Continuous soil attribute modeling and mapping: Regression Kriging (1)

Soil Security Laboratory

2018

1 Regression kriging

In the previous sections we looked at a few soil spatial prediction functions which at the most fundamental level, target the the correlation between the target soil variable and the available covariate information. We fitted a number of models which included simple linear functions to non-linear functions such as regression trees to other more complicated data mining techniques (Cubist and Random Forest). In this section we will extend upon this DSM approach from what are called deterministic models to also include the spatially correlated residuals that result from fitting these models.

The approach we will now concentrate is a hybrid approach to modelling, whereby the predictions of the target variable are made via a deterministic method (regression model with covariate information) and a stochastic method where we determine the spatial auto-correlation of the model residuals with a variogram. The deterministic model essentially "detrends" the data, leaving behind the residuals for which we need to investigate whether there is additional spatial structure which could be added to the regression model predictions. These residuals are the random component of the $scorpan + emodel$. This method is described as regression kriging and has formally been described in Odeh et al. (1995) and is synonymous with universal kriging (Hengl et al., 2007), which is the formal linear model procedure to this soil spatial modeling approach. The purpose of this exercise is to introduce some basic concepts of regression kriging. You will have already had some experience in regression models. We have also investigated briefly the fundamental concepts of kriging for which the variogram is central to.

1.1 Universal kriging

The universal kriging function in `R` is found in the `gstat` package. It is useful from the view that both the regression model and variogram modeling of the residuals are handled together. Using universal kriging, one can efficiently derive prediction uncertainties by way of the kriging variance. A limitation of universal kriging in the true sense of the model parameter fitting is that the model is linear. The general preference is DSM studies is to used non-linear

and recursive models that do not require strict model assumptions and assume a linear relationship between target variable and covariates.

One of the strict requirements of universal kriging in `gstat` is that the CRS (coordinate reference system) of the point data and covariates must be exactly the same.

First get the data and perform the covariate data intersection

```r
library(ithir)
library(raster)
library(rgdal)
library(sp)

# point data
data(HV_subsoilpH)

# Start afresh round pH data to 2 decimal places
HV_subsoilpH$pH60_100cm <- round(HV_subsoilpH$pH60_100cm, 2)

# remove already intersected data
HV_subsoilpH <- HV_subsoilpH[, 1:3]

# add an id column
HV_subsoilpH$id <- seq(1, nrow(HV_subsoilpH), by = 1)

# re-arrange order of columns
HV_subsoilpH <- HV_subsoilpH[, c(4, 1, 2, 3)]

# Change names of coordinate columns
names(HV_subsoilpH)[2:3] <- c("x", "y")


# grids (covariate raster)
data(hunterCovariates_sub)
```

Perform the covariate intersection.

```r
coordinates(HV_subsoilpH) <- ~x + y

# extract
DSM_data <- extract(hunterCovariates_sub, HV_subsoilpH, sp = 1, method = "simple")
DSM_data <- as.data.frame(DSM_data)
str(DSM_data)

## 'data.frame': 506 obs. of  15 variables:
##  $ id                     : num  1 2 3 4 5 6 7 8 9 10 ...
##  $ x                      : num  340386 340345 340559 340483 340734 ...
##  $ y                      : num  6368690 6368491 6369168 6368740 6368964 ...
##  $ pH60_100cm             : num  4.47 5.42 6.26 8.03 8.86 7.28 4.95 5.61 5.39 3.44 ...
##  $ Terrain_Ruggedness_Index: num  1.34 1.42 1.64 1.04 1.27 ...
##  $ AACN                   : num  1.619 0.281 2.301 1.74 3.114 ...
```

```
##  $ Landsat_Band1      : num   57 47 59 52 62 53 47 52 53 63 ...
##  $ Elevation          : num   103.1 103.7 99.9 101.9 99.8 ...
##  $ Hillshading        : num   1.849 1.428 0.934 1.517 1.652 ...
##  $ Light_insolation   : num   1689 1701 1722 1688 1735 ...
##  $ Mid_Slope_Positon  : num   0.876 0.914 0.844 0.848 0.833 ...
##  $ MRVBF              : num   3.85 3.31 3.66 3.92 3.89 ...
##  $ NDVI               : num   -0.143 -0.386 -0.197 -0.14 -0.15 ...
##  $ TWI                : num   17.5 18.2 18.8 18 17.8 ...
##  $ Slope              : num   1.79 1.42 1.01 1.49 1.83 ...
```

Often it is handy to check to see whether there are missing values both in the target variable and of the covariates. It is possible that a point location does not fit within the extent of the available covariates. In these cases the data should be excluded. A quick way to assess whether there are missing or NA values in the data is to use the `complete.cases` function.

```
which(!complete.cases(DSM_data))
```

```
## integer(0)
```

```
DSM_data <- DSM_data[complete.cases(DSM_data), ]
```

```
set.seed(123)
training <- sample(nrow(DSM_data), 0.7 * nrow(DSM_data))
cDat <- DSM_data[training, ]
```

```
coordinates(cDat) <- ~x + y
crs(cDat) <- "+proj=utm +zone=56 +south +ellps=WGS84 + datum=WGS84 +units=m +no_defs"
crs(hunterCovariates_sub) = crs(cDat)
```

```
# check
crs(cDat)
```

```
## CRS arguments:
##  +proj=utm +zone=56 +south +ellps=WGS84 +units=m +no_defs
```

```
crs(hunterCovariates_sub)
```

```
## CRS arguments:
##  +proj=utm +zone=56 +south +ellps=WGS84 +units=m +no_defs
```

Now lets parametise the universal kriging model, and we will use selected covariates as we were using before.

```
library(gstat)
vgm1 <- variogram(pH60_100cm ~ AACN + Landsat_Band1 + Elevation + Hillshading +
    Mid_Slope_Positon + MRVBF + NDVI + TWI, cDat, width = 200, cutoff = 3000)
mod <- vgm(psill = var(cDat$pH60_100cm), "Exp", range = 3000, nugget = 0)
model_1 <- fit.variogram(vgm1, mod)
model_1
```

```
##   model     psill     range
## 1   Nug 0.8140845    0.0000
## 2   Exp 0.6168554 426.2951
```

```r
# Universal kriging model
gUK <- gstat(NULL, "pH", pH60_100cm ~ AACN + Landsat_Band1 + Elevation + Hillshading +
    Mid_Slope_Positon + MRVBF + NDVI + TWI, cDat, model = model_1)
```

Using the validation data we can assess the performance of universal kriging using the `goof` function.

```r
vDat <- DSM_data[-training, ]
coordinates(vDat) <- ~x + y
crs(vDat) <- "+proj=utm +zone=56 +south +ellps=WGS84 + datum=WGS84 +units=m +no_defs"

crs(vDat)

## CRS arguments:
##   +proj=utm +zone=56 +south +ellps=WGS84 +units=m +no_defs

# make the predictions
UK.preds.V <- as.data.frame(krige(pH60_100cm ~ AACN + Landsat_Band1 + Elevation +
    Hillshading + Mid_Slope_Positon + MRVBF + NDVI + TWI, cDat, model = model_1,
    newdata = vDat))

## [using universal kriging]

goof(observed = DSM_data$pH60_100cm[-training], predicted = UK.preds.V[, 3])

##          R2 concordance      MSE     RMSE      bias
## 1 0.399111   0.5490607 1.091746 1.044867 0.1559361
```

The universal kriging model performs a little better than the MLR model that was fitted earlier.

Applying the universal kriging model spatially is facilitated through the `interpolate` function from `raster`. One can also use the `clusterR` function used earlier in order to speed things up a bit by applying the model over multiple compute nodes. Kriging results in two main outputs: the prediction and the prediction variance. When using the `interpolate` function we can control the output by changing the `index` parameter. The below script results in the maps on Figure 1.

```r
par(mfrow = c(1, 2))
# predictions
UK.P.map <- interpolate(hunterCovariates_sub, gUK, xyOnly = FALSE, index = 1)
plot(UK.P.map, main = "Universal kriging predictions")

# prediction variance
UK.Pvar.map <- interpolate(hunterCovariates_sub, gUK, xyOnly = FALSE, index = 2)
plot(UK.Pvar.map, main = "Universal kriging prediction variance")
```

References

Hengl, T., G. B. M. Heuvelink, and D. G. Rossiter
    2007. About regression kriging: From equations to case studies. *Computers*
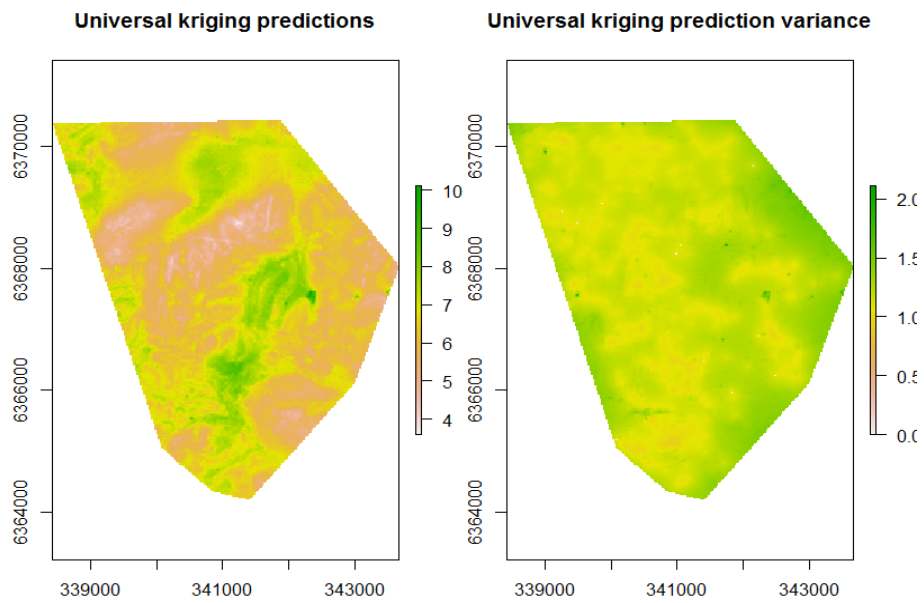
Figure 1: Universal kriging prediction and prediction variance of Hunter Valley soil pH (60-100cm).

& Geosciences, 33:1301–1315.

Odeh, I. O. A., A. B. McBratney, and D. J. Chittleborough
    1995. Further results on prediction of soil properties from terrain attributes: heterotopic co-kriging and regression kriging. *Geoderma*, 67:215–226.