Digital soil assessment: Homosoil

Soil Security Laboratory

2018

1 Homosoil: a procedure for identifying areas with similar soil forming factors.

In many places in the world, soil information is difficult to obtain and can be non-existent. When no detailed maps or soil observations are available in a region of interest, we have to interpolate or extrapolate from other parts of the world. When dealing with global modeling at a coarse resolution, we can interpolate or extrapolate soil observations available from other similar areas (that are geographically close) or by using spatial interpolation or a spatial soil prediction function.

Homosoil, is a concept proposed by Mallavan et al. (2010) which assumes the homology of predictive soil-forming factors between a reference area and the region of interest. These include: climate, parent materials, and physiography of the area. We created the `homosoil` function to illustrate the concept. It is relatively simple whereby, given any location (latitude, longitude) in the world, the function will determine other areas in the world that share similar climate, lithology and topography. Shortly we will unpack the function into its elemental components. First we will describe the data and how to measure similarity between sites.

1.1 Global climate, lithology and topography data.

The basis is a global $0.5° × 0.5°$ grid data of climate, topography, and lithology.

For climate, this consists of variables representing long-term mean monthly and seasonal temperature, rainfall, solar radiation and evapotranspiration data. We also use the DEM representing topography, and lithology, which gives broad information on the parent material. The climate data come from the ERA-40 reanalysis and Climate Research Unit (CRU) dataset. More details on the datasets are available on the website `http://www.ipcc-data.org/obs/get_30yr_means.html`. For each of the 4 climatic variables (rainfall, temperature, solar radiation and evapotranspiration), we calculated 13 indicators: annual mean, mean for the driest month, mean at the wettest month, annual range, driest quarter mean, wettest quarter mean, coldest quarter mean, hottest quarter mean, lowest ET

quarter mean, highest ET quarter mean, darkest quarter mean, lightest quarter mean, and seasonality. From this analysis and including the acquired data, 52 global climatic variables were composed.

The DEM is from the Hydro1k dataset supplied from the USGS (`https://lta.cr.usgs.gov/HYDRO1K`), which includes the mean elevation, slope, and compound topgraphic index (CTI).

The lithology is from a global digital map (Durr et al., 2005) with 7 values which represent the different broad groups of parent materials. The lithology classes are: non- or semi-consolidated sediments, mixed consolidated sediments, silic-clastic sediments, acid volcanic rocks, basic volcanic rocks, complex of metamorphic and igneous rocks, and complex lithology.

This global data is available as a `data.frame` from the `ithir` package.

```r
library(ithir)
data(homosoil_globeDat)
```

1.2   Estimation of similarity.

The climatic and topographic similarity between two points of the grid is calculated by a Gower similarity measure (Gower, 1971). The Gower distance measures the similarity S_{ij} between sites i and j, each with p number of variables, standardized by the range of each variable:

$$S_{ij} = \frac{1}{p} \sum_{k=1}^{p} \left(1 - \frac{\mid x_{ik} - x_{jk} \mid}{rangek} \right) \tag{1}$$

Where p denotes the number of climatic variables, $\mid x_{ik} - x_{jk} \mid$ represents the absolute difference of climate variables between site i and j. The similarity index has a value between 0 and 1 and is applicable for continuous variables. For categorical variables such as those for lithology, we simply just want to match category for category between sites i and j.

Considering the scale and the resolution of this study and the available global data ($0.5° \times 0.5°$), the climatic factor is probably the most important and reliable soil forming factor. This is inspired by the study of Bui et al. (2006) who showed at continental extent, the state factors of soil formation form a hierarchy of interacting variables, with climate being the most important factor, and different climatic variables dominate in different regions. This is not to say that climate to be the most important factor at all scales. Their results also show that lithology is almost equally important in defining broad scale spatial patterns of soil properties, and shorter-range variability in soil properties appears to be driven more by terrain variables.

In the `homosoil` function, we first is to identify homoclimes around the world, and within the homoclime, we find areas with similar lithology (homolith), within which we then find similar topography (homotop). For climate and topography, we arbitrarily select the top x% similarity index as areas of homologue. In the following example we select the top 15%. So lets

look at the internals of the function before running an example.

1.3 The `homosoil` function.

Essentially from below we are creating the `homosoil` function. This function takes three inputs: `grid.data`, which is the global environmental dataset, and `recipient.lon` and `recipient.lat`, which correspond the coordinates of the sites to which we want to find soil homologues. For brevity we will call this the recipient site. Inside the `homosoil` function, we first encounter another function which is an encoding of the Gower's similarity measure as defined previously. This is followed by a number of indexation steps (to make the following steps clearer to implement), where we explicitly make groupings of the data, for example the object `grid.climate` is composed of all the global climate information from the `grid.data` object. Finally make the object `world.grid` which is a `data.frame` for putting outputs of the function into. Ultimately this object will get returned at the end of the function execution.

```
homosoil <- function (grid_data,recipient.lon,recipient.lat) {
#Gower's similarity function
gower <- function(c1, c2, r) 1-(abs(c1-c2)/r)

#index global data
grid.lon <-grid_data[,1] #longitude
grid.lat <-grid_data[,2] #latitude
grid.climate <-grid_data[,3:54]   #climate data
grid.lith <-grid_data[,58] #lithology
grid.topo <-grid_data[,55:57] #topography

#data frame to put outputs of homosoil function
world.grid<- data.frame(grid_data[, c("X", "Y")],
 fid = seq(1, nrow(grid_data), 1), homologue = 0, homoclim=0, homolith=0, homotop=0)
```

We then want to find which global grid point is the closest to the recipient site based on the Euclidean distance of the coordinates. We then want to extract the climate, lithological, and topographical data that is recorded for the nearest grid point.

```
# find the closest recipient point
dist = sqrt((recipient.lat - grid.lat)^2 + (recipient.lon - grid.lon)^2)
imin = which.min(dist)

# climate, lithology and topography for recipient site
recipient.climate <- grid.climate[imin, ]
recipient.lith <- grid.lith[imin]
recipient.topo <- grid.topo[imin, ]
```

Starting with climate, we want to estimate Gower's similarity measure. Firstly we estimate the range of values for each variable. Note the use of the `apply` function, which facilitates an efficient way to estimate ranges for each

variable. Then using our `gower` function, we perform the similarity calculation. This is done for each variable. The `mapply` function allows us to do this without the need to use a `for` loop. We then take the mean of these values, to which corresponds to the Gower's similarity measure and which is saved to the object `Sr`.

```
# range of climate variables
rv <- apply(grid.climate, 2, range)
rr <- rv[2, ] - rv[1, ]

# calculate similarity to all variables in the grid
S <- (mapply(gower, c1 = grid.climate, c2 = recipient.climate, r = rr))

Sr <- apply(S, 1, mean)  # take the average
```

We can then determine which grid points are most similar to the recipient site. Here we use an arbitrarily selected cutoff of 0.85 which corresponds to the top 15% of grid data similar to the recipient site. Lastly, we save the results of the homocline analysis to the `world.grid` object.

```
# row index for homoclime with top X% similarity.
iclim = which(Sr >= quantile(Sr, 0.85), arr.ind = TRUE)

# save homocline result
world.grid$homologue[iclim] <- 1
world.grid$homoclim[iclim] <- 1
```

Now we want to find within the areas we have defined as homocline, areas that are homolith. We simply want to find the lithology match between the recipient site and global lithology. We can do this for the entire globe, and then index those sites that also correspond to homoclines. Again, we save the results of the homolith analysis to the `world.grid` object.

```
# find within homoclime, areas with homolith
ilith = which(grid.lith == recipient.lith, arr.ind = TRUE)  #global comparison

# homolith in areas of homocline
clim.match <- which(world.grid$homologue == 1)
climlith.match <- clim.match[clim.match %in% ilith]

# save homolith result
world.grid$homologue[climlith.match] <- 2
world.grid$homolith[climlith.match] <- 1
```

Now we want to find within the areas we have defined as homolith, areas that are homotop. This analysis can be initiated by doing estimating the Gower's similarity measure for the whole globe. These steps below are just the same as before for the climate date, except now we are using the topographic data. Again we are also using the arbitrarily selected threshold value of 15%.

```
# range of topographic variables
rv <- apply(grid.topo, 2, range)
```

```
rt <- rv[2, ] - rv[1, ]

# calculate similarity of topographic variables
Sa <- (mapply(gower, c1 = grid.topo, c2 = recipient.topo, r = rt))
St <- apply(Sa, 1, mean)   # take the average

# row index for homotop
itopo = which(St >= quantile(St, 0.85), arr.ind = TRUE)
```

Now we want to determine those areas that are homotop within the areas that are homolith.

```
top.match <- which(world.grid$homologue == 2)
lithtop.match <- top.match[top.match %in% itopo]

# save homotop result
world.grid$homologue[lithtop.match] <- 3
world.grid$homotop[lithtop.match] <- 1
```

That more-or-less completes the homosoil analysis. The last few tasks are to create a raster object of the soil homologues.

```
# homologue raster object
r1 <- rasterFromXYZ(world.grid[, c(1, 2, 4)])
r1 <- as.factor(r1)
rat <- levels(r1)[[1]]
rat[["homologue"]] <- c("", "homocline", "homolith", "homotop")
levels(r1) <- rat
```

Followed by directing the `homosoil` function to save the relevant outputs which here are the `world.grid` object and the raster object of the soil homologues. Then finally we close the function.

```
retval <- list(r1, world.grid)
return(retval)}
```

1.4   Example of finding soil homologues.

With the `homosoil` function now established, lets put it to use. The coordinates below correspond to a location in Jakarta in Indonesia.

```
recipient.lat = -(6 + 10/60)
recipient.lon = 106 + 49/60
```

Now we run the `homosoil` function.

```
result <- homosoil(grid_data = homosoil_globeDat, recipient.lon = recipient.lon,
    recipient.lat = recipient.lat)
```

Then we plot the result. Here we want to use the map object that was created inside the `homosoil` function. We also specify colors to correspond the non-homologue areas, homoclines, homoliths, and homotops. Because of the

hierarchical nature of the homosoil analysis, essentially the homotops are the soil homologues to the recipient site (Figure 1.

```
# plot
area_colors <- c("#EFEFEF", "#666666", "#FFDAD4", "#FF0000")
levelplot(result[[1]], col.regions = area_colors, xlab = "", ylab = "")
 + layer(sp.points(dats,
     col = "green", pch = 20, cex = 2))
```
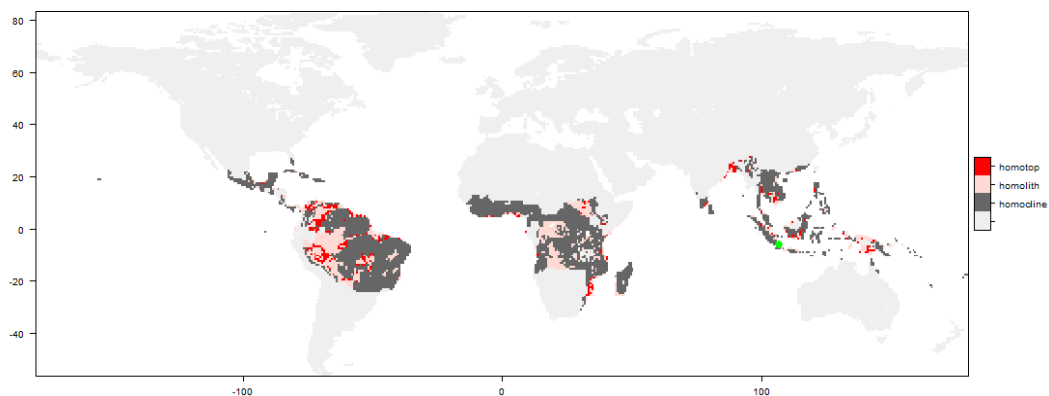

Figure 1: Soil homologues to an area of Jakarta, Indonesia (green dot on map).

Using the other object that is returned from the `homosoil` function (which was the `world.grid` data frame used to put the analysis outputs into) we can also map out the homologues individually, for example, we may just want to map the homoclines. That you can work out to do in your own time.

References

Bui, E. N., B. L. Henderson, and K. Viergever
  2006. Knowledge discovery from models of soil properties developed through data mining. *Ecological Modelling*, 191:431–446.

Durr, H. H., M. Meybeck, and S. H. Durr
  2005. Lithologic composition of the earths continental surfaces derived from a new digital map emphasizing riverine material transfer. *Global Biogeochemical Cycles*, 19:GB4S10.

Gower, J. C.
  1971. A general coefficient of similarity and some of its properties. *Biometrics*, 27:857–871.

Mallavan, B. P., B. Minansy, and A. B. McBratney
  2010. *Digital Soil Mapping: Bridging Reseach, Environmental Application,*

and Operation, chapter Homosoil: A methodology for quantitative extrapolation of soil information across the globe, Pp. 137–149. New York: Springer.