Copernicus Global Land Service Resampling Tool Using R

Xavier Rotllan-Puig1,2,✉, Tim Jacobs3, Federico Gianoli1, Pier Lorenzo Marasco1, and Michael Cherlet1

22/06/2020

1 Joint Research Centre – European Commission. Directorate D – Sustainable Resources. Unit D6 – Knowledge for Sustainable Development & Food Security Unit. Via Enrico Fermi 2749. I-21027 Ispra (VA), ITALY  
2 ASTER-Projects. Barri Reboll, 9, 1r. 08694 Guardiola de Berguedà (Barcelona), SPAIN  
3 VITO NV. Boeretang 200. BE-2400 Mol, BELGIUM

✉ Correspondence: [Xavier Rotllan-Puig <xavier.rotllan.puig@aster-projects.cat>](mailto:xavier.rotllan.puig@aster-projects.cat)

**Abstract**

# Introduction

The Copernicus Global Land Service (CGLS; <https://land.copernicus.eu/global/>) is a component of the Land Monitoring Core Service (LMCS) of Copernicus, the European flagship programme on Earth Observation. CGLS systematically produces and distributes time series of global bio-geophysical products on the status and evolution of the land surface, at different spatial resolutions. These products are used to monitor the vegetation, the water cycle, the energy budget and the terrestrial cryosphere.

The CGLS vegetation-related products (i.e. NDVI, LAI, FAPAR…), based on PROBA-V observations, have been distributed at 1km and 333m until June, 2020. However, as of July, 2020, all Near Real Time (NRT) production of the vegetation biophysical variables, based on Sentinel-3 observations, are no longer provided at 1km resolution. Nonetheless, the users who might be interested on continuing their 1km time series can use a resample of the new 333m products.

The science and production teams of the Global Land service, in support to the 1km users, provide different tools to make their own resampling exercises from the new 333m products to a 1km resolution, corresponding to the usual 1km grid. A Notebook with R code and some explanations can be found at <https://nbviewer.jupyter.org/github/VITObelgium/notebook-samples/blob/master/datasets/probav/ResampleTool_R_notebook.ipynb>. In addition, Python code to run in QGIS can be found in …

In this document we present a comparison of different resampled products using these two tools with the original CGLS products at 1km resolution for different areas. In addition, a comparison of the results obtained with both tools is also provided.

# Materials and methods

## Data

The analysis was done using different subsets of several 10-daily CGLS vegetation-related global products (i.e. Normalized Difference Vegetation Index-NDVI and Leaf area index-LAI). The selected images (i.e. 1km and 333m resolution) corresponded to the May 10, 2019, 10-day period, and the assessments were made in two different areas. On the one hand, NDVI resampled product was tested for Europe and North Africa (coordinates in Decimal Degrees xmin = -18.58, xmax = 51.57, ymin = 28.5, ymax = 62.95) in order to have a wide representation of different landscapes. On the other hand, the LAI product was tested both using the same European-North African window and also on a subset from a tropical area (Amazonia; coordinates xmin = -70, xmax = -63, ymin = -5.5, ymax = -0.2) in order to have as well a good representation of evergreen broadleaf forests (EBF). While Figure 1 shows the 333m NDVI working map used for the resampling in Europe-North Africa, Figure 2 shows the 333m LAI working map used for the Amazonian area.

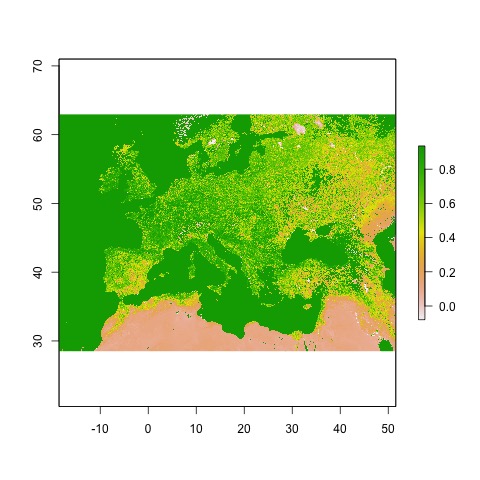


Figure 1: NDVI map at 333m resolution for the European-North African working extent

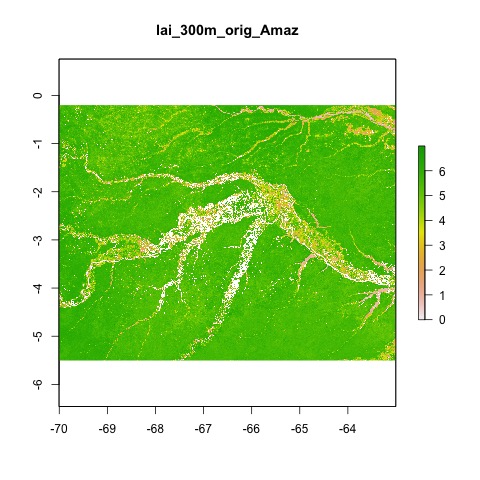


Figure 2: LAI map at 333m resolution for the Amazonian working extent

The original Global Land product files usually can be downloaded as a netCDF4 file. They can often contain specific values for invalid pixels (flagged values), which need to be dealt with. In the case of the NDVI products, digital values in the netCDF (DN) > 250 are flagged and need to be converted into NA. When the netCDF files are read in as a *raster* object, the digital values are scaled into real NDVI values automatically. Therefore, after reading the files, all pixels with NDVI values bigger than 0.92 (= DN 250 x scale + offset) were set to NA. In the same way, LAI values bigger than 7 and smaller than 0, if any, were converted to NAs.

The information regarding flagged values, as well as other supporting information, can be seen in the netCDF file metadata. In addition, more details on the CGLS global products can be found in their Product User Manual at <https://land.copernicus.eu/global/products/>.

## Resample method

There are several approaches to resample data from a finer to a coarser resolution. They can be grouped into area-based aggregation methods and point-based interpolation methods (e.g. Bilinear and Nearest Neighbour), and can be applied depending on the data type and other considerations. Preliminary tests run on NDVI products, although not showed in this document, gave nearly equal results for both approaches.

The area-based aggregation method used in this assessment groups rectangular matrix of pixels of the finer resolution image to create a new map with larger cells. In this case, as we wanted to resample from 333m to 1km, a factor of 3 was implemented (i.e. a matrix of 3×3 pixels).

On the one hand, for the R-based method (R Core Team, 2019), the function *aggregate()* of the package *raster* (Hijmans, 2019) was used. *aggregate()* can perform the calculation using different functions. While the default is the average (*mean()*) it can work also with *modal()*, *max()*, *min()* or even with *ad hoc* functions programmed by the user. For both the NDVI and the LAI products the most suitable method is the average. Table 1 shows a recommendation of the best suited method for each product/layer. In addition, as it is also recommended in the tool, it was included the condition that at least 5 out of the 9 pixels had to have valid values (i.e. not NA) to return a valid value for the resampled pixel.

Table 1: Best suited method recommended for each product/layer.

|  |  |  |  |
| --- | --- | --- | --- |
| Product | Version | Data.layer.in.file | Resample.method |
| NDVI 300m | v1 | NDVI | mean |
| NDVI 300m | v1 | NDVI\_unc | uncertainty\_propagation |
| LAI 300m | v1 | LAI | mean |
| LAI 300m | v1 | RMSE | mean |
| LAI 300m | v1 | LENGTH\_AFTER | \* |
| LAI 300m | v1 | LENGTH\_BEFORE | \* |
| LAI 300m | v1 | NOBS | \* |
| LAI 300m | v1 | QFLAG | \* |
| FAPAR 300m | v1 | FAPAR | mean |
| FAPAR 300m | v1 | RMSE | mean |
| FAPAR 300m | v1 | LENGTH\_AFTER | \* |
| FAPAR 300m | v1 | LENGTH\_BEFORE | \* |
| FAPAR 300m | v1 | NOBS | \* |
| FAPAR 300m | v1 | QFLAG | \* |
| FCOVER 300m | v1 | FCOVER | mean |
| FCOVER 300m | v1 | RMSE | mean |
| FCOVER 300m | v1 | LENGTH\_AFTER | \* |
| FCOVER 300m | v1 | LENGTH\_BEFORE | \* |
| FCOVER 300m | v1 | NOBS | \* |
| FCOVER 300m | v1 | QFLAG | \* |
| DMP 300m | v1 | DMP | mean |
| DMP 300m | v1 | QFLAG | \* |
| GDMP 300m | v1 | GDMP | mean |
| GDMP 300m | v1 | QFLAG | \* |

**Note:** \* QFLAG, LENGTH\_BEFORE/AFTER and NOBS cannot be resampled due to different implementations for 1km-v2 and 300m-v1 products. For example, LAI-NOBS ranges are 0-120 for 1km-v2 and 0-40 for 300m-v1, or LAI/FAPAR/FCOVER-LENGTH\_BEFORE go up to 60 days and up to 210 days, respectively for both products.

On the other hand, for the Python-based method (Van Rossum, G. & Drake, F.L., 2009) implemented for QGIS (QGIS Development Team, 2009)….

## Metrics and plots

In order to assess the performance of the resampling methods, besides mapping the results, three well known and widely used metrics and a scatterplot were produced. The metrics were:

* Pearson correlation coefficient (Pearson’s *r*)
* Root-mean-square error (RMSE)
* Mean absolute error (MAE)

In addition, a map representing the spatial distribution of the largest (> 95th percentile) absolute errors (|original1km – resampled1km|) was also generated to observe possible spatial patterns of the errors.

The R code used to perform the assessment reported in this document can be seen at <https://github.com/xavi-rp/NDVI_resample>.

# Results

In this document we present five different assessments depending on the CGLS product analysed, the area of study and the resampling tool in use. While three of them compare the resampling results obtained with the R-based tool (1 NDVI and 2 LAI) against the original 1km product, one analyses the QGIS/Python results (NDVI) against the same 1km product and another one compares the results of both tools (R vs QGIS/Python).

Table 2 summarizes the three metrics calculated in each assessment, so that an overview of all the cases is provided. In the following subsections each case will be developed separately and in more detail.

Table 2: Pearson’s r, Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) of each case study

|  |  |  |  |
| --- | --- | --- | --- |
| Assessment | Pearson’s *r* | RMSE | MAE |
| NDVI\_Europe/NorthAfrica\_R | 0.982 | 0.048 | 0.03 |
| NDVI\_Europe/NorthAfrica\_QGIS/Python | 0.972 | 0.058 | 0.037 |
| NDVI\_Eur/NorthAf\_R\_vs\_QGIS/Python | 0.982 | 0.047 | 0.031 |
| LAI\_Europe/NorthAfrica\_R | 0.942 | 0.469 | 0.31 |
| LAI\_Amazonia\_R | 0.705 | 0.903 | 0.659 |

## R-based tool vs the original 1km product: NDVI – Europe/North Africa

To have a first impression of the results of the resample tool, Figure 3 shows both the original NDVI map at 1km resolution for the region of study (Europe and North Africa) and the resampled one to 1km using the R-based tool.

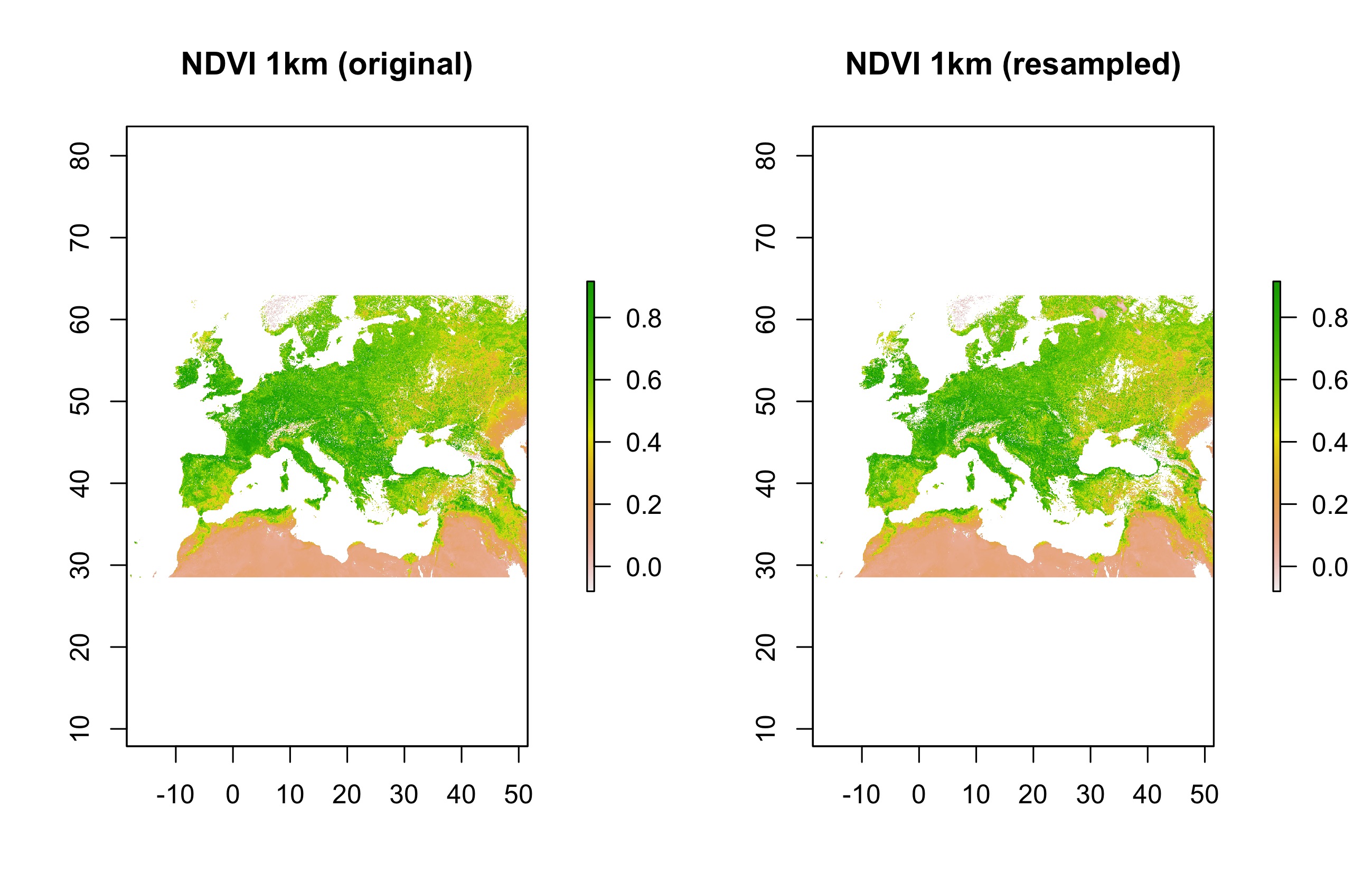


Figure 3: Original NDVI map at 1km resolution and the resampled one using the R-based tool for Europe/North Africa

As it can be seen in the scatterplot (Figure 4), and corroborated by the Pearson correlation coefficient (Pearson’s *r* = 0.982; Table 2), there was a good level of correlation among the original 1km map and the resampled one using the R tool. In addition, considering that NDVI values ranged from -0.08 to 0.92, also RMSE and MAE reported good levels of error among the two maps (0.048 and 0.03, respectively).

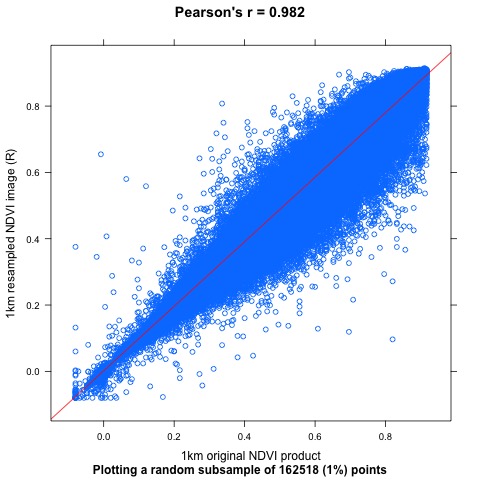
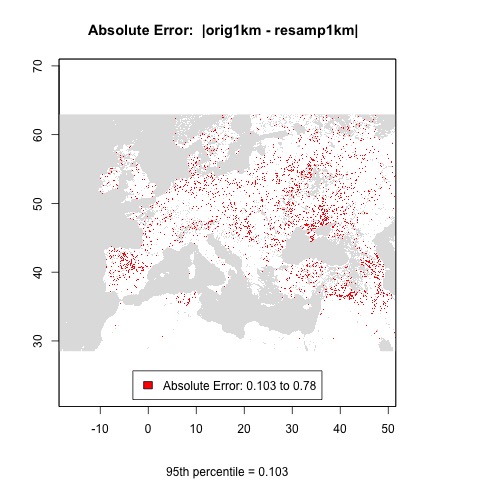


Figure 4: Scatter plot displaying a random subset (1%) of pixel values of the 1km original NDVI product against the values of the same pixels of the resampled map using the R tool (blue points) for Europe/North Africa. Also the regression (red) line

Finally, Figure 5 shows how not a particular spatial pattern of the largest (> 0.103) absolute errors could be observed in the study area.

Figure 5: Spatial pattern of the Absolute Errors larger than their 95th percentile for Europe/North Africa

## QGIS/Python tool vs the original 1km product: NDVI-Europe/North Africa

Figure 6 shows respectively the original NDVI map at 1km resolution and the resampled one to 1km using the QGIS/Python-based tool for Europe/North Africa.

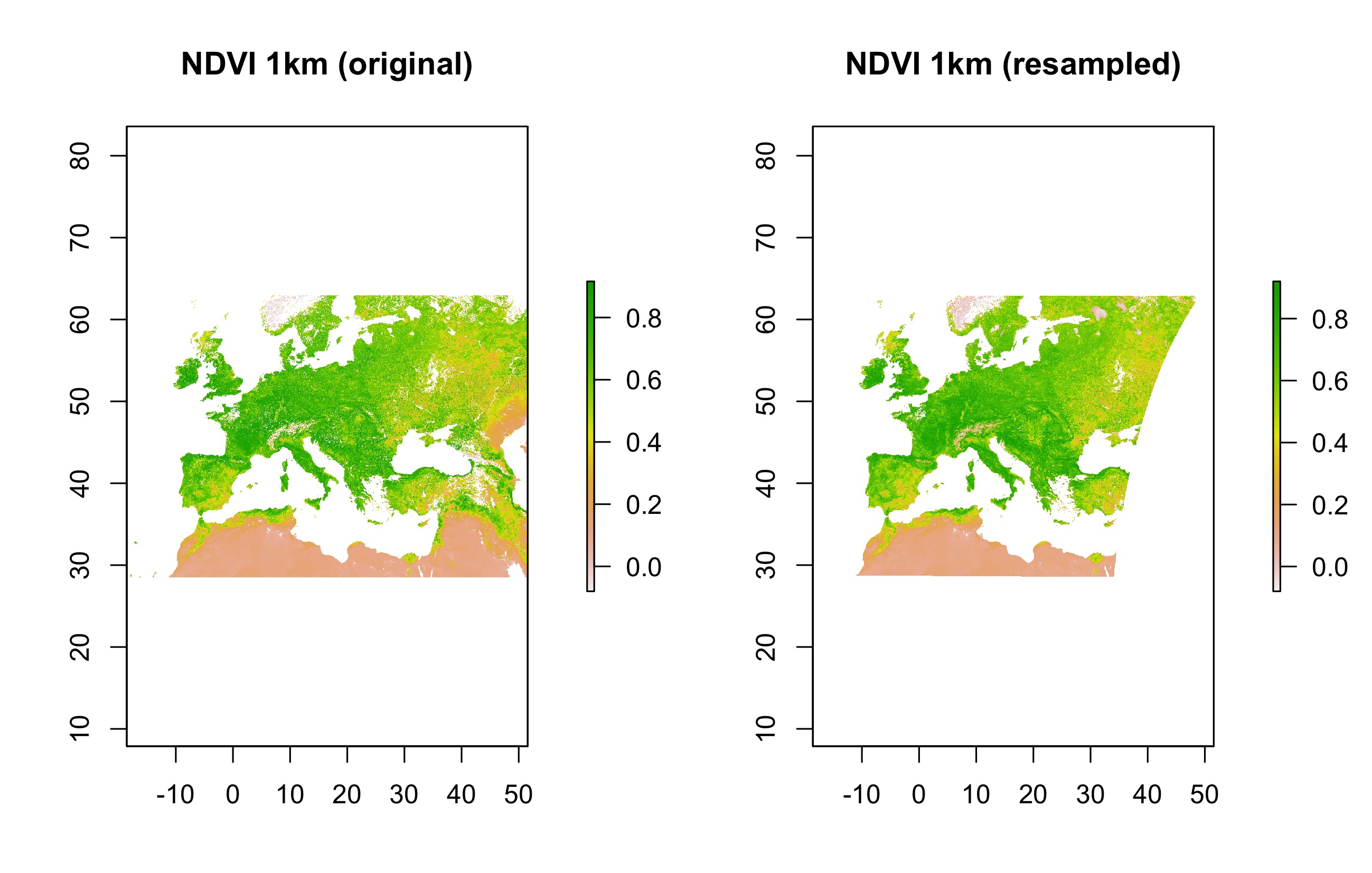


Figure 6: Original NDVI map at 1km resolution and the resampled one using the QGIS/Python-based tool for Europe/North Africa

Similarly to the performance of the R tool, the following scatterplot (Figure 7) together with the correlation coefficient (Pearson’s *r* = 0.972; Table 2), also showed a good level of correlation among the original 1km map and the resampled one using the QGIS/Pyhton tool. In addition, also the reported RMSE and MAE gave similar, therefore good, levels of error among the two maps (0.058 and 0.037, respectively).

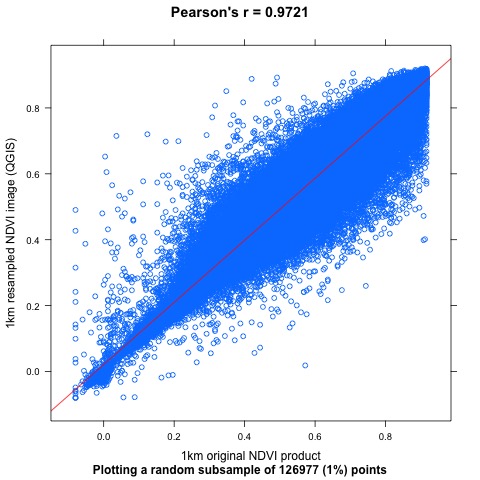


Figure 7: Scatter plot displaying a random subset (1%) of pixel values of the 1km original NDVI product against the values of the same pixels of the resampled map using the QGIS/Python tool (blue points) for Europe/North Africa. Also the regression (red) line

## R-based tool vs QGIS/Python-based tool: NDVI-Europe/North Africa

The following scatterplot (Figure 8) represents the correlation between the resampled map values using the two tools presented in this document and showed in the previous subsections. It could be seen a very high correlation between them, reporting a correlation coefficient (Pearson’s *r*) of 0.982 (Table 2). Additionally, RMSE and MAE were 0.047 and 0.031, respectively.

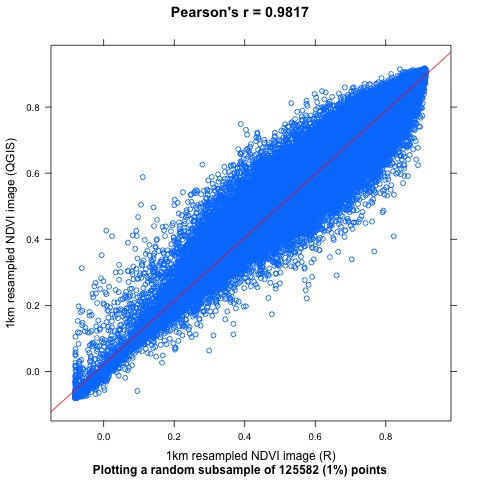


Figure 8: Scatter plot displaying a random subset (1%) of pixel values of the 1km NDVI product resampled using the R-based tool against the ones resampled using the QGIS/Python tool (blue points) for Europe/North Africa. Also the regression (red) line

In light of the two assessments made on both the R-based and the QGIS/Python-based tools, as well as the results showed in this subsection, it can be confirmed that both tools can be equally used, only depending on user’s preferences.

## R-based tool vs the original 1km product: LAI - Europe/North Africa

The resulting resampled map of the R-based tool for LAI in the study area of Europe and North Africa, together with the original 1km map, can both be seen in Figure 9.

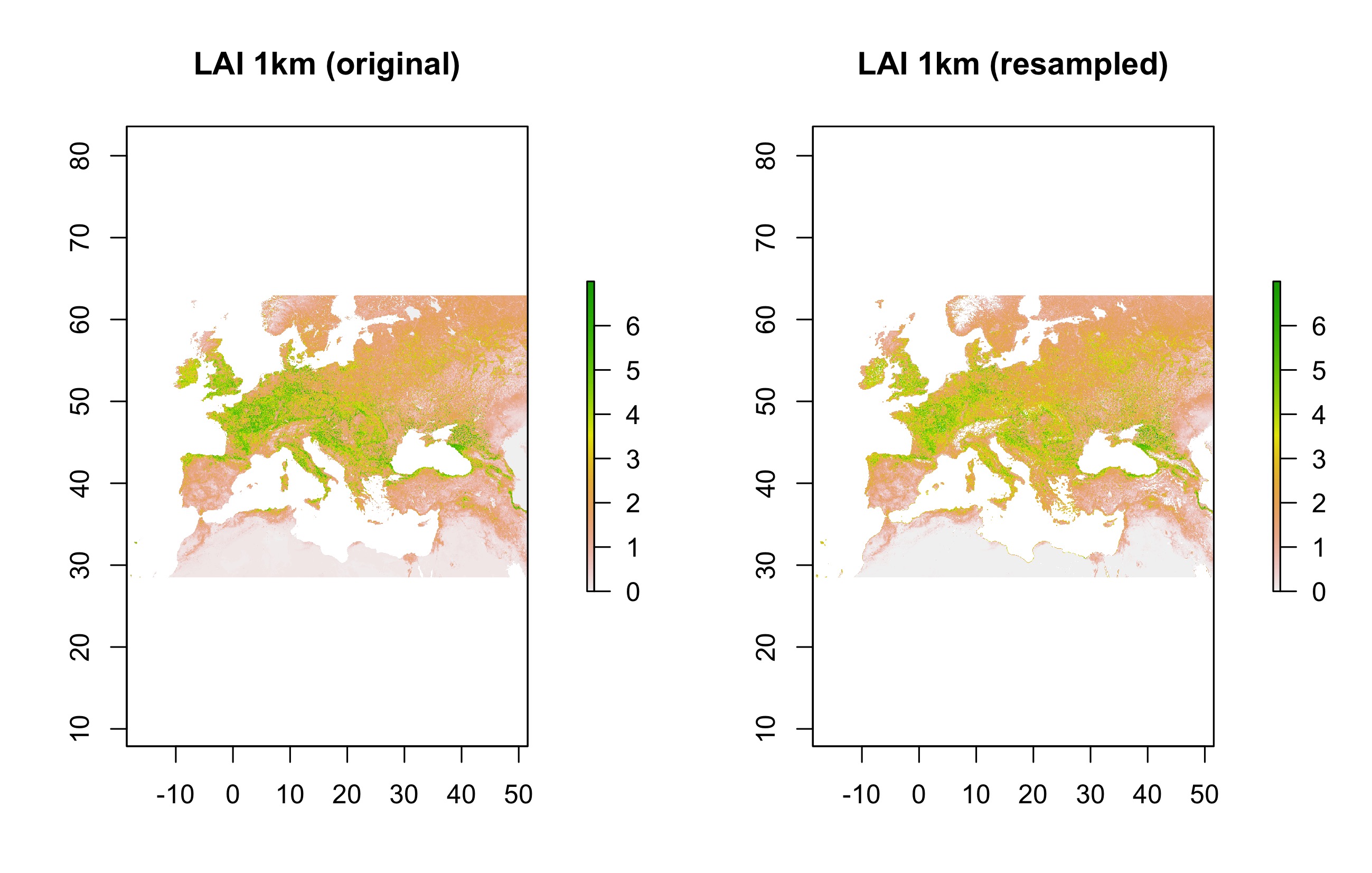


Figure 9: Original LAI map at 1km resolution and the resampled one using the R-based tool for Europe/North Africa

As it can be observed, the scatterplot (Figure 10) and the correlation coefficient (Pearson’s *r* = 0.942; Table 2) showed a good level of correlation among the original 1km map and the resampled one using the R-based tool, although slightly lower than the results of the NDVI product in the same region. In addition, considering that LAI values ranged from 0 to 7, also RMSE and MAE reported good values between the two maps (0.469 and 0.31, respectively).

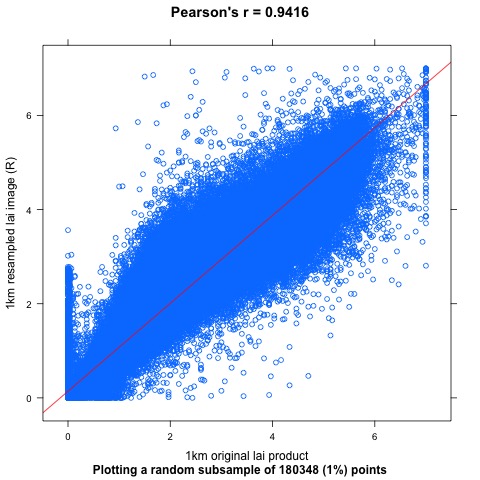


Figure 10: Scatter plot displaying a random subset (1%) of pixel values of the 1km original LAI product against the values of the same pixels of the resampled map using the R tool (blue points) for Europe/North Africa. Also the regression (red) line

## R-based tool vs the original 1km product: LAI - Amazonia

In Figure 11, showing both the original 1km map and the resampled one using the R-based tool for LAI in the Amazonian study area, it can already be seen that the tool’s performance for the tropical areas (EBF) is lower than for non-EBF areas.

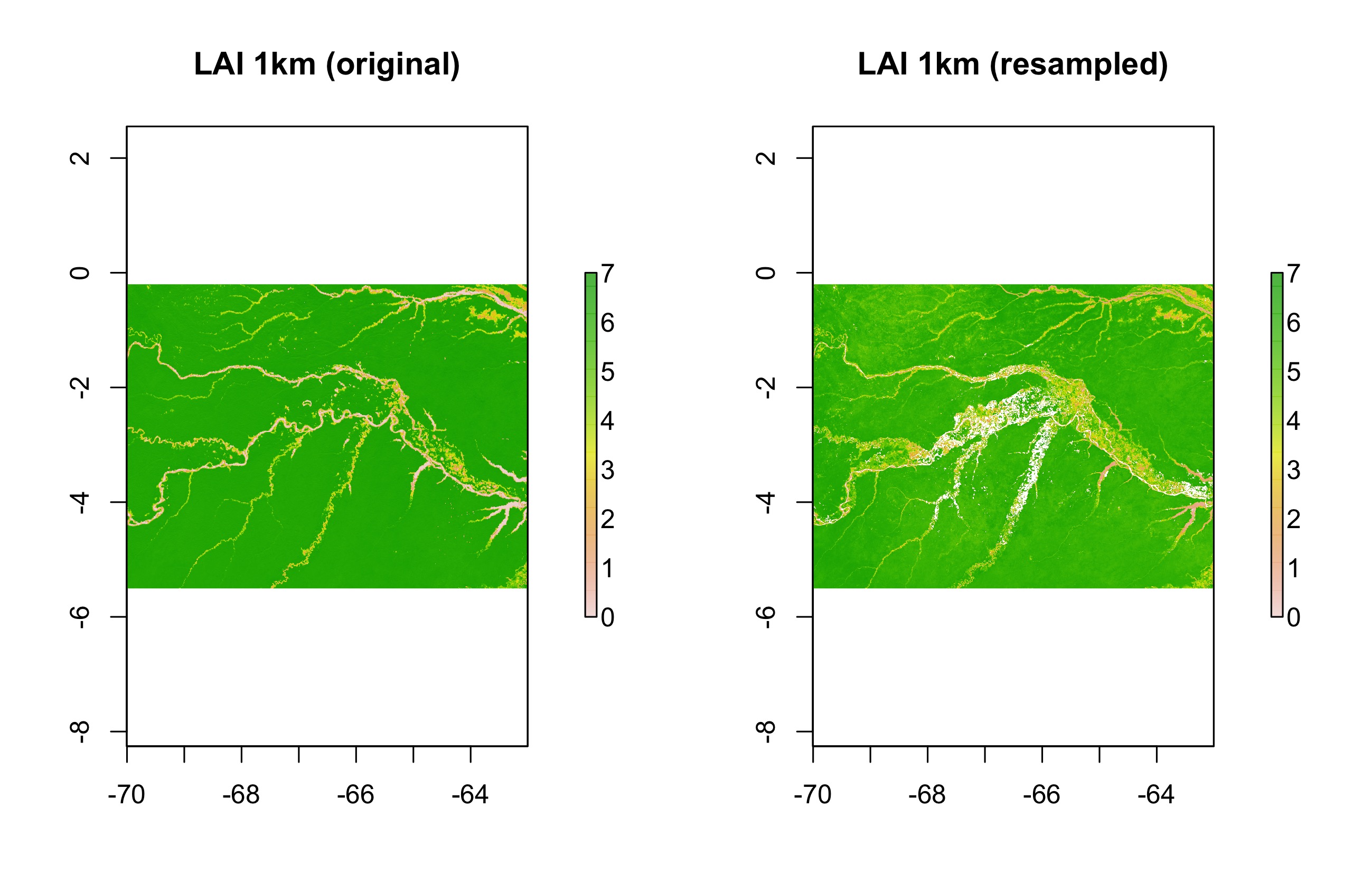


Figure 11: Original LAI map at 1km resolution and the resampled one using the R-based tool for the Amazonian study area

Both the scatterplot (Figure 12) and the statistics corroborated the worse results than for other areas and/or products. While Pearson’s *r* was 0.705, RMSE and MAE were 0.903 and 0.659, respectively (Table 2).

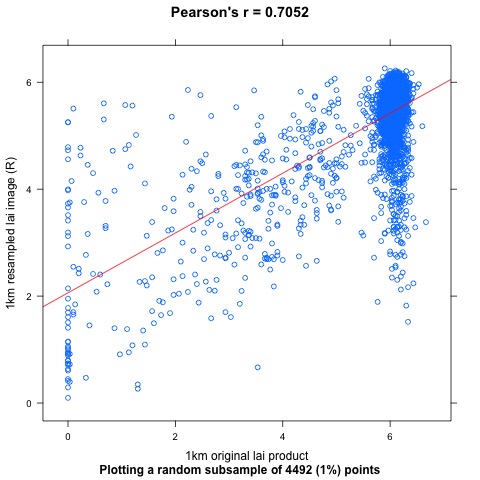


Figure 12: Scatter plot displaying a random subset (1%) of pixel values of the 1km original LAI product against the values of the same pixels of the resampled map using the R tool (blue points) for the Amazonian region. Also the regression (red) line

These poorer results might be mainly due to the differences in the temporal composition and the cloud gap filling method used for the production of the 1km and the 300m products. In this sense, for the former, a more complex algorithm was implemented in the processing chain in order to improve the final generated 10-day vegetation-related global product, especially in EBFs, such is the case for the Amazonian rainforest.

# Conclusions

After the discontinuity on the supply of NRT vegetation-related products at 1km resolution, the Global Land service has made available two resample tools based on different programming languages (i.e. R and QGIS/Python) in order to ease the users to keep producing their own time series at that resolution. In this document we show several assessments made on the performance of these two tools in resampling different 300m products (i.e. NDVI and LAI) and in different landscape typologies (i.e. EBF and non-EBF).

The results showed similar and good performance of the two tools in non-EBF landscapes, both for NDVI and for LAI products. In contrast, the evaluation of the resample results of LAI in an EBF area in the Amazonia gave poorer results. This fact is likely due to the different algorithm implemented in the production of the 10-day vegetation-related global products at 1km and 300m resolution. In light of this, the users must be aware of these differences when using these two tools or any other resampling approach.

# References

Hijmans, R.J. 2019. raster: Geographic Data Analysis and Modeling. R package version 3.0-2. <https://CRAN.R-project.org/package=raster>

QGIS Development Team. 2009. QGIS Geographic Information System. Open Source Geospatial Foundation. URL [http://qgis.org](http://qgis.org/)

R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Van Rossum, G. & Drake, F.L. 2009. Python 3 Reference Manual, Scotts Valley, CA: CreateSpace.