**species distribution modeling of *Triticum aestivum,* Bo Bode, s2162261, 11-12-2019**

**introduction**

*Triticum aestivum* is also known as common wheat is an annual plant that belongs to the grasses grass family Poaceae, tribe Triticeae, and subtribe Triticineae. It is thought to have originated on the Eurasian continent, a starting point from which man spread it throughout the world, including China and central Europe. Wheat is one of the earliest domesticated crop plants in the Pre-Pottery eolithic Near East. Wheat is one of the most cultivated food crops in the world and provides the most grain (Haider, 2012). This means that wheat plays an important role in the food supply and that unforeseen changes in the ability to cultivate this crop can have a large impact on food supply and economy. This could hurt countries all over the world as it cultivated in Europe, Africa, Asia, the Americas and Australia (Börner *et al.*, 2005). One of the potential treads for this crop and any other crop is changes in climate. To get a better understanding of how climate changes could affect the distribution of *Triticum aestivum* prediction models can be made based on multiple environmental variables. By comparing these models to the current distribution data it is possible to visualize the potential effects of climate change.

**Method**

To analyze the potential effects of climate change on the distribution of wheat the data had to be collected. The occurrence data selected were only the human observation with coordinates for *Triticum aestivum* L. (GBIF.org 4 December 2019). With these settings, 16880 occurrences ware stored in a CSV file form witch only the species name, longitude and latitude in decimals were used. To visualize the data an overlay was made in ArcMap (version 10.6.1) with the WGS84 projection on a global scale. As there were no points visible in the ocean nor at other unlikely places the data remained unchanged. The environmental data were downloaded from https://www.worldclim.org/version1 for the current conditions (WorldClim 1.4: Current conditions 1960-1990) and the predicted conditions of 2050 based on the bioclimatic variables. The model used was the General Circulation Models (GCM), HadGEM2-ES with Representative Concentration Pathways (RCPs) 45 which is a moderate change in climate that could happen if steps are taken to reduce climate change.

After getting the data it was processed with R studio (R version 3.6.1, 2019-07-05 ) and packages: sp, rgdal, raster, and biomod2, to prepare the climate variables by sorting, cropping it and test for autocorrelation. By plotting the occurrence data on one of the environmental data maps the frame can be made to trim the map so that only the parts that have wheat remain in the exports (figure 1). After setting the boundaries for the models, the environmental data was processed and models of each variable were made. To select the relevant environmental variables for the rest of the analysis two tests were used. The first test was the Pearson pairwise correlation was used to measure the statistical association of dependence between two variables where a score of less than 0.7 was acceptable (Table 1). The second test was the Variance Inflation Factor (VIF) to measure the statistical dependence of one variable to the combination of the other variables. This score had to be lower than 10 to be accepted (Table 2).

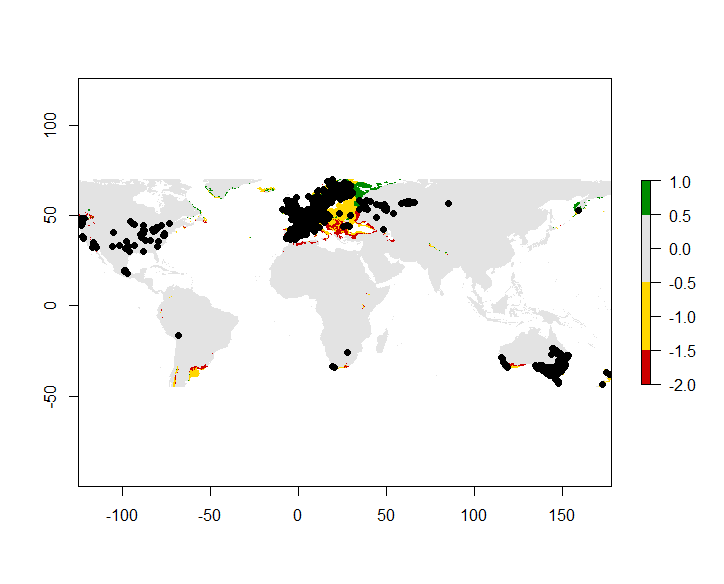
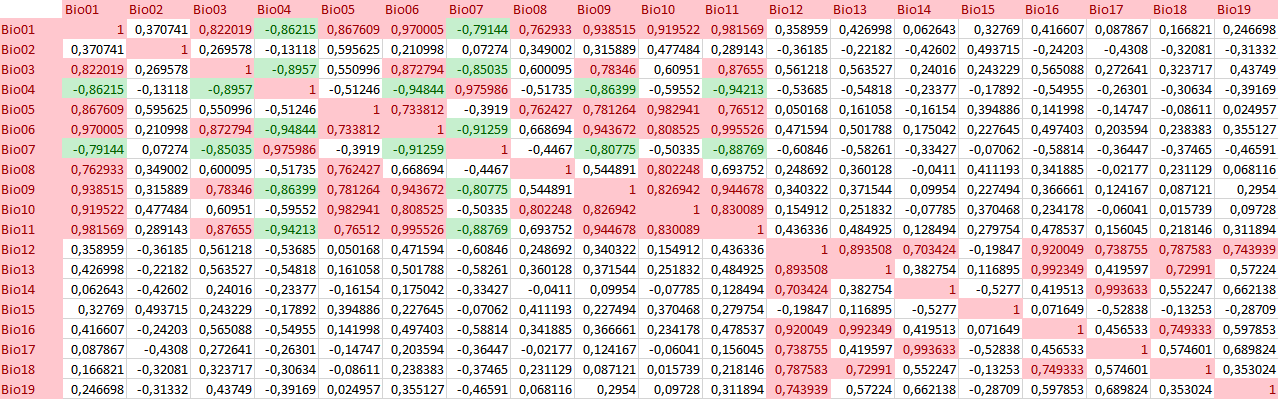


Figure 1: cropped data with the occurrence of wheat in black on a heatmap

Table 1: Pearson pairwise correlation



The cell marked in red is above the critical value of 0.7 and cells in green are below the critical value of -0.7

Table 2: Variance Inflation Factor after selection

|  |  |  |
| --- | --- | --- |
| Variables | Definition | VIP |
| Bio01 | mean annual temperature | 3.456814 |
| Bio07 | temperature annual range | 4.080325 |
| Bio12 | annual precipitation | 9.402523 |
| Bio15 | precipitation seasonality | 1.381307 |
| Bio18 | precipitation of the warmest quarter | 4.183588 |
| Bio19 | precipitation of the coldest quarter | 3.506086 |

Critical cutoff (>10)

The variables were chosen based on the previous test and are shown in table 1. To make the models of the future predictions these variables were used in (MaxEnd figure 2).

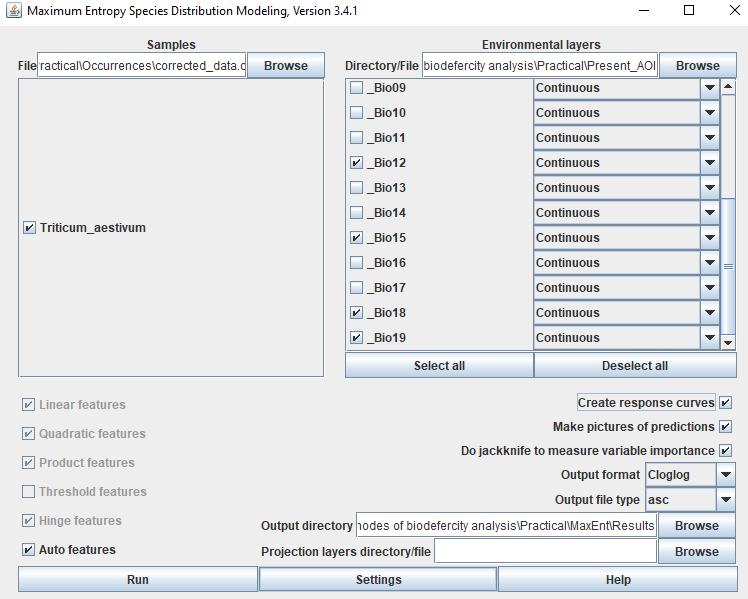


Figure 2 maxent settings: the projection used the future climate variables: Future WLD and AOI.

For the basic settings all boxes except “skip if output exists” were checked and only the number of replicates was changed to 10 replicates. The full overview of the settings can be found in appendix 1.

**Maxent output**

After running maxent the diagnostics of the experiment were represented in the analysis of omission a commission. The average omission and predicted Area for wheat were both following the predictions closely indicating that the simulations outside of the initial learning set closely follow the predicted values (figure 3A). This is also supported by the receiver operating characteristic (ROC) where the sensitivity of the models is plotted against 1- sensitivity. The area under the curve (AUC) can be used to calculate a measure of the model performance. This measure gives the probability that a random positive instance and a random negative instance are correctly ordered by the classifier (Phillips, Anderson and Schapire, 2006). As the AUC value of the models is 0.819 with a standard deviation of 0.005 it indicates that the species distribution modeling was accurate (>0.7)as is shown in figure 4.

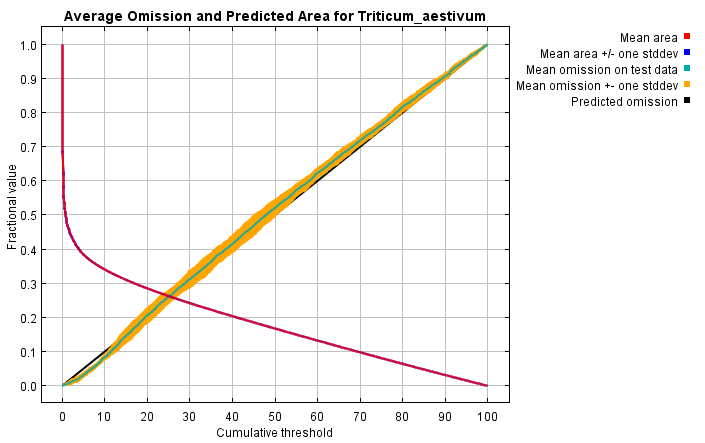


Figure 3: average omission and predicted Area for *Triticum aestivum*. The red line indicates the mean area, the blue zone indicates + or – the standard deviation of the mean area, the green line is the mean omission of the test data, the yellow area indicates the standard deviation of the omission and the black line is the predicted omission.

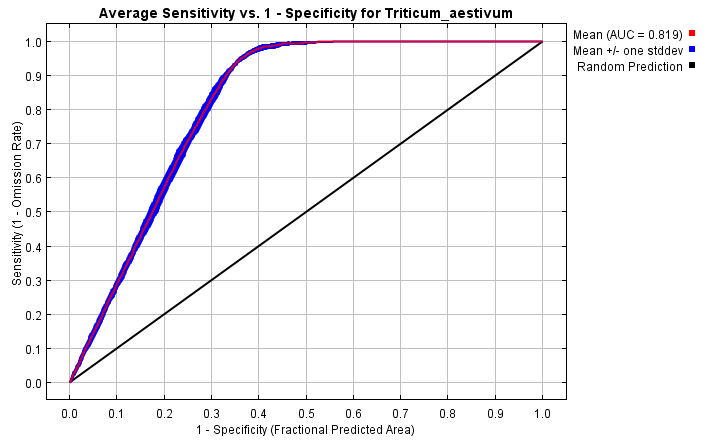


Figure 4: ROC and AUC test. The red line indicates the area under the curve (AUC = 0.819 on a standard deviation of 0.005) with the standard deviation in bleu. The black line is the random prediction if the differences in de models were only caused by random effects.

When looking at the combined maps of the predictions the same trend holds and there are no unexpected points in the maps. This counts for the occurrence data, the present data, future data with uncropped environmental variables (AOI) and the cropped environmental data (WLD). In the figures can be seen that there is a slight decrease when comparing the occurrence of the data of the present and the future (figure 5 B and C)

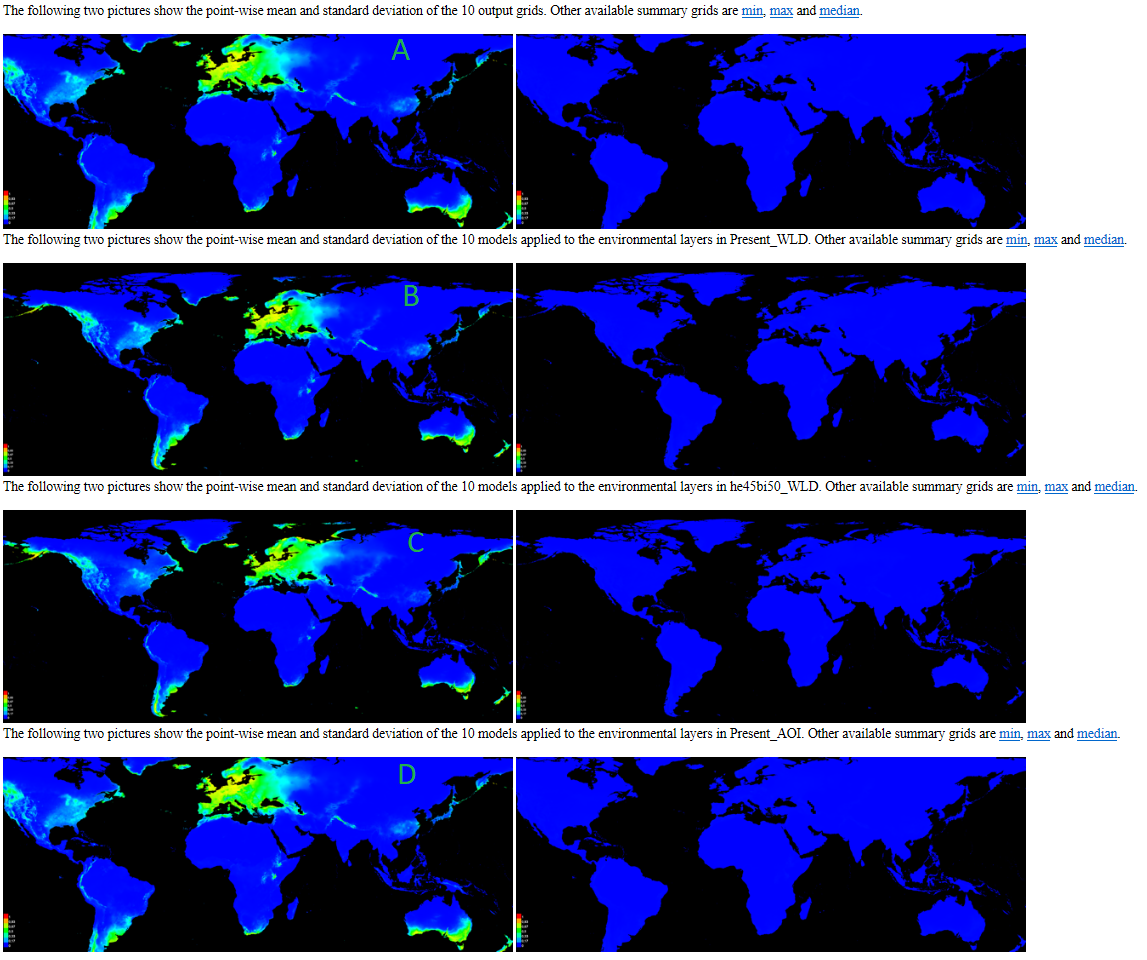
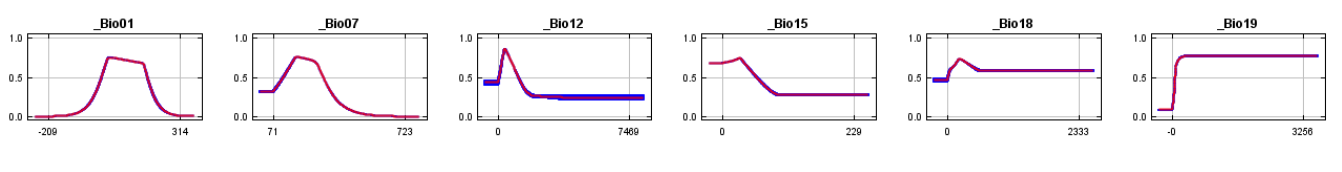


Figure 5: combined data of all samples. A shows the mean and standard error of the output grits, B shows the combined present plots, C shows the future plots and D shows the present plots when cropped.

When looking at the individual variables of figure 6 and table 2 it can be seen that the main effects on the occurrence are caused by the mean annual temperature, temperature annual range, and the annual precipitation.

A

B

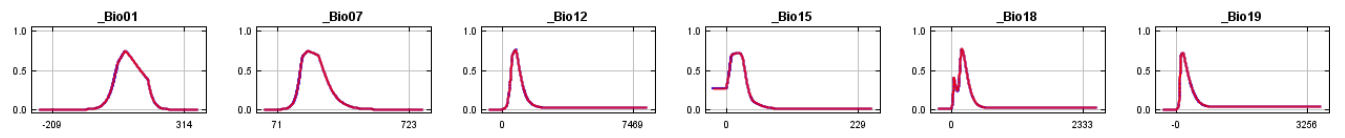


Figure 6 response curves of the different variables. Response curves of the individual variables while still being affected by the other variables. B response curves of the different variables without the interactions of the other variables. The red lines represent the average of the 10 replicates and the blue shade represents the standard deviation.

Table 2 Analysis of variable contribution

|  |  |  |  |
| --- | --- | --- | --- |
| Variables | Definition | Percent contribution | Permutation importance |
| Bio19 | mean annual temperature | 49.8 | 13.4 |
| Bio01 | temperature annual range | 26.5 | 34.9 |
| Bio07 | annual precipitation | 19.2 | 36.4 |
| Bio12 | precipitation seasonality | 2.1 | 7.7 |
| Bio15 | precipitation of the warmest quarter | 1.4 | 5.5 |
| Bio18 | precipitation of the coldest quarter | 1 | 2.1 |

The percentage estimates of contribution based on the training data, the permutation percentages are based on the training AUC

**Discussion**

Climate is changing and this could have consequences for all life on the planate. In these cases, the focus was on *Triticum aestivum* which is an important food crop all over the globe. By linking the occurrence data of the present with the climate models of the present it is possible to compare it to future climate models. This way a prediction of the occurrence can be made for future habitats. in the models made form the data of RCP 45 already have a small decrease in Europe and the VS. From the prediction, the factors that determine where wheat can be cultivated are the precipitation and the mean annual temperatures. This was however based on the moderate model that expects people to take actions against climate change. Another important aspect to take into account is that the occurrence data is based on human observations and this will give bias in the data. This was also observed in the organization of the data points in eastern Europe where the data points were in a symmetrical cluster that did not match the satellite images from google earth (data not shown). Considering the drawbacks and the relatively small change in occurrences in future predictions this could already have an impact on the crop yield and economic effects. However, for further improvements, additional scenarios of climate change need to be tested to give more understanding of how impactful climate change can be and potentially show the importance of preventing further climate change.

Börner, A. *et al.* (2005) ‘ Associations between geographical origin and morphological characters in bread wheat ( Triticum aestivum L.) ’, *Plant Genetic Resources*, 3(3), pp. 360–372. doi: 10.1079/pgr200589.

Haider, N. (2012) ‘Evidence for the origin of the B genome of bread wheat based on chloroplast DNA’, *Turkish Journal of Agriculture and Forestry*, 36(1), pp. 13–25. doi: 10.3906/tar-1011-1394.

Phillips, S. J., Anderson, R. P. and Schapire, R. E. (2006) ‘Maximum entropy modeling of species geographic distributions’, *Ecological Modelling*, 190, pp. 231–259. doi: 10.1016/j.ecolmodel.2005.03.026.

GBIF.org (04 December 2019) GBIF Occurrence Download https://doi.org/10.15468/dl.wvi9td

Appendix 1 settings maxent

