Logistic regression

Zofie Cimburova

28 juni 2017

## Libraries

library(gstat)  
library(sp)  
library(raster)  
library(rgdal)   
library(nnet)

# 1. Import observations

* Import rasters of explanatory variables (temperature, TPI, latitude, slope, solar radiation) and response variable (7 - forest, 8 - open land, 9 - forest line):

bio11 <- raster("BIO11\_kriged\_sample.tif", band=1)  
projection(bio11) <- "+proj=utm +zone=33 +ellps=GRS80 +towgs84=0,0,0,0,0,0,0 +units=m +no\_defs"  
  
r.explanatory <- brick(bio11,  
 raster("TPI1010.tif", band=1),  
 raster("lat.tif", band=1),  
 raster("slope.tif", band=1),  
 raster("srad10.tif", band=1),  
 raster("input\_binomial\_sample.tif", band=1))  
names(r.explanatory) <- c('BIO11','TPI1010','lat','slope','srad10','response')

* Generate random observation points (200) across the whole space. Exclude those where some explanatory variable is NaN. Reclass response variable to 0 (open land) and 1 (forest + forest line)

observations <- sampleRandom(r.explanatory,200, na.rm=TRUE, xy=TRUE)  
  
observations <- as.data.frame(observations)  
  
observations$response[observations$response == 8] <- 0  
observations$response[observations$response == 9] <- 0  
observations$response[observations$response == 7] <- 1

# 2. Fit binomial logistic regression

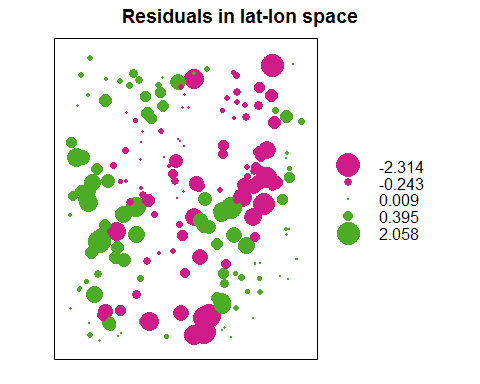
* Solar radiation excluded due to low significance.

mod<- glm(response ~ BIO11 + TPI1010 + lat + slope,family=binomial(link='logit'),data=observations)  
summary(mod)

##   
## Call:  
## glm(formula = response ~ BIO11 + TPI1010 + lat + slope, family = binomial(link = "logit"),   
## data = observations)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.31360 -0.24314 0.00853 0.39549 2.05754   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 1.140e+04 2.063e+03 5.528 3.24e-08 \*\*\*  
## BIO11 5.500e-01 9.383e-02 5.861 4.60e-09 \*\*\*  
## TPI1010 4.677e-02 1.128e-02 4.147 3.37e-05 \*\*\*  
## lat -1.795e+02 3.248e+01 -5.525 3.30e-08 \*\*\*  
## slope 7.904e-02 2.255e-02 3.506 0.000456 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 275.64 on 199 degrees of freedom  
## Residual deviance: 104.14 on 195 degrees of freedom  
## AIC: 114.14  
##   
## Number of Fisher Scoring iterations: 7

## Check spatial autocorrelation in residuals

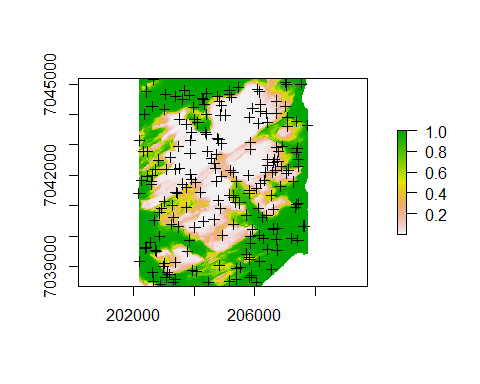
* apparently clusters of similar values exist



## Predict the whole area

* Predict probabilities of forest, and then set treshold for forest / open land.

newdata <- rasterToPoints(r.explanatory)  
newdata <- as.data.frame(newdata)  
  
newdata$predict <- predict(mod,newdata=newdata,type='response')  
newdata$predict10 <- ifelse(newdata$predict > 0.5,1,0)  
  
predict.raster <- rasterFromXYZ(newdata)

* Plot probabilities and observation points 
* Plot forest / open land and original forest line 