

W11-1 Satellite-supported predictions

This worksheet revisits the prediction using linear models but this time a satellite and hence area-wide independent variable is used for the prediction of the animal activity. After completing this worksheet you should know how to use raster data sets for predictions with multiple linear models.

Things you need for this worksheet

- R — the interpreter can be installed on any operation system. For Linux, you should use the r-cran packages supplied for your Linux distribution. If you use Ubuntu, [this](#) is one of many starting points. If you use windows, you could install R from the official [CRAN](#) web page.
- R Studio — we recommend to use R Studio for (interactive) programming with R. You can download R Studio from the official [web page](#).
- your script and data from [W10-1 GIS data I/O and some visualizations](#)

What's the plan?

What we want to do in this worksheet is more or less identical to what we did in [W05-1 Predictions](#). In that worksheet, we used the vegetation coverage recorded at all of your field survey plots to predict the animal activity for all those plots where we have vegetation coverage records but no animal records. Hence, we predicted animal activity for selected locations (i.e. our field survey plots).

In this worksheet, we will do exactly the same but this time we will use satellite derived NDVI and height values instead of the field recorded vegetation coverage. Since the satellite derived data sets cover the entire island with pixel sizes of about 30 meter by 30 meter, we can also predict the animal diversity for each of those 30 meter by 30 meter grid cells (which are also called pixels). In order to use both NDVI and height information, we will not use a simple linear regression model (i.e. one which uses one independent variable) but a multiple linear regression (i.e. one which uses two independent variables in our example).

Please note that the prediction of the animal activity for the entire island is quite questionable since we base this prediction on 50 (agricultural) field survey plots where we did one recording. So please understand this example as a finger exercises not as solid science.

Learning log assignments

🚨 First things first: the following analysis is build on top of your script from [W10-1](#). Please copy your script "W10-1.R", rename the copy to "W11-1.R" and use it for the programming tasks of this worksheet.

😓 Please load the shape and raster data sets into memory.

In order to compute the multiple linear regression model. we need a data set which shows us the value of the NDVI and DEM data sets along with the animal activity information for each location where we recorded the latter.

😊 Please extract the values of the DEM and NDVI data set for all locations where animal activity information has been recorded and combine the three variables in one data set. Have a look at the `extract()` function which will help you by this task.

😞 Compute a multiple linear regression model using NDVI and DEM information as independent and animal activity as dependent variable and check the goodness of the model fit.

😞 Please compute an area-wide prediction of the animal activity based on your linear model. Please note that you can not use the `predict()` function in this case. Therefore, please extract the coefficients of your linear model (i.e. offset and slopes) and combine them in a linear equation where you can use your variables which hold the NDVI and DEM data as if they would just consist of one value. Plot your results and feel happy!

From:
<http://bisfogo.environmentalinformatics-marburg.de/> - **BIS-Fogo**

Permanent link:
<http://bisfogo.environmentalinformatics-marburg.de/doku.php?id=en:learning:schools:s01:worksheets:ba-ws-11-1>

Last update: **2015/09/22 15:22**

