University of Osnabrück

IANNwtF Final Project

March 2021

Predicting future crop yields

Unsing ANNs to predict future crop yields under different climate change scenarios

---

by:

Leonhard Urs Bürger
Matriculation Number: 965753
Noel Sommerkamp
Matriculation Number: 974562

Contents

1 Introduction **2**

2 Data **2**

3 CNN **3**

4 Evaluation **4**

References **7**

1 Introduction

Due to rising population, climate change and ever increasing need of food the future of the earth relies on good prediction of crop yields, in order to use the limited resources efficiently. Without predictions about the fertility of the land one may plan ahead to compensate possible dry seasons or desertification. But predicting such factors is very complicated to do 'by hand', as there are many factors involved, where we might not even know their degree of their interactions to the full extend. This seems to be a perfect problem for a neural network.

2 Data

The first thing to do is quite obvious: We need to think about what kind of data we want to use. For that a distinct class was implemented. As we don't exactly know, which factors play a role for ground to be fertile for certain crops, we first wanted to include as many factors as possible, that might even only play a small role. For our dataset we needed some kind of map, where we could map yearly crop yields of certain crops to a specific location. Intuition suggests to just take a map of the world, quite literally, and map the yields to each pixel on this map. We used the "Global dataset of historical yields" [1] containing four different crops: maize, rice, wheat and soybean. For our prediction we mainly relied on the two of the three most used criteria for crop yield prediction[5]: temperature and percipitation [6]. The criterion we did not use is the soil type. We chose not to for two reasons, on the one hand is worldwide soil data not easily available and on the other hand are soil types not likely to change due to the pressures of climate change. Therefore we settled on the ERA5 data from the EU's Copernicus Programm [4] with hourly surface temperature and percipitation values.
For ideal computational efficiency and to avoid crop yields at sea, a mask is used that describes water pixels as zero and land pixels as one. Every water pixel is gets ignored when constructing our dataset.

Figure 1: The masking, where each water pixel renders 0 and each ground pixel 1

We then implemented a generator, that yields 17x17 snippets from our dataframe map by looking at every possible latitude and longitude value combination within a certain range (in this case we have a box width of 8 because our snippet size is 17x17). Before yielding, it checks whether or not the corresponding pixel has a mask value of one to

avoid generating unnecessary data on water (see figure 1). With every sample yielded, the generator also yields the corresponding crop yields seperatly as targets.
Before we feed this dataset in our network, we have to think about one last important preperation step: We need to normalize our data around zero to tackle the well-known gradient problem. For that we need to subtract the mean off the data (i.e. the mean temparature or percipitation over the whole time period rather than one month) and than divide the result by the standard deviation. To get these normalization parameters we implemented a seperate method in our data class, that reshapes our data to get access to the needed values to compute.
For testing the fitness of our model, we chose to introduce three different possible future climate scenarios[3] and build up their datasets, which we will come back to later in the evaluation section.

3 CNN

For our model, we decided to use a convolutional neural network, which was first founded by Yann LeCun in his paper 'Backpropagation Applied to Handwritten Zip Code Recognition'[2]. This choice was quite easy, as the dataset already suggested to use some kind of convolution, as we mapped our data in a picture-like structure. The only major difference here is that an image usually has three channels for RGB, while our 'images' contain 26 channels due to the number of needed data for the specific location on the map. We also needed to adjust the output, because we do not have a single target, but 4 different crop yields as target. Therefore the dense layer that gives the output has to have 4 units.
We now had to think of the architecture of our model. Thinking about the task, we decided, that few layers may be sufficient, because with repeating convolution one may think of the network detecting shapes evolving in complexity. This task suggests, that there is no complex shape, but more like a pattern on lower levels. And the other factor is, that with this gigantic dataset conserning the amount of channels it would just train too long to be feasible. We therefore decided to use just four convolutional layers.
For optimaziation we used dropout-layers, that helped improving the generality of the fitness, i.e. preventing overfitting to the dataset we used for training. We also implemented kernel regulizers for the kernels of the convolutional layers, namely L2-regulizers.
The trained weights of the model got saved and then used to predict yields on different scenarios.

4 Evaluation

After training our model for 7 epochs, we yielded a loss value of 0.5, which might seem a bit high at first glance, but seemed to be suffieciently trained for the task at hand. The loss problem might also be an indicator that the variables percipitation and temperature might not be sufficient enough and only inherit a certain significance that can be learned. Some changes in yields seemingly aren't attributable to only temperature and percipitation, i.e. when a ground type suffers from desertification.
For testing the quality of our model's predictions, we compared the prediction for 2017 generated by our model with the actual yield map of 2016. We also then checked the predictoin of 2100 out of curiosity.

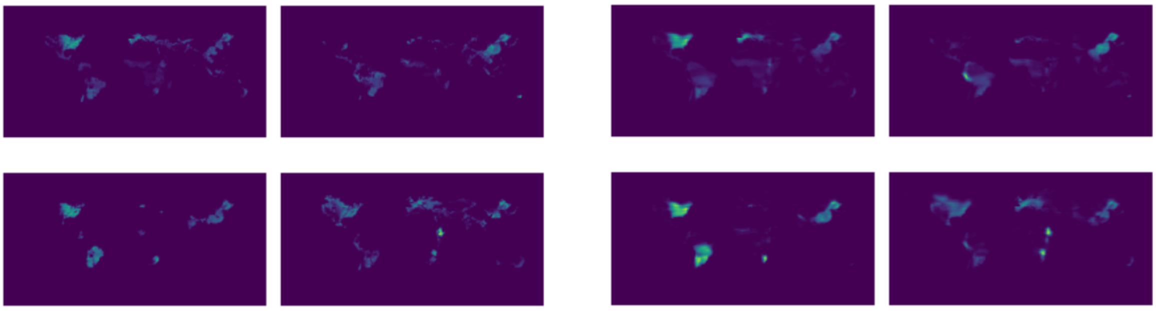

Figure 2: Crop yield (maize, rice, soybean, wheat) actual (left) and predicted (right) distribution for the Year 2016.

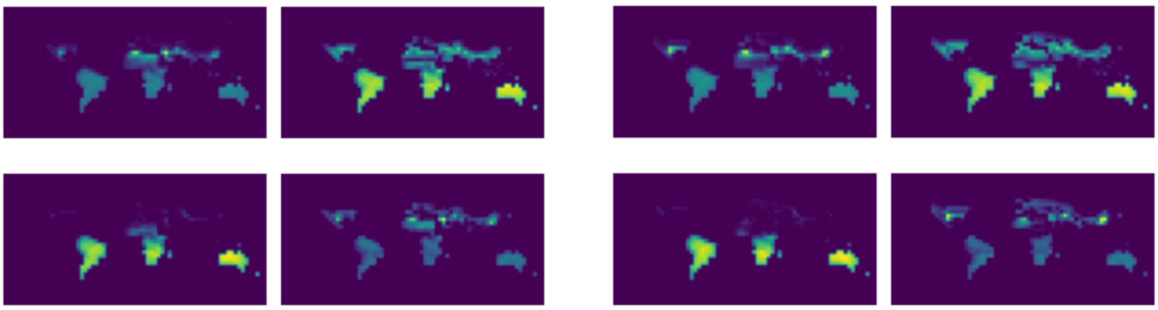

Figure 3: Crop yield (maize, rice, soybean, wheat) predictions for the year 2050 (left) and 2100 (right).

This seemed promising and we continued by using constructed artificial scenarios that climate change may cause in future.
Here we used the Max Plank Institute CMIP6 predictions [3] containing three different climate change sceanrios. The three scenarios "SSP126", "SSP370" and "SSP585" differ in their CO_2 pathway and a higher SSP number indicates a generally more severe climate change.
One can see, that the crop yield predictions indeed changed depending on the scenario we fed into the model. For example, maize yields are predicted to increase with rising temperature and percipitation, but wheat seems to be uneffected for the most part, it only slightly increases (see figure 4 and 5).
This proves that our model was able to learn at least to some degree the dependencies

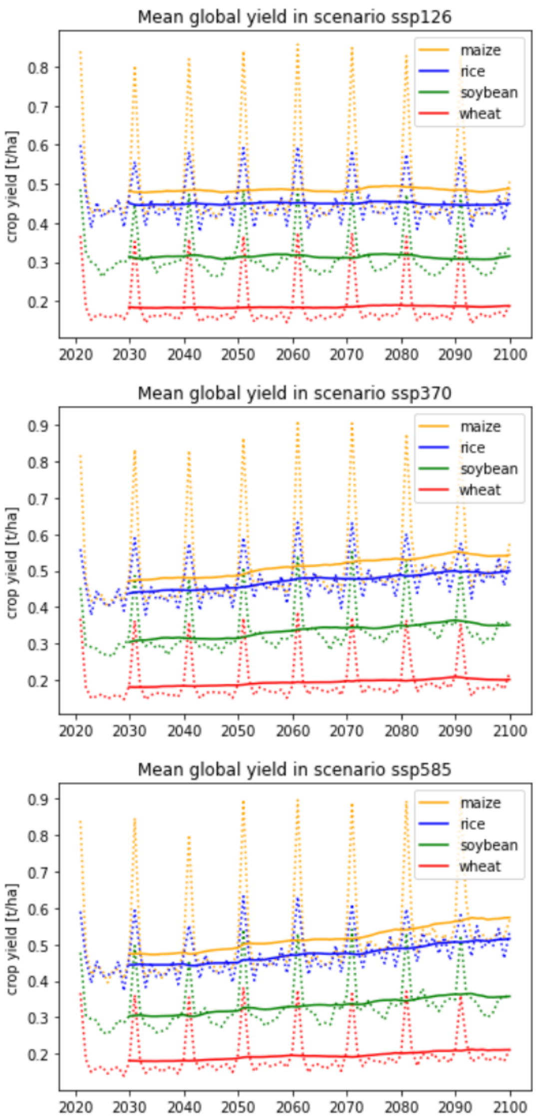

Figure 4: Yearly crop yields (dotted) and 10-year running average yields (solid) for the three different MPI climate change scenarios.

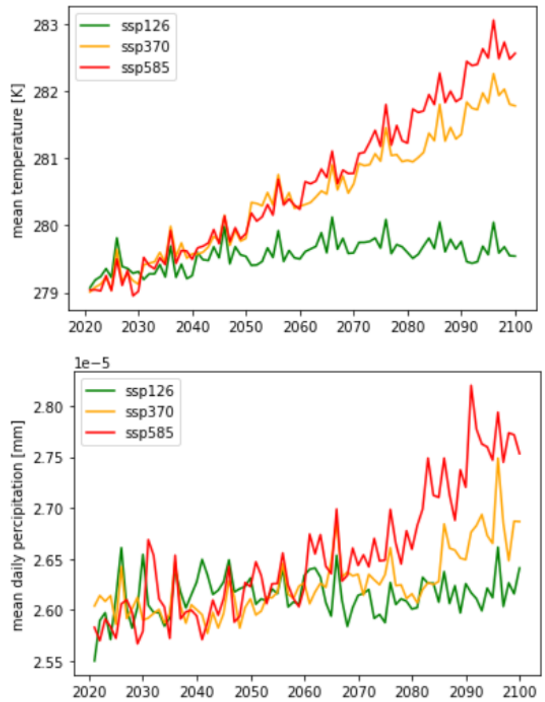

Figure 5: Predicted CMIP6 mean temperatures and daily percipitations for the three different climate change scenarios.

and interactions of crop yields, temperature and percipitation. The only bevorehand already mentioned underlying problem here is still that rain and percipitation data might not be enough to predict the yields in a usable manner, but for using additional data one might need access to more GPU's, because of the computational costs this dataset will bring with it. But luckily our code is structured modular, such that one only has to add a little bit of code, if one indeed has the access to multiple GPU's.

References

[1] Iizumi, T. Global dataset of historical yields v1.2 and v1.3 aligned version. Publisher: PANGAEA type: dataset.

[2] LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. Backpropagation applied to handwritten zip code recognition. *Neural Computation 1*, 4 (Winter 1989), 541–551.

[3] Müller, W. A., Jungclaus, J. H., Mauritsen, T., Baehr, J., Bittner, M., Budich, R., Bunzel, F., Esch, M., Ghosh, R., Haak, H., et al. A higher-resolution version of the max planck institute earth system model (mpi-esm1. 2-hr). *Journal of Advances in Modeling Earth Systems 10*, 7 (2018), 1383–1413.

[4] Muñoz Sabater, J. ERA5-land hourly data from 1981 to present, 2019.

[5] Reddy, V. R., and Pachepsky, Y. A. Predicting crop yields under climate change conditions from monthly GCM weather projections. 79–86.

[6] van Klompenburg, T., Kassahun, A., and Catal, C. Crop yield prediction using machine learning: A systematic literature review. 105709.