Similarity Tool

User Manual

Code written by Marius Gilbert (ULB), adjusted by Catherine Pfeifer (ILRI)

User manual and input data: Catherine Pfeifer (ILRI), Tim Robinson (ILRI)

Contact : Catherine Pfeifer (c.pfeifer@cgiar.org)

Version : November 2016

This tool was funded by Humidtropics program.

Contents

1 Introduction

1.1 Objective

Bringing a best practice to a new location, does not automatically lead to successful adoption. It is therefore important to understand the context and assess the success factors before trying to replicate it elsewhere. Scaling out potential refers to other location where replication is likely to lead to a successful implementation of a best practice.

The objective of the similarity analysis tool is to give its users a quick and approachable access to similarity measure, allowing to assess the spatial out scaling potential of a best practice that has worked in one study site to other locations, based on user-defined success factors.

The tool comes with readily prepared success factors, as spatial layers, from which the user can select the relevant ones. These success factors reflect both the bio-physical and the socio-economic context. This user manual explains not only how to use the tool but also how to add other success factor.

1.2 Similarity measures

In its simplest term, similarity analysis tries to compare different geographical layers and assess to what extent a point in space is similar to a reference point, also referred to as distance. This computation is then done for every point is space (pixel) in order to get a map. Different methods exist to compare these layers. For this report three of them have been retained, namely (i) Euclidian similarity (ii) Mahalanobis similarity and (iii) Multivariate Environmental Similarity Surfaces.

Euclidian similarity, assess the distance by using the absolute difference, whereas Mahalanobis similarity, takes the distribution of the data into account by measuring how many standard deviations away a given value is away from the reference point. For both methods, the reference point is in fact the average of randomly sampled points within the area for which similarity is computed. Also, as the different layer could be correlated, a principal component analysis is performed and is used as input for Euclidian similarity.

Finally, Multivariate Environmental Similarity Surfaces (MESS) is the most sophisticated method. It is an index that compares the value of a point in space with the distribution of the randomly sampled reference points. It allows negative values – these are areas where at least one variable has a value that is outside the distribution of the reference points. These areas are referred to as novel environments. To assess similarity we are therefore interested in positive values of this index.

2 Set up of the tool

2.1 Structure

The similarity tool runs in the R-environment and comes with the four following folders:

- 0_Code : contains all the code needed to run the similarity analysis
- 0_Covariate : contains all the readily available geo-data that can be selected as success factor
- 1_Intputs : contains all files that need user input, namely the set up code, the study area definition as well as the set of success factors, referred as variables in the tool, to select from
- 2_Workspace : folder used to store intermediary data
- 3_ Output : folder where similarity maps are found after running the tool
- 4_AncillaryData : folder that contain other useful data to add new spatial layers.

2.2 Readily available global dataset defining success factor

The following spatial layers are ready for use

Variable name	description	Source	link
mask		Gridded livestock of the world	internal
aridity	aridity	FAO	http://www.fao.org/geonetwork/
access25	Access to place populated with more than 25000 people	Nelson 2008	http://forobs.jrc.ec.europa.eu/products/gam/
access50	with more than 50000 people	Nelson 2008	http://forobs.jrc.ec.europa.eu/products/gam/
access100	with more than 100000 people	Nelson 2008	http://forobs.jrc.ec.europa.eu/products/gam/
access500	with more than 500000 people	Nelson 2008	http://forobs.jrc.ec.europa.eu/products/gam/
cattle	Cattle distribution	Gridded livestock of the world	http://livestock.geo-wiki.org/
chicken	Chicken distribution	Gridded livestock of the world	http://livestock.geo-wiki.org/
crop	Cropping areas	IIAZA hybrid crop	http://livestock.geo-wiki.org/
goat	Goat distribution	Gridded livestock of the world	http://livestock.geo-wiki.org/
hapctnnp	Human Appropriation of Net Primary Productivity	Ciesien	http://sedac.ciesin.columbia.edu/data/collection/hanpp/maps/gallery/search?facets=region:global

lgp	Length of growing period	Thornton et al	ILRI
ndviA0	Average NDVI	Modis	ILRI
ndviA1	Standard deviation of NDVI	Modis	ILRI
pig	Pig distribution	Gridded livestock of the world	http://livestock.geo-wiki.org/
pop	population	worldpop	www.worldpop.org
rain	Mean annual precipitation	Humid tropic precipitation	Received from CIP
rainWCavg	Average annual precipitation	Worldclim 2	http://www.worldclim.org/
rainWCmax	Maximum monthly precipitation	Worldclim 2	http://www.worldclim.org/
rainWCmin	Minimum monthly precipitation	Worldclim 2	http://www.worldclim.org/
tminWCavg	Average annual minimum temperature	Worldclim 2	http://www.worldclim.org/
tminWCmax	Maximum monthly minimum temperature	Worldclim 2	http://www.worldclim.org/
tminWCmin	Minimum monthly minimum temperature	Worldclim 2	http://www.worldclim.org/
tmaxWCavg	Average annual maximum temperature	Worldclim 2	http://www.worldclim.org/
tmaxWCmax	Maximum monthly maximum temperature	Worldclim 2	http://www.worldclim.org/
tmaxWCmin	Minimum monthly maximum temperature	Worldclim 2	http://www.worldclim.org/
tavgWCavg	Average annual average temperature	Worldclim 2	http://www.worldclim.org/
tavgWCmax	Maximum monthly average temperature	Worldclim 2	http://www.worldclim.org/
tavgWCmin	Minimum monthly average temperature	Worldclim 2	http://www.worldclim.org/

rpop	Rural population	Worldpop	Population from Worldpop on non-urban area defined by Schneider https://nelson.wisc.edu/sage/data-and-models/schneider.php
sheep	Sheet distribution	Gridded livestock of the world	http://livestock.geo-wiki.org/
srtm	Elevation	CGIAR-CSI 1.4	http://www.cgiar-csi.org/data/srtm-90m-digital-elevation-database-v4-1
stunting	Percent of stunting children below 5 years	FAO	http://www.fao.org/geonetwork/

Note that base worldclim data layer are made available, but need to be added manually to the tool. These data come with the following name structure wc2.0_30s_prec_01, suggesting this is world clim2, prec suggests the precipitation and 01 stands for January. Other options are tavg for average temeperatrue, tmax for maximum temperature and tmin for minimum temperature, the number refers to the months, 01 is January and 12 is December.

3 Running the tool

3.1 Software to install

In order to run the tool, R (https://cran.r-project.org/bin/windows/base/) and R Studio (https://www.rstudio.com/products/rstudio/download3/) must be installed. The code is opened in R Studio.

The first time after installation, packages need to be installed. This is done by copy pasting the following command lines into the console.

```
install.packages("raster")
install.packages("maptools")
install.packages("rgdal")
install.packages("dismo")
install.packages("biomod2")
```

3.2 Define user input (all in 1_input folder)

This is the only folder where the user should make changes.

3.2.1 Variable.csv

The Variable.csv contains all the readily available spatial layers (success factors) to select from. This sheet allows to define if the variable is used (by putting a 1 in the respective column) and with which weight it should enter the similarity component.

PATH	VARNAME	USE	Mask	WEIGHT	EXTRACTION
0_Covariates/Af_8kACCESS.rst	Access	1	0	1	1
0_Covariates/Af_8kAfriPop.rst	Hpop	1	1	1	1

The file also allows to select a mask layer. This mask will define which layer should be used to standardize all other layers when extracted for the tool. Any layer can be used as a mask, but only one should be defined as a mask.

The code will then pick the spatial layers in 0_Covariate selected by setting USE =1, and extract and standardized them to the layer defined as mask and save the layer in 2_workspace. This extraction takes time. If there is a speed problem, extraction can be done once and then avoided for the subsequent runs by setting EXTRACTION = 0.

3.2.2 Study area definition

The reference study area for which the similarity is run needs to be inputted as a shapefile. This shapefile needs to be saved into the 1_Inputs/StudyArea folder.

3.2.3 Background definition

The code allows to put a meaningful background shapefile, which aims at helping the viewer what the map represents. The code comes with the world countries boundaries (GAUL) as default. But this can be changed by changing the shapefile in 1_Inputs/MapBackground/ folder.

3.2.4 Set up code

Open the set up code in Rstudio and adjust the parameters.

Sets the general path

Set the path to the tool folder as follows :

setwd("D:/xxxxx/xxxx")

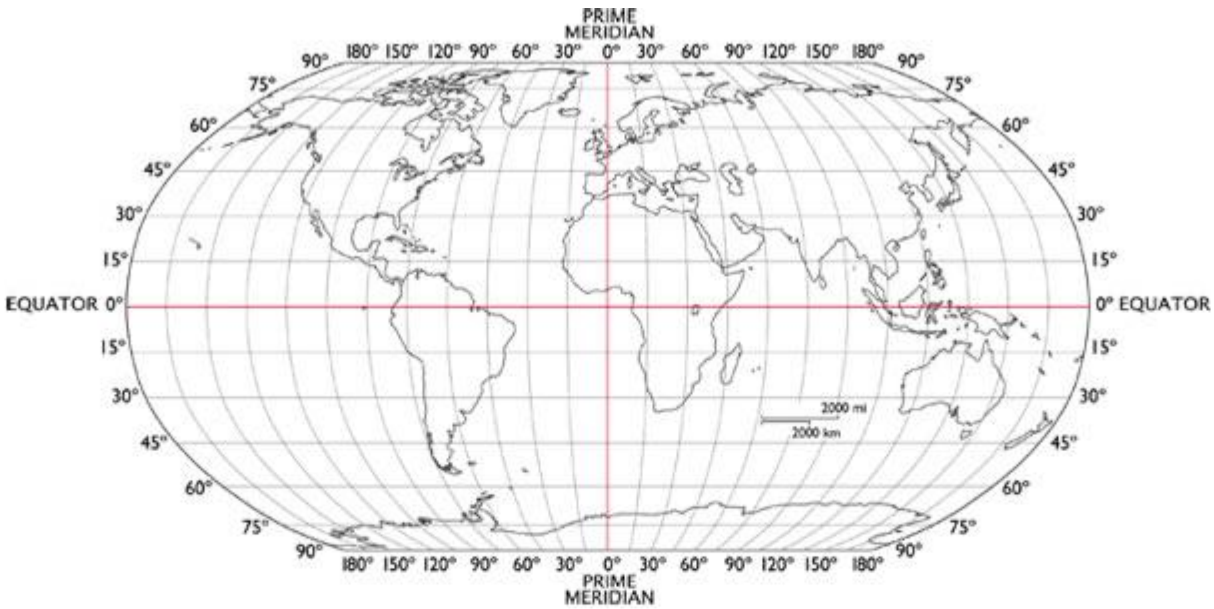
Note that R only reads /. If you copy the path from Windows explorer, it will come with \, which need to be replaced with /.

1.0 General name of the run

Define the name of the run that will appear in the output pdf

1.1 Sets the bounding box of the study region

Set here extent for which the similarity should be computed. This is in decimal degree that can be derived from the map below.



1.3 What is the name of the Shape file containing the area from which similarity should be mapped?

Make sure that the path "1_Inputs/StudyArea/STUDYAREA.shp", links to the study area definition

1.4 Path of the general shape file used for reporting maps

Make sure that the path "1_Inputs/MapBackground/country.shp" links to the background definition

3.2.5 Run the code

Select the whole set up code and press run or CLT + R, to run the whole. After it has run the 3_Ouputs folder has 6 different files:

- 0_StudyArea.pdf shows the data used for the similarity analysis
- 1_Similarity.pdf shows the maps for the three similarity measures
- 3 .tif files (geotiff) that are the raster data (in geo-tif format) for the maps presented in 1_Similarity.pdf
- 1 highres.tif, which provides a high resolution image of the 3 maps in publication quality (300 dpi)

4 Add new dataset

4.1 Prepare your spatial layer

It is important that all raster layers have the same projections, the same resolution and same extent. A mask layer is provided in 0_Covariates.

Then enter the following code lines into the R-Studio console

Newlayer <-raster ('path to the data'))-Covariates/mask, sep="/")

NameOfNewlayer<-resample(Newlayer,Mask)

Save you new layer into you

writeRaster(NameOfNewlayer, paste('path to the tool box', '0_Covariates/NameofNewlayer.tif' ask, sep="/")

Your new layer is now ready to be used in the 0_Covariate folder.

Note : if the tool is run with new data only, i.e. not in combination with the provided data, it is smart to choose the layer with the biggest extent and lowest resolution as mask, and make sure that all layers have the same format as described above.

4.2 Adjust you variable.csv file

In order to read the new variable into the tool, add a line to the variable.csv file and define the path as

0_Covariate/NameofNewlayer.tif