Selection of methodological approach and methodological improvement suggestions

Selection of modelling approach

The modelling of crop growth and crop yields has a long history and extensive literature on crop modelling approaches exist (for an overview see Van Ittersum and Donatelli 2003). Generally, climate impacts on crop yields are assessed by using process-based models (Olesen and Bindi 2002, Easterling et al. 2007) and there has been work on estimating the effects of a nuclear winter scenario with a DSSAT model (Xia and Robock, 2013). Process-based crop models simulate the physical plant growth of a crop in great detail. However, these models are not very well suited for the use case presented in this work. Firstly, calibrating and running the models is too time consuming and data intensive for a first assessment and therefore, beyond the scope of the work. Furthermore, the crop models are currently not well adapted to the requirements of the analysis at hand as management techniques and pesticide influences are sparsely implemented. A second approach is the modelling of effects based on empirical relationships between yield and the explaining variables as reported in the literature. This method, however, does not allow for the level of spatial detail sought in this assessment. Multiple regression or generalized linear models were considered as the third alternative. Statistical regression models have previously been applied to assess factors that influence spatial variability in observed yield (Bakker et al. 2005, Kaufmann and Snell 1997, Reidsma et al. 2007) and in predictive crop modelling contexts (Ferraro et al. 2009). The approach was chosen as the best fit for the proposed use case.

The elected methodological approach proved suitable for the use case in this thesis. The fitted generalized linear models showed good agreement with the data (ρ^2 between 0.34 and 0.47) and almost all chosen variables had a significant effect on the crop yield ($p < 0.05$). The models provide a solid basis to further develop the statistical approach by including more relevant variables or applying specific variants of the regression methodology (see below).

Suggestions for improving and refining the approach

The generalized linear model approach provided good results for a first spatial estimate. Nonetheless, as discussed above, there is room to improve the fit of the generalized linear model beyond the scope of this thesis. Several limitations are listed above, namely the misalignment of the spatial value distribution between datasets, the differing weight of each yield value according to the harvested area and autocorrelation between values in close vicinity. All these aspects can be addressed by applying special forms of the regression model.

A weighted regression can be employed to ensure that each yield value is included into the model according to the size of the harvested area in the same cell. The same concept is also useful to handle autocorrelation in the data. In a geographically weighted regression, the cells are weighted according to their proximity. The implementation is more difficult, however, as the calculation of the weights is more complex.

The confounding effect of misaligned data can be eased in two different ways. First a different resolution can be chosen by choosing coarser input datasets. Datasets with a coarser resolution generally have a lower level of uncertainty. Ideally, datasets are retrieved from one source. In the use case at hand a country-level analysis could be sensible as most of the variables can be obtained from the FAOSTAT database.

Apart from managing the resolution, the fuzzy regression is a good tool to correct for high variability and mismatch in the data. Huang et al. (2010) recommend the application of soft computing like fuzzy regression for agricultural predictions as real-world data tends to be too imprecise to be analyzed with hard computing methods. In contrast to “hard” methods like a regular generalized linear model, fuzzy logic does not decide between true and false but rather estimates the degree of truth for a value. Based on this logic it is better equipped to handle imprecise relationships and confounding factors.

Next to the numerous special regression forms, accounting for missing variables is another method to improve model accuracy. Missing variables can be additional explaining factors but also interactions between already included variables. The present model does not include interactions. The categorical variables were coded as dummies to be considered in the generalized linear model, but this type of codification does not allow for a correct fitting of the interactions. Instead, effect coding must be applied.

References

Bakker, M.M., Govers, G., Ewert, F., Rounsevell, M., Jones, R. (2005). Variability in regional wheat yields as a function of climate, soil and economic variables: assessing the risk of confounding. Agriculture, Ecosystems and Environment, 110(3-4), 195–209.

Easterling, W.E., Aggarwal, P.K., Batima, P., Brander, K.M., Erda, L., Howden, S.M., Kirilenko, A., Morton, J., Soussana, J.-F., Schmidhuber, J., Tubiello, F.N., 2007. Food, fibre and forest products. In: Parry, M.L., Canziani, O.F., Palutikof, J.P., van der Linden, P.J., Hanson, C.E. (Eds.), Climate Change 2007: Impacts Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (pp. 273–313). Cambridge, UK: Cambridge University Press.

Ferraro, D. O., Rivero, D. E., Ghersa, C. M. (2009). An analysis of the factors that influence sugarcane yield in Northern Argentina using classification and regression trees. Field Crops Research, 112(2-3), 149-157. https://doi.org/10.1016/j.fcr.2009.02.014

Huang, Y., Lan, Y., Thomson, S. J., Fang, A., Hoffmann, W. C. (2010). Development of soft computing and applications in agricultural and biological engineering. Computers and Electronics in Agriculture, 71(2), 107-127. https://doi.org/10.1016/j.compag.2010.01.001 Manheim, D. (2020). The Fragile World Hypothesis: Complexity, Fragility, and Systemic Existential Risk. Futures, 122, 102570. https://doi.org/10.1016/j.futures.2020.102570

Kaufmann, R.K., Snell, S.E. (1997). A biophysical model of corn yield: integrating climatic and social determinants. *American Journal of Agricultural Economics*, *79*(1), 178–190.

Olesen, J.E., Bindi, M. (2002). Consequences of climate change for European agricultural productivity, land use and policy. *European Journal of Agronomy 16*(4), 239–262.

Reidsma, P., Ewert, F., Oude Lansink, A. (2007). Analysis of farm performance in Europe under different climate and management conditions to improve understanding of adaptive capacity. *Climatic Change 84*(3), 403–422.

Van Ittersum, M., Donatelli, M. (Eds.). (2003). Modelling Cropping Systems: Science, Software and Applications [Special Issue]. *European Journal of Agronomy*, *18*(3-4).

Xia, L., Robock, A. (2013). Impacts of a nuclear war in South Asia on rice production in Mainland China. *Climate Change*, *116*(2), 357-372. https://doi.org/10.1007/s10584-012-0475-8