

## Exercise 3:

# Derived Environmental Variables

### Aim:

- DEM derived variables treatment
- Landcover derived variables
- Temporal series derived variables
- Familiarize with the *raster* package in R

## INTRODUCTION

A common issue in landscape genomics experiment is the characterization of a peculiar gradient at sufficient extent and resolution. For the most widely studied variables (for ex. temperature and rainfall), there are public databases with overall means, global extent and various degrees of resolution. For other variables, it is necessary to perform some calculations. In this exercise we will tackle some of these issues.

### *Data Access*

Continue working on the QGIS project we used during previous exercises. In case you missed a step, you can download the required files from the course website.

### *Software Access*

Download R from the course website. It is strongly suggested to download R-studio as well. R-studio is a free environment to develop R codes and facilitates libraries management.

## EXERCISE

### *Digital Elevation Models*

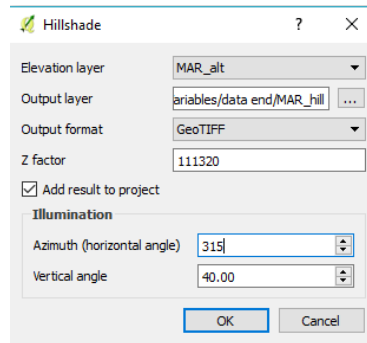
Digital elevation models (DEM) are one of the most employed variables in the production of derived descriptors. DEM contain the basic information to describe a terrain, from which we can extrapolate other variables of interest. Before starting, make sure that the Morocco altitude layer is loaded on the QGIS project, then click on:

**>> Raster > Terrain Analysis**

As you can see, 5 different options are possible. We will work on two of them. Select:

## > Hillshade

And check as follows:

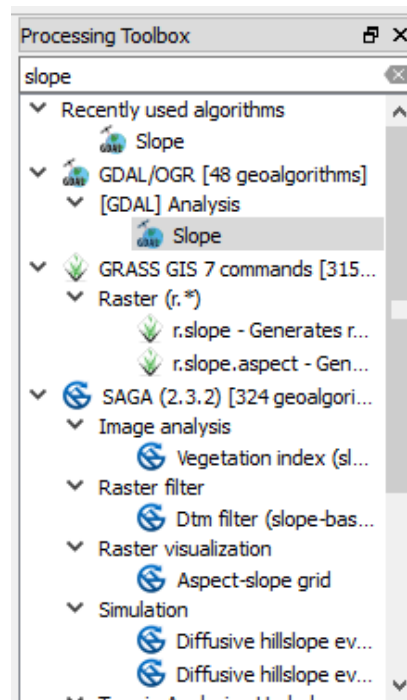
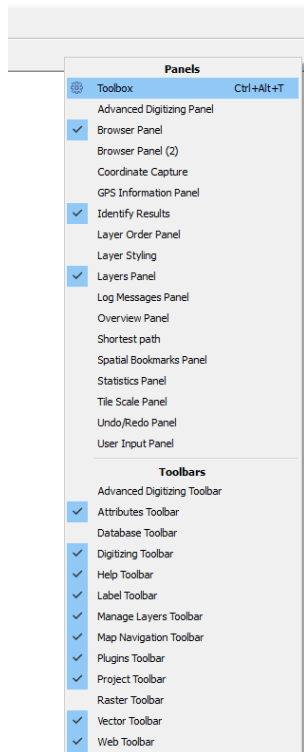


For historical reasons, the shadow azimuth is always defined as 315° (to north-west). Click OK. The new layer should appear in the map. The scale parameter represents the ratio between the horizontal distance unit (defined by the coordinates reference system, in our case WGS84 uses decimal degrees) and vertical units (in our altitude layer, in meters). At the equator, 1 decimal degree = 111320 m.



Q1) What does the Hillshade layer represent?

The second derived variable we want to obtain is more trivial to understand. This time we will use another tool for calculating slope, included in the GDAL tool package. This package comes with QGIS and contains several tools to work with raster data. To find the package you can search for it in the *Processing Toolbox* window. You can obtain it, as shown here below on the left, by right clicking on the toolbar and selecting *Toolbox* (for windows users, you can also obtain it with ctrl+alt+T). The toolbox will be added to the interface. You can text search the tool you need, in our case the Slope tool of the GDAL package (here below on the right). Open it.

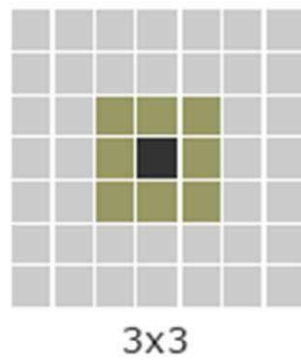


Specify:




- Input layer: MAR\_alt
- Slope: > Save to file as MAR\_slopes.tif
- Scale: 111320

Leave the other parameters as default.

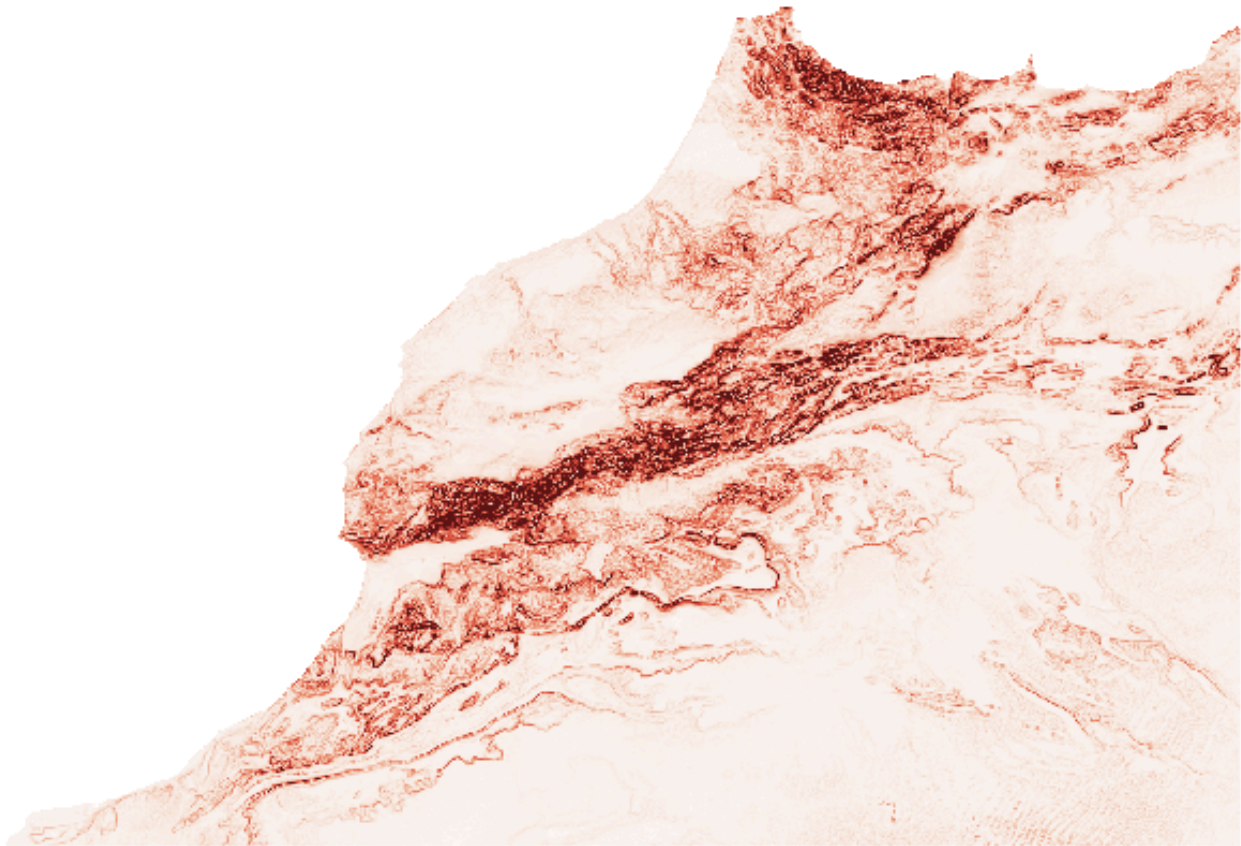
NOTE: By default, QGIS calculates the slope of each pixel based on the altitude (for each stored pixel in the layer) of all neighbor pixel in the window 3x3:



Légende:

-  Pixel to calculate
-  Pixel used for the calculation
-  Pixel not used for the calculation

The slope layer should now be in your map. Visualize it on a red palette.




Q2) *What does the slope value represent?*

### **Landcover**

Landcover maps represent the distribution of different types of land use (for ex. urban areas, forests, freshwaters, etc.). This data can come (as in our case) as a raster layer, where each

possible value is associated to a type of land use, or as a vector file where the land use is explained in the attribute table. Landcover data can be extremely useful since it can generate proxy variables of interesting selective pressures.

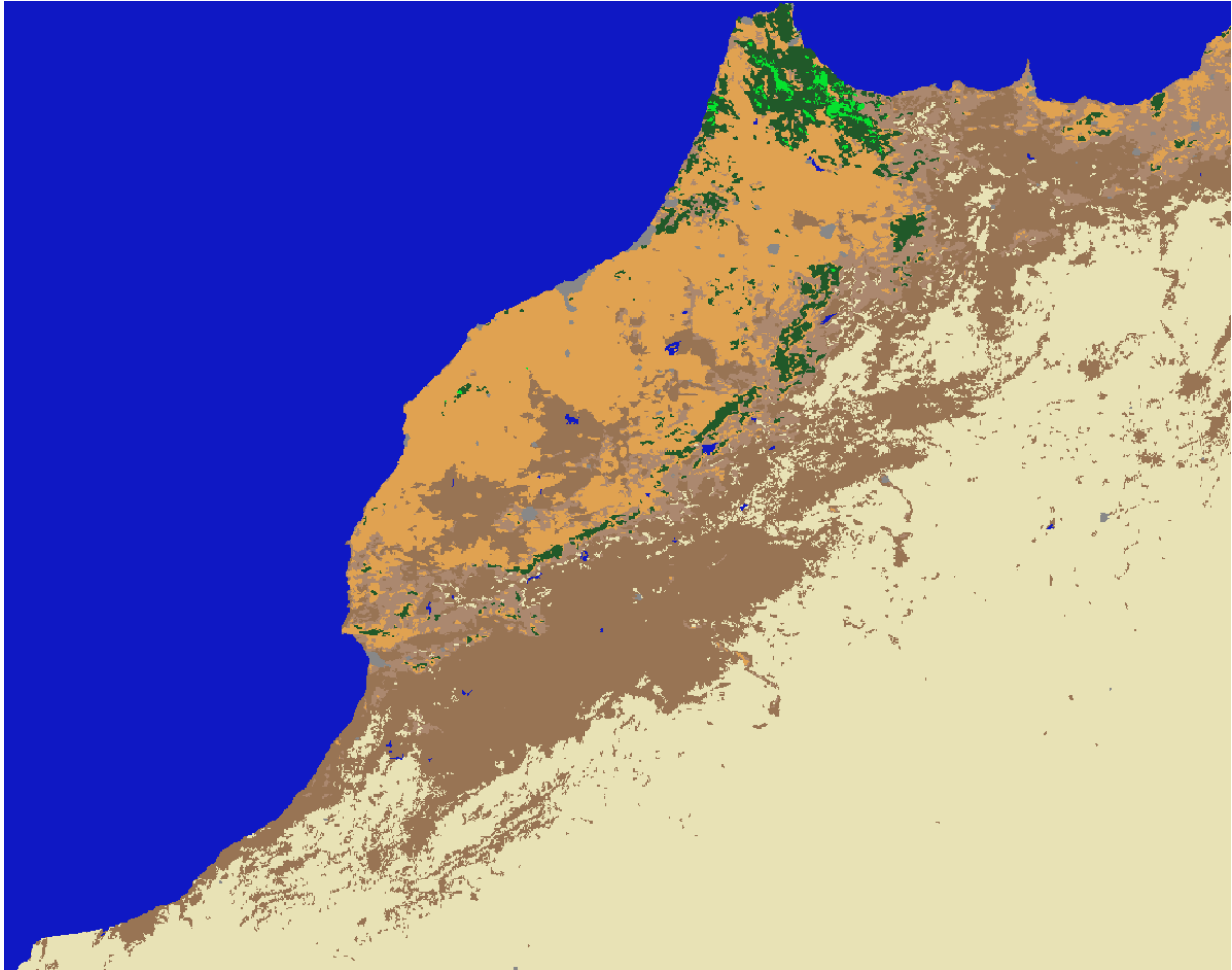
Visualize the MAR\_cov layer on the QGIS project. Use the  tool to investigate the pixel values on the map.

Q3) What is the value associated to ocean? What is the value associated to desert?

You will probably realize that for most of the pixel value it is hard to understand what it is the meaning. For this reason, we have to retrieve the metadata corresponding to the classification. The landcover data we downloaded from DIVA-GIS belongs to the [Global Land Cover 2000 project](#) that provides a landcover classification with world-wide extent at 1km of resolution. Here below you can find the legend of the codes:

GLC Global Class (according to LCCS terminology)	
1.	<b>Tree Cover, broadleaved, evergreen</b> <i>LCCS &gt;15% tree cover, tree height &gt;3m</i> (Examples of sub-classes at regional level*: <i>closed &gt; 40% tree cover; open 15-40% tree cover</i> )
2.	<b>Tree Cover, broadleaved, deciduous, closed</b>
3.	<b>Tree Cover, broadleaved, deciduous, open</b> ( <i>open 15-40% tree cover</i> )
4.	<b>Tree Cover, needle-leaved, evergreen</b>
5.	<b>Tree Cover, needle-leaved, deciduous</b>
6.	<b>Tree Cover, mixed leaf type</b>
7.	<b>Tree Cover, regularly flooded, fresh water (&amp; brackish)</b>
8.	<b>Tree Cover, regularly flooded, saline water,</b> (daily variation of water level)
9.	<b>Mosaic:</b> <b>Tree cover / Other natural vegetation</b>
10.	<b>Tree Cover, burnt</b>
11.	<b>Shrub Cover, closed-open, evergreen</b> (Examples of sub-classes at reg. level *: (i) sparse tree layer)
12.	<b>Shrub Cover, closed-open, deciduous</b> (Examples of sub-classes at reg. level *: (i) sparse tree layer)
13.	<b>Herbaceous Cover, closed-open</b> (Examples of sub-classes at regional level *: (i) natural, (ii) pasture, (iii) sparse trees or shrubs)
14.	<b>Sparse Herbaceous or sparse Shrub Cover</b>
15.	<b>Regularly flooded Shrub and/or Herbaceous Cover</b>
16.	<b>Cultivated and managed areas</b> (Examples of sub-classes at reg. level *: (i) terrestrial; (ii) aquatic (=flooded during cultivation), and under terrestrial: (iii) tree crop & shrubs (perennial), (iv) herbaceous crops (annual), non-irrigated, (v) herbaceous crops (annual), irrigated)
17.	<b>Mosaic:</b> <b>Cropland / Tree Cover / Other natural vegetation</b>
18.	<b>Mosaic:</b> <b>Cropland / Shrub or Grass Cover</b>
19.	<b>Bare Areas</b>
20.	<b>Water Bodies</b> (natural & artificial)
21.	<b>Snow and Ice</b> (natural & artificial)
22.	<b>Artificial surfaces and associated areas</b>

Now that we know what pixel value stands for, we can visualize the landcover map with a more convenient color scale. Right click on the MAR\_cov layer in the panel and open the **Properties** window. Go to **Style** and select Singleband Pseudocolor as render type. Set minimum value as 1 and maximum value as 22. Represent the map using random colors. Click on apply. You will see that the map of landcover will assign a different color to each land-use class. You can also improve the rendering by assigning color in accordance to the meaning of the land-use code (for ex. water in blue, forest in green, etc.). To do so, simply double-click on the color box next to the respective code. You will obtain something like this:



*Q4) Which types of land-cover classes are predominant across our study area? What kind of environments might be interesting for studying local adaptation?*

In landscape genomics studies, environmental variables usually have continuous values. For this reason, for a variable of interest, we need to transform a discrete information (the landcover map) into a gradient. We will do it with the 'artificial surfaces' class (coded 22).

*Q5) Do you think this variable might be pertinent to the study? Which selective pressure might be associated to this variable?*

First of all, we need to create a binary map in which we will assign 1 to artificial areas and 0 to all the other areas. To do so, click on:

## >> Raster > Raster Calculator

Set the parameters as shown here below:

The screenshot shows the QGIS Raster Calculator dialog box. On the left, under 'Raster bands', a list of available bands is shown, with 'MAR\_cov@1' selected. On the right, the 'Result layer' section is configured: 'Output layer' is set to 'Derived Variables/data end/MAR\_urb\_bin', 'Output format' is 'GeoTIFF', and 'Current layer extent' is defined with X min: -13.30000, XMax: -0.90000, Y min: 27.60000, Y max: 36.00000, Columns: 1488, and Rows: 1008. The 'Output CRS' is 'Selected CRS (EPSG:4326, WGS 84)'. The 'Add result to project' checkbox is checked. Below these sections is an 'Operators' panel with various mathematical and logical operators. At the bottom, the 'Raster calculator expression' field contains the text '"MAR\_cov@1" = 22'. A status bar at the bottom left indicates 'Expression valid'. 'OK' and 'Cancel' buttons are at the bottom right.

**Raster bands**

- MAR\_alt@1
- MAR\_cov@1
- MAR\_hill@1
- OpenStreetMap@1
- OpenStreetMap@2
- OpenStreetMap@3
- Slope@1

**Result layer**

Output layer: Derived Variables/data end/MAR\_urb\_bin

Output format: GeoTIFF

Current layer extent

X min: -13.30000 XMax: -0.90000

Y min: 27.60000 Y max: 36.00000

Columns: 1488 Rows: 1008

Output CRS: Selected CRS (EPSG:4326, WGS 84)

☒ Add result to project

**Operators**

+ \* sqrt cos sin tan log10 (

- / ^ acos asin atan ln )

< > = != <= >= AND OR

**Raster calculator expression**

"MAR\_cov@1" = 22

Expression valid

OK Cancel

The expression used will select only the pixel with a value of 22 (artificial areas). Click ok, a new raster layer will appear in the map. Set its rendering (from the properties window) so that 0 is the minimum value and 1 the maximum value. You should observe a binary map with one associated to each artificial area, as the one here below:

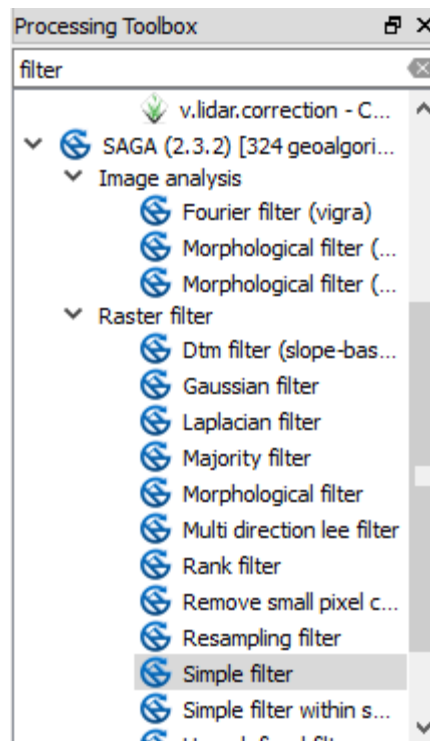




Now we want to transform this binary map into a gradient. There are two ways of doing this:  
 1) Calculate, for each pixel, the distance from the closest pixel with the value 1; 2) Calculate, for each pixel, the mean of its neighborhood.

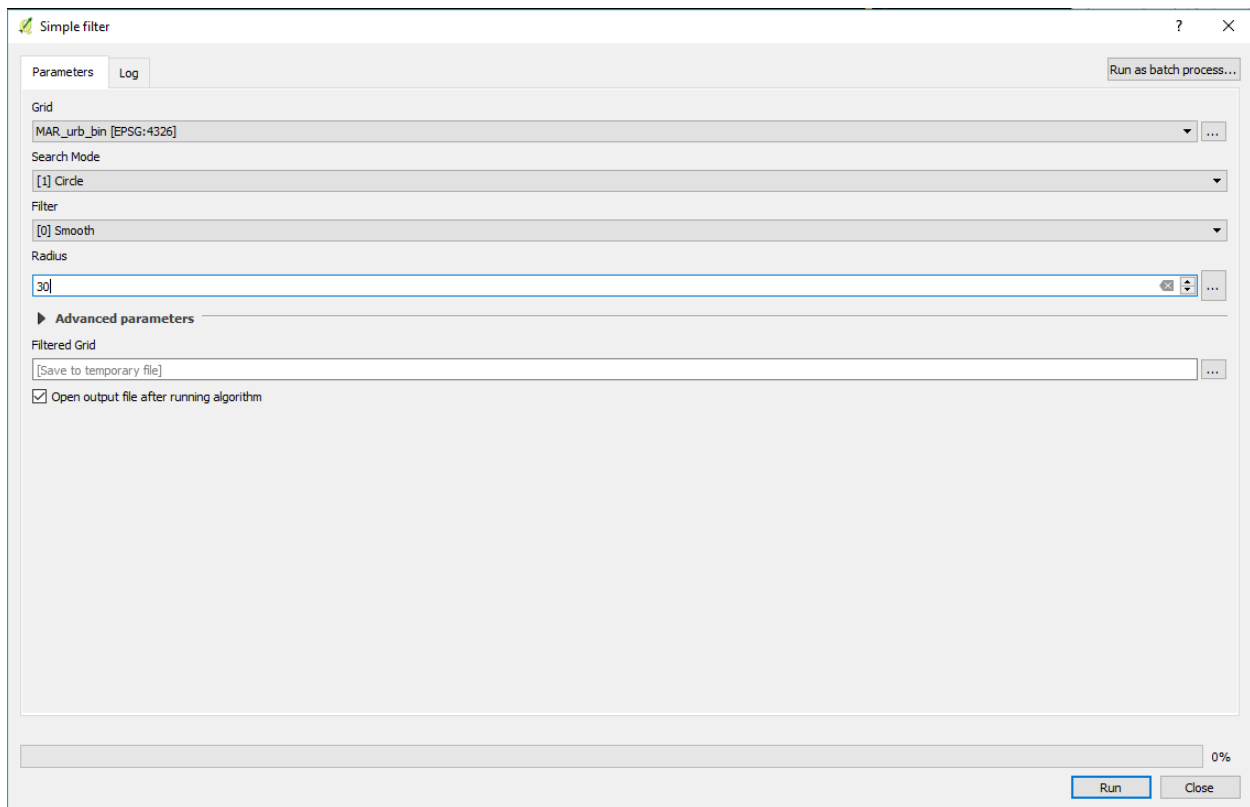
*Q6) What method would you use? What are the drawbacks of each method?*

In this exercise, we will use the method 2, creating a variable that represents the frequency of artificial areas in a radius of 30 km. In the Processing Toolbox, look for the Filter tool of the SAGA package.

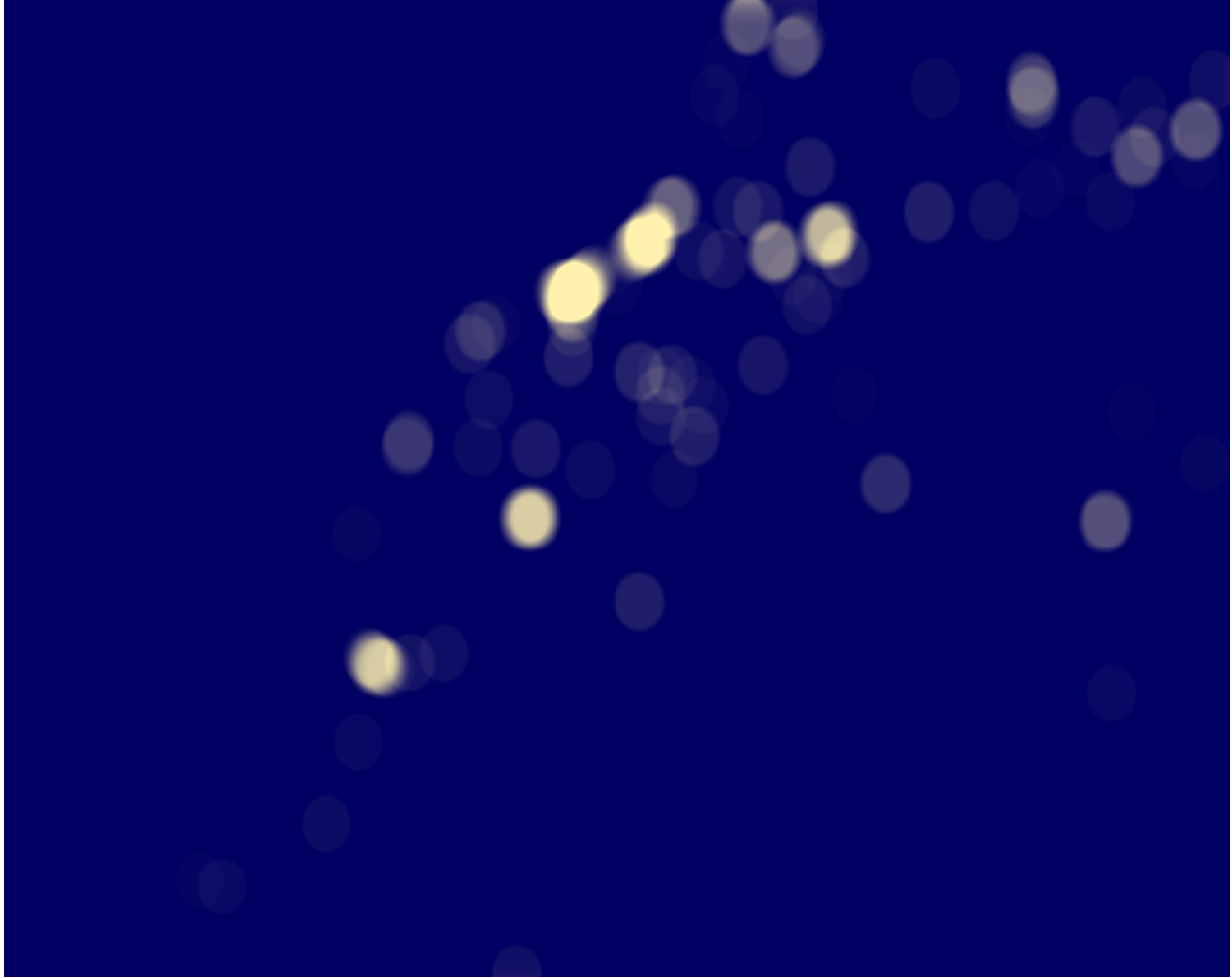



Next, set the parameters as shown here below:





This function calculates, for each pixel, the mean of the neighboring pixels. The search mode indicates the shape of the neighborhood (square or circle), the radius its extent (here 30 units of the raster layer, i.e. 30 km since resolution is  $\sim 1$ km). The filter represents the function to apply to calculate the value, the Smooth option calculates a mean of the neighborhood. Click RUN and wait for the result (it may take some time). If this returns an error, select the grid input using the “...” button and selecting the file from your file system. If everything worked correctly, you will obtain a map of this kind:



Use the  to observe that the closer you move towards an artificial area the higher pixel value you will get. Remember to save the raster you just created (you can call it *MAR\_urb\_30km.tiff*).

This kind of maps can be obtained from any data with discrete values. For example, you can start from a vector file and transform it to a raster layer (use the **Rasterize** tool). From then on, follow the workflow shown here above.

### ***Temporal Series***

Another common issue when looking for environmental data is the lack of products representing overall trends. For example, for the Cloud Fraction data we obtained from the NASA Earth Observation website, monthly means are available year-by-year. For a better of characterization of a landscape, it is usually more representative to have a mean value over several years. This is what we will calculate in this section. Load the cloud fraction monthly means for August from 2000 to 2016 on the QGIS project. Then, click on:

**>> Raster > Raster Calculator..**

Set the parameters as shown here below:

**Raster calculator**

**Raster bands**

- MODAL2\_M\_CLD\_OT\_2002-08@1
- MODAL2\_M\_CLD\_OT\_2003-08@1
- MODAL2\_M\_CLD\_OT\_2004-08@1
- MODAL2\_M\_CLD\_OT\_2005-08@1
- MODAL2\_M\_CLD\_OT\_2006-08@1
- MODAL2\_M\_CLD\_OT\_2007-08@1
- MODAL2\_M\_CLD\_OT\_2008-08@1
- MODAL2\_M\_CLD\_OT\_2009-08@1
- MODAL2\_M\_CLD\_OT\_2010-08@1
- MODAL2\_M\_CLD\_OT\_2011-08@1
- MODAL2\_M\_CLD\_OT\_2012-08@1
- MODAL2\_M\_CLD\_OT\_2013-08@1
- MODAL2\_M\_CLD\_OT\_2014-08@1
- MODAL2\_M\_CLD\_OT\_2015-08@1
- MODAL2\_M\_CLD\_OT\_2016-08@1
- OpenStreetMap@1

**Result layer**

Output layer: Derived Variables/data end/CLD\_mean

Output format: GeoTIFF

**Current layer extent**

X min: -15.00000 XMax: 0.00000

Y min: 25.00000 Y max: 40.00000

Columns: 150 Rows: 150

Output CRS: Selected CRS (EPSG:4326, WGS 84)

☒ Add result to project

**Operators**

+	*	sqrt	cos	sin	tan	log10	(
-	/	^	acos	asin	atan	ln	)
<	>	=	!=	<=	>=	AND	OR

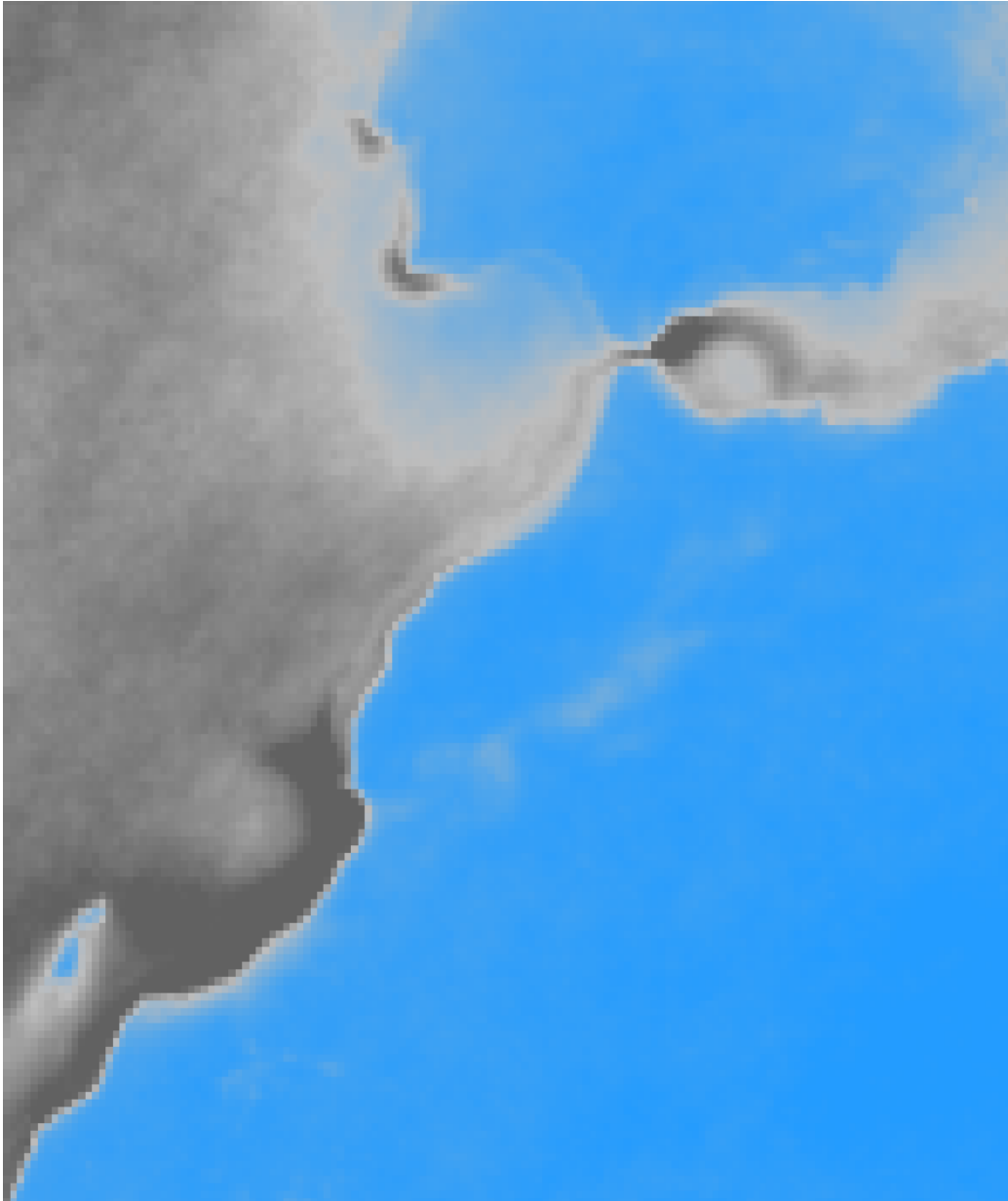
**Raster calculator expression**

```
("MODAL2_M_CLD_OT_2000-08@1" + "MODAL2_M_CLD_OT_2001-08@1" + "MODAL2_M_CLD_OT_2002-08@1" +  
"MODAL2_M_CLD_OT_2003-08@1" + "MODAL2_M_CLD_OT_2004-08@1" + "MODAL2_M_CLD_OT_2005-08@1" + "MODAL2_M_CLD_OT_2006-08@1" +  
"MODAL2_M_CLD_OT_2007-08@1" + "MODAL2_M_CLD_OT_2008-08@1" + "MODAL2_M_CLD_OT_2009-08@1" +  
"MODAL2_M_CLD_OT_2010-08@1" + "MODAL2_M_CLD_OT_2011-08@1" + "MODAL2_M_CLD_OT_2012-08@1" + "MODAL2_M_CLD_OT_2013-08@1" +  
"MODAL2_M_CLD_OT_2014-08@1" + "MODAL2_M_CLD_OT_2015-08@1" + "MODAL2_M_CLD_OT_2016-08@1") / 17
```

Expression valid

OK Cancel

The expression here below calculates the sum (pixel-by-pixel) of seventeen years of Cloud Fraction monthly means of August and then divides it by the number of years (in other words, calculates the mean). Make sure that the resolution and extent matches the one of the input layers. Click OK. You will obtain a map as the one here below:

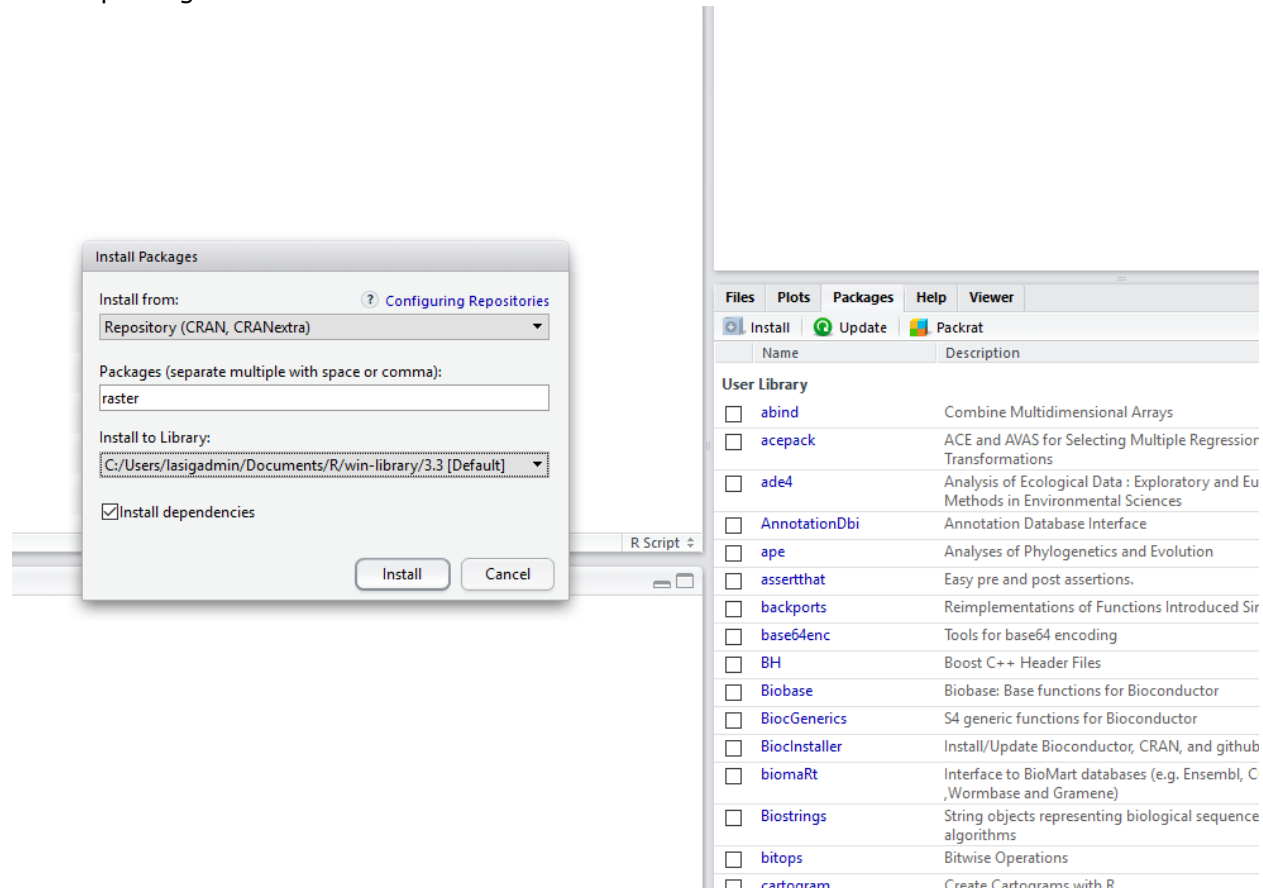


*Q7) How do cloud distribution in August distributes across the study area?*

### ***The R raster package***

All the procedures that we applied to environmental data in QGIS during this exercise can be performed in the R environment. QGIS has the advantage of having tools 'ready-to-use' and a user friendly interface that allows to visualize environmental layers on-the-fly. On the other hand, QGIS can become quite uncomfortable when complex calculations need to be made, or when we want to parallelize a single process for different layers. R provides us an environment where we can easily perform customized calculations on environmental data, but forces us to renounce to an intuitive data exploration interface. Here below we briefly introduce the R raster package and show some basic computations.

Open Rstudio and create a New Project in the folder where the environmental data of the exercise is stored. Open the `raster_computation.R` script, which contains the code we will use. For this exercise you will need to Download the *raster* package for R. To do so, click on the Packages tab of R studio and then on Install. When the install window open, look for the raster package as here below and click on Install.



Do the same for a package called "rgdal".

Next, follow the instructions contained in the `raster_computation.R` script.

## ***Calculate your own environmental variables***

In the last part of the exercise, calculate two derived variables that you think might be interesting for a landscape genomics study. You can create new variables from DEM, from other Landcover classes, temporal series or even look for other sources.