RFsp-ok-sm-2012

mg

April 20, 2018

Objective: Use RFsp (from Hengl et al. 2018 https://github.com/thengl/GeoMLA) to predict 10 days of simulated soil moisture point data (0-5cm, 2D+time) across a study site in the great plains of continental US

```
library(ranger)
library(rgdal)
```

```
## Loading required package: sp
```

```
## rgdal: version: 1.2-16, (SVN revision 701)
##  Geospatial Data Abstraction Library extensions to R successfully loaded
##  Loaded GDAL runtime: GDAL 2.1.0, released 2016/04/25
##  Path to GDAL shared files: /usr/share/gdal
##  GDAL binary built with GEOS: TRUE
##  Loaded PROJ.4 runtime: Rel. 4.9.2, 08 September 2015, [PJ_VERSION: 492]
##  Path to PROJ.4 shared files: (autodetected)
##  Linking to sp version: 1.2-5
```

```
library(raster)
library(GSIF)
```

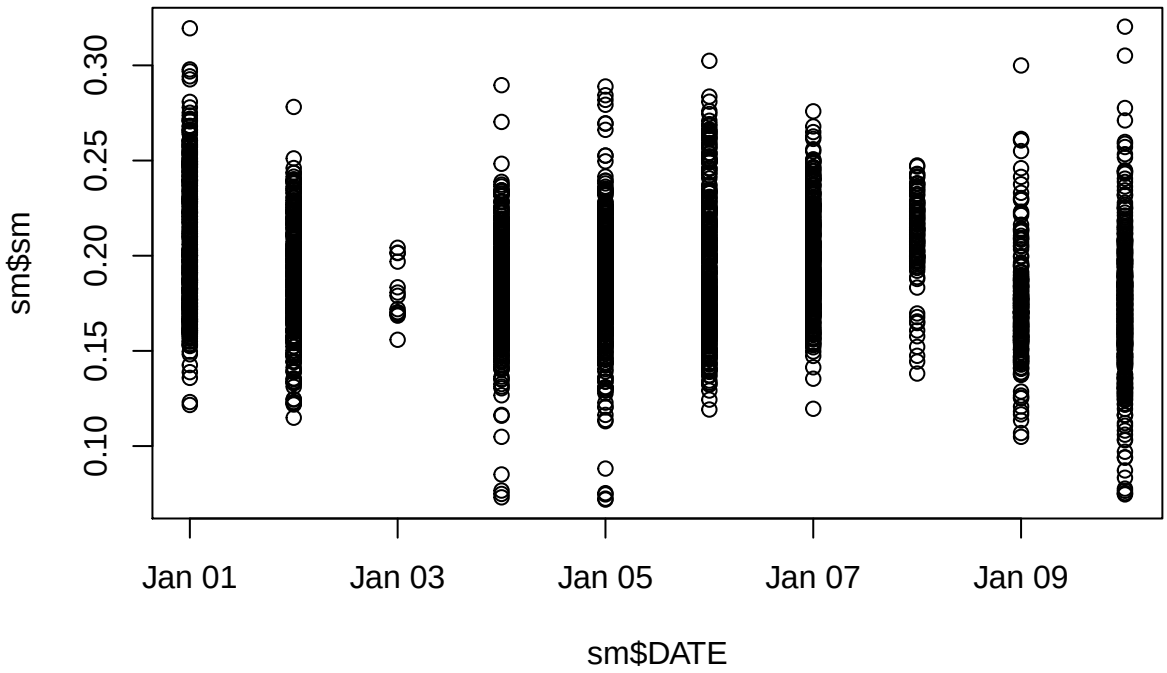
```
## GSIF version 0.5-4 (2017-04-25)
```

```
## URL: http://gsif.r-forge.r-project.org/
```

```
setwd("~/work/2012")
sm <- readRDS('sm2012.rds')
str(sm)
```

```
## 'data.frame':    2900 obs. of  5 variables:
##  $ sm       : num  0.18 0.177 0.228 0.242 0.293 ...
##  $ STATION  : int  7 8 9 10 11 12 13 14 15 16 ...
##  $ LATITUDE : num  37.1 37.1 37.1 37.1 37.1 ...
##  $ LONGITUDE: num  -102 -101 -101 -101 -101 ...
##  $ DATE     : Date, format: "2012-01-01" "2012-01-01" ...
##  - attr(*, "na.action")=Class 'omit'  Named int [1:187] 1 2 3 4 5 6 23 24 25 33 ...
##   .. ..- attr(*, "names")= chr [1:187] "1" "2" "3" "4" ...
```

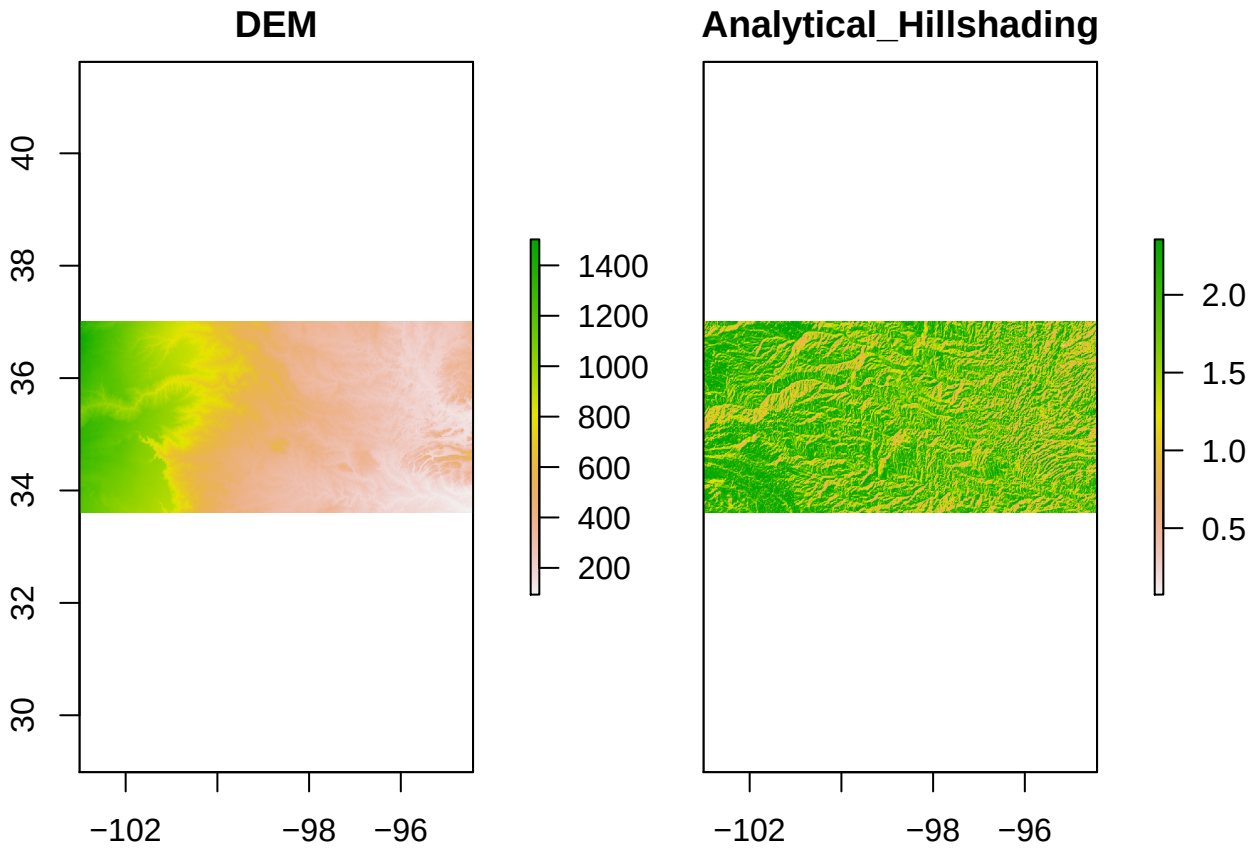
```
plot(sm$DATE, sm$sm)
```




```
(predictors <- stack(readRDS("co_grids.rds")))
```

```
## class       : RasterStack
## dimensions  : 407, 1030, 419210, 2  (nrow, ncol, ncell, nlayers)
## resolution  : 0.008333247, 0.008333857  (x, y)
## extent      : -103.0074, -94.42417, 33.61154, 37.00342  (xmin, xmax, ymin, ymax)
## coord. ref. : +proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0,0,0
## names       :         DEM, Analytical_Hillshading
## min values  : 94.00000000,             0.01868829
## max values  : 1514.000000,               2.356164
```

```
plot(predictors)
```




```
#tom code
co_prec <- sm
co_grids <- predictors
co_prec$cdate = floor(unclass(as.POSIXct(as.POSIXct(paste(co_prec$DATE), format="%Y-%m-%d")))/86400)
co_prec$doy = as.integer(strftime(as.POSIXct(paste(co_prec$DATE), format="%Y-%m-%d"), format = "%j"))
co_locs.sp = co_prec[!duplicated(co_prec$STATION),c("STATION","LATITUDE","LONGITUDE")]
coordinates(co_locs.sp) = ~ LONGITUDE + LATITUDE
proj4string(co_locs.sp) = CRS(projection(predictors))
co_grids = as(co_grids, "SpatialPixelsDataFrame")
co_locs.sp = spTransform(co_locs.sp, co_grids@proj4string)
sel.co <- over(co_locs.sp, co_grids[1])
co_locs.sp <- co_locs.sp[!is.na(sel.co$DEM),]
#grid.distP <- GSIF::buffer.dist(co_locs.sp["STATION"],
#co_grids[1],as.factor(1:nrow(co_locs.sp)))
#saveRDS(grid.distP, file='grid-distP.rds')
grid.distP <- readRDS('grid-distP.rds')
dnP <- paste(names(grid.distP), collapse="+")
ov.prec <- do.call(cbind, list(co_locs.sp@data, over(co_locs.sp, grid.distP), over(co_locs.sp, co_grids
rm.sm <- plyr::join(co_prec, ov.prec)
```

```
## Joining by: STATION
```

```
rm.sm <- na.omit(rm.sm)
```

Please note the differences in the out-of-bag cross validation form

```
#TWO MODELS
fm1 <- as.formula(paste("sm ~ ", paste(names(co_grids), collapse='+'),'+', dnP))
fm2 <- as.formula(paste("sm ~ cdate + doy +", paste(names(co_grids), collapse='+'),'+', dnP))
```

```
#NOTE THE IMPROVEMENT HERE
m1.sm <- ranger(fm1, rm.sm, importance = "impurity", num.trees = 150, mtry = 180, quantreg=TRUE)
m2.sm <- ranger(fm2, rm.sm, importance = "impurity", num.trees = 150, mtry = 180, quantreg=TRUE)
(m1.sm);(m2.sm)
```

```
## Ranger result
##
## Call:
##  ranger(fm1, rm.sm, importance = "impurity", num.trees = 150,      mtry = 180, quantreg = TRUE)
##
## Type:                             Regression
## Number of trees:                  150
## Sample size:                      2556
## Number of independent variables:  463
## Mtry:                             180
## Target node size:                 5
## Variable importance mode:         impurity
## OOB prediction error (MSE):       0.000448709
## R squared (OOB):                  0.5544961
```

```
## Ranger result
##
## Call:
##  ranger(fm2, rm.sm, importance = "impurity", num.trees = 150,      mtry = 180, quantreg = TRUE)
##
## Type:                             Regression
## Number of trees:                  150
## Sample size:                      2556
## Number of independent variables:  465
## Mtry:                             180
## Target node size:                 5
## Variable importance mode:         impurity
## OOB prediction error (MSE):       0.0002481261
## R squared (OOB):                  0.7536463
```

Here is where the problem is I can predict with no errors for model 1 but only one of the days #how to predict for all days at the same time? I can not predict for model 2 because cdate and doi are not in the covariate space, # is there any automatic RFsp way to include those covariates?

```
#works but gives only one map
sm.rfd <- predict(m1.sm,  cbind(co_grids, grid.distP)@data, type="quantiles",
                  quantiles=c(0.025, 0.5, 0.975 ))$predictions
#does not work
#sm.rfd <- predict(m2.sm,  cbind(co_grids, grid.distP)@data, type="quantiles", #quantiles=quantiles=c(0
```

please and thank you mg