```
#########################################################################################
############
## Script to read raster datasets of several variables and extract gridded values to the arsenic
## points (n=3208) for running Random Forest model to map arsenic at the national scale.
#########################################################################################
############

require(ncdf4); require(reshape); require(fields); require(maps); require(mapdata); require(zoo);
require(maptools); require(raster); require(sp); require(RColorBrewer); require(lubridate);
require(rasterVis); require(Kendall); require(modifiedmk); require(spatialEco); require(Hmisc);
require(corrplot);

# source("BD_arsenic_150m_covariate_data_extraction_29Mar20.r")

plot.pdf.file <- paste("D:\\Shams data and workshop\\All GIS files",
                 "\\Bangladesh_RF_covariate_data\\Raster data for Charlie",
                 "\\BD_arsenic_150m_covariate_data_extraction_29Mar20.pdf", sep="")

# pdf(file=plot.pdf.file)

# rm(list=ls(all=TRUE))

# Load groundwater arsenic data points

arsen.data <- read.csv("BD_shallow150m_arsenic_data.csv", header=T, sep=",")
# print(head(arsen.data))

arsen.xpos <- arsen.data$Lon
arsen.ypos <- arsen.data$Lat

grid.pts <- data.frame(cbind(X=arsen.xpos, Y=arsen.ypos))
coordinates(grid.pts) = ~ X + Y
proj4string(grid.pts) <- "+proj=longlat +datum=WGS84"

grid.pts$Arsenic <- as.numeric(arsen.data$Arsenic)

# Create some colour palletes for plotting raster datasets

my.col1 <- colorRampPalette(c("gray98","yellow","red","blue","navy"), space="rgb")
```

```
my.col2 <- rev(tim.colors(50))

# Load the country and arsenic shapefiles:

bd.outln <- readShapePoly("BD_country_polygon.shp")
proj4string(bd.outln) <- "+proj=longlat +datum=WGS84"

bd.arsen <- readShapePoints("BD_arsenic_depth_150m_data.shp", verbose=T)
proj4string(bd.arsen) <- "+proj=longlat +datum=WGS84"
bd.extn <- extent(bd.arsen)

plot(bd.arsen, pch=20, cex=0.5, col="blue")
plot(bd.outln, add=T)

# Load the gridded raster data (.img format) of covariates to extract values at arsenic points

bd.dem <- raster("bd_merit_dem_2p5km.img")                    # MERIT DEM data (m)
# print(bd.dem)
# plot(bd.dem, col=my.col1(50))

bd.rain <- raster("BD_ann_rainfall_2p5km.img")               # Annual rainfall (mm)
# print(bd.rain)
# plot(bd.rain, col=my.col1(50))

bd.dry.gwd <- raster("bd_dry_gwd_2p5km.img")                 # Depth to dry-season GW levels (m)
# print(bd.dry.gwd)

# plot(bd.dry.gwd, col=my.col1(50))

bd.sand <- raster("bd_sand_percent_2p5km.img")               # Sand percent in soil
# print(bd.sand)

# plot(bd.sand, col=my.col1(50))

bd.loam <- raster("bd_loamy_soil_grid_2p5km.img")            # Silt (loam) percent in soil
# print(bd.loam)

# plot(bd.loam, col=my.col1(50))
```

```
bd.hcond <- raster("bd_kx_ok_2p5km.img")                    # Hydraulic conductivity (m/day)
# print(bd.hcond)

# plot(bd.hcond, col=my.col1(50))

bd.usc <- raster("bd_usc_2p5km.img")                        # Surficial silt-clay deposits (m)
# print(bd.usc)

# plot(bd.usc, col=my.col1(50))

bd.rchg <- raster("mrech_85_99_2p5km.img")                  # Groundwater mean recharge (mm)
# print(bd.rchg)

# plot(bd.rchg, col=my.col1(50))

bd.rchg.trnd <- raster("rechtrnd_8599_2p5km.img")           # Groundwater recharge trend (mm/yr)
# print(bd.rchg.trnd)

# plot(bd.rchg.trnd, col=my.col1(50))

bd.irrig <- raster("gw_irri_8599_2p5km.img")                # Groundwater-fed irrigation (mm/yr)
# print(bd.irrig)

# plot(bd.irrig, col=my.col1(50))

# There are inconsistency in grid extent though the grid resolution is the same for all data

xmin <- max(bbox(bd.dem)[1,1], bbox(bd.rain)[1,1], bbox(bd.dry.gwd)[1,1], bbox(bd.sand)[1,1],
        bbox(bd.loam)[1,1], bbox(bd.hcond)[1,1], bbox(bd.usc)[1,1], bbox(bd.rchg)[1,1],
        bbox(bd.rchg.trnd)[1,1], bbox(bd.irrig)[1,1])
xmax <- min(bbox(bd.dem)[1,2], bbox(bd.rain)[1,2], bbox(bd.dry.gwd)[1,2], bbox(bd.sand)[1,2],
        bbox(bd.loam)[1,2], bbox(bd.hcond)[1,2], bbox(bd.usc)[1,2], bbox(bd.rchg)[1,2],
        bbox(bd.rchg.trnd)[1,2], bbox(bd.irrig)[1,2])
ymin <- max(bbox(bd.dem)[2,1], bbox(bd.rain)[2,1], bbox(bd.dry.gwd)[2,1], bbox(bd.sand)[2,1],
        bbox(bd.loam)[2,1], bbox(bd.hcond)[2,1], bbox(bd.usc)[2,1], bbox(bd.rchg)[2,1],
        bbox(bd.rchg.trnd)[2,1], bbox(bd.irrig)[2,1])
ymax <- min(bbox(bd.dem)[2,2], bbox(bd.rain)[2,2], bbox(bd.dry.gwd)[2,2], bbox(bd.sand)[2,2],
```

```
            bbox(bd.loam)[2,2], bbox(bd.hcond)[2,2], bbox(bd.usc)[2,2], bbox(bd.rchg)[2,2],
            bbox(bd.rchg.trnd)[2,2], bbox(bd.irrig)[2,2])

newextent <- c(xmin, xmax, ymin, ymax)

# Extent or crop does not seem to copy the raster extent beyond the data extent; try resample()

var.rain <- resample(bd.rain, bd.dem, method="bilinear")     # resample is basically aggregate

# check grid resolution: res(var.rain) == res(bd.dem)
# check extent: extent(var.rain) == extent(bd.dem)

var.dem <- bd.dem
var.rain <- resample(bd.rain, bd.dem, method="bilinear")
var.gwd <- resample(bd.dry.gwd, bd.dem, method="bilinear")
var.sand <- resample(bd.sand, bd.dem, method="bilinear")
var.loam <- resample(bd.loam, bd.dem, method="bilinear")
var.hcond <- resample(bd.hcond, bd.dem, method="bilinear")
var.usc <- resample(bd.usc, bd.dem, method="bilinear")
var.rchg <- resample(bd.rchg, bd.dem, method="bilinear")
var.rtrnd <- resample(bd.rchg.trnd, bd.dem, method="bilinear")
var.irrig <- resample(bd.irrig, bd.dem, method="bilinear")

# Now create a multi-band raster dataset (i.e. raster stack or brick)

all.vars <- stack(var.dem, var.rain, var.gwd, var.sand, var.loam, var.hcond, var.usc, var.rchg,
                  var.rtrnd, var.irrig)

names(all.vars) <- c("DEM", "RAIN", "GWD", "SAND", "LOAM", "HCOND", "USC", "RCHG", "RTRND", "IRRIG")

# writeRaster(all.vars, filename="BD_hydro_variables_raster.tif", options="INTERLEAVE=BAND",
              format="GTiff", overwrite=T)

# plot(all.vars, col=tim.colors(50))

arsen.covar <- extract(all.vars, grid.pts, method="bilinear", na.rm=T)
arsen.cov.df <- data.frame(arsen.data, arsen.covar)
# head(arsen.cov.df)
```

```
# write.csv(arsen.cov.df, "BD_arsenic_150m_n10_covariates_29Mar20.csv", row.names=F)

cur.df <- arsen.cov.df[,c(11:21)]                    # only arsenic and 10 covariate data
# head(cur.df)

# plot(arsen.cov.df$Arsenic ~ arsen.cov.df$USC, pch=20, cex=0.8, type="p")

flattenCorrMatrix <- function(cormat, pmat) {
  ut <- upper.tri(cormat)
  data.frame(
  row = rownames(cormat)[row(cormat)[ut]],
  column = rownames(cormat)[col(cormat)[ut]],
  cor  =(cormat)[ut],
  p = pmat[ut]
  )
}

data.corr <- rcorr(as.matrix(cur.df))

flattenCorrMatrix(round(data.corr$r, 6), round(data.corr$P, 6))

corrplot(data.corr$r, type="upper", order="hclust",  tl.col="black", tl.srt=45)

# plot(arsen.cov.df$Arsenic ~ arsen.cov.df$DEM, pch=20, cex=0.8, type="p")


# file.nm1 <- paste("BD_GW_arsenic_covariates_data_plots_29Ma20.pdf", sep="")
dev.dims <- dev.size(units=c("in"))
# dev.copy(pdf, file.nm1, width=dev.dims[1], height=dev.dims[2])
# dev.off()

# R codes were written by Shams at the University of Sussex to process groundwater arsenic data.
# Last modified date: 29 Mar 2020.
```