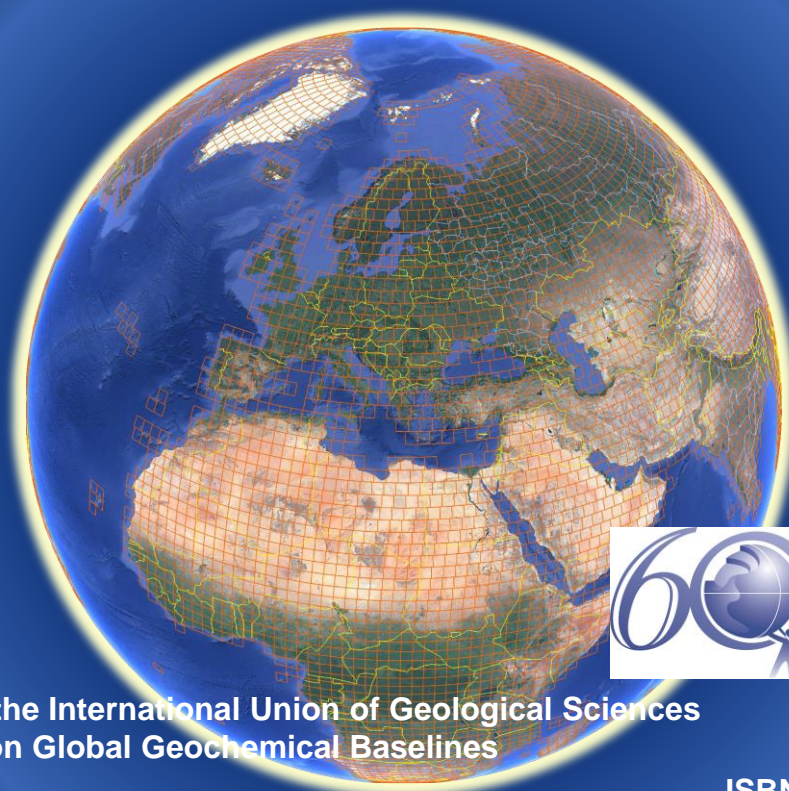


R-scripts for Generation of 5, 8 and 16 Random Sampling Points Within Predefined Rectangles

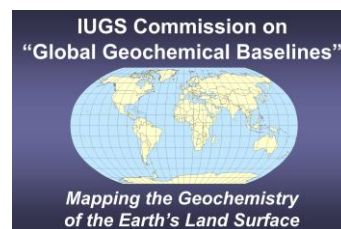
Juanxia He and Xiaoyuan Geng

International Union of Geological Sciences
Commission on Global Geochemical Baselines
Special Publication
No. 3



Published by the International Union of Geological Sciences
Commission on Global Geochemical Baselines

ISBN: 978-618-85049-2-9



**International Union of Geological Sciences
Commission on Global Geochemical Baselines**

**R-scripts for Generation of 5, 8 and 16 Random
Sampling Points Within Predefined Rectangles**

Juanxia He and Xiaoyuan Geng

Canadian Soil Information Service (CanSIS), Science and Technology Branch, Agriculture and Agri-Food Canada
IUGS Commission on Global Geochemical Baselines

2022

Published by
The International Union of Geological Sciences
Commission on Global Geochemical Baselines

Front and back cover images designed by Alecos Demetriades, Institute of Geology and Mineral Exploration, Athens, Hellenic Republic & IUGS Commission on Global Geochemical Baselines

It is recommended that reference to this publication should be made in the following way:

He, J. & Geng, X., 2022. *R-scripts for Generation of 5, 8 and 16 Random Sampling Points Within Predefined Rectangles*. IUGS Commission on Global Geochemical Baselines, Athens, Hellenic Republic, Special Publication, **3**, vi, 14 pp., 5 Figures, 3 Appendices, Supplementary material, ISBN: 978-618-85049-2-9.

This publication presents the R-scripts for generating 5, 8 and 16 random points within predefined rectangles for planning sampling campaigns.

Editorial handling: Alecos Demetriades
Institute of Geology and Mineral Exploration, Athens, Hellenic Republic
IUGS Commission on Global Geochemical Baselines

Published by
The International Union of Geological Sciences
Commission on Global Geochemical Baselines
P.O.Box 640 37, Zografou
GR-157 10 Athens
Hellenic Republic
<http://www.globalgeochemicalbaselines.eu/>
<http://www.iugs.org/>

National Library of Greece Cataloguing in Publication Data

A catalogue record for this electronic book is available from the
National Library of Greece

ISBN: 978-618-85049-2-9



Synopsis

The Global Geochemical Reference Network consists in total of 19,833 grid cells of 160x160 km that cover the whole globe. The land surface of the Earth is covered by 7356 grid cells (Global Terrestrial Network, GTN), and the marine surface by 12,477 grid cells (Global Marine Network, GMN). Planning the sampling campaigns in both realms requires the generation of random sampling points within each of the 160x160 km grid cell. For this purpose, three R-scripts were written for the generation of 5-, 8- and 16-random sampling points within each 160x160 km grid cell. The three R-scripts can be used to generate random sampling points within any other size rectangular grid cell. This publication provides the instructions for using the three R-scripts, the format of the input and output files, and the availability of the executable files, as well as the output files, from the website of the IUGS Commission on Global Geochemical Baselines.

Keywords: R-scripts, random sampling points, global terrestrial network, global marine network

Blank back page

Contents

Synopsis.....	iii
1. Introduction	1
2. Methodology.....	3
2.1. Data input file	3
2.2. Random point generation procedure.....	3
2.2.1. Five random points.....	3
2.2.2. Eight random points	3
2.2.3. Sixteen random points.....	4
2.3. Results.....	4
2.3.1. Five random points output file	4
2.3.2. Eight and sixteen random points output files	4
3. Global Marine Network.....	4
Supplementary material	5
References	8
Appendix 1: R-script for generating 5 random sampling points	9
Appendix 2: R-script for generating 8 random sampling points	11
Appendix 3: R-script for generating 16 random sampling points	13

Blank back page

1. Introduction

The R-scripts were developed for the generation of random sampling points within the Global Geochemical Reference Network grid cells of 160x160 km (Darnley *et al.*, 1995; Darnley, 1997; Salminen, Tarvainen *et al.*, 1998; Demetriades *et al.*, 2022). Three different R-scripts were designed to generate 5-, 8- and 16-random sampling points within each Global Terrestrial Network (GTN) grid cell of 160x160 km (Figs. 1, 2, 3, 4). Further, these R-scripts can be used to generate random sampling points within each Global Marine Network (GMN) grid cell of 160x160 km. Refer to the Supplementary material for information about the R-scripts and output files.

The R-scripts can also be used to generate random sampling points within grids (rectangles) of any size.

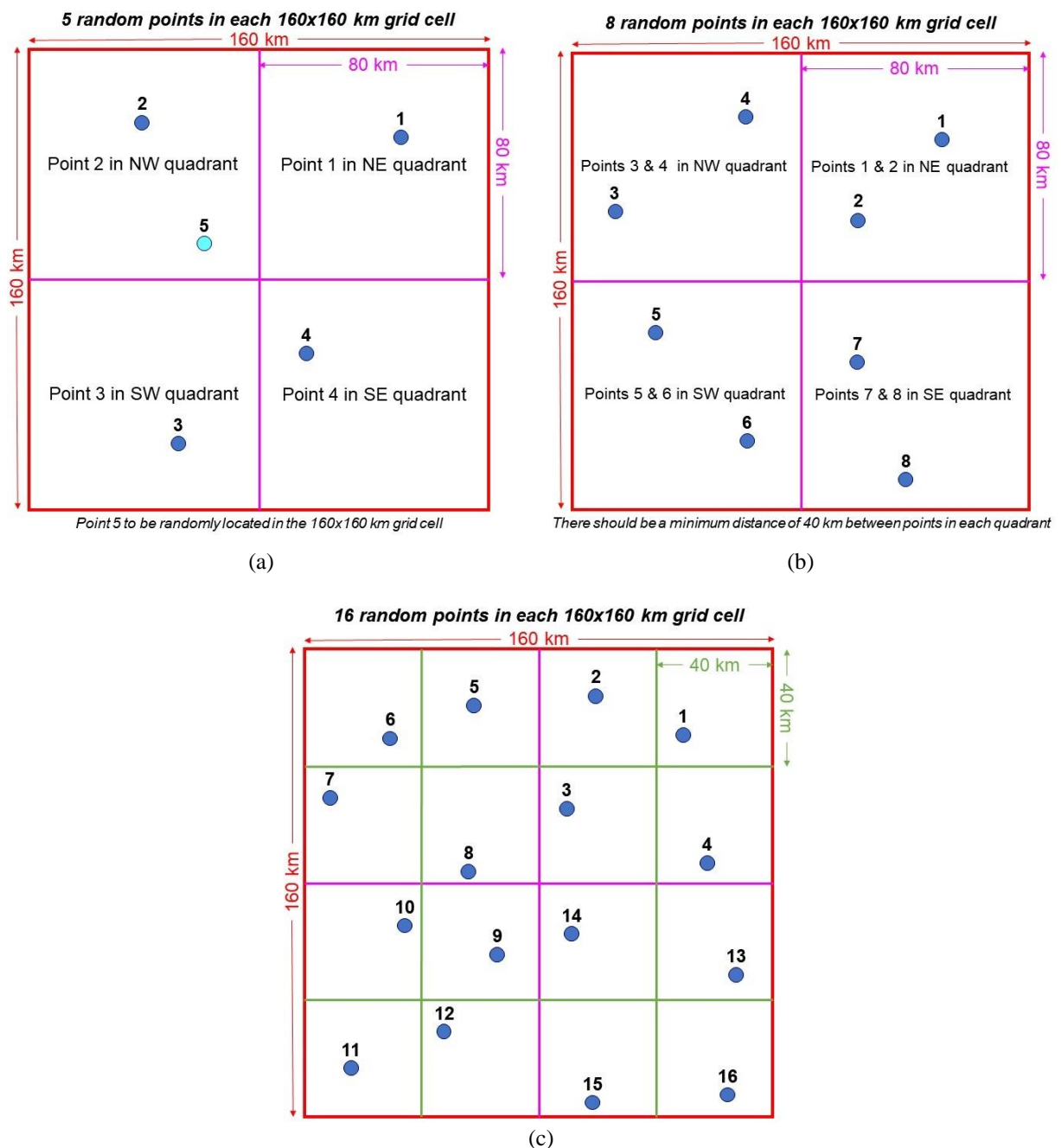


Figure 1. Random sampling schemes in each 160x160 km GTN grid cell: (a) 5-random point sampling scheme; (b) 8-random point sampling scheme, and (c) 16-random point sampling scheme.

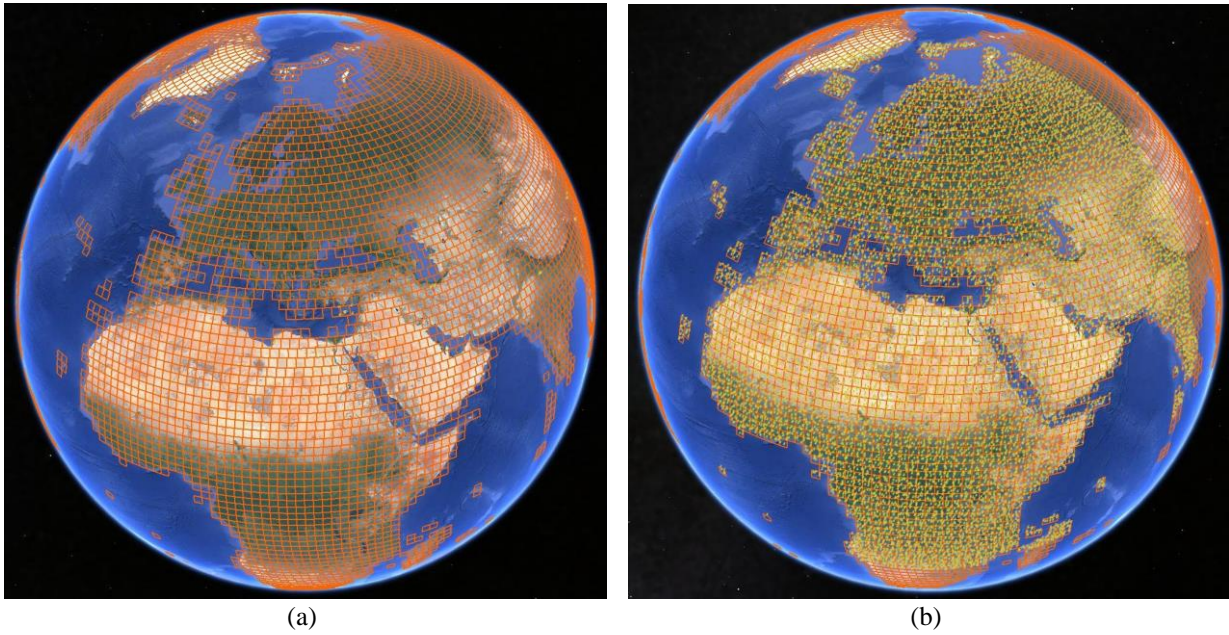


Figure 2. Google Earth images: (a) GTN grid cells 160x160 km, and (b) GTN grid cells 160x160 km with 5-random sampling points (see Fig. 3 for better definition). See also Figure 4.

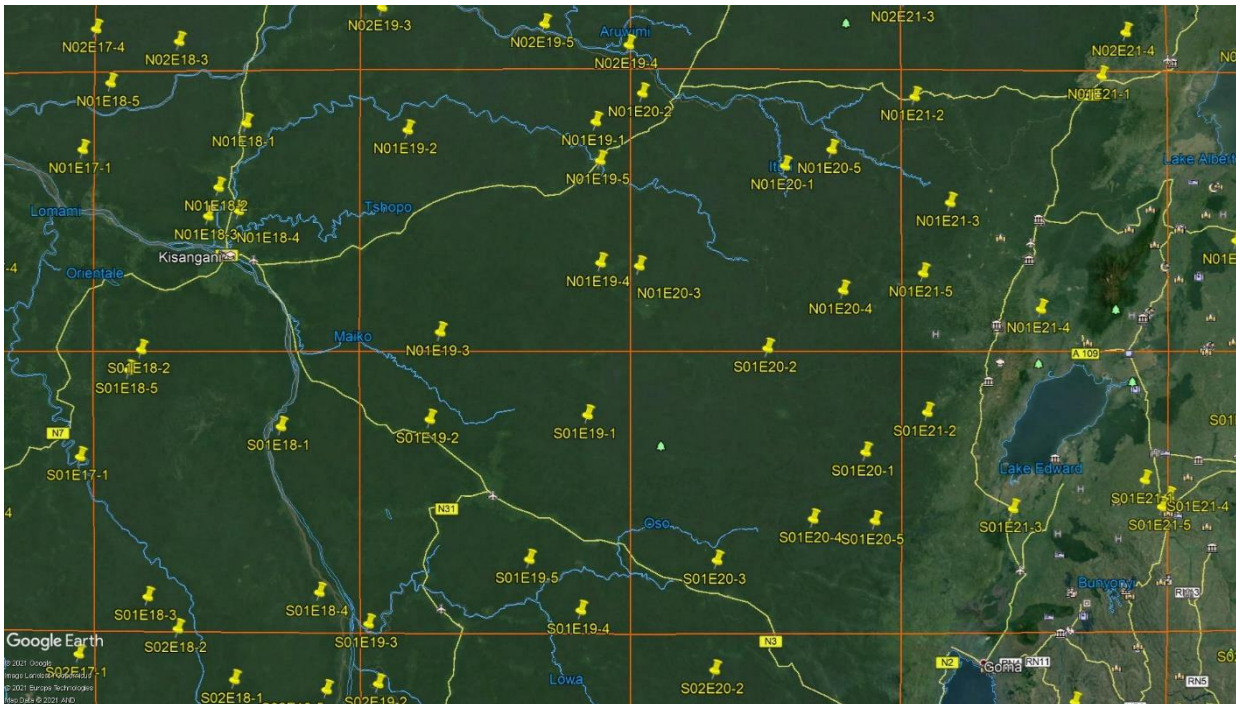


Figure 3. Google Earth image of part of the Democratic Republic of Congo in Africa showing the 5-random sampling points in each 160x160 km GTN grid cell with their sample number. Note that random sampling point No. 5 can fall in any quadrant of the 160x160 km GTN grid cell.

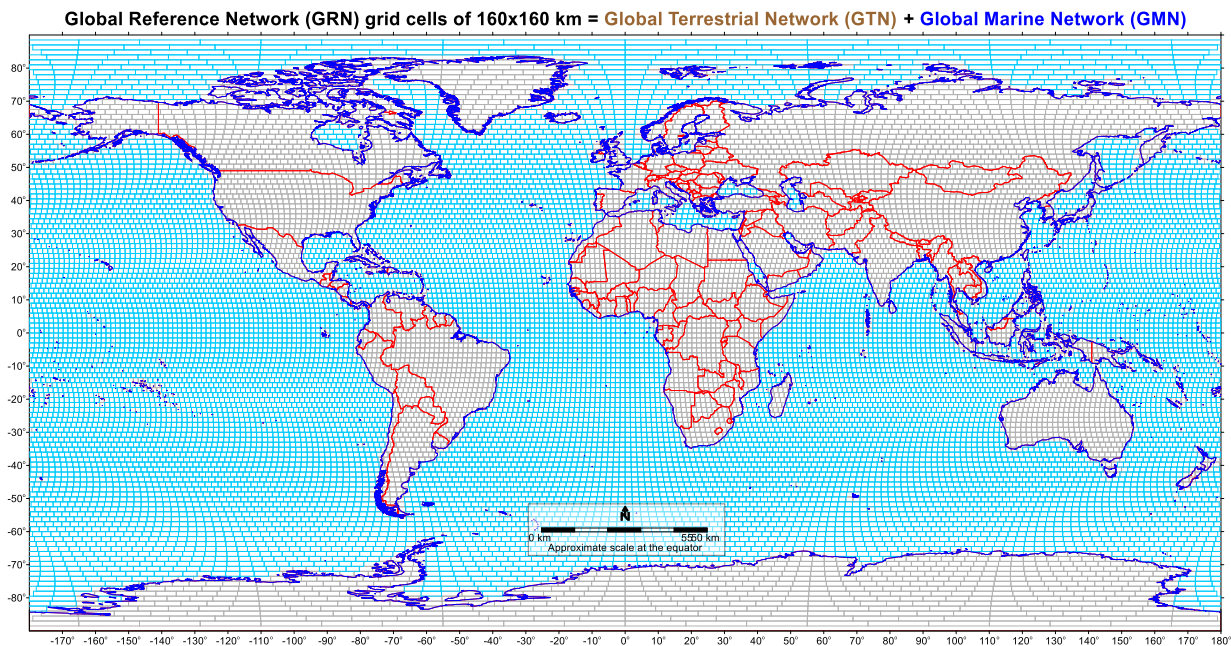


Figure 4. Map showing the 19,833 grid cells of the Global Reference Network (GRN), which consists of the 7356 Global Terrestrial Network (GTN) grid cells and the 12,477 Global Marine Network (GMN) grid cells, each of 160x160 km.

2. Methodology

2.1. Data input file

The original grid cell information is captured within a KML file. In this case, the Global Terrestrial Network (GTN) file with 7356 grid cells of 160x160 km can be downloaded from the website of the IUGS Commission on Global Geochemical Baselines, namely zip file “*IUGS-CGGB_7356_GTN_grid_cells_160x160km.zip*” (refer to Supplementary material). When unzipped the file name is: *IUGS-CGGB_7356_GTN_160x160km_grid_cells.kml*. As already noted, any KML file with a rectangular grid of any size can be the input file.

2.2. Random point generation procedure

2.2.1. Five random points

The following procedure (R script: *randomSampling5PointsOnGrids.R*) is used for the generation of five random points within each 160x160 km grid cell (Fig. 1a):-

- Load the GTN grid KML file.
- Divide the 160x160 km grid cell into 4 quadrants of 80x80 km.
- Randomly sample one point in each of the four 80x80 km quadrants, and
- Randomly sample one point in each of the 160x160 km grid cells (this is the fifth random sampling point that can fall in any quadrant).

2.2.2. Eight random points

The following procedure (R script: *randomSampling8PointsOnGrids.R*) is used for the generation of eight random points within each 160x160 km grid cell (Fig. 1b):-

- Load the GTN grid KML file.

- Divide the 160x160 km grid cell into 4 quadrants of 80x80 km, and
- Randomly sample two points (40 km apart) in each of the four 80x80 km quadrants.

2.2.3. Sixteen random points

The following procedure (R script: *randomSampling16PointsOnGrids.R*) is used for the generation of sixteen random points within each 160x160 km grid cell (Fig. 1c):-

- Load the GTN grid KML file.
- Divide the 160x160 km grid cell into 16 sub-cells of 40x40 km, and
- Randomly sample one point in each of the 16 sub-cells of 40x40 km.

2.3. Results

In the following sections, only the generation of random sampling points in the Global Terrestrial Network 160x160 km grid cells is described. The generation of random sampling points in the Global Marine Network grid cells is similar and reported in the Supplementary material.

2.3.1. Five random points output file

The original Terrestrial KML file has 7356 GTN grid cells of 160x160 km. So, in total 36,780 (=7356x5) random sampling points are generated, and are stored in three files with different format:

- a KML file named *sample_5_per_grid.kml*
- a shapefile named *sample_5_per_grid.shp*, and
- a CSV table named *sample_5_per_grid.csv*.

The coordinate system of the two output files is WGS 84, same as the KML file has.

The columns in the CSV table and the attribute table of the shapefile include:

- ID: the name of the 160x160 km GTN grid cell (*e.g.*, N21E39)
- Name: country name
- X: Longitude of a random sampling point
- Y: Latitude of a random sampling point, and
- Random point number.

2.3.2. Eight and sixteen random points output files

The output files for the 8- and 16-random points in each 160x160 km GTN grid cell have the same format as that of the 5-random point output file. The only difference is the total number of random points generated, *i.e.*, 58,848 and 117,696 points for the 8- and 16-random point sampling schemes, respectively.

3. Global Marine Network

Some GMN grid cells are located in more than one country's marine jurisdiction. Since not all countries have defined their continental shelves and Exclusive Economic Zones, country names are not given for each grid cell and random point.

Of interest may be the Marine Regions Organisation's database for the Global Exclusive Economic Zones according to the 1982 United Nations Convention on the Law of the Sea over which a sovereign state has special rights regarding the exploration and use of marine resources,

including energy production from water and wind. For further information consult the Marine Regions Organisation website: <https://www.marineregions.org/>.

Supplementary material

The following files can be downloaded from different web pages of the IUGS Commission on Global Geochemical Baselines (<http://www.globalgeochemicalbaselines.eu/>): -

(a) Executable R-scripts:

- randomSampling5PointsOnGrids.R (R script for the generation of 5 random points).
- randomSampling8PointsOnGrids.R (R script for the generation of 8 random points).
- randomSampling16PointsOnGrids.R (R script for the generation of 16 random points).

The source code of the 5, 8 and 16 random sampling points are in Appendices 1, 2 and 3, respectively. The three executable R-scripts for the generation of 5-, 8- and 16-random points in predefined rectangles on the terrestrial surface are in a zip file: 'IUGS-CGGB_R-scripts_Generation_5-8-16_random_points.zip', which can be downloaded from the '[Publications/Software programs](#)' web page.

(b) Global Reference Network grid cells of 160x160 km files:

- IUGS-CGGB_19833_GRN_160x160km_grid_cells.xlsx (Microsoft™ Excel workbook file with the 19,833 grid cells of the Global Reference Network each of 160x160 km).
- IUGS-CGGB_GRN_19833_160x160km_grid_cells.zip (same as above but in a Google Earth kml file format). When unzipped the file name is: *IUGS-CGGB_19833_GRN_160x160km_grid_cells.kml*.

The above two files can be downloaded from the '[Sampling/GTN 160x160 km](#)' web page.

(c) Global Terrestrial Network grid cells of 160x160 km files:

- IUGS-CGGB_7356_GTN_160x160km_grid_cells.xlsx (Microsoft™ Excel workbook file with the 7356 grid cells of the Global Terrestrial Network each of 160x160 km).
- IUGS-CGGB_7356_GTN_grid_cells_160x160km.zip (same as above but in a Google Earth kml file format). When unzipped the file name is: *IUGS-CGGB_7356_GTN_grid_cells_160x160km.kml*.

The above two files can be downloaded from the '[Sampling/GTN 160x160 km](#)' web page.

(d) Randomly selected sampling points in each Global Terrestrial Network grid cell:

- IUGS-CGGB_5-random-points_per_GTN_160x160km-grid-cell.xlsx (Microsoft™ Excel workbook file with 5 random points in each of the 7356 grid cells of 160x160 km totalling 36,780 random sampling points).
- IUGS-CGGB_5-random-points_per_GTN_160x160km-grid-cell.zip (same as above but in a Google Earth kml file format). When unzipped the file name is: *IUGS-CGGB_5-random-points_per_GTN_160x160km-grid-cell.kml*.
- IUGS-CGGB_8-random-points_per_GTN_160x160km-grid-cell.xlsx (Microsoft™ Excel workbook file with 8 random points in each of the 7356 grid cells of 160x160 km totalling 58,848 random sampling points).

- [IUGS-CGGB_8-random-points_per_GTN_160x160km-grid-cell.zip](#) (same as above but in a Google Earth kml file format). When unzipped the file name is: *IUGS-CGGB_8-random-points_per_GTN_160x160km-grid-cell.kml*.
- [IUGS-CGGB_16-random-points_per_GTN_160x160km-grid-cell.xlsx](#) (Microsoft™ Excel workbook file with 16 random points in each of the 7356 grid cells of 160x160 km totalling 117,696 random sampling points). *It is noted that this sampling density is not recommended for the Global Geochemical Reference Network.*
- [IUGS-CGGB_16-random-points_per_GTN_160x160km-grid-cell.zip](#) (same as above but in a Google Earth kml file format). When unzipped the file name is: *IUGS-CGGB_16-random-points_per_GTN_160x160km-grid-cell.kml*.

The above six files can be downloaded from the ‘[Sampling/Sampling Design](#)’ web page.

(e) Global Marine Network grid cells of 160x160 km:

- [IUGS-CGGB_12477_GMN_160X160km_grid_cells.xlsx](#) (Microsoft™ Excel workbook file with the 12,477 grid cells of the Global Marine Network each of 160x160 km).
- [IUGS-CGGB_12477_GMN_160x160km_grid_cells.zip](#) (same as above but in a Google Earth kml file format. When unzipped the file name is: *IUGS-CGGB_12477_GMN_160x160km_grid_cells.kml*).

The above two files can be downloaded from the ‘[Sampling/GMN 160 x 160 km](#)’ web page.

The following files are necessary for the planning of the Global Marine Network random sampling campaign:

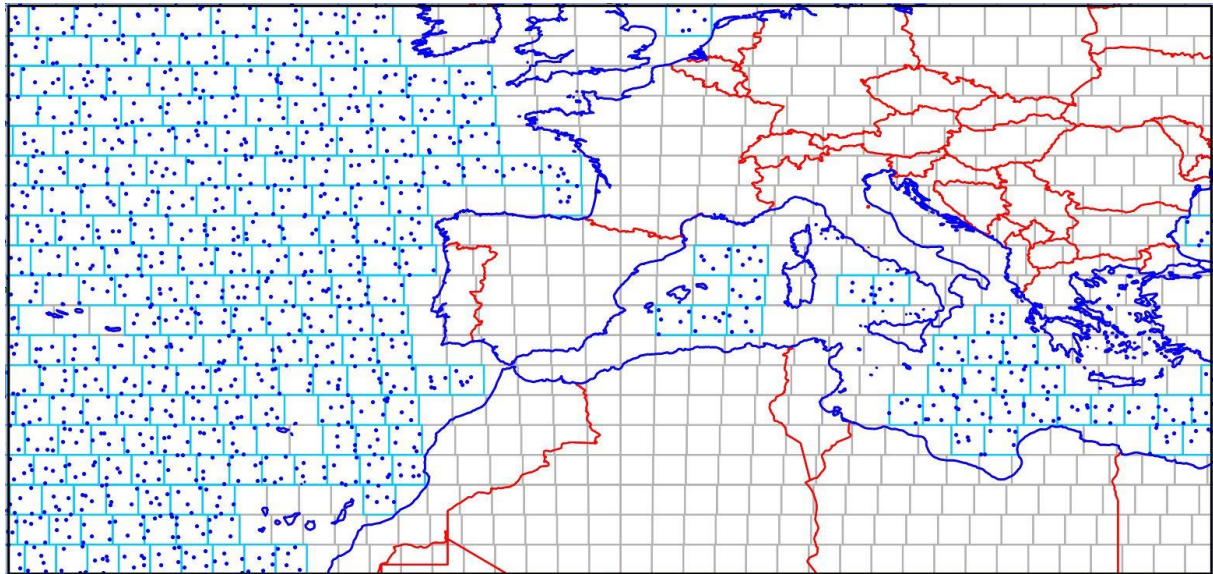
- [IUGS-CGGB_19833_GRN_160x160km_grid_cells.xlsx](#) (File is the same as in (b)).
- [IUGS-CGGB_19833_GRN_160x160km_grid_cells.zip](#) (File is the same as in (b)).
- [IUGS-CGGB_7356_GTN_160x160km_grid_cells.xlsx](#) (File is the same as in (c)).
- [IUGS-CGGB_7356_GTN_grid_cells_160x160km.zip](#) (File is the same as in (c)).
- [IUGS-CGGB_5-random-points_per_GTN_160x160km-grid-cell.xlsx](#) (File is the same as in (d)).
- [IUGS-CGGB_5-random-points_per_GTN_160x160km-grid-cell.zip](#) (File is the same as in (d)).
- [IUGS-CGGB_8-random-points_per_GTN_160x160km-grid-cell.xlsx](#) (File is the same as in (d)).
- [IUGS-CGGB_8-random-points_per_GTN_160x160km-grid-cell.zip](#) (File is the same as in (d)).

The reason that the Global Terrestrial Network files are required for the planning of the Global Marine Network random sampling is pictorially explained in Figure 5. Figure 5a shows the five random sampling points in the marine realm, and Figure 5b the five random sites within the 160x160 km grid cells of both the marine and terrestrial parts. It is obvious from these two figures that some of the random sampling points in coastal GTN grid cells fall in the sea (or water bodies in general). As these coastal GTN grid cells are not in the GMN database, the only way to have a complete set of random sampling points in the marine real is to plot on the same map the two databases.

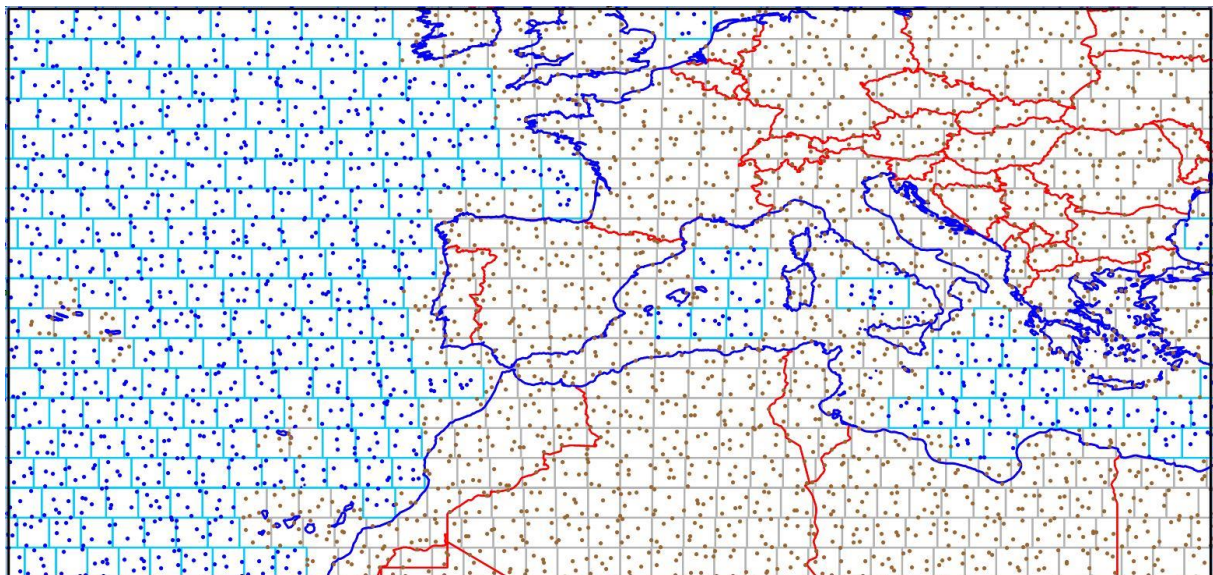
(f) Randomly selected points in each Global Marine Network grid cell:

- [IUGS-CGGB_5-random-points_per_GMN_160x160km-grid-cell.xlsx](#) (Microsoft™ Excel workbook file with 5 random points in each of the 12,477 grid cells of 160x160 km totalling 62,385 random sampling points).
- [IUGS-CGGB_8-random-points_per_GMN_160x160km-grid-cell.xlsx](#) (Microsoft™ Excel workbook file with 8 random points in each of the 12,477 grid cells of 160x160 km totalling 99,816 random sampling points).

The above two files can be downloaded from the '[Sampling/GMN 160 x 160 km](#)' web page.



(a)



(b)

Figure 5. (a) The map shows the 5 random sampling points (blue dots) within the 160x160 km GMN grid cells in the marine realm, and (b) the random sampling points (brown dots) falling on land and sea in coastal GTN grid cells.

References

- Darnley, A.G., 1997. A global geochemical reference network: the foundation for geochemical baselines. *Journal of Geochemical Exploration*, 60(1),1–5; [https://doi.org/10.1016/S0375-6742\(97\)00020-4](https://doi.org/10.1016/S0375-6742(97)00020-4).
- Darnley, A.G., Björklund, A., Bølviken, B., Gustavsson, N., Koval, P.V., Plant, J.A., Steenfelt, A., Tauchid, M., Xie, X., Garrett, R.G. & Hall, G.E.M., 1995. A Global Geochemical Database for Environmental and Resource Management. Recommendations for International Geochemical Mapping – Final Report of IGCP Project 259. Earth Science Report 19. UNESCO Publishing, Paris, 122 pp.; https://www.globalgeochemicalbaselines.eu/datafiles/file/Blue_Book_GGD_IGCP259.pdf
- Demetriades, A., Johnson, C.C., Smith, D.B. & Batista, M.J., 2022. Global Terrestrial Network Grid Cells, Selection of Sample Sites and Sample Types to be Collected. Chapter 2 In: Demetriades, A., Johnson, C.C., Smith, D.B., Ladenberger, A., Adánez Sanjuan, P., Argyraki, A., Stouraiti, C., Caritat, P. de, Knights, K.V., Prieto Rincón, G. & Simubali, G.N. (Editors), [International Union of Geological Sciences Manual of Standard Methods for Establishing the Global Geochemical Reference Network](#). IUGS Commission on Global Geochemical Baselines, Athens, Hellenic Republic, Special Publication, 2, 11–25; available from the [Publications](#) web page of the IUGS Commission on Global Geochemical Baselines.
- Salminen, R., Tarvainen, T., Demetriades, A., Duris, M., Fordyce, F.M., Gregorauskiene, V., Kahelin, H., Kivisilla, J., Klaver, G., Klein, P., Larson, J.O., Lis, J., Locutura, J., Marsina, K., Mjartanova, H., Mouvet, C., O'Connor, P., Odor, L., Ottonello, G., Paukola, T., Plant, J.A., Reimann, C., Schermann, O., Siewers, U., Steenfelt, A., Van Der Sluys, J. & Williams, L., 1998. FOREGS Geochemical Mapping Field Manual. Geological Survey of Finland, Espoo, Guide 47, 36 pp.; http://tupa.gtk.fi/julkaisu/opas/op_047.pdf.

Appendix 1: R-script for generating 5 random sampling points

```
(randomSampling5PointsOnGrids.R)
```

```
## Acknowledgements:
```

```
## Juanxia He and Xiaoyuan Geng (Canadian Soil Information Service,  
## Science and Technology Branch, Agriculture and Agri-Food Canada,  
## http://sis.agr.gc.ca/) developed this script for the generation  
## of 5 random points within the 160x160 km GTN & GMN grid cells.
```

```
## The script can be used to randomly sample one point  
## from each polygon of a selected KML file (via a pop-up window)  
## and one point from each of the four quadrants of each polygon in the KML file
```

```
## input:
```

```
## kml.filename: a KML file
```

```
## output:
```

```
## a KML file in the folder where the input kml file is located
```

```
## a CSV file
```

```
## a shapefile
```

```
## four columns in the attribute table: GTNID, country, long (i.e., Longitude), lat (i.e., Latitude),  
## and ptNum (i.e., Random point number)
```

```
# the required R packages.
```

```
require(rgdal)
```

```
require(maptools)
```

```
# function to divide a rectangle to four quadrants
```

```
subPolygons <- function(xmin, xmax, ymin, ymax){
```

```
  xmid <- (xmin + xmax) / 2
```

```
  ymid <- (ymin + ymax) / 2
```

```
  p.1234.list <- list()
```

```
  p.1234.list[[1]] <- Polygon(cbind(c(xmid, xmax, xmax, xmid), c(ymax, ymax, ymid, ymid))) #  
  NE
```

```
  p.1234.list[[2]] <- Polygon(cbind(c(xmin, xmid, xmid, xmin), c(ymax, ymax, ymid, ymid))) #  
  NW
```

```
  p.1234.list[[3]] <- Polygon(cbind(c(xmin, xmid, xmid, xmin), c(ymid, ymid, ymin, ymin))) #  
  SW
```

```
  p.1234.list[[4]] <- Polygon(cbind(c(xmid, xmax, xmax, xmid), c(ymid, ymid, ymin, ymin))) #  
  SE
```

```
  return(p.1234.list)
```

```
}
```

```
kml.filename <- file.choose() # set the kml via a popup window
```

```
setwd(dirname(kml.filename))
```

```
kml.filename
```

```
grid160 <- readOGR(kml.filename, stringsAsFactors = F)
```

```
for (i in 1: nrow(grid160@data)){
```

```

poly1 <- grid160@polygons[[i]]

ext <- bbox(poly1) # extent of the polygon/rectangle/grid
poly1.grid <- subPolygons(ext[1, 1], ext[1, 2], ext[2, 1], ext[2, 2])

# please note that although the coordinate of the grids is longitude/latitude, here we assume that
  the coordinate is planar
sample.1 <- lapply(poly1.grid, function(x){
  spsample (x, n = 1, "random")
})

sample.1 <- do.call("rbind", sample.1)
repeat {
  sample.2 <- spRbind(sample.1, spsample(poly1, n = 1, "random"))
  # to ensure that the sampled point from the grid is not overlapped with the other 4 sampled
  points
  if (sum(duplicated(sample.2@coords)) == 0) {break}
}

sample.2$GTNID <- as.character(rep(grid160@data[i, c("Name")], 5)) # names of the grid160
  polygons
sample.2$country <- as.character(grid160@data[i, c("Description")], 5) # country names
sample.2$long <- sample.2@coords[, 1]
sample.2$lat <- sample.2@coords[, 2]
sample.2$ptNum <- seq(1, 5)

# put all the samples from all the polygons to one
if (i == 1) {
  sample.all <- sample.2
} else {
  sample.all <- spRbind(sample.all, sample.2)
}
cat(i)
cat("-")
}

sample.all@proj4string <- CRS(proj4string(grid160)) # set coordinate system, same as the kml
  grids
# output to KML
writeOGR(sample.all, dsn = "sample_5_per_grid.kml", layer= "sample_5_per_grid.kml", driver
  = "KML")
# output to csv table
write.csv(sample.all@data, file = "sample_5_per_grid.csv", quote = F, row.names = F)
# output to ESRI shapefile
writeOGR(sample.all, dsn = getwd(), layer = "sample_5_per_grid", driver = "ESRI Shapefile" )
cat("done")

```

Appendix 2: R-script for generating 8 random sampling points

```
(randomSampling8PointsOnGrids.R)
```

```
## Acknowledgements:
```

```
## Juanxia He and Xiaoyuan Geng (Canadian Soil Information Service,  
## Science and Technology Branch, Agriculture and Agri-Food Canada,  
## http://sis.agr.gc.ca/) developed this script for the generation  
## of 8 random points within the 160x160 km GTN and GMN grid cells.
```

```
## The script can be used to randomly sample 8 points  
## from each polygon of a selected KML file (via a pop-up window)  
## (two points from each 80*80km quadrant & minimum 40 km distance between these two  
## points)
```

```
## input:
```

```
## kml.filename: a KML file
```

```
## output:
```

```
## a KML file in the folder where the input kml file is located  
## a CSV file  
## a shapefile  
## four columns in the attribute table: GTNID, Country, long (i.e., Longitude), lat (i.e.,  
## Latitude), and ptNum (i.e., Random point number)
```

```
# the required R packages.
```

```
require(rgdal)  
require(maptools)  
require(geosphere)
```

```
# function to divide a rectangle to four quadrants
```

```
subPolygons <- function(xmin, xmax, ymin, ymax){  
  xmid <- (xmin + xmax) / 2  
  ymid <- (ymin + ymax) / 2  
  p.1234.list <- list()  
  p.1234.list[[1]] <- Polygon(cbind(c(xmid, xmax, xmax, xmid), c(ymax, ymax, ymid, ymid))) #  
  NE  
  p.1234.list[[2]] <- Polygon(cbind(c(xmin, xmid, xmid, xmin), c(ymax, ymax, ymid, ymid))) #  
  NW  
  p.1234.list[[3]] <- Polygon(cbind(c(xmin, xmid, xmid, xmin), c(ymid, ymid, ymin, ymin))) #  
  SW  
  p.1234.list[[4]] <- Polygon(cbind(c(xmid, xmax, xmax, xmid), c(ymid, ymid, ymin, ymin))) #  
  SE  
  return(p.1234.list)  
}
```

```
kml.filename <- file.choose() # set the kml via a popup window  
setwd(dirname(kml.filename))  
kml.filename # show the kml file name & path in R Console window
```

```

grid160 <- readOGR(kml.filename, stringsAsFactors = F)

for (i in 1: nrow(grid160@data)){
  poly1 <- grid160@polygons[[i]]

  ext <- bbox(poly1) # extent of the polygon/rectangle
  poly1.grid <- subPolygons(ext[1, 1], ext[1, 2], ext[2, 1], ext[2, 2])

  # please note that although the coordinate of the grids is longitude/latitude, here we assume that
  # the coordinate is planar
  sample.1 <- lapply(poly1.grid, function(x){
    dis.t <- 0
    while (dis.t <= 40000) {
      sample.t <- spsample(x, n = 2, "random")
      dis.t <- distGeo(sample.t@coords[1,], sample.t@coords[2,]) # distance should be > 40km
      (geodesic distance)
    }
    return(sample.t)
  })

  sample.1 <- do.call("rbind", sample.1)

  sample.1$GTNID <- as.character(rep(grid160@data[i, c("Name")], 8)) # names of the grid160
  # polygons
  sample.1$country <- as.character(grid160@data[i, c("Description")], 8) # country names, the
  # field name could be different
  sample.1$long <- sample.1@coords[, 1]
  sample.1$lat <- sample.1@coords[, 2]
  sample.1$ptNum <- seq(1, 8) #

  if (i == 1) {
    # put all the samples from all the polygons to one
    sample.all <- sample.1
  } else{
    sample.all <- spRbind(sample.all, sample.1)
  }
  cat(i)
  cat("-")
}

sample.all@proj4string <- CRS(proj4string(grid160)) # set coordinate system, same as the input
# KML
# output to KML
writeOGR(sample.all, dsn="sample_8_per_grid.kml", layer= "sample_8_per_grid", driver =
  "KML")
# output to csv table
write.csv(sample.all@data, file = "sample_8_per_grid.csv", quote = F, row.names = F)
# output to ESRI shapefile
writeOGR(sample.all, dsn = getwd(), layer = "sample_8_per_grid", driver = "ESRI Shapefile" )
cat("done")

```

Appendix 3: R-script for generating 16 random sampling points

```
(randomSampling16PointsOnGrids.R)
```

```
## Acknowledgements:
```

```
## Juanxia He and Xiaoyuan Geng (Canadian Soil Information Service,  
## Science and Technology Branch, Agriculture and Agri-Food Canada,  
## http://sis.agr.gc.ca/) developed this script for the generation  
## of 16 random points within the 160x160 km GTN and GMN grid cells.
```

```
## The script can be used to randomly sample 16 points  
## from each polygon of a selected KML file (via a pop-up window)  
## (one point from each 40*40km cell)
```

```
## input:
```

```
## kml.filename: a KML file
```

```
## output:
```

```
## a KML file in the folder where the input kml file is located  
## a CSV file  
## a shapefile  
## four columns in the attribute table: GTNID, Country, long (i.e., Longitude), lat (i.e.,  
## Latitude), and ptNum (i.e., Random point number)
```

```
# the required R packages.
```

```
require(rgdal)
```

```
require(maptools)
```

```
# function to divide a rectangle to four quadrants
```

```
subPolygons <- function(xmin, xmax, ymin, ymax){
```

```
  xmid <- (xmin + xmax) / 2
```

```
  ymid <- (ymin + ymax) / 2
```

```
  p.1234.list <- list()
```

```
  p.1234.list[[1]] <- Polygon(cbind(c(xmid, xmax, xmax, xmid), c(ymax, ymax, ymid, ymid))) #  
NE
```

```
  p.1234.list[[2]] <- Polygon(cbind(c(xmin, xmid, xmid, xmin), c(ymax, ymax, ymid, ymid))) #  
NW
```

```
  p.1234.list[[3]] <- Polygon(cbind(c(xmin, xmid, xmid, xmin), c(ymid, ymid, ymin, ymin))) #  
SW
```

```
  p.1234.list[[4]] <- Polygon(cbind(c(xmid, xmax, xmax, xmid), c(ymid, ymid, ymin, ymin))) #  
SE
```

```
  return(p.1234.list)
```

```
}
```

```
kml.filename <- file.choose() # set the kml via a popup window
```

```
setwd(dirname(kml.filename))
```

```
kml.filename # show the kml file name & path in R Console window
```

```
grid160 <- readOGR(kml.filename, stringsAsFactors = F) # converting
```

```
for (i in 1: nrow(grid160@data)){
```

```

poly1 <- grid160@polygons[[i]]

ext <- bbox(poly1) # extent of the polygon/rectangle
poly1.grid <- subPolygons(ext[1, 1], ext[1, 2], ext[2, 1], ext[2, 2]) # divide a grid to four
quadrants (the result is 80km by 80 km grids)

# please note that although the coordinate of the grids is longitude/latitude, here we assume that
the coordinate is planar
sample.1 <- lapply(poly1.grid, function(x) {
  ext <- bbox(x)
  poly.t <- subPolygons(ext[1, 1], ext[1, 2], ext[2, 1], ext[2, 2]) # further divide a 80*80 grid
cell to four 40*40 cells
  lapply(poly.t, function(y) {
    spsample(y, n = 1, "random") # sample one point in a 40*40 cell
  })
})

sample.2 <- do.call("rbind", sample.1[[1]])
sample.2 <- rbind(sample.2, do.call("rbind", sample.1[[2]]))
sample.2 <- rbind(sample.2, do.call("rbind", sample.1[[3]]))
sample.2 <- rbind(sample.2, do.call("rbind", sample.1[[4]]))

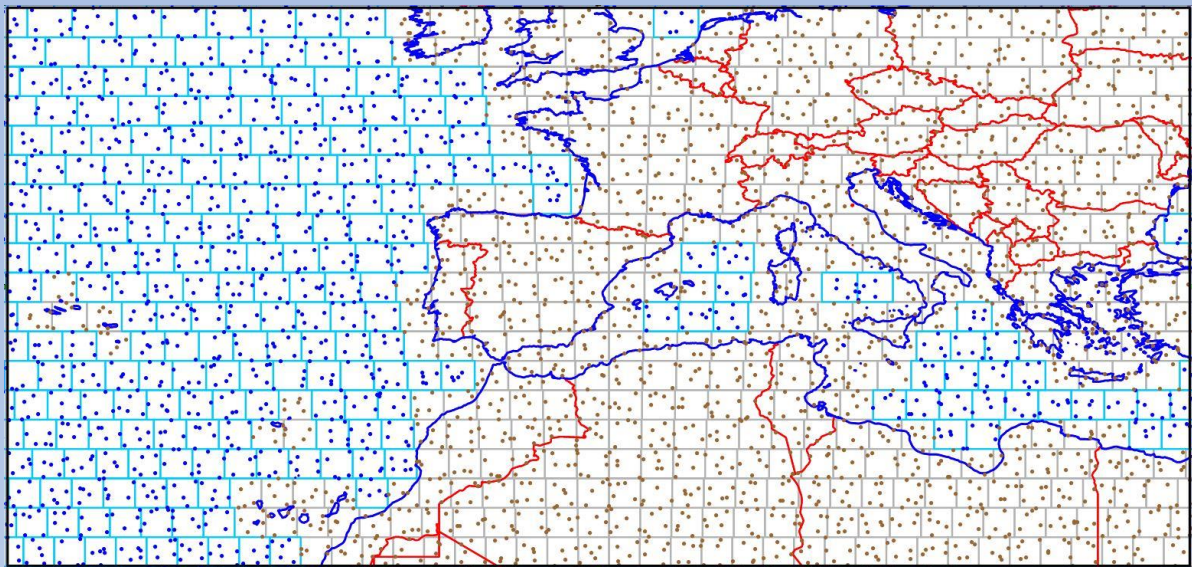
sample.2$GTNID <- as.character(rep(grid160@data[i, c("Name")], 16)) # names of the grid160
polygons
sample.2$country <- as.character(grid160@data[i, c("Description")], 16) # country names, the
field name could be different
sample.2$long <- sample.2@coords[,1]
sample.2$lat <- sample.2@coords[, 2]
sample.2$ptNum <- seq(1,16)

# put all the samples from all the polygons to one
if (i == 1) {
  sample.all <- sample.2
} else{
  sample.all <- spRbind(sample.all, sample.2)
}
cat(i)
cat("-")
}

sample.all@proj4string <- CRS(proj4string(grid160)) # set coordinate system, same as the input
KML
# output to KML
writeOGR(sample.all, dsn="sample_16_per_grid.kml", layer= "sample_16_per_grid",
driver="KML")
# output to csv table
write.csv(sample.all@data, file = "sample_16_per_grid.csv", quote = F, row.names = F)
# output to ESRI shapefile
writeOGR(sample.all, dsn = "sample_16_per_grid.shp", layer = "sample_16_per_grid", driver =
"ESRI Shapefile" )
cat("done")

```


The Global Geochemical Reference Network consists in total of 19,833 grid cells of 160x160 km that cover the whole globe. The land surface of the Earth is covered by 7356 grid cells (Global Terrestrial Network, GTN), and the marine surface by 12,477 grid cells (Global Marine Network, GMN). Planning the sampling campaigns in both realms requires the generation of random sampling points within each of the 160x160 km grid cell. For this purpose, three R-scripts were written for the generation of 5-, 8- and 16-random sampling points within each 160x160 km grid cell. The three R-scripts can be used to generate random sampling points within any other size rectangular grid cell. This publication provides the instructions for using the three R-scripts, the format of the input and output files, and the availability of the executable files, as well as the output files, from the website of the IUGS Commission on Global Geochemical Baselines.



ISBN: 978-618-85049-2-9

