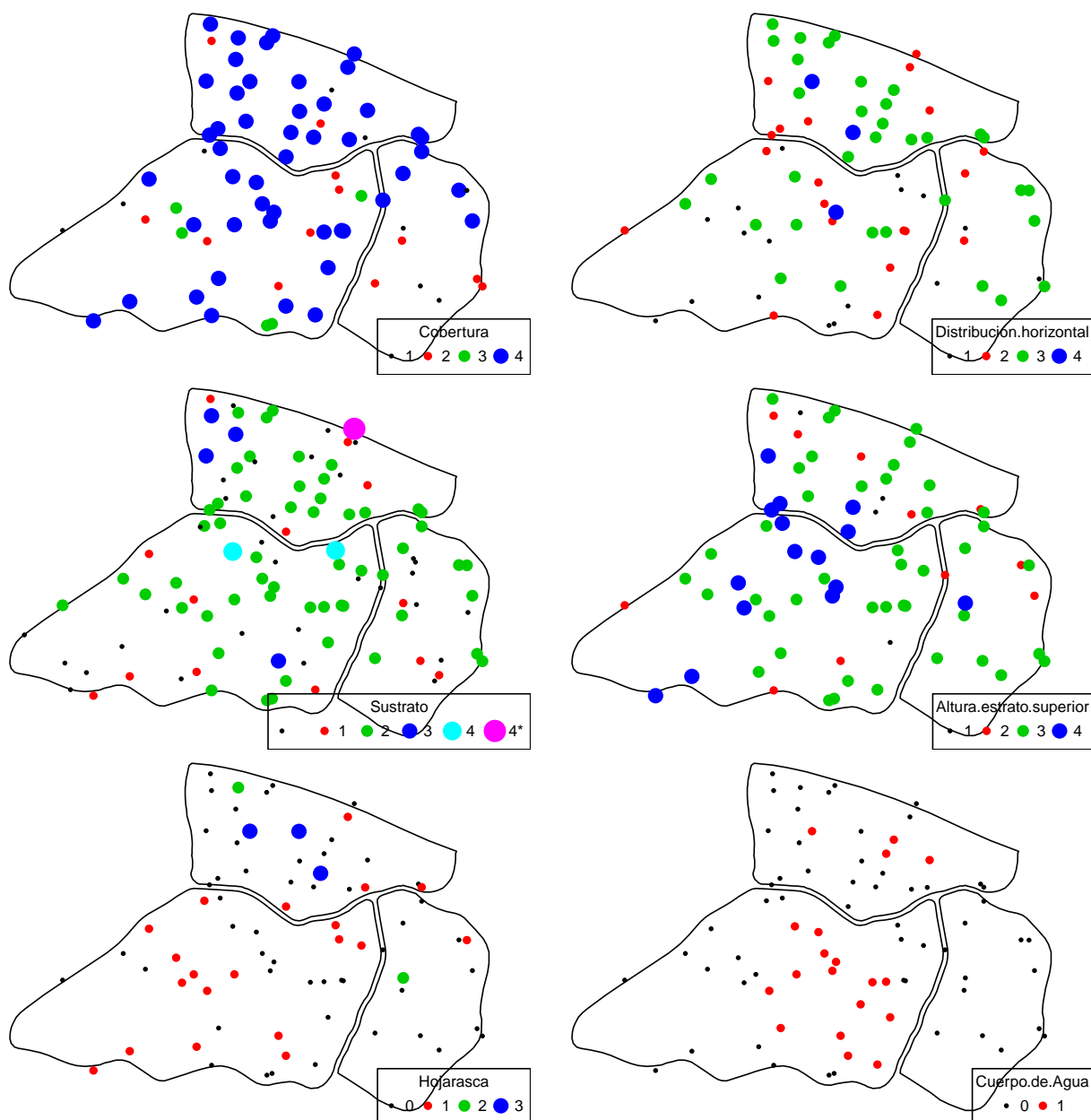
Para visualizar la incertidumbre (expresada en metros), vamos primero a proyectar los datos a UTM (función `spTransform`), hacer el gráfico con la función `plot`, y añadir círculos proporcionales a la incertidumbre con la función `draw.arc` del paquete *plotrix*. Los círculos oscuros fueron puntos caracterizados durante la actividad de campo, y los grises fueron los puntos que no pudimos caracterizar (columna `Anotador` con valor “No muestreado”).



2. Caracterización de los puntos

Para visualizar las variables caracterizadas usamos las opciones de la función `points` para modificar el tamaño y color de los puntos según la variable de interés. Posteriormente se pueden simplificar las categorías para hacer más sencillo el análisis.

En el caso de la cobertura hay cuatro niveles de la variable. En la columna de sustrato hay un punto caracterizado como “4*”. En hojarasca la escala empieza con cero. Para Distribución.horizontal hay cuatro categorías. Altura.estrato.superior En Cuerpo.de.Agua solo hay dos categorías.



3. Predicción espacial

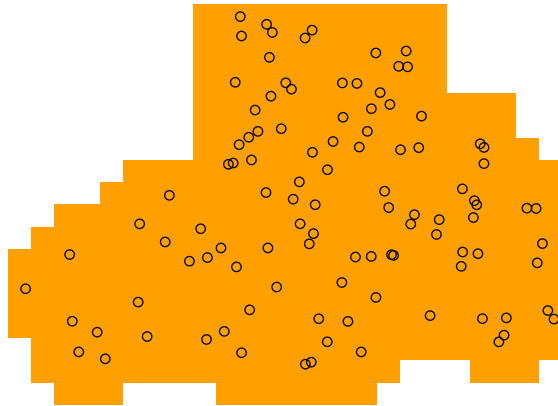
A partir de los puntos caracterizados realizamos una predicción espacial para todo el área de estudio. Primero generamos una capa espacial alrededor de los puntos muestreados:

```
> rnull <- raster(extent(JBM.xy[JBM.xy@data$id %in% c(7,8,11),]),
+                 nrow=18, ncol=24)
> mask <- distanceFromPoints(rnull,pts)<50
> ##coordinates(mask)[values(mask)==1,]
>
```

```

> JBM.grd <- SpatialPixelsDataFrame(points=xyFromCell(mask,1:ncell(mask))[values(mask)
+                                     data=data.frame(values(mask)[values(mask)==1]),
+                                     proj4string=JBM.xy@proj4string)
> image(JBM.grd)
> points(pts)

```



Para la predicción espacial vamos a utilizar el método de “krigging” de una variable indicadora. Vamos a usar las funciones en el paquete *gstat*.

```

> require(gstat)

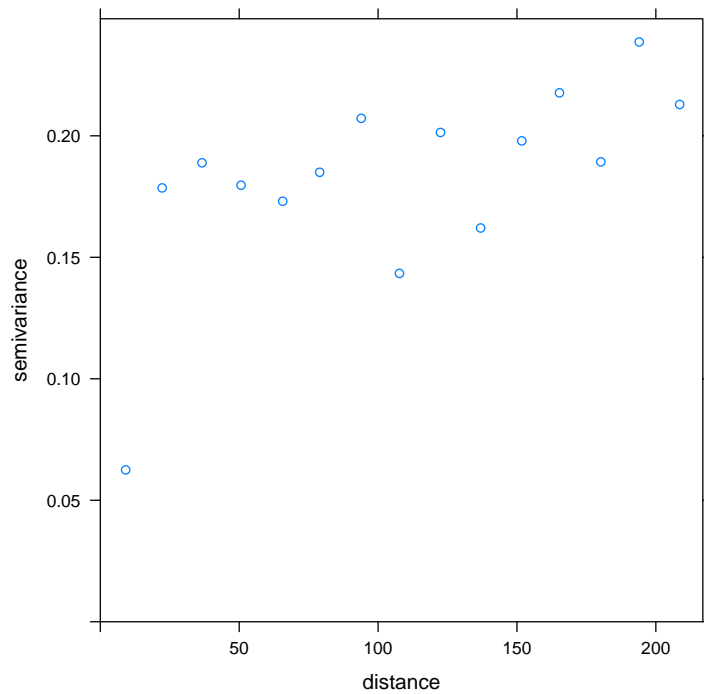
```

Vamos a simplificar nuestras variables para considerar solamente dos categorías. Por ejemplo en la Cobertura, vamos a considerar valores mayores a dos. Aplicamos la función `variogram` para observar la dependencia espacial de los datos en forma de un semi-variograma. Usamos un filtro (`ss`) para excluir las observaciones sin valores.

```

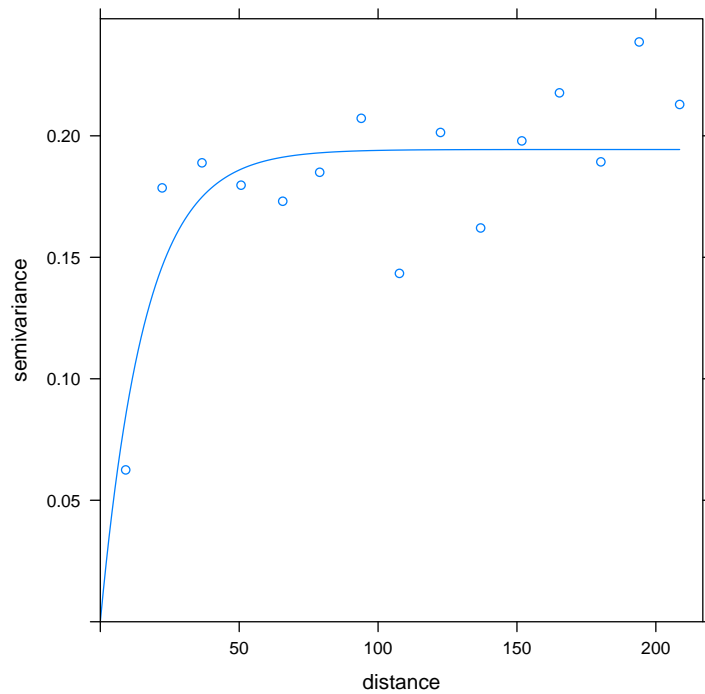
> ss <- !is.na(pts@data$Cobertura)
> v = variogram(I(Cobertura > 2)~1,pts[ss,])
> plot(v)

```



Observamos que la semi-varianza es baja cuando la distancia es menor a 20 metros, y luego se mantiene alrededor de 0,15 y 0,25. Para hacer la predicción espacial necesitamos definir la estructura espacial en terminos de un modelo. Aplicamos las funciones `fit.variogram` y `vgm` para ajustar el modelo a partir de unos parámetros iniciales. En este caso utilizamos un modelo exponencial.

```
> vm = fit.variogram(v, vgm(1, "Exp", 70, .1))  
> plot(v,vm)
```



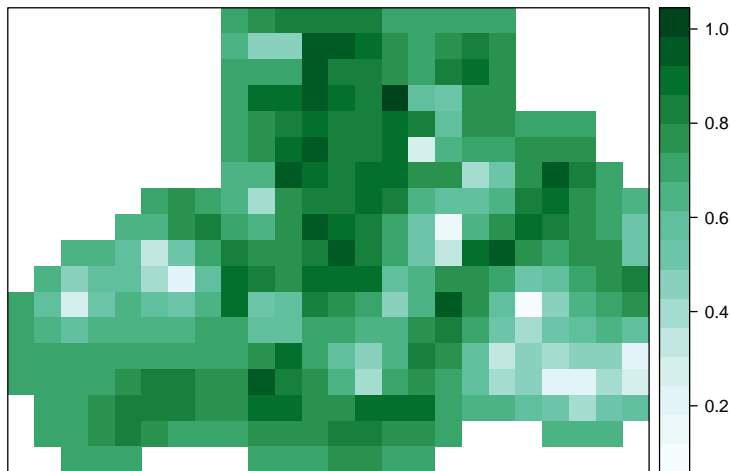
Ahora podemos utilizar la función `krige` para la interpolación espacial basada en “kriging” de una variable indicadora según el modelo de variograma estimado anteriormente.

```
> ik = krige(I(Cobertura > 2)~1, pts[ss,], JBM.grd, vm)
```

```
[using ordinary kriging]
```

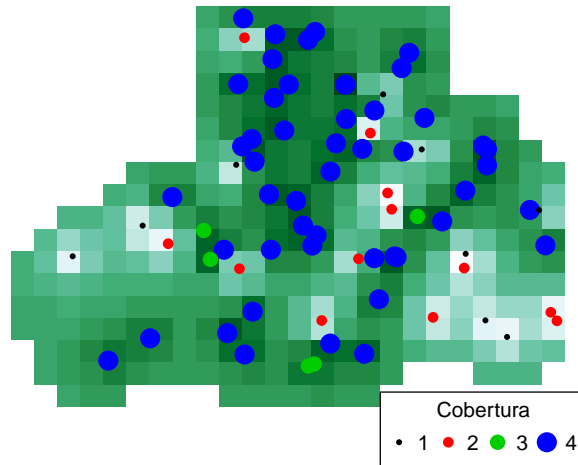
```
> colores <- colorRampPalette(brewer.pal(9, "BuGn"))
```

```
> print(spplot(ik[1], col.regions=colores(30)))
```

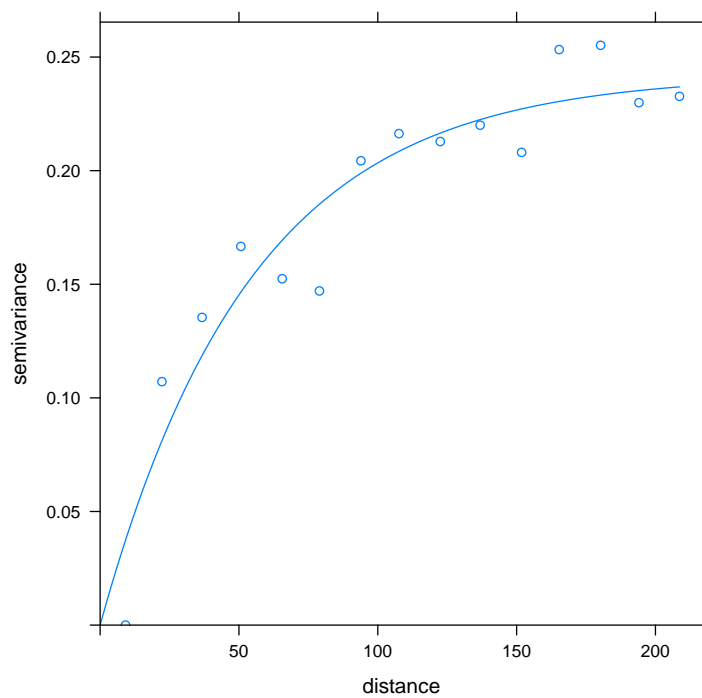



El resultado indica la probabilidad de observar valores de cobertura mayores a dos según los puntos caracterizados en campo. En el siguiente gráfico podemos ver la correspondencia entre los valores observados y los predichos. Notese que la predicción en lugares alejados de los sitios muestreados toma un valor promedio.

```
> image(ik[1],col=colores(30))
> points(pts,col=pts@data$Cobertura,cex=pts@data$Cobertura/2,pch=19)
> legend("bottomright",legend=1:4,col=1:4,pch=19,pt.cex=c(1:4)/2,title="Cobertura",
```

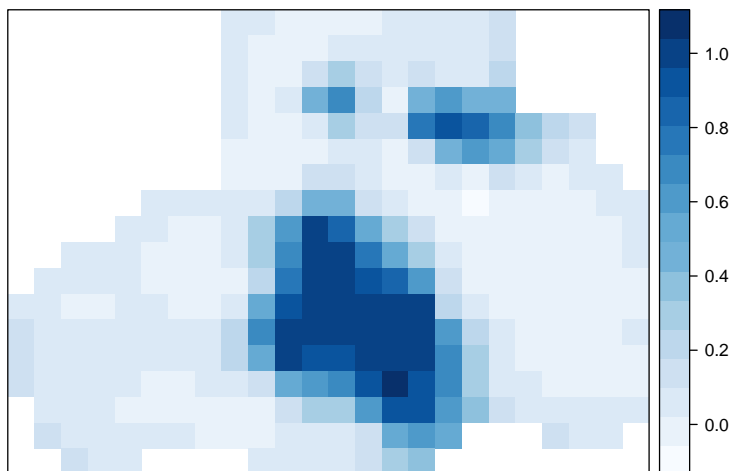


Aplicamos este mismo procedimiento a otras variables, por ejemplo la variable Cuerpo.de.Agua. En este caso el semi-variograma muestra una tendencia creciente.



Aplicamos la función de `krige` y observamos la predicción espacial.

[using ordinary kriging]



En este caso observamos valores extremos ligeramente menores a cero y mayores a uno. Redondeamos estos valores para ajustarlos al intervalo de probabilidades.

```
> summary(ik.Ag[[1]])
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-0.05728	0.01546	0.05686	0.21380	0.27420	1.04000

```
> ik.Ag[[1]][ik.Ag[[1]]<0] = 0
```

```
> ik.Ag[[1]][ik.Ag[[1]]>1] = 1
```

```
> summary(ik.Ag[[1]])
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00000	0.01546	0.05686	0.21470	0.27420	1.00000

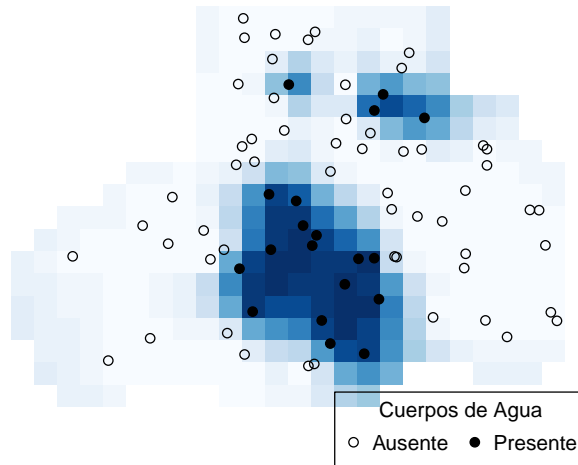
```
>
```

El mapa resultante nos indica la ubicación de los cuerpos de agua en el área de estudio.

```
> image(ik.Ag[1],col=colores(30))
```

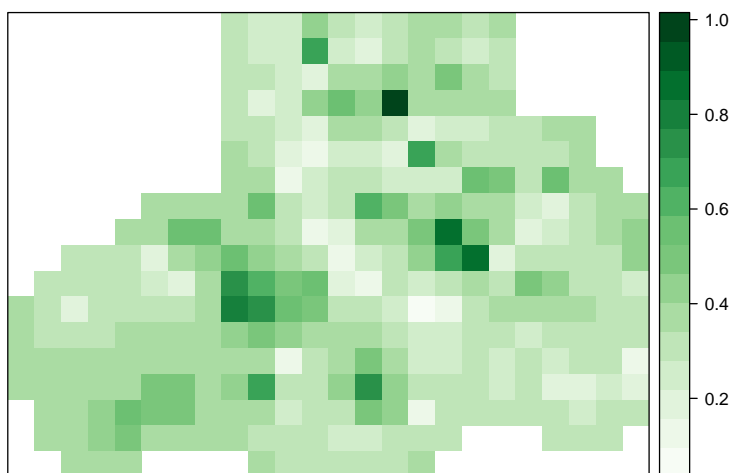
```
> points(pts,col=1,pch=c(1,19)[pts@data$Cuerpo.de.Agua+1])
```

```
> legend("bottomright",legend=c("Ausente","Presente"),col=1,pch=c(1,19),pt.cex=1,ti
```



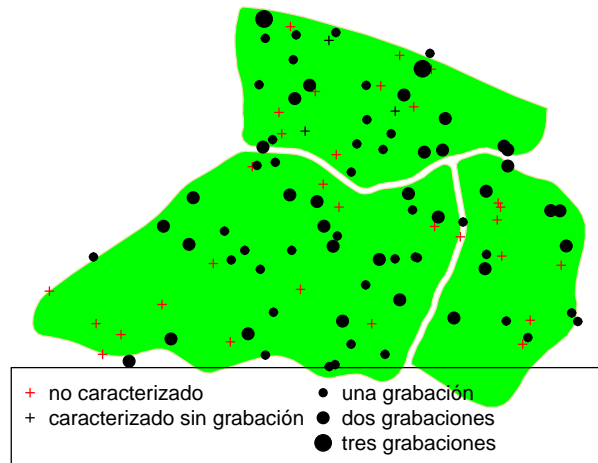
Seguimos el mismo procedimiento para la hojarasca (ver siguiente gráfico), y podemos continuarlo para otras variables (pendiente por hacer).

[using ordinary kriging]



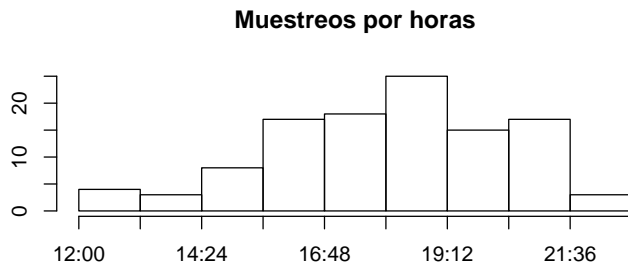
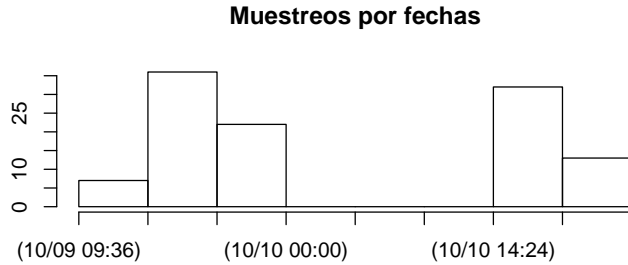
4. Puntos muestreados y grabaciones

De los 105 puntos seleccionados originalmente se caracterizaron y muestrearon 75, pero sólo se cuentan con grabaciones para 72 puntos, pues hubo algunas grabaciones que no se realizaron correctamente.



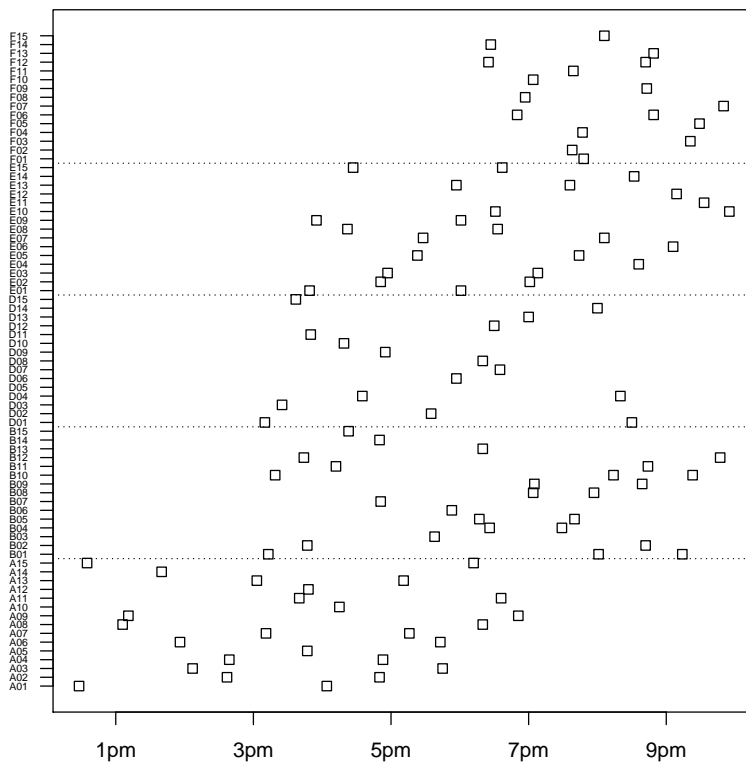
Para ver la distribución del esfuerzo de muestreo entre los dos días de trabajo de campo utilizamos la función `chron` del paquete *chron*

```
> require(chron)
> fechas <- chron(dates.=as.character(Muestreo$Fecha),
+               times.=sprintf("%s:00",as.character(Muestreo$Hini)),
+               format = c(dates = "ymd", times = "h:m:s"))
> horas <- chron(times.=sprintf("%s:00",as.character(Muestreo$Hini)),
+               format = c(dates = "ymd", times = "h:m:s"))
> layout(1:2)
> hist(fechas)
> title(main="Muestreos por fechas")
> hist(horas)
> title(main="Muestreos por horas")
```



También podemos ver a que horas se realizó cada muestreo, por ejemplo el grupo A fue muestreado desde la 1 hasta las 7, y el grupo F sólo fue muestreado de 6 a 10, pero los otros tres grupos fueron muestreados entre las 3 y 10 pm.

```
> par(mar=c(2,2,0,0))
> stripchart(horas~Muestreo$Punto, axes=F)
> axis(1, c(13,15,17,19,21)/24, c("1pm", "3pm", "5pm", "7pm", "9pm"))
> axis(2, 1:75, levels(Muestreo$Punto),
+       las=2, cex.axis=.5)
> box()
> abline(h=15*c(1,2,3,4)+.5, lty=3)
```



En las planillas se anotaron también las condiciones del muestreo:

```
> table(Muestreo$Viento)
```

```
0  1  2
73 35  2
```

```
> table(Muestreo$Ruido)
```

```
0  1  2
28 49 33
```

```
> table(Muestreo$Cielo)
```

```
0  1  2
49 44 17
```

La posición y sustrato del micrófono no se anotó consistentemente:

```
> table(Muestreo$Posición.mic)
```

	F	F	I	N/I no muestreado
	64	1	3	42
No muestreado				37
	7			

```
> table(Muestreo$Sustrato.mic)
```

```
      A    C N/I
44  12  56  42
```

La humedad y temperatura se anotaron originalmente con las unidades de medición, lo cual no es recomendable, además para la temperatura se mezclan grados Fahrenheit y Centígrados. Parte de estos errores fueron corregidos.

```
> table(Muestreo$Humedad)
```

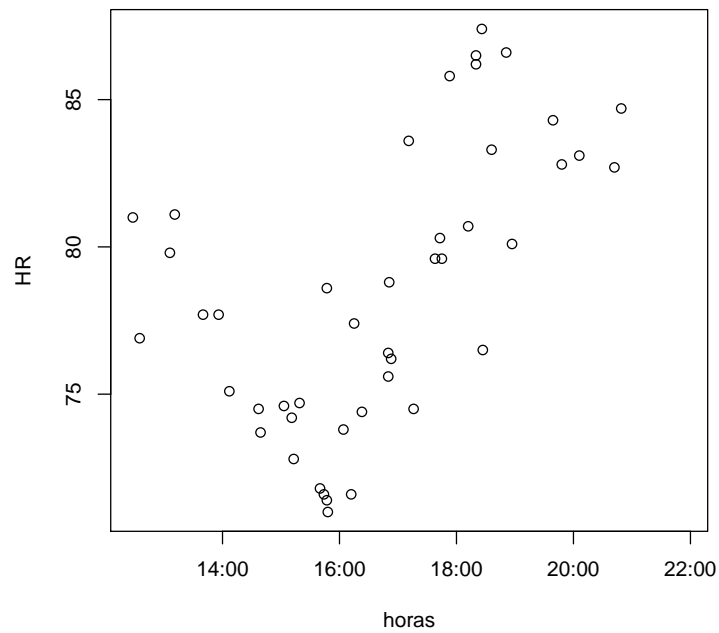
```
      71,0 71,4 71,6 71,8 72,8 73,7 73,8 74,2 74,4 74,5 74,6 74,7 75,1 75,6 76,2
44      1      1      2      1      1      1      1      1      2      1      1      1      1      1
76,4 76,5 76,9 77,4 77,7 78,6 78,8 79,6 79,8 80,1 80,3 80,7 81,0 81,1 82,7 82,8
      1      1      1      1      2      1      1      2      1      1      1      1      1      1      1
83,1 83,3 83,6 84,3 84,7 85,8 86,2 86,5 86,6 87,4 N/A
      1      1      1      1      1      1      1      1      1      1      65
```

```
> table(Muestreo$Temperatura)
```

```
      22,7 25,6 25,9 27,0 27,1 27,2 27,4 27,8      28 28,1 28,3 28,8
44      1      1      2      1      1      3      2      1      1      1      1      1
29,2 29,3 29,5 29,5 29,8 30 30,1 30,2 30,6 30,7 30,8 31 31,2
      1      1      1      1      1      1      1      2      1      4      1      2      4
31,3 31,5 31,9 32,2 32,3 33,3 87,1 N/A
      1      2      1      1      1      1      1      65
```

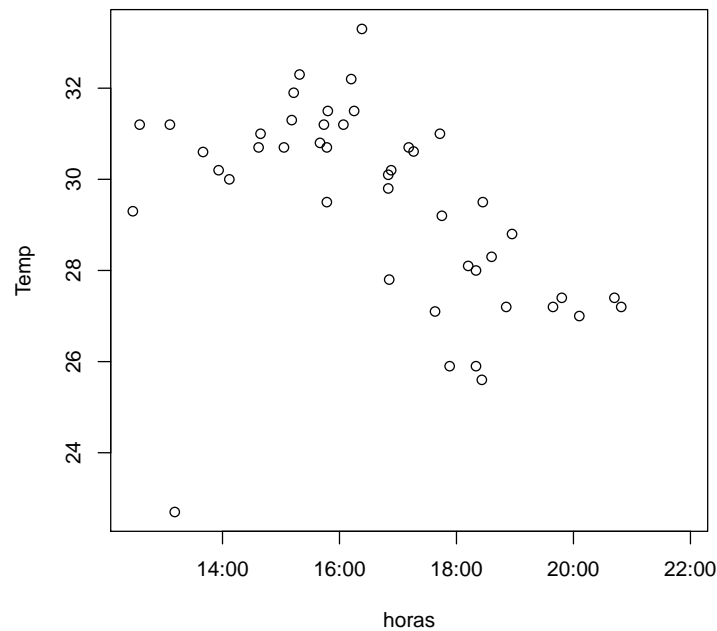
Hacemos una variable nueva para la humedad relativa (HR) removiendo las anotaciones de porcentaje y cambiando las comas por puntos, y luego graficamos su comportamiento con respecto a la hora de muestreo:

```
> HR <- as.numeric(sub(",", ".", sub("%", ""), Muestreo$Humedad)))
> plot(HR~horas)
>
```

Hacemos lo mismo con la temperatura, aplicando una transformación de Farenheit a Celsius para las que tienen valores muy altos:

```
> Temp <- as.numeric(sub(",", ".", sub("[CF]", "", Muestreo$Temperatura)))
> Temp[!is.na(Temp) & Temp>50] <- (Temp[!is.na(Temp) & Temp>50]-32)*5/9
> plot(Temp~horas)
```



5. Especies detectadas

La lista de especies conocidas del jardín botánico es la siguiente (datos de Larreal y Cardozo, presentados en congresos pero no publicados):

- *Rhinella marina* (Linnaeus, 1758)
- *Rhinella humboldti* (Gallardo, 1965)
- *Ceratophrys calcarata* (Boulenger, 1890)
- *Dendropsophus microcephalus* (Cope, 1886)
- *Hypsiboas pugnax* (Schmidt, 1857)
- *Scinax ruber* (Laurenti, 1768)
- *Pseudis paradoxa* (Linnaeus, 1758)
- *Leptodactylus bolivianus* (Boulenger, 1898)
- *Leptodactylus fragilis*
- *Pleurodema brachyops* (Cope, 1869)
- *Engystomops pustulosus* (Lynch, 1970)

Por ahora analizaremos los datos de las especies detectadas a través de una inspección manual (realizada por Arlene Cardozo) de un primer fragmento de 5 minutos de grabación en cada punto de muestreo. Falta procesar dos fragmentos adicionales de 5 minutos para cada punto de muestreo.

En estos fragmentos de grabaciones analizados se han detectado seis especies conocidas y una nueva aún no identificada:

Lf *Leptodactylus fragilis*

Hp *Hypsiboas pugnax*

Dm *Dendropsophus microcephala*

Rh *Rhinella humboldti*

Rm *Rhinella marina*

Pb *Pleurodema brachyops*

Ep *Engystomops pustulosus*

SpA Especie “A” (no identificada)

SpB Especie “B” (no identificada)

A continuación se muestra el número total de detecciones para cada especie:

```
> colSums(VM[,3:11]>0)
```

	Lf	Hp	Dm	Rh	Rm	Pb	SpA	SpB	Ep
	104	127	20	97	22	39	1	14	2

Este valor incluye detecciones basadas en la inspección visual del espectrograma, y casos en los que se escuchan las vocalizaciones en la pista de audio aunque no se observa en el espectrograma. En estos últimos casos puede que se trate de vocalizaciones muy alejadas del micrófono, y por ello las vamos a excluir de los análisis.

El número de detecciones basadas en el espectrograma sería el siguiente:

```
> colSums(VM[,3:11]>.5)
```

	Lf	Hp	Dm	Rh	Rm	Pb	SpA	SpB	Ep
	56	87	13	73	16	15	1	14	1

5.1. Riqueza de especies

El primer objetivo es verificar el esfuerzo muestral necesario para estimar riqueza. Usamos un modelo de acumulación de especies que considere el número de puntos de muestreo y el esfuerzo de muestreo por punto (número de minutos de grabaciones/número de puntos).

El modelo sugiere que el número de trampas es más importante que el número de repeticiones o tiempo de grabación (pero falta incluir los otros dos periodos de 5m). Se necesitan unos 40 puntos de muestreo para garantizar observar 6 de las 7 especies presentes. Si se consideran solamente las grabaciones posteriores a las 5pm se puede conseguir ese objetivo con 25 puntos.

Formula: $spps \sim C + (\alpha * \log(prds) + 1) + (\beta * \log(sfrzs) + 1)$

Parameters:

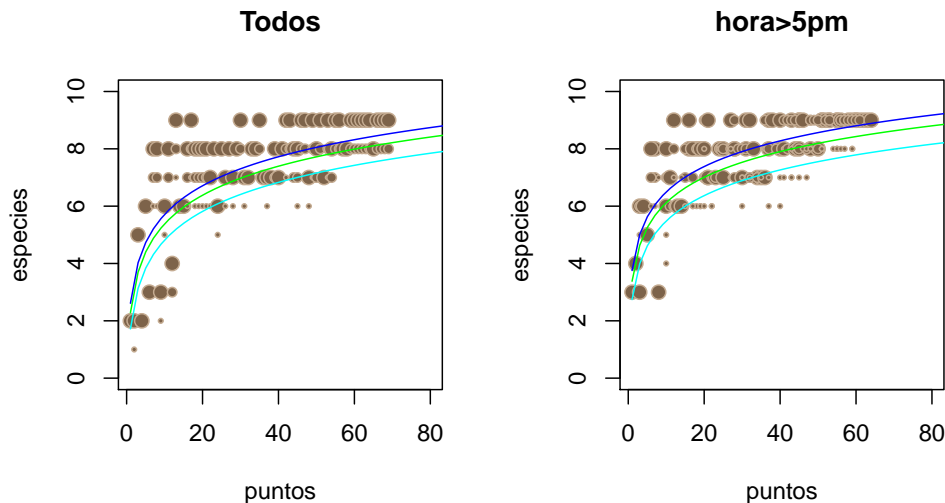
	Estimate	Std. Error	t value	Pr(> t)	
C	-1.5932	0.4303	-3.702	0.000275	***
alpha	1.4735	0.0775	19.014	< 2e-16	***
beta	0.8142	0.1542	5.281	3.28e-07	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.006 on 204 degrees of freedom

Number of iterations to convergence: 1

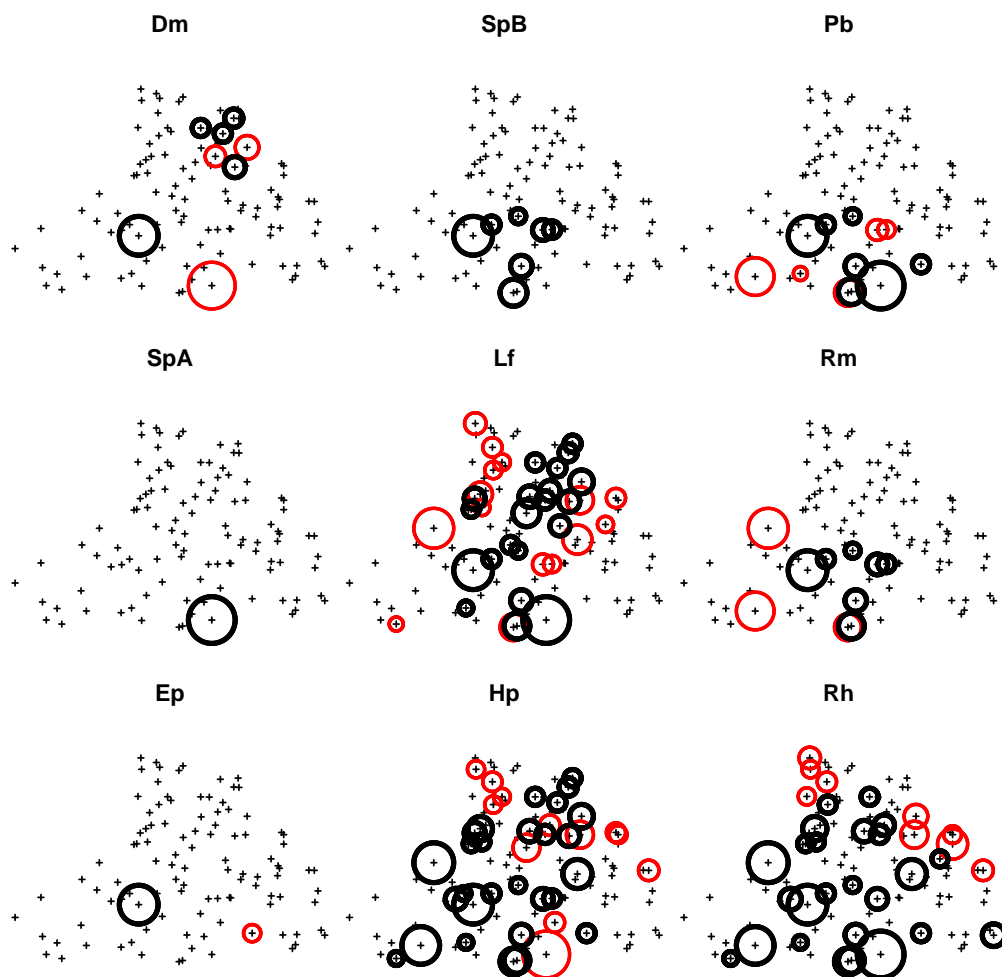
Achieved convergence tolerance: 1.124e-07



5.2. Área ocupada por las especies

5.2.1. Predicción espacial

Comparar la distribución de detecciones basadas en el espectrograma (negro) con los registros escuchados pero no visibles (rojo). ¿Hace falta comparar predicciones con ambos tipos de datos? actualmente sólo usamos las que están basadas en el espectrograma (negro).

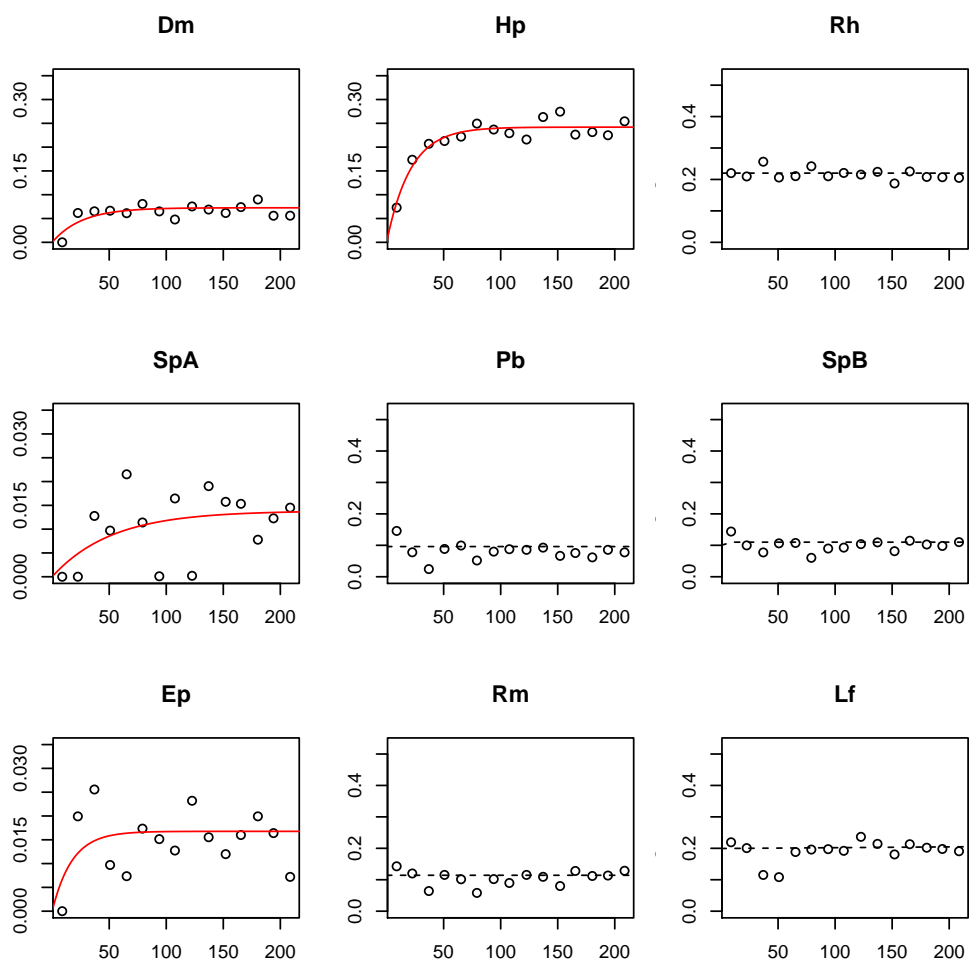


Primero verificamos si hay un patrón de dependencia espacial calculando el semi-variograma y ajustando un modelo, en cuatro casos hay una dependencia espacial notoria (línea roja), en los otros tres la semi-varianza es constante e independiente de la distancia (línea negra segmentada). Para seis especies usamos una escala común para comparar la magnitud de los patrones, con la excepción de SpA que tiene un patrón con una magnitud 10 veces menor que el resto de las especies:

Warning: singular model in variogram fit

Warning: singular model in variogram fit

Warning: singular model in variogram fit



Aplicamos la función de `krige` para realizar la predicción espacial. En los casos que tienen un patrón de autocorrelación espacial marcado especificamos el modelo de semivariograma para aplicar universal Kriging, en caso contrario la función aplica interpolación ponderada por el inverso de distancia.

```
> ik.Dm = krige(Dm~poly(x,y,degree=2), pts[ss,], JBM.grd, vm.Dm)
```

```
[using universal kriging]
```

```
> ik.SpA = krige(SpA~y, pts[ss,], JBM.grd, vm.SpA)
```

```
[using universal kriging]
```

```
> ik.Hp = krige(Hp~poly(x,y,degree=1), pts[ss,], JBM.grd, vm.Hp)
```

```
[using universal kriging]
```

```
> ik.Ep = krige(Ep~y, pts[ss,], JBM.grd, vm.Ep)
```

```

[using universal kriging]
> ik.Pb = krige(Pb~1, pts[ss,], JBM.grd)

[inverse distance weighted interpolation]
> ik.Rm = krige(Rm~1, pts[ss,], JBM.grd)

[inverse distance weighted interpolation]
> ik.Rh = krige(Rh~1, pts[ss,], JBM.grd)

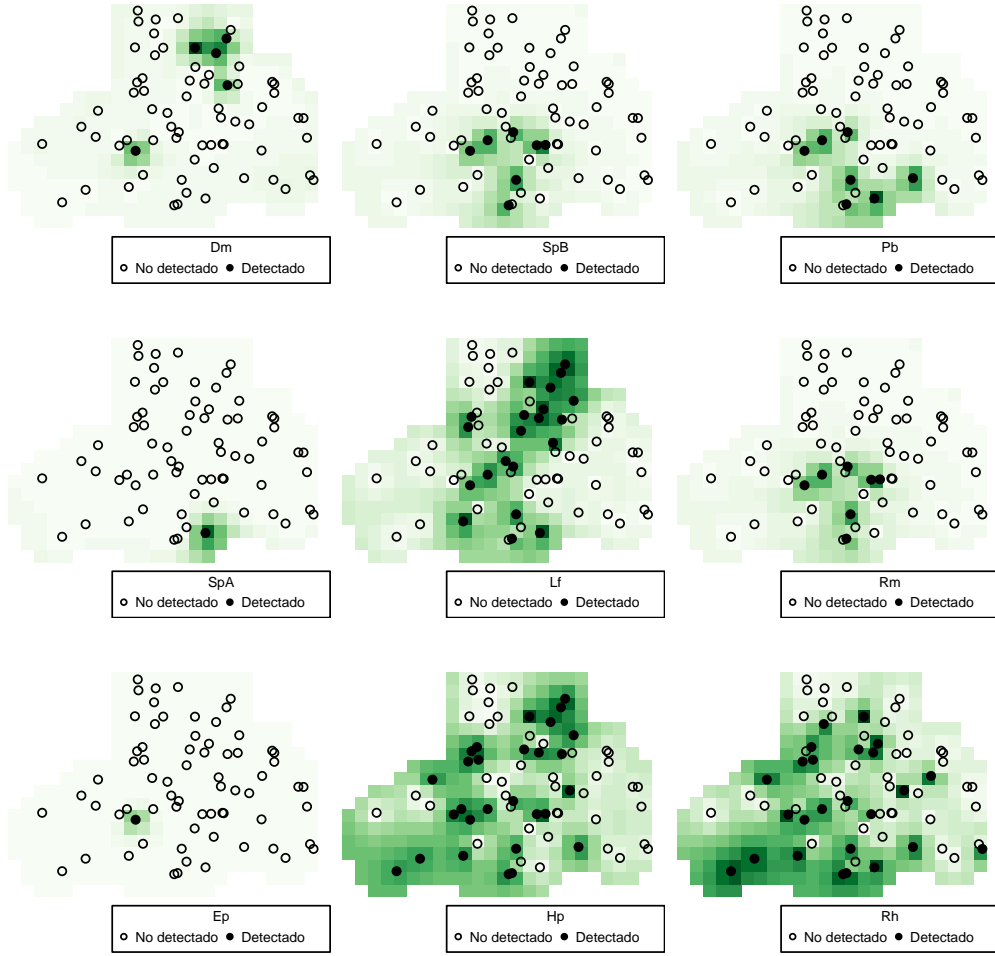
[inverse distance weighted interpolation]
> ik.SpB = krige(SpB~1, pts[ss,], JBM.grd)

[inverse distance weighted interpolation]
> ik.Lf = krige(Lf~1, pts[ss,], JBM.grd)

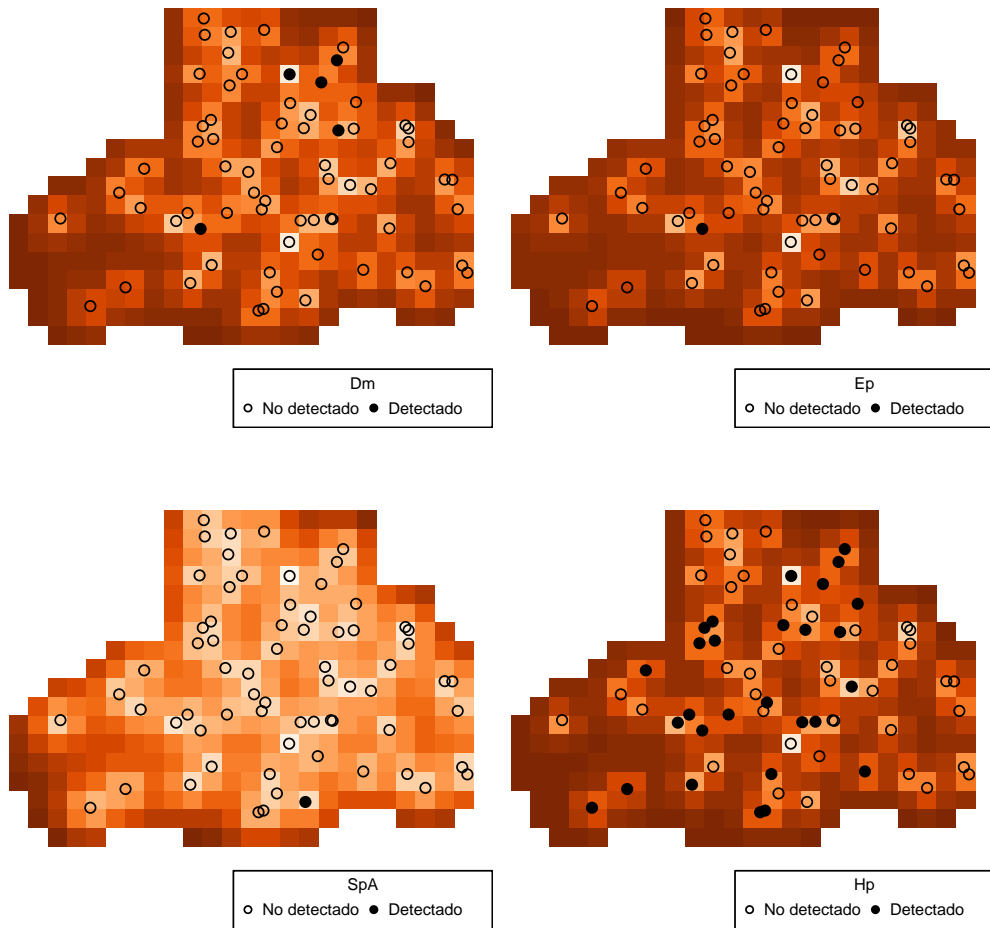
[inverse distance weighted interpolation]
> ##ik.Hp = krige(Hp~y, pts[ss,], JBM.grd)

```

El resultado de la predicción espacial con estos modelos se presenta a continuación utilizando una escala de valores común entre 0 (ausencia) y 1 (presencia).



El modelo de kriging universal calcula también la variación espacial en la varianza de la predicción. Aquí presentamos la varianza de las cuatro especies con modelo de kriging en una escala común de 0 a 0.3, exceptuando a SpA con una escala entre 0 y 0.03.

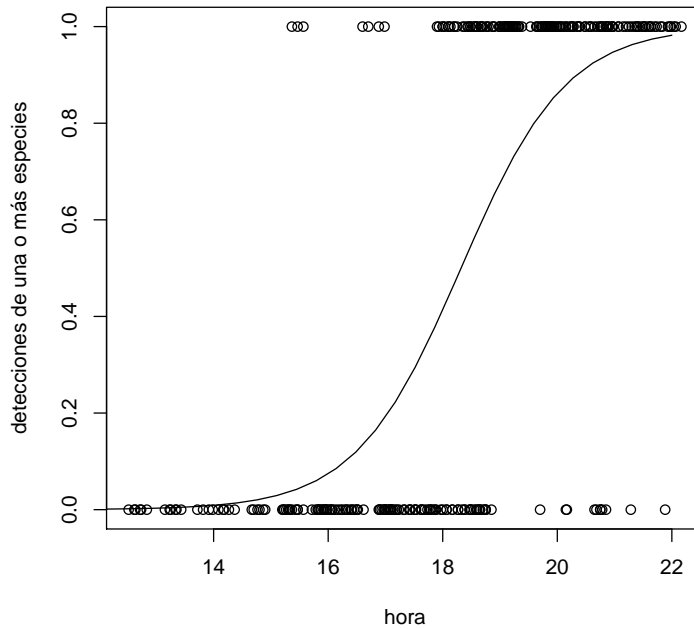


Podemos interpretar las predicciones como la probabilidad de presencia de la especie en cada celda del mapa bajo el supuesto de que la detección haya sido perfecta. La suma de esas probabilidades multiplicada por el área de las celdas nos permite calcular un estimado del área ocupada por cada especie (en el caso de kriging se pueden añadir intervalos de confianza para estimar la incertidumbre del estimado, pendiente):

	especie	area.m2.krig
1	Dm	15527.384
2	SpA	4959.176
3	SpB	18816.658
4	Ep	3076.303
5	Lf	56194.057
6	Hp	76004.136
7	Pb	15527.384
8	Rm	18665.725
9	Rh	74740.185

5.2.2. Modelo lineal ponderado

El estimado anterior se basa en el patrón espacial sin tomar en cuenta todas las variables ambientales ni el efecto de la detectabilidad. Sabemos que las detecciones son mayores en la noche cuando las vocalizaciones son más frecuentes:

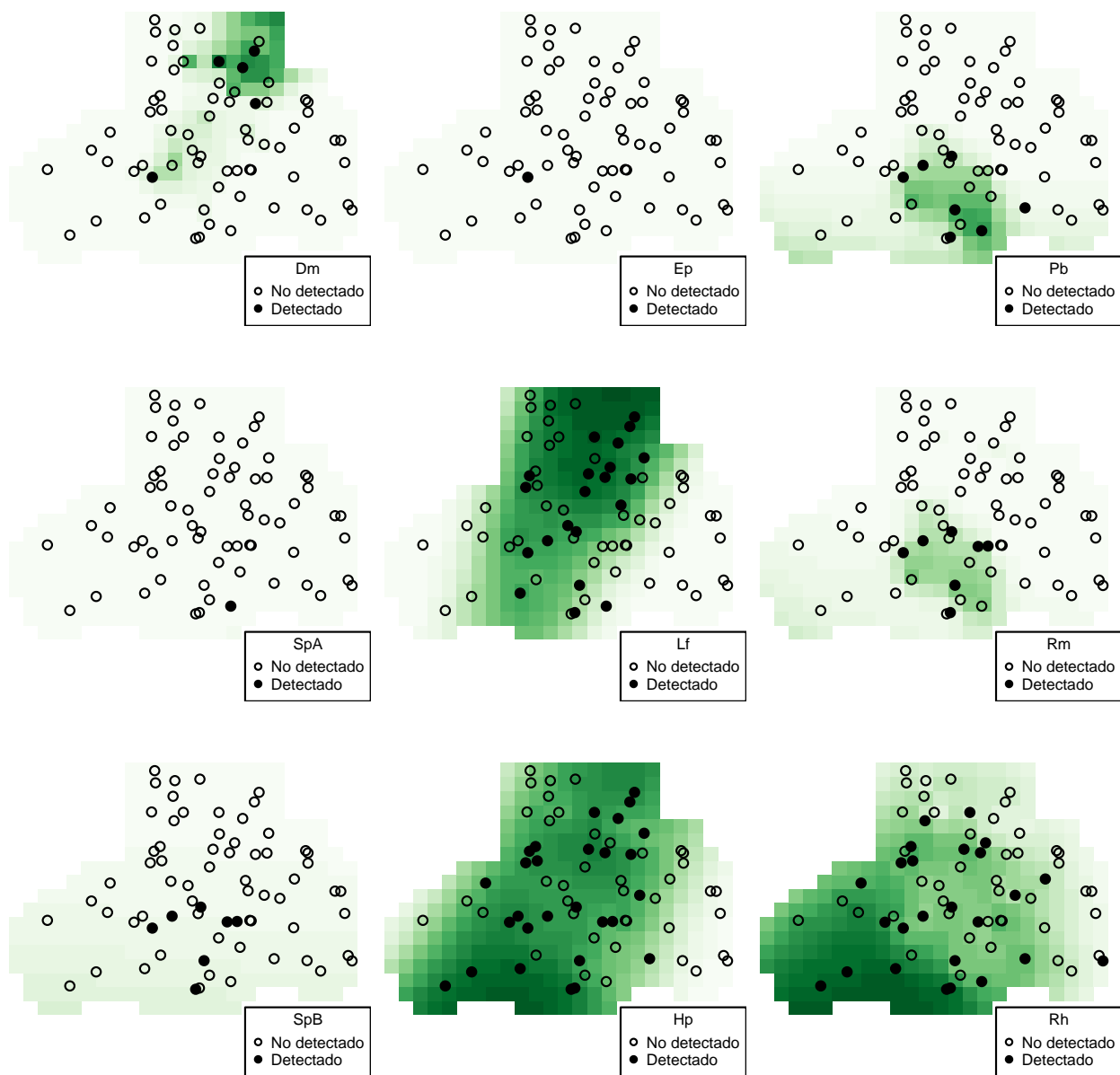


Podemos utilizar una regresión logística ponderada para estimar la probabilidad de presencia corrigiendo por el efecto de la hora. Para estimar los pesos de las observaciones calculamos la frecuencia de detección por hora (curva del gráfico anterior).

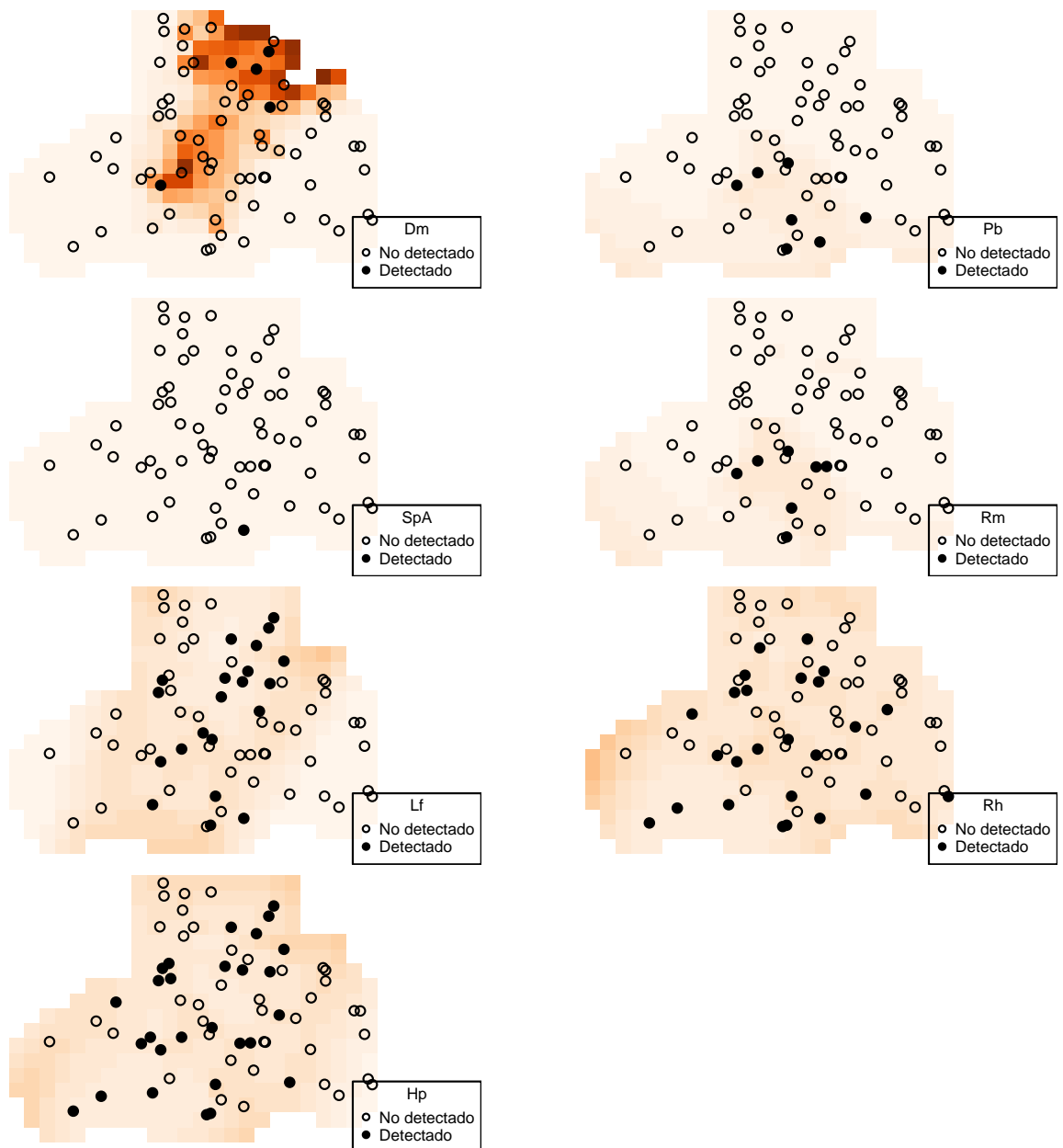
Utilizamos un modelo generalizado lineal con una distribución quasibinomial para manejar los pesos fraccionales de las observaciones.

```
> mdl00.Dm <- glm(Dm~poly(x,y,degree=2)+Hojarasca+Cuerpo.de.Agua,VM.df,family=quasi
> mdl00.Lf <- glm(Lf~poly(x,y,degree=2)+Hojarasca+Cuerpo.de.Agua,VM.df,family=quasi
> mdl00.Hp <- glm(Hp~poly(x,y,degree=2)+Hojarasca+Cuerpo.de.Agua,VM.df,family=quasi
> mdl00.Rh <- glm(Rh~poly(x,y,degree=2)+Hojarasca+Cuerpo.de.Agua,VM.df,family=quasi
> mdl00.SpA <- glm(SpA~poly(x,y,degree=1),VM.df,family=quasibinomial(logit),weight=
> mdl00.SpB <- glm(SpB~poly(x,y,degree=1),VM.df,family=quasibinomial(logit),weight=
> mdl00.Ep <- glm(Ep~poly(x,y,degree=1),VM.df,family=quasibinomial(logit),weight=wd
> mdl00.Pb <- glm(Pb~poly(x,y,degree=1)+Cuerpo.de.Agua,VM.df,family=quasibinomial(1
> mdl00.Rm <- glm(Rm~poly(x,y,degree=1)+Cuerpo.de.Agua,VM.df,family=quasibinomial(1
>
>
```

Podemos visualizar la predicción espacial basada en el modelo generalizado lineal.

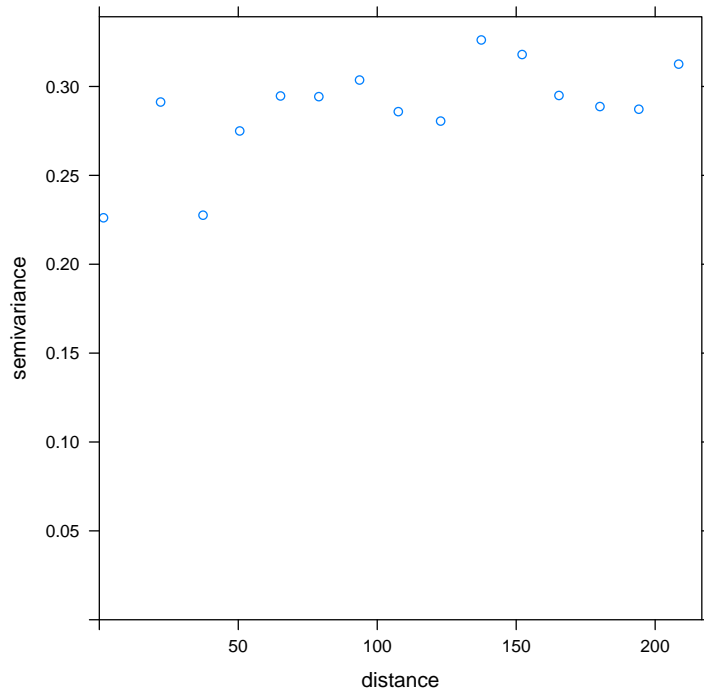


Igualmente visualizamos el error estandar de la predicción



Es importante verificar que no hayan problemas por la auto-correlación espacial, para ello debemos inspeccionar los residuales del modelo y verificar que no haya estructura espacial residual (Pendiente).

```
> v.resid = variogram(resid mdl00.Lf)~1, VM.df)
> plot(v.resid)
>
>
```



Estos estimados nos indican la probabilidad de presencia suponiendo que la probabilidad de detección depende exclusivamente de la hora del día, que la misma es similar para todas las especies, y que puede ser estimada indirectamente a través de la ponderación escogida. Comparamos los estimados de área ocupada con el método de glm y con kriging y vemos que en el glm los estimados son superiores (mayor área ocupada) para Pb y Rh, pero considerablemente menores para el resto de las especies.

	especie	area.m2.krig	area.m2.glm
1	Dm	15527.384	15979.201
2	SpA	4959.176	142.794
3	SpB	18816.658	13223.089
4	Ep	3076.303	1457.704
5	Lf	56194.057	78932.066
6	Hp	76004.136	107608.590
7	Pb	15527.384	20528.088
8	Rm	18665.725	15698.373
9	Rh	74740.185	92438.632

5.2.3. Modelos mixtos de probabilidad de presencia y detección

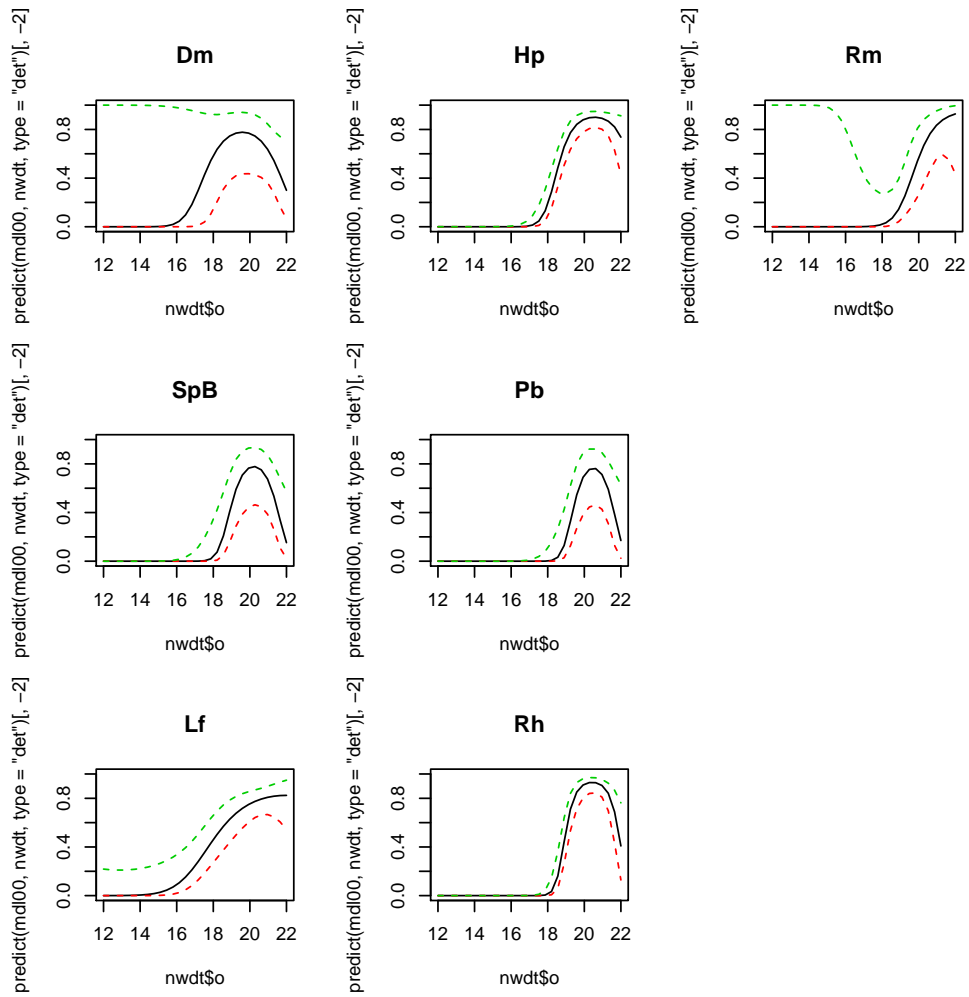
Los modelos mixtos de probabilidad de presencia y detección (“occupancy models”) se basan en una regresión logística pero modelan explícitamente el proceso de detección y permiten obtener estimados independientes de probabilidad de presencia y probabilidad de detección. En este caso se requieren de historias de detección por punto de muestreo (varias

visitas u observaciones independientes por punto) para estimar la probabilidad de detectar una especie en caso de estar presente.

Ajustamos los modelos, con los datos actuales aún incompletos no tenemos suficiente datos para ajustar todos los modelos. Para una especie no converge ningún modelo y para dos no logramos convergencia con el modelo completo pero si con modelos más sencillos

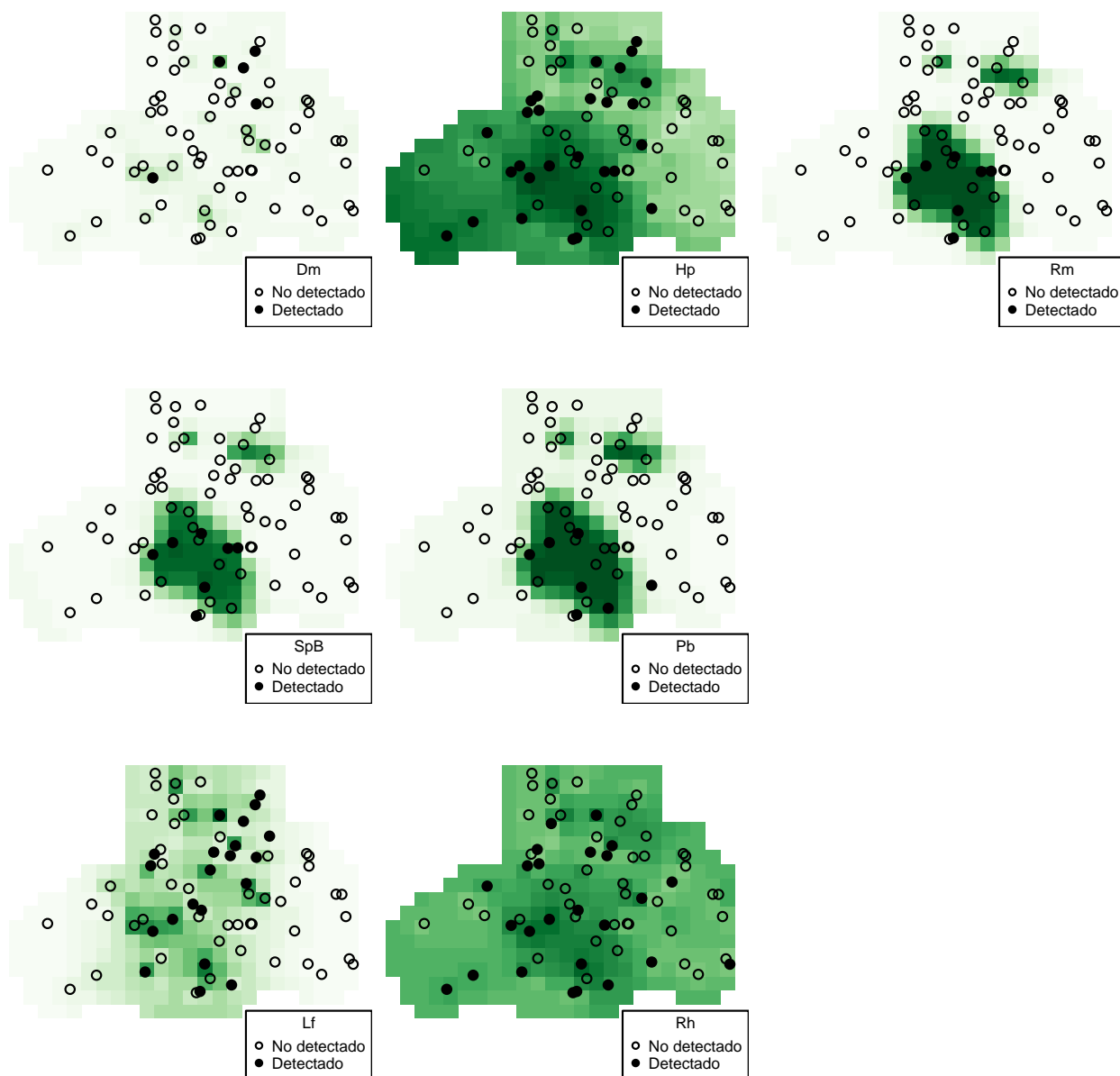
```
> ##Convergen
> mdlhA.Pb <- occu(~hora+I(hora^2) ~ poly(x,degree=2)+Cuerpo.de.Agua+Hojarasca,
+                  uF.Pb)
> mdlhA.Lf <- occu(~hora+I(hora^2) ~ poly(x,degree=2)+Cuerpo.de.Agua+Hojarasca,
+                  uF.Lf)
> mdlhA.Hp <- occu(~hora+I(hora^2) ~ poly(x,y,degree=1)+Hojarasca+Cuerpo.de.Agua,
+                  uF.Hp)
> mdlhA.Dm <- occu(~hora+I(hora^2) ~ poly(x,degree=1)+Hojarasca+Cuerpo.de.Agua,
+                  uF.Dm)
> ## no converge el modelo completo
> mdlhA.Rh <- occu(~hora+I(hora^2) ~ Hojarasca+Cuerpo.de.Agua,
+                  uF.Rh)
> mdlhA.Rm <- occu(~hora+I(hora^2) ~ Hojarasca+Cuerpo.de.Agua,
+                  uF.Rm)
> mdlhA.SpB <- occu(~hora+I(hora^2) ~ Hojarasca+Cuerpo.de.Agua,
+                  uF.SpB)
> ## no converge
> mdlhA.SpA <- occu(~hora ~ 1, uF.SpA)
> mdlhA.Ep <- occu(~hora ~ 1, uF.Ep)
>
```

Verificamos primero los estimados de probabilidad de detección para cada especie en función de la hora y vemos que hay diferencias importantes entre especies lo cual justifica la necesidad de ajustar modelos específicos.



Para SpA y Pb los estimados de probabilidad de detección son poco confiables (intervalos de confianza muy amplios). Para el resto de las especies los estimados muestran una clara tendencia positiva, aunque en unas especies es más marcado (intervalos de actividad más estrechos) que en otras (intervalos de actividad más amplios).

Vemos ahora la predicción espacial:



Igualmente visualizamos el error estandar de la predicción:



Podemos comparar ahora los estimados de área ocupada entre la predicción espacial, el modelo logístico (quasibinomial) y el modelo mixto. Los estimados del modelo mixto son considerablemente superiores a los anteriores, especialmente para Dm, y esto sugiere que la falta de detección puede estar sesgando los estimados de área ocupada. Sin embargo los estimados son todavía muy imprecisos (pendiente: intervalos de confianza de los otros dos métodos).

	especie	area.m2.krig	area.m2.glm	area.m2.occ	lower.occ	upper.occ
1	Dm	15527.384	15979.201	9714.851	1854.393	45732.04
2	SpA	4959.176	142.794	0.000	0.000	0.00
3	SpB	18816.658	13223.089	31237.317	9290.276	71167.82
4	Ep	3076.303	1457.704	0.000	0.000	0.00

5	Lf	56194.057	78932.066	42877.504	16704.691	97152.68
6	Hp	76004.136	107608.590	120385.780	59474.140	158165.14
7	Pb	15527.384	20528.088	44462.300	2766.239	113189.59
8	Rm	18665.725	15698.373	34180.600	10135.711	77837.77
9	Rh	74740.185	92438.632	117530.000	71414.330	152974.82

6. Conclusiones (preliminares)

- ¿Esfuerzo mínimo para el inventario de especies? al menos 40 puntos (con grabaciones de 5 minutos...) o 20 puntos si restringimos el trabajo después de las 5pm
- ¿más puntos, más visitas o grabaciones más largas? más puntos, por ahora el número de visitas no da significativo
- ¿qué tan difíciles de detectar? Según los modelos mixtos cada especie tiene una ventana de tiempo muy definida con alta detectabilidad, pero Lf y Rh tiene pendiente menos pronunciadas
- la estrategia optima es utilizar la mañana y tarde para caracterizar la mayor cantidad de punto y poder generar mapas de indicadores ambientales, luego en la noche aprovechar las horas de máxima probabilidad de detección para hacer un muestreo óptimo. Si el objetivo es estimar riqueza de especies 20 o 30 puntos son suficientes. Para hacer predicción espacial se requieren más puntos, si nos basamos en el espectrograma la densidad de puntos necesaria es mayor, pero si consideramos además la detección auditiva podemos ampliar el área cubierta por cada punto.
- Comparar ventajas y desventajas de los tres métodos (“kriging”, modelo generalizado lineal y modelo mixto o compuesto) Calcular esfuerzo mínimo necesario (costo) para cada estrategia, los riesgos de sesgos y la precisión de los estimados finales para considerar si vale la pena invertir más o menos esfuerzo en el muestreo.

Referencias

- HIJMANS, R. J. & VAN ETTEN, J., 2012. *raster: Geographic analysis and modeling with raster data*. URL <http://CRAN.R-project.org/package=raster>. R package version 2.0-08.
- JAMES, D. & HORNIK, K., 2011. *chron: Chronological Objects which Can Handle Dates and Times*. URL <http://CRAN.R-project.org/package=chron>. R package version 2.3-42. S original by David James, R port by Kurt Hornik.
- KEITT, T. H., BIVAND, R., PEBESMA, E. & ROWLINGSON, B., 2012. *rgdal: Bindings for the Geospatial Data Abstraction Library*. URL <http://CRAN.R-project.org/package=rgdal>. R package version 0.7-16.

- LEMON, J., 2006. *Plotrix: a package in the red light district of r*. R-News, 6:8–12.
- NEUWIRTH, E., 2011. *RColorBrewer: ColorBrewer palettes*. URL <http://CRAN.R-project.org/package=RColorBrewer>. R package version 1.0-5.
- PEBESMA, E. & BIVAND, R., 2005. *Classes and methods for spatial data in r*. R News, 5. URL <http://cran.r-project.org/doc/Rnews/>.
- PEBESMA, E. J., 2004. *Multivariable geostatistics in s: the gstat package*. Computers & Geosciences, 30:683–691.
- R DEVELOPMENT CORE TEAM, 2010. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>. ISBN 3-900051-07-0.