**Carl von Ossietzky Universität Oldenburg**
**Institut für Chemie und Biologie des Meeres**
Conducted at International Institute for Applied System Analysis

**Master Thesis in Environmental Modelling**

# Risk management for sustainable food security in West Africa

Presented by:
Debbora Mara Leip
Born on 06. September 1995 in Büdingen
Matriculation Number: 4587207
Tannenweg 14, 63683 Ortenberg, Germany
debbora.leip@uni-oldenburg.de

First Examiner: P.D. Dr. Jan Freund
Second Examiner: Dr. Matthias Wildemeersch

Ortenberg, 14th August 2020

# Acknowledgement

When I decided to write my master thesis at IIASA on the topic “Risk management for sustainable food security in West Africa”, it was already clear that this would be a rather ambitious project. Looking back, it was even more work than I expected, and there is still much more that we could have done if there would have been more time. But despite the high work load, I am very grateful to have had this opportunity and for the people that shaped this experience.

I want to thank Jan Freund, for his constant support throughout my time studying at the Institute for Chemistry and Biology of the Marine Environment of the University of Oldenburg.

Special thanks go to Matthias Wildemeersch, who supervised my master thesis at IIASA. Thank you for putting so much effort into the supervision of this project. Thank you for always finding the time for a Skype call whenever I ran into problems, despite my emails being on short notice most of the times. Thank you for patience on a level I could not have asked for, putting up with my endless discussions about details in the last months. Your commitment turned this master thesis project into a great experience, way beyond my expectations, and allowed me to learn so much on so many levels.

I also want to thank my best friends, Kirsten Fischer and Max Kanold, for their support in the most stressful phase of my master thesis and for reading through the final thesis to point out possible improvements.

And finally, my deepest gratitude goes to my parents for always supporting and believing in me. In particular, I want to thank my father for finding the time for discussions about the content of this project and for great advice, and my mother for her encouragement whenever my frustration level would spike.

# Abstract

In this study, we developed a stylized two-stage stochastic optimization model to analyze sustainable food security in West Africa. To this end, we subdivided West Africa into clusters, based on which a simplified spatial dependence structure of crop yields was defined. Crop yield distributions are the main input and the source of uncertainty in the model. The model includes food availability in terms of a certain food demand, and the sustainability dimension was implemented by increasing farmers resilience to extreme events through a large scale insurance scheme financed by taxes on agricultural profits. Objectives in the model are meeting the food demand in each year, and solvency of the insurance scheme after payouts to ensure a successful operation over time. As such objectives in general cannot be met for every realization of the uncertain parameters, they were included as probabilistic constraints that only need to be satisfied with certain probabilities. To ensure that the required probabilities are met, second stage penalties for violation of the constraints were included. The model output informs on optimal crop area allocations for each crop in each cluster and year to minimize the sum of first stage crop cultivation costs and second stage penalties that arise if objectives are not met, respecting limits given by available arable area in each cluster. It thereby decides on the best balance between risk prevention through first stage investments and post-disaster payments according to second stage penalties. Results show that in theory the available arable area in West Africa is currently sufficient for a high probability of food security. For the insurance scheme, we find risk pooling to be a key strategy to allow for high probabilities of solvency, making a supranational approach essential. Population growth will put additional stress on food security, as the current positive trends in crop yields cannot keep up with the the expected increase in population size. Therefore, apart from a large scale insurance scheme, an approach to sustainable food security in West Africa should include actions such as investments in closing the yield gap through sustainable intensification of agricultural production, long-term investments in disaster preparedness, or investment in farmer advisory systems to enhance the application of best management practices.

# Contents

# List of Figures

# List of Tables

# List of Algorithms

# Introduction

In 1996, at the World Food Summit organized by the Food and Agriculture Organization of the United Nations (FAO), food security was defined as the state in which "all people, at all times, have physical and economic access to sufficient, safe and nutritious food to meet their dietary needs and food preferences for an active and healthy life"[1]. Four "pillars of food security" were added to this definition at the World Food Summit 2009, namely "availability, access, utilization and stability"[2], which recently were supplemented by two additional dimensions, "agency and sustainability", in the latest report on food security and nutrition by the High Level Panel of Experts on Food Security and Nutrition (HLPE), 2020[3]. In 2015, the United Nations (UN) addressed food security in their second Sustainable Development Goal (SDG) for 2030, which is to "end hunger, achieve food security and improved nutrition and promote sustainable agriculture"[4], in general referred to as *Zero Hunger*[5]. The risk of hunger is often associated with the first of the six food security dimensions, i.e. food availability[6]. But while increasing humanitarian aid and improving agricultural technologies work towards reducing the risk of hunger in many areas of the world, climate change and rapid population growth are counteracting these efforts[7].

According to the 2018 *Africa Regional Overview of Food Security and Nutrition* by FAO and the United Nations Economic Commission for Africa (ECA), Africa is not on track to meet SDG2[8]. The prevalence of undernourishment in Africa is rising since 2015 and with 257 million people, one fifth of the African population was affected in 2018. Nearly half of the increase in undernourished people since 2015 occurred in West Africa[8]. Among the major causes of hunger in Africa are poverty, pre- and post-harvest losses due to high incidence of pests and diseases, as well as conflicts, wars, and corruption[9]. But already now, poor climatic conditions and climate variability and extremes play a key role in increasing food insecurity[8]. Future impacts of climate change on African ecosystems are expected to be high[10]. Sub-Saharan Africa is especially vulnerable, as it is strongly relying on rainfed agriculture, struggles with high poverty rates and poor infrastructure, and faces the fastest population growth world wide[7,8].

During the 21$^{st}$ century, the rise in temperature in Africa for each season will most likely exceed the global average temperature increase[10] and the continent will increasingly suffer from extreme temperature, tropical storms, droughts and floods[8]. Temperatures exceeding a certain threshold during the development phase of crops, even if only for a short period, are likely to reduce yields[11]. Changes in seasonal rainfall patterns affect the growth of crops as well. Furthermore, rising climate variability and extremes

can also have effects on the occurrence and distribution of pests and diseases, thereby additionally damaging harvests.

Droughts have led to humanitarian crises before, such as the 2012 Sahel crisis [12], and are predicted to increase in frequency and severity [8]. Regionally or countrywide, droughts can cause higher food insecurity and even famines [13], which in turn might negatively affect the world food market and lead to civil uprising [14]. At the household level, a single extreme event can destroy the livelihood of farmers, who often lack the resources and resilience to overcome such events. This leads to migration and urbanization on one hand [15], and can change agricultural practices on the other. Specifically, the risk of extreme events can cause farmers to choose safer crops with low return, accepting a loss of up to 20% of potential income, which can be seen as an implicit insurance premium [8]. Both of these effects on household level put additional strain on food security by decreasing agricultural production.

Often, catastrophic natural disasters such as floods or droughts are mainly met by post-disaster actions [16], to provide food, shelter, medicine, and other essential facilities to the affected people and to rebuild homes or livelihoods. While this is definitely a necessary action in the immediate aftermath of an extreme event, it will not reduce the risk for similar disasters in the future. Even though it is broadly recognized that risk reduction and increase of resilience is required for sustainable development, in Africa less than 5% of yearly humanitarian funding has gone to disaster preparedness and prevention in recent years [8].

For a path towards a sustainable future and to reach SDG2, all the above mentioned aspects and more have to be taken into account. To increase the resilience to extreme events and narrow down the yield gap[1] as much as possible, investments in modern agricultural technology, drought resistant seeds, irrigation, and other infrastructure need to be made. Appropriate crops have to be cultivated and best management practices applied. For that, farmers need to have the necessary knowledge as well as the resources and security to apply this knowledge. To provide the needed financial resilience to extreme events and make practices as the above described implicit insurance policy obsolete, real insurance against yield losses due to extreme events should be available to farmers. Currently, this kind of insurance is still rare in sub-Saharan Africa, and if it is available, farmers generally cannot afford the high premiums without subsidization [8]. Such high insurance premiums arise when payouts are subject to correlated risks. If a drought leads to reduced yields, it will most likely affect a large area and thus many farmers at the same time, leading to huge simultaneous payouts. One approach to mitigate this problem is risk pooling: instead of focusing on a single region, insurances cover areas under different agro-climatic conditions, for which the risks are not correlated.

Obviously, these options are restrained by availability of money and other resources. To decide on the best possible path forward, short- and long-term approaches as well

[1] The yield gap is the difference between actual yields and potential yields under given climatic circumstances.

as possibilities of risk reduction and preparedness have to be assessed and the best balance of actions within the given limitations has to be identified. As the trade-off between different investments is complicated, optimization models are a useful method in the decision making process. Such models can inform on combinations of actions that minimize the sum of direct and indirect costs while ensuring objectives as food security and farmers resilience, under given constraints such as total agricultural area, available water, or sustainability requirements.

Weather conditions, and especially extremes such as floods and droughts, as well as other aspects affecting crop yields have a high uncertainty. Focusing only on mean changes and omitting variability would underestimate the impact of climate change and other factors[8]. A stochastic modeling approach including the uncertainties and risks related to agricultural yields can give more insight and better advice than deterministic models focusing on specific scenarios. Instead of giving an optimal result for one scenario, a stochastic optimization model can give a general solution depending on the distribution of possible scenarios. For the given situation, a two-stage stochastic optimization model can be developed, consisting of strategic actions in the first stage, and adaptive actions in the second[17]. Strategic actions have to be chosen while the actual conditions are still uncertain, whereas the second stage actions follow after the outcome is known. The latter are necessary to counteract potential shortcomings with respect to the objectives and their consequences. A food security objective can be included by a certain caloric demand which has to be supplied by domestic agricultural production. If the demand is not met, costs for second stage actions will arise, which can correspond to direct costs of crop imports or indirect costs related to foreign aid, as well as indirect costs of socio-economic consequences in the case of hunger or famine.

In this project, we pursue the above described approach and present a proof-of-concept by developing a stylized stochastic optimization model for sustainable food security in West Africa. Initially, we only include a constraint on the total agricultural area and a second stage cost in the case that a designated caloric demand is not met. The output is an allocation of agricultural area in a given year to the different considered crops, such that the expected costs under the given uncertainty in possible yields are minimized. Subsequently, the model is extended to include a time dimension and a government fund, which is built up by farmers paying taxes on their profits. The fund is then used to pay a guaranteed income to farmers in case of an extreme event. This makes farmers resilient against years with bad harvests, thus adding a sustainability dimension to the food security objective. Solvency of the government after payouts is included as an additional objective. If this objective is not met, the negative consequences of missing financial resources will lead to additional second stage costs. These can correspond to interests on loans, indirect costs through political dependence following foreign aid, or socio-economic consequences if the guaranteed income can no longer be provided.

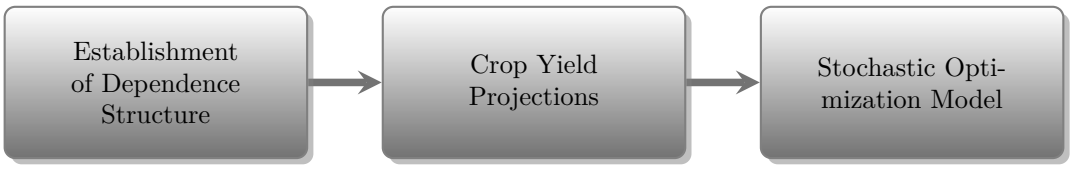
The optimization model thereby takes the point of view of the government and does not include farmers' direct objectives as profit maximization. It is a decision supporting

model that identifies optimal crop area allocations given specific circumstances, which can be used by the government to set up prescriptive policies in order to reach sustainable food security. The model however does not capture costs and constraints that might be associated with the implementation of such policies, as this aspect is beyond the scope of this project.

We focus on government interventions such as tax rates and the level of guaranteed income, developments in demography and agricultural productivity, and the effect of spatial correlation of yields and analyze their respective impact on food security and solvency of a catastrophe fund. We try to understand how government decisions and general developments are reflected in sustainable food security, and what impact a large scale insurance system might have on the road to achieving *Zero Hunger* in West Africa. More specifically, we address the following research questions: First, to what extent is it feasible to ensure food security and contribute to its sustainability by supporting farmers' resilience to extreme events through a large scale insurance system? How can the government regulate this feasibility by policy decisions concerning the financing mechanism and the government fund? Second, given the structure of the model, how does spatial correlation affect the overall costs of this approach to sustainable food security? To what extent can pooling of uncorrelated risks enhance the feasibility of an insurance system to build up farmers' resilience to extreme events? And finally, how will medium-term developments in agricultural productivity and population growth impact sustainable food security?

# 1. Data Analysis

The development of any model is preceded by extended data analysis, used to build an understanding of the situation at hand, to develop the structure and relations within the resulting model, and to generate input data. The main input to the stochastic optimization model for sustainable food security are yield values for different crops within West Africa. Yields vary with space and time but also show correlation in both of these dimensions. Therefore, we first build an understanding of the dependence structures and develop a way to include them in our model, while taking into account the trade-off between accurate representation of the true correlations and model limitations arising both from available data and computational resources. The applied method is based on cluster analysis using a high resolution data set of a historic drought index describing weather patterns in West Africa over space and time. Once the underlying dependence structure is defined, a yield model needs to be established to project yield distributions for the time period covered by the stochastic optimization model. Several approaches were tested using yield data simulated by different crop growth models and data on related variables such as fertilizer application and climate indicators. The final yield model relies solely on extrapolation of trends using a historic yield data set. Finally, data on food demand, energy content of crops, and both direct and indirect costs are




| | | |
|---|---|---|
| Input: | | |
| · historic drought index | · historic yield data<br>· subdivision of West Africa into clusters<br>· assumed spatial dependence of yields | · crop yield realizations<br>· direct & indirect costs<br>· additional parameters such as available arable land and food demand |
| Output: | | |
| · spatial subdivision of West Africa into clusters<br>· definition of a dependence structure based on the clusters | · crop yield distributions per cluster and year | · allocation of arable land to different crops |

Figure 1.1.: Schematic overview of the steps involved in setting up the stochastic model

included into the optimization framework. A schematic overview of the different steps of setting up the stochastic optimization model is given in Figure 1.1.

The first section of this chapter (Section 1.1) reviews in detail the main data sets used for data analysis within the scope of this project. In Section 1.2 we analyze different options for spatial and temporal dependence within the model, discussing their advantages and disadvantages, and substantiate the decision to use a dependence structure based on clusters as subregions of West Africa. The clustering algorithm and its application are described in detail. Finally, we address different approaches to project yield distributions in Section 1.3. All of the data analysis was conducted using Python 3.7 [18] (see Appendix A for details on the implementation).

## 1.1. Data Sets

### 1.1.1. Standardized Precipitation-Evapotranspiration Index

As the weather condition and in particular water availability has a strong impact on crop growth, we include a drought index in the data analysis of this project. Quantifying droughts in terms of intensity and temporal or spatial magnitude is a difficult task and much research has been done on developing techniques for drought analysis and monitoring, with objective indices being the most commonly used method [19]. Several different drought indices have been proposed over time, with different strengths and weaknesses. The drought index used for the analysis of climate patterns in this project is the Standardized Precipitation-Evapotranspiration Index (SPEI), developed and first presented by Vicente-Serrano et al. in 2010 [19]. In this section we describe the two primarily used indices before 2010 and the reasons why the SPEI is generally seen as an improvement, the details of the computation of the SPEI, certain relevant properties of the SPEI, and the openly accessible database SPEIbase.

#### The sc-PDSI and SPI in comparison to the SPEI

The Palmer Drought Severity Index (PDSI) was first developed in 1965 by Palmer [20] and was an important step in the development of drought indices [19]. It is based on a soil water balance equation that takes into account different factors influencing the occurrence of droughts such as precipitation, evapotranspiration and soil moisture condition, and accounts for climatic differences between locations and seasons of the year to allow comparison over time and space [21]. The resulting PDSI has a fixed temporal scale of six to nine months [22] and ranges from negative values describing dry periods to positive values describing wet periods [23]. Based on progress in computing resources, the PDSI was later improved and an updated version called self-calibrating Palmer Drought Severity Index (sc-PDSI) was introduced by Wells et al. in 2004 [21]. However, droughts are so-called multiscalar phenomenons, as the responses of different ecosystems to dry

periods is apparent on different timescales [24]. Due to this, it is possible that in the same area one ecosystem shows severe drought conditions (e.g. low river flows) while another exhibits normal conditions (e.g. crops) [24]. Thus, the timescale is relevant to identify different types of droughts, e.g. the more quickly arising hydrological drought or the long-term socio-economic drought [25]. The fixed temporal timescale of the PDSI and sc-PDSI of six to nine months does not allow a distinction between such types of droughts and short-term droughts might not be detected at all. As a typical time period for crop growth ranges between three and four months, this shortcoming is a reason not to use the PDSI or sc-PDSI for data analysis within this project.

The Standardized Precipitation Index (SPI), first introduced by McKee et al. in 1993 [26], in contrast is able to differentiate between different time scales. It uses a moving average of monthly precipitation data with respect to a time window of either 3, 6, 12, 24, or 48 months. These time series are fitted to a gamma distribution and then standardized to represent the deviation from normal conditions, with mean zero and a standard deviation of one. But while allowing to investigate droughts on different time scales and thereby overcoming one of the shortcomings of the sc-PDSI, the SPI misses a detailed consideration of different factors influencing droughts by just taking into account precipitation.

The later developed SPEI adds to the multiscalar approach of the SPI by including not only precipitation but also evapotranspiration and hence improves the index' ability to monitor droughts. Since its development it has been widely used, e.g. by Miah et al., 2017 [27] or Tam et al., 2018 [28].

#### Computation of the SPEI

The methodology underlying the computation of the SPEI is described in detail by Vicente-Serrano et al., 2010 [19], on which the following paragraphs are based.

The first step is the calculation of the Potential Evapotranspiration (PET), for which the Thornthwaite method is used [29]. This method only requires monthly mean temperatures as data input but the performance of resulting drought indices is similar to results using more elaborated PET calculations [30]. Apart from the data on monthly mean temperatures, the approach uses a coefficient $K$, which is given by a function of latitude and month, to correct for location and seasonality. After obtaining the time series $PET_i^{(g)}$ for each location $g$ and months $i$, the monthly water surplus or deficit can be calculated: $D_i^{(g)} = P_i^{(g)} - PET_i^{(g)}$, where $P_i^{(g)}$ is the monthly precipitation. According to the selected timescale $k$, the values of $D_i^{(g)}$ are then aggregated as

$$X_i^{(g,k)} = \sum_{n=0}^{k-1} D_{i-n}^{(g)} \tag{1.1}$$

In the next step, a probability function is fit to each of these time series of aggregated water balances. A variety of different distributions was tested, and ultimately a three-parameter log-logistic distribution was chosen, as it exhibits the best behavior, in particular at the most extreme values. The cumulative distribution function, with parameters estimated separately from each time series, is given by

$$F^{(g,k)}(x) = \left(1 + \left(\frac{\alpha^{(g,k)}}{x - \gamma^{(g,k)}}\right)^{\beta^{(g,k)}}\right)^{-1} \tag{1.2}$$

The last step is the standardization of the time series $X_i^{(g,k)}$ using the distribution given by $F^{(g,k)}(x)$ and a numerical approximation method. This results in a drought index following approximately a Gaussian distribution for each location and time scale, with mean zero and a standard deviation of one. Periods drier than the average are indicated by negative values, while wetter periods have a positive SPEI. Note that the SPEI is standardized to local normal climatic conditions, and therefore a similar SPEI value can correspond to very different values in terms of absolute water deficit depending on the given location[13].

### Comparability of the SPEI in different locations

The standardization of the SPEI can increase comparability between different locations. As ecosystems adapt to their normal circumstances, a certain deficit in precipitation might have very different consequences on locations with different normal conditions. Hence by normalizing each location separately, the index quantifies the deviation from normal and thus allows to compare the stress put on ecosystems by the respective conditions[24]. However, in our case we are not looking at natural ecosystems, but at crops cultivated by local farmers. Of course farmers decide on which crops to grow depending on the given climatic conditions, but if one crop type, e.g. wheat, is grown in two different locations, better comparability with regards to weather conditions affecting the wheat yields might be given by an indicator such as water availability. Therefore, we also include the unstandardized SPEI, i.e. water deficit calculated as the difference between precipitation and PET, in the regression analysis discussed in Section 1.3.1.

### The SPEIbase

The SPEIbase was developed as a scientific and openly accessible database of gridded SPEI at $0.5° \times 0.5°$ resolution[24], and is continuously updated to cover more recent years. At the time of writing, it covers the period from January 1901 to December 2018 on a monthly basis, which is referred to as version 2.6 of the data set and is the version used in the context of this thesis. Calculation of the SPEI is based on temperature and precipitation data from the Climatic Research Unit (CRU), more specifically the CRU

TS v. 4.03 data set [31]. Theoretically, the SPEI can attain any real value $x \in \mathbb{R}$ [13], but it typically ranges from $-2.5$ to 2.5, which corresponds to the 98.8% confidence interval [24]. The SPEI uses the same categories for classifying the intensity of droughts as the SPI: values below $-1$ indicate the presence of a drought, with droughts corresponding to a value below $-2$ considered as extreme events [26].

The SPEIbase was used for regression analysis in order to set up a simplified data-driven deterministic yield model. The original intention was to use this to calculate yield realizations from SPEI values which are drawn from given SPEI distributions. Thereby, the source of uncertainty in the stochastic optimization model would be the climatic conditions described by the SPEI. This approach had to be discontinued due to insufficient quality of regressions between crop yields and the SPEI (see Section 1.3.1). However, the SPEIbase is still used for cluster analysis to divide West Africa into subregions, which is the basis of the spatial dependence structure implemented in the resulting stochastic optimization model (see Section 1.2).

### 1.1.2. The GGCMI phase 1 crop yield data set of the AgMIP project

One of the two crop yield data sets used for data analysis within this project was established in the context of the Agricultural Model Intercomparison and Improvement Project (AgMIP), which was founded in 2010 by international agricultural modelers to link climate, crop, and economic modeling communities. The aim of AgMIP is to improve agricultural models and scientific assessment of the sustainability of agricultural systems, e.g. regarding food security or impacts of climate change [32,33]. As there exist many different Global Gridded Crop Models (GGCMs) with a wide range of model outputs, the Global Gridded Crop Model Intercomparison (GGCMI) phase 1 was set up to better understand the skill of different models by comparing their ability to reproduce historic yield data [34]. In this process a large data set of global crop model outputs on historic yields was accumulated. The following two paragraphs are based on information on this data set given by Müller et al., 2019 [34].

In the GGCMI phase 1 modeling exercise, 14 different GGCM were considered, each using up to 11 alternative weather data sets as main input source. The models were run for three harmonization levels: the default version for each model respectively, a version with crop growing seasons and fertilizer input harmonized between the different models, and a version with harmonized growing seasons and unlimited nutrient supply. For each version, a fully irrigated and a fully rainfed scenario were implemented.

All model outputs are given as annual time series for a spatial resolution of $0.5° \times 0.5°$, but cover different periods of time. The main output variable is crop yield in tonnes per hectare, given for four major crops by most models, i.e. rice, maize, wheat and soybean, and for some additional secondary crops by a few models. Other output variables range from characteristics of crop growth (such as planting day or date of crop maturity), to model specific weather data (such as accumulated precipitation or solar radiation

during crop growth), and technical data (such as applied irrigation water or fertilizer). Whether a particular output variable is reported depends on the model and settings chosen.

A strength of the AgMIP GGCMI phase 1 data set is the vast availability of yield data including corresponding information about other relevant aspects such as fertilizer application. However, even though this data represents historic yields, different models can vary strongly in their output, indicating high uncertainty in the data. By choosing one of the models, yields are directly based on specific relations to input variables, as GGCMs in general are deterministic models. Understanding where differences come from and how the models can be improved was part of GGCMI phase 1, but further research is necessary and encouraged by the involved scientists [34]. Within this thesis, the data set was used for regression analysis of crop yields and the SPEI (see Section 1.3.1), but is not included in the final stochastic optimization model.

### 1.1.3. Global Dataset of Historic Yields (GDHY)

Limitations of the GGCMI phase 1 data set suggest to include a historic yield data set for higher reliability. Although crop yield observations are important in many aspects of agricultural, environmental, and sustainability research, data availability at high spatial resolution is very limited [35]. A main source is the global yield data set of FAO, which reports yield statistics at country level [36]. A global gridded data set of historic yields was published in 2008 by Monfreda et al., but only covering the year 2000 [37]. In 2014, Iizumi et al. published a data set combining the high spatial resolution and the temporal coverage, presenting a global gridded data set for major crops from 1981 to 2011 at $1.25° \times 1.25°$ spatial resolution [35]. Even though the data set is called Global Dataset of Historic Yields (GDHY), it relies on modeling yields using among others the aforementioned data sets by FAO and Monfreda et al. as input. The modeling procedure is described in detail by Iizumi et al. [35], and the following paragraph gives an overview based on this description.

In a first step, time and length of crop growing periods per crop and grid cell are modeled by fitting normal distributions to planting and harvest dates. From each distribution, 500 realizations are generated and the accumulated Net Primary Production (NPP)[1] in the growing season is calculated from daily satellite derived NPP values. In some regions, maize, rice, or wheat can be grown twice a year. Where this is the case, FAO country yields are assumed to represent the main cropping system, and yields for the secondary harvest are calculated using the ratio of NPP between the two cropping systems. It is then assumed, that the harvest index, i.e. the fraction of yield in total aboveground biomass, follows political boundaries. Therefore, to get gridded yield values, the yearly country level yield given by FAO is scaled by the ratio of mean NPP in each cell to

[1] NPP refers to the amount of biomass an ecosystem accumulates, excluding losses from the process of respiration.

the average NPP of that country. Summarizing, the spatial variation in the modeled yield data follows that of the NPP, while the temporal variation is mainly given by the FAO data set. Thus while the GDHY data set is also a model output, a pivotal difference to data sets as described in Section 1.1.2 is that no assumptions on the relation between yields and climate variables or other factors are needed, as the approach relies on downscaling instead.

The GDHY was later updated to a higher spatial resolution and to cover a longer time period. The current version, which is used in this project, provides gridded yield data for maize (main and secondary), wheat (spring and winter), rice (main and secondary), and soybean at $0.5° \times 0.5°$ spatial resolution with yearly data from 1981 to 2016 [38]. Out of the given crops, only maize major and rice major provide data for a significant amount of cells within West Africa, and thus analysis including GDHY data is reduced to these two crops. Ultimately, the GDHY data set is used to predict yield distributions (see Section 1.3.2) which are a direct input to the stochastic optimization model.

## 1.2. Spatial dependence of crop yields

The stochastic optimization model relies on yield realizations for different crops for up to 25 years starting from 2017, covering the area of West Africa between 19° West and 10.5° East and between 3° and 18.5° North (see Figure 1.2 (A)). As none of the major crops are grown in the entire region, only the two widest spread crops of the main cropping system, i.e. maize and rice , were used when introducing GDHY data on crop yields to the model. The area was then slightly reduced to the subset of cells for which at least one of them reports data (see Figure 1.2 (B)).

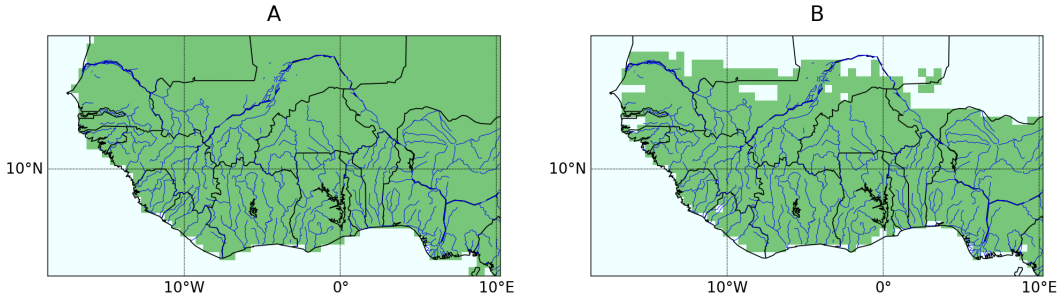



Figure 1.2.: Visualization of the considered area of West Africa covered by the SPEI data set in (A) and covered by GDHY data on rice and/or maize in (B)

Crop yields can vary greatly over the considered range of space and time. There are many aspects that can influence yields: from quality of equipment to soil fertility or fertilization and of course weather related variables. Different factors are subject to different correlation structures, as e.g. fertilization might depend on governmental subsidies and thus follow political borders, while weather will rather respect climate zones then man-made partitions. We are mainly interested in the weather-related aspect of

uncertainty in crop yields, therefore we used the SPEI data set described in Section 1.1.1 to set up a dependence structure instead of directly analyzing yield data. This dependence structure can be utilized to predict SPEI distributions which subsequently can be used as input for a data-driven yield model. The approaches in the first part of this section (Section 1.2.1) were considered with this motive. However, due to insufficient quality of regressions between crop yields and SPEI, no simple crop yield model based on SPEI could be established (see Section 1.3.1). Therefore, we later chose to directly use projected yield distributions as input to the stochastic optimization model (see Section 1.3.2). Nonetheless, the dependence structure results from SPEI values (Section 1.2.2), underlining that our main focus lies on weather-related aspects of uncertainty in crop yields.

### 1.2.1. Considered approaches to cover spatial and temporal dependence

With a resolution of $0.5° \times 0.5°$, the SPEIbase gives about 1 400 cells within the considered area in West Africa and covers 118 years. While for each single cell we could analyze trends and make predictions of the separate SPEI distributions, this would not take into account the dependence between yields in different cells: It is highly unlikely that one cell faces a severe drought while its neighboring cells exhibit excessive rainfall. As explained in detail in Section 1.1.1, the SPEI per grid cell is approximately normal distributed with mean zero and standard deviation of one. Hence, a straightforward way to deal with spatial correlation would be estimating the parameters of a multivariate normal distribution. In order to include temporal changes, for each cell a linear trend can be subtracted from the SPEI to then estimate the distribution of the residuals. Projections of these trends can be added to the realizations of the estimated distribution to generate SPEI values of future years which respect given spatial dependences and temporal trends. However, if we let $S$ be the number of cells and $T$ the number of covered years, for a multivariate normal distribution $\frac{S^2+S}{2}$ correlation coefficients have to be estimated, with only $T$ available high dimensional data points. In the given case, this amounts to almost a million parameters to estimate out of 1416 $S$-dimensional data points. Estimating correlation coefficients could be avoided by splitting SPEI values into different scenarios according to thresholds, and then using a multidimensional histogram to describe the joint distribution. But this approach runs into problems due to limited data availability as well: Even considering only two scenarios (dry or wet conditions) already gives $2^S$ possible combinations.

Even if an accurate multidimensional distribution of the SPEI values in West Africa could be found, more problems would occur when running the stochastic optimization model. The higher the complexity of the underlying uncertainties, the higher also the sample size needed to achieve a sufficient level of accuracy of the model output. This quickly exceeds available computational resources. In order to relax computational requirements, the spatial dimension of the food security problem needs to be reduced.

This decreases the spatial accuracy of the model by simplifying the dependence structure, but the resulting simplified structure can be analyzed with higher accuracy due to the relatively bigger number of observations.

Reducing the spatial dimension can be done by clustering cells into subregions of West Africa according to the behavior of the SPEI. There are different algorithms to accomplish this task, the method chosen in this project is described in the next section (Section 1.2.2). Once clusters are obtained, dependence within and between clusters still has to be quantified. For the former, we again considered using histograms, this time representing the distributions within each cluster: the lower number of cells $S'$ within a cluster reduces the number of possible scenario-combinations, while the number $T$ of observations remains the same. As the clustering algorithm we use is based on Pearson Correlation between the SPEI time series of the grid cells, we expect the cells within a cluster to be highly correlated, while correlation to cells of other clusters should be low. Hence, we can assume that for each point in time the values within a cluster do not vary much. Therefore, the complexity of the cluster-specific distributions used for SPEI generations could be further reduced by only considering combinations with all cells showing similar scenarios (e.g. corresponding to adjacent bins in the histogram). But even in this reduced form the number of possible combinations exceeds the number of observations already for very small cluster sizes.

We therefore assume full spatial dependence within clusters and work with average yield values per crop in the stochastic optimization model. Between different crops within the same cluster we assume full dependence regarding the presence of extreme yields. Yield distributions are subdivided in extreme and non-extreme yields by a given quantile. Hence, in a certain year either all crops within a cluster exhibit yields drawn from the lower quantile of the respective distributions, or all crops exhibit yields drawn from the remaining upper part of the respective distributions, depending on whether the cluster faces an extreme event or not (for more details see Section 3.1).

Correlations between separate clusters can be captured by different techniques. As we now work with a single distribution per crop and cluster, we could estimate $k$-dimensional multivariate distributions given $k$ clusters. Another wide spread method to connect different distributions are copulas. This method is based on Sklar's theorem, which states that the joint distribution of any two random variables $X$ and $Y$ with marginal cumulative distribution functions $F_X(x)$ and $F_Y(y)$ can be written as

$$H(x, y) = C\left(F_X(x), F_Y(y)\right), \; x, y \in \mathbb{R}, \tag{1.3}$$

where $C : [0, 1]^2 \to [0, 1]$ is called copula. Sklar proved that $C$ is uniquely defined for continuous $F_X(x)$ and $F_Y(y)$ [39]. By linking two distributions through a copula, the relation between them can be chosen independently from the marginal probability distributions, which is the method's main advantage [40]. Due to this characteristic, copulas are also suited to analyze tail dependencies, i.e. dependence between extreme

events, which can be very relevant in risk management [40,41]. Hence, copulas can be a fitting technique for dependence between clusters in this project.

However, the focus of this master thesis is to analyze a stylized stochastic optimization model as a proof-of-concept and spending more time on e.g. implementing copulas would have shifted this focus and exceeded the time frame of a master thesis. Therefore, in the scope of this thesis, the effects of changing correlation within West Africa are only analyzed by using a changing number of clusters as model input while assuming full independence between the given clusters. We acknowledge that this as well as full dependence within clusters are strong assumptions. Using more advanced statistical methods might have made a higher spatial resolution possible and copulas or other methods for dependence between different clusters should definitely be included in future work.

### 1.2.2. Cluster analysis

Cluster analysis in general is the task to divide a set of objects into subgroups called clusters, according to some sort of similarity measure. Objects assigned to the same cluster should be similar, while there should be dissimilarity between elements of different clusters. Cluster analysis has applications in a broad range of scientific fields, spanning from market research to archeology [42], and different clustering methods have been developed to fit the respective needs [43].

Hierarchical clustering algorithms start either with a single cluster made up of all objects which is then divided successively into smaller clusters (top-down), or with individual clusters for each object which are then merged successively into larger clusters (bottom-up). Other then for hierarchical algorithms, in partitional clustering the number of clusters is decided beforehand, and all clusters are determined in parallel [43]. Further groups of clustering algorithms include density-based [44], graph-based [45], and model-based methods [46].

#### The $k$-Medoids algorithm

We chose to use partitional clustering, as it allows to work with a changing number of clusters for a fixed set of objects. One of the most common partitional clustering algorithms is $k$-Means, which finds $k$ clusters in such a way, that the total sum of the euclidean distance between any object and the mean of its respective cluster is minimized. If we interpret the SPEI time series given for each grid cell as points in a $T$-dimensional euclidean space, $k$-Means could be applied. However, this would not reflect well the kind of similarity on which we want to base the clustering. We are interested in similar weather patterns within clusters, thus wanting a high correlation of the SPEI time series within a cluster. One approach would be to use Principal Component Analysis (PCA) to transform the time series and then apply $k$-Means on the resulting data, as done by

Diasso and Abiodun, 2015 [22]. However, intuitive interpretation of the data and resulting clusters is lost by this transformation. We therefore use the $k$-Medoids clustering method instead. The $k$-Medoids algorithm, also called PAM algorithm (Partitioning Around Medoids), is similar to $k$-Means, but instead of minimizing distance to the cluster mean, existing objects act as representatives of the clusters, i.e. the medoids, and the distance to those are minimized. By this adaptation, any distance function defined on the set of objects can be used [47]. The algorithm was first described by Kaufmann and Rousseeuw in 1987 [47], and the following paragraph is based on their description.

The $k$-Medoids algorithm starts by choosing $k$ objects as initial medoids. This can either be done randomly, or by using a greedy algorithm: As first medoid the object that minimizes the sum of distances between itself and all other objects is selected. Then in each iteration the object that most reduces the sum of distances between each object and its closest medoid, i.e. the total costs of the current configuration, is chosen as additional medoid, until $k$ initial medoids are found. Then for all combinations of medoid $m$ and non-medoid $o$ the new total costs in case of swapping $m$ and $o$, such that $o$ is a medoid and $m$ is not, are calculated. The swap with the lowest new total costs is performed if it improves the previous configuration. This process is iterated until no improving swap can be found, which terminates the algorithm. The clusters are then given by associating each object to the medoid it is closest to. A pseudocode description of $k$-Medoids is given by Algorithm 1.

```
Input  : Distance matrix of objects which are to be clustered and number of
         clusters k
Output: Allocation of objects to k different medoids, thus forming clusters
Initialize k objects as medoids
Calculate total costs (i.e. sum of distances between cells and associated
  medoids)
while total costs decrease do
    forall medoids m and non-medoids o do
        Compute new costs in case m and o were switched
        if this is the best change so far (i.e. giving the lowest costs) then
            Save m and o as m_best and o_best
        end
    end
    if swapping m_best and o_best decreases the total costs then
        Update the medoids and total costs accordingly
    end
end
return final medoids and the association of each object
```

**Algorithm 1:** Pseudocode of k-Medoids clustering algorithm

The $k$-Medoids algorithm is a greedy algorithm, choosing the locally optimal swap in each iteration. This strategy does not necessarily lead to a globally optimal solution [47]. It is a heuristic approach towards finding an optimum, as trying out all possible combi-

nations of $k$ medoids in general has too high computational costs[2]. A naive implementation of the $k$-Medoids algorithm will have a runtime of $\mathcal{O}(k(n-k)^2)$ per iteration [49]. Optimizations exist, e.g. by not calculating costs from scratch for each possible swap, as well as approximative algorithms with similar performance but lower runtime [48].

### Application of $k$-Medoids to SPEI data

The main motivation for cluster analysis in this project is to reduce the spatial dimension and work with single yield values per cluster. We therefore want high correlation between the cells within a cluster, making a metric based on the Pearson correlation coefficient the obvious choice as distance for the $k$-Medoids algorithm.

**Definition 1.1.** *Let $X = (x_1, \ldots, x_T)$ and $Y = (y_1, \ldots, y_T)$ be samples of two random variables, with sample means $\bar{x}$ and $\bar{y}$. Then the sample Pearson correlation coefficient is defined as*

$$r_{X,Y} := \frac{\sum_{i=1}^{T}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{T}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{T}(y_i - \bar{y})^2}} \tag{1.4}$$

**Lemma 1.2.** *Let $X = (x_1, \ldots, x_T)$ and $Y = (y_1, \ldots, y_T)$ be samples of two random variables, and let $r_{X,Y}$ be the sample Pearson correlation coefficient. Then*

$$d_{X,Y} := \sqrt{\frac{1}{2}(1 - r_{X,Y})} \tag{1.5}$$

*defines a distance metric based on the Pearson correlation coefficient.*

A proof of Lemma 1.2 is given by van Dongen and Enright, 2012 [50]. The resulting Pearson distance assumes values between 0 and 1, setting anti-correlated objects, i.e. objects with Pearson correlation $-1$, the furthest apart, while fully correlated, i.e. identical objects, have distance 0 [50].

Our main interest is in extreme dry events that can negatively affect yields. These might correlate over larger distances than normal weather driven patterns, thus clusters could change when focusing on correlation of tail events of the distribution. Patterns could also change over time, leading to different clusters depending on the considered time window. Therefore, except for the original version of the SPEI data set, several different modifications were used to calculate a Pearson distance matrix:

1. For each pair of time series only those years are taken into account for which at least one of them exhibits a dry event indicated by a corresponding SPEI value below a given threshold.

2. For a given threshold to define dry events, the data is modified to only encode whether such an event has happened (value 1) or not (value 0).

3. The data set is reduced to the most recent 30 years.

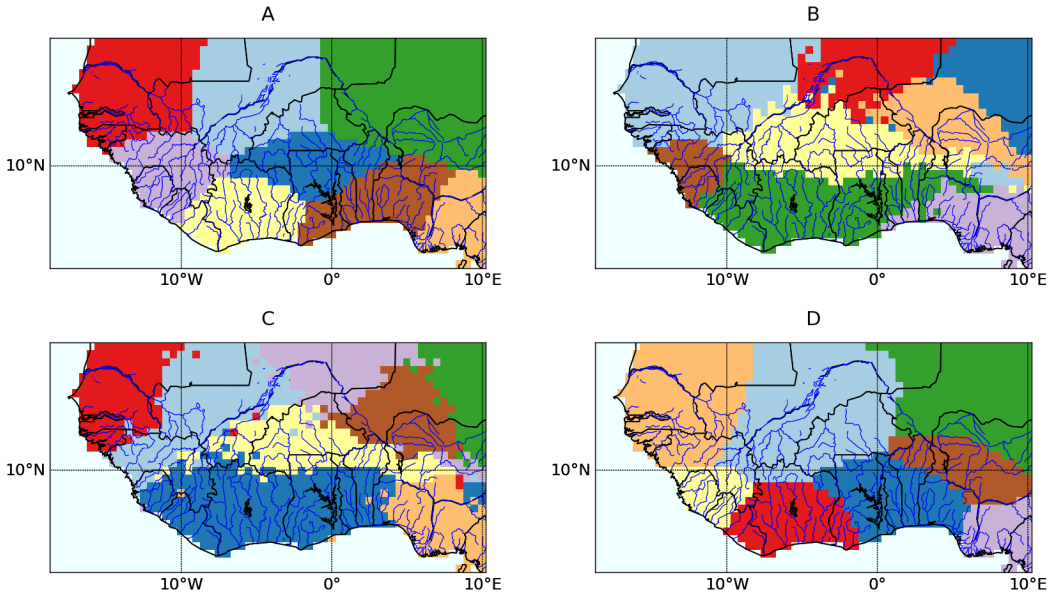



Figure 1.3.: Visualization of clusters resulting from the $k$-Medoids algorithm for $k = 8$. (A) uses the full SPEI data set, (B) uses time series reduced to occurrence of extreme events (modification 1), (C) uses a boolean data set encoding whether an extreme event took place (modification 2), and (D) uses the data set reduced to the most recent 30 years (modification 3).

However, reducing the data set to information on extreme events leads to frayed clusters as can be seen in Figure 1.3 (B) and 1.3 (C), which is an undesired effect. Using only the most recent 30 years gives similar kinds of clusters as using the full data set, but with a slightly different pattern as can be seen in Figure 1.3 (D). It is interesting to see that these patterns might change over time. However, it is not clear whether these changes reflect an actual change of weather patterns over time, or if it is just an effect of fluctuations within the data. In order to include all information embedded in the full data set, we use the original SPEI data as basis of the cluster analysis.

### Determining the optimal number of clusters

While the ideal clusters would have full dependence within a cluster and maximum distance to others, in empirical situations one typically does not get this kind of result. Therefore, to decide which number $k$ of clusters gives the best clustering, the trade-off between intra-cluster similarity and inter-cluster dissimilarity has to be analyzed.

The $k$-Medoids algorithm is run for $k$ in the range of 2 to 20, and measures for intra-cluster similarity and inter-cluster dissimilarity are calculated. For the former, the average distance of a cell to its corresponding medoid is taken, while the distance between different clusters is approximated by the average distance between a medoid and the closest other medoid, as medoids are seen as representatives of the clusters. Both

[2] Finding the global optimum of the $k$-Medoids problem is NP hard [48].

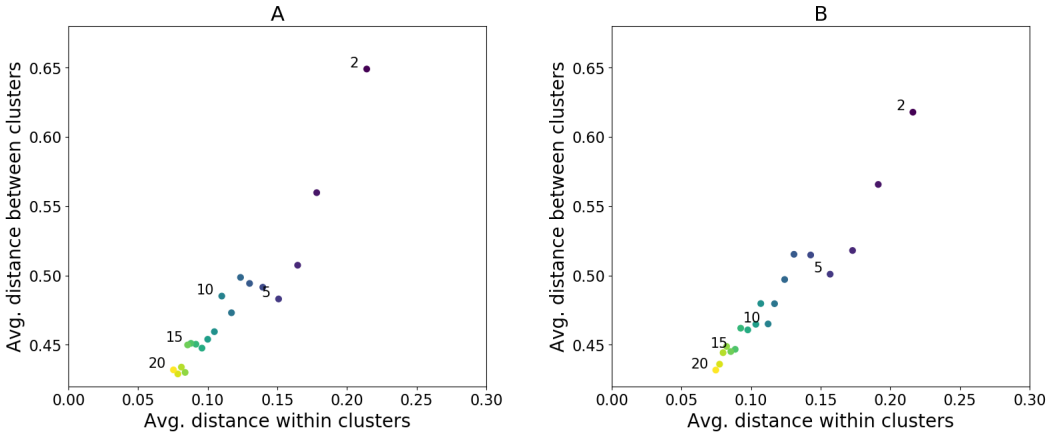



Figure 1.4.: Average intra-cluster distances plotted against average inter-cluster distances for the number of clusters $k$ between 2 and 20. $k$ increases as the scatter-points get lighter. (A) uses the full area, while (B) uses the area reduced to grid cells that report rice and/or maize yields in the GDHY data set.

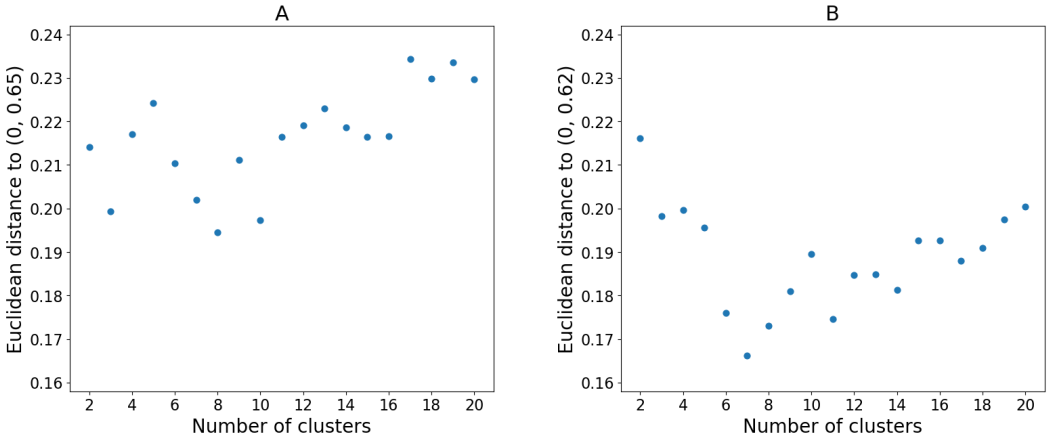



Figure 1.5.: Euclidean distance between the point $(x, y)$ = (average intra-cluster distance, average inter-cluster distance) and the respective reference point for different numbers of clusters $k$ as measure of the clustering quality. Lower distance denotes better clustering. (A) uses the full area with $(0, 0.65)$ as reference point, while (B) uses the area reduced to grid cells that report rice and/or maize yields in the GDHY data set with $(0, 0.62)$ as reference point.

measures are shown as a scatter plot in Figure 1.4 (A). As already mentioned at the beginning of this section (Section 1.2), the considered area was later slightly reduced, and the cluster analysis was repeated for the new area. This gives the results shown in Figure 1.4 (B). Ideal clustering would result in the point $(0, 1)$ on the scatter plot. While for an increasing number of clusters the average distance within clusters will tend towards zero, the average distance between clusters will never reach one in a realistic setting using Pearson distance, as medoids would have to be perfectly anti-correlated. We therefore use the maximal observed average distance between medoids as reference distance, which generally and also in both cases of our analysis is given for $k = 2$ (see Figure 1.4). Hence we use $(0, 0.65)$ and $(0, 0.62)$ as points of reference for clusters defined on the full region of West Africa and the reduced region respectively. We rank

the performance of different $k$ by calculating the euclidean distance between the corresponding points on the scatter plot of intra- and inter-cluster distances and the point of reference. The results for both areas are shown in Figure 1.5. According to this quantification of clustering performance, the best $k$ when using the whole area is 8, while for the reduced area it is 7.

## 1.3. Generation of crop yield distributions

As the project focuses on climate change induced risks in sustainable food security, it was first intended to relate crop yields to the drought indicator SPEI as source of uncertainties in the stochastic optimization model. This can be done by setting up a mechanistic crop yield model, which however was not feasible within this project, since it requires expert knowledge on crop growth and huge amounts of data. An alternative is to use one of the many already existing GGCMs, but these come with high computational costs and generally do not take SPEI as input. Instead, our initial objective was to utilize a less complex regression model with significant correlation between SPEI or absolute water deficit/surplus and yields. By projecting SPEI distributions into the future, one could then obtain expected yield distributions. Unfortunately, in West Africa the link between SPEI (or water stress related climate variables in general) and yield seems to be weakened by nutrient stress and other factors [51]. Therefore, in the final setup we worked directly with yield distributions coming from the GDHY data set, which indirectly include uncertainties coming from nutrient stress, weather or climate related factors, pests, and other agriculturally relevant variables. This section first gives an overview of the attempts to set up a significant relation between SPEI and yields (Section 1.3.1), and then describes the projection method used to approximate future yield distributions (Section 1.3.2).

### 1.3.1. Regression analysis of SPEI and yields

The first approach in the regression analysis uses the GGCMI phase 1 data set from the AgMIP project, as it has the same spatial resolution as the SPEIbase and provides a large quantity of data. It consists of global gridded output data on yields and other variables for 14 different GGCMs with default and harmonized as well as irrigated and rainfed scenarios, each using up to 11 alternative climate data sets. A more detailed description can be found in Section 1.1.2. A big advantage of using this data source is the completeness and length of the time series. Furthermore, many of the GGCMs also include data on nitrogen application in the output, which is an important factor in crop development. The consistency between yield and fertilization data is an advantage when performing regression analysis. As we are not interested in the intercomparison aspect of the data set, and models are expected to be closest to reality in their default modus which was calibrated to give the best results, we use the default scenario for

our regression analysis where possible. Agriculture in West Africa is predominantly rainfed[8], so we further focus on rainfed scenarios within the data set. A drawback of using model output is that the underlying model setup is already determining the relation between yields and different relevant factors, which therefore changes from model to model. Within the GGCMI data set the yields from different models are not significantly correlated, and possibly none of the models is representing the actual relations in West Africa reasonably. This would suggest to use historical data.

When looking for historical data, it is hard to find complete records, especially consistent over a large area as West Africa and on the desired spatial scale. Iizumi et al. developed the global historic yields data set GDHY, based on downscaling FAO yield data on country level using satellite-derived NPP data on grid level[35]. Even though these are not directly historic observed yield values, they are shown to approximate real crop yields quite well[35], and can be expected to have better quality than time series given by GGCMs. The global data set consists of yearly yield data on $0.5° \times 0.5°$ spatial resolution from 1981 to 2016, and is described in more detail in Section 1.1.3. The advantages of this data set are countered by negative properties, including the reduced length of the time series, no associated fertilizer applications rates or other explanatory variables, and no specification of yearly growing seasons, due to which the growing season had to be kept constant at values taken from the crop calendar of 2000[52] for all years in the regression.

For both data sets, linear regressions were done independently for each cell as well as clusterwise. In the latter case, we either used time series of cluster averages, or data of all cells within a cluster were joined to form a bigger sample. This sample was used directly on the one hand, or reduced to data points with SPEI below a given threshold on the other, as the relation between SPEI and yields was expected to be higher when water stress is present. As SPEI, the 3-month value starting from the month in which the crops were planted was used, as this time scale is suited for studying the impact of droughts on agriculture[22]. Within the AgMIP data, the GIS-based Environmental Policy Integrated Climate Model (GEPIC)[53] was mainly used, as it covers the longest time period, includes fertilizer application as output variable, and covers the whole region of West Africa without missing cells. It also includes output on the planting date of crops which is needed to identify the corresponding SPEI value, and the accumulated precipitation during growing season. Furthermore, a constant was included in all regressions, as well as a time coefficient as proxy for any improvements in technology or agricultural practices and other unaccounted trends. Even though the linear regression models partially showed significance for a large part of grid cells or clusters, when using AgMIP data the $R^2$-values were very low ($< 0.5$) in most cases[3] and SPEI was never a significant independent variable within the regression. The same regressions were done using 3-month water deficit averages instead of SPEI,

[3] The coefficient of determination $R^2$ is a statistical measure describing the proportion of variance in the dependent variable that can be explained by the independent variables.

with comparable results. Using GDHY crop yield data did give slightly better results, but only for around half of the cells.

As this approach was strongly simplifying the relation between crop yields and influencing factors, it was not surprising that it could not capture the complex interrelations affecting crop growth. Folberth et al. worked on a similar problem in the context of downscaling yield data[54]: As GGCMs normally have a quite coarse spatial resolution, yield data at a higher resolution would be very useful for many projects working on regional or local scales. Furthermore, GGCMs take the average or dominant characteristics of each region, hence assumptions may not actually match the farmed land. Setting up GGCMs at a higher spatial resolution is generally not feasible, due to low data availability and high computational demand. Earlier downscaling approaches focused on simplified GGCMs and purely statistic methods, while Folberth et al. used machine learning trained on yield output of the International Institute for Applied System Analysis (IIASA) version of the Environmental Policy Integrated Climate Model (EPIC), creating a meta-model which then can be applied to regional data on higher spatial resolution. This method thereby is scale-free, free from a priori assumptions on any relations between indicators and crop yield, and completely data driven. The machine learning approach decides which of the provided covariates to use and what the respective relevances are. Folberth et al. tested their approach on maize yields without limitations by nutrients or pests, both for a rainfed and a completely irrigated scenario, and reported very good results. The top three most relevant covariates in the rainfed case were the total precipitation during growing season, the total precipitation during the calender year, and the Potential Heat Units (PHU)[54]. As a second approach for the regression we focused on the relation established by Folberth et al., hoping to get results of a similar quality. However, significance levels and $R^2$-values did not differ greatly from the first approach, again suggesting that relative to other parameters as nutrient availability or pests and diseases, climate variables have no strong impact on crop yields in the considered area.

Regression approaches between yields and SPEI were aborted, as it was acknowledged that the failures most likely in large part do not stem from poor quality of the used regressions, but from the fact that in West Africa the relation between water stress and yields is muted by nutrient stress and other factors.

### 1.3.2. Projection of yield distributions

As regressions between SPEI and yield data turned out to have insufficient statistical quality, we instead worked directly with yield data. As the disadvantage of the GDHY data set having no associated data on fertilizer application and growing season is only relevant in regards to regressions, this leaves the shorter temporal coverage as only disadvantage compared to the AgMIP yields. We assumed that the higher accuracy and independence of model internal structures outweigh this aspect and decided to

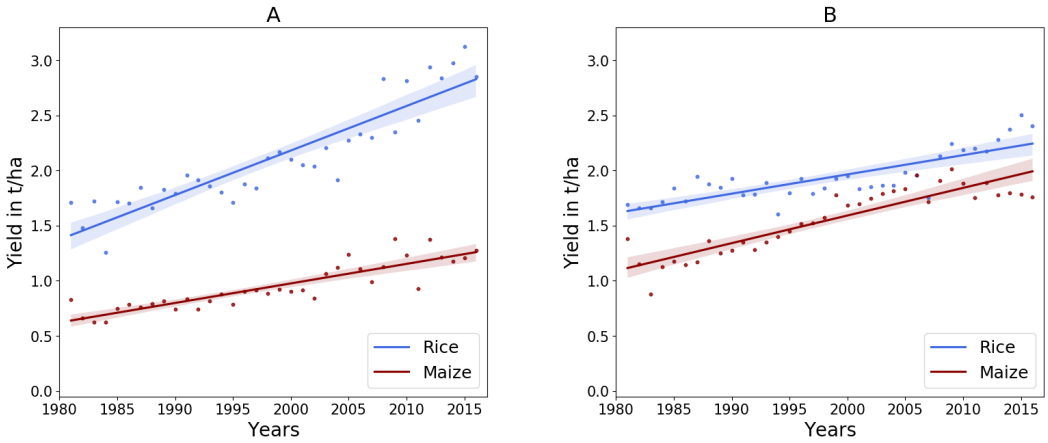



Figure 1.6.: For $k = 2$ clusters, GDHY yield data (points) and the linear trend (line) including 95% confidence interval (shaded area) are given for both of the considered crops. (A) shows the first cluster, (B) the second.

use the GDHY data set for the projection of yield distributions. As the input to the stochastic model will be single yield values per cluster, we further decided to directly analyze cluster averages.

Ideally, one could work with moving time windows to get time depending and normally distributed samples of cluster average yields for which mean and variance could be estimated. This would not only allow to analyze the linear trend in average yields, but also account for possible changes in the frequency of extreme events by including the trend in standard deviation. However, reducing the time series to smaller segments does not leave enough data points to make reliable estimations, as they only have a length of 36 years to begin with. We considered to use bootstrapping to increase the sample size, by repeatedly choosing 80% of the cells within a cluster and for each sample calculating the average, but yield distributions within a cluster differ depending on the year. Therefore resulting bootstrap samples will be coming from different distributions, possibly even showing multi-modal behavior, rendering this approach unfeasible.

Instead, we directly calculate a linear trend from the time series of cluster average yields. According to analysis of Q-Q plots and the Shapiro-Wilk normality test, the resulting residuals follow approximately a normal distribution for most of the clusters and most $k$. For $k = 2$ clusters the trends including 95% confidence intervals are shown in Figure 1.6. Projecting the linear trend into the future will give the mean of future yield distributions which are assumed to be Gaussian. The variance of the residuals is used as variance of the yield distributions in every year. Unfortunately, possible trends in variance of crop yields cannot be analyzed by this technique. However, keeping in mind the limited time and the focus of the project as a proof-of-concept to study the functionality and dynamics of a stochastic approach to sustainable food security, we considered the simplified yield model as appropriate to analyze the feasibility of a large scale insurance scheme in this setting.

# 2. Theoretic analysis of a time-invariant food security model

The aim of this project is to build an optimization model as a decision supporting tool for achieving food security in West Africa. Given information about agricultural area, direct and indirect costs, food demand, and yield distributions, we want to find a solution for the best actions to take considering the uncertainties. These should reflect the cheapest feasible combination to reach designated goals such as food security while meeting other constraints. Possible actions include investments in expansion of agricultural area, irrigation systems, or improved fertilization, and the selection of specific crop types for cultivation in specific areas. In our model, we focus on crop allocation as the only decision variable. In a first time-invariant version, we consider just a single time step, and reaching a certain food supply is the only objective. If not met by own agricultural production following the decisions of crop allocation, the food supply has to be ensured by food import or international aid, or negative socio-economic consequences of a food shortage have to be dealt with. As a subsequent step in the next chapter, we introduce a time dimension into the model, allowing to additionally set up a government based insurance system.

In this chapter, we focus on the theoretical background of stochastic optimization (Section 2.1) and the development of the time-invariant food security model (Section 2.2), before providing an analytical interpretation of the model's behavior (Section 2.3) and discussing different methods to solve such optimization problems (Section 2.4).

## 2.1. Background on stochastic optimization

Optimization is present in numerous aspects of everyday life and models are set up to systematically find optimal solutions to complex problems in economics, environmental research, engineering and many other fields. Stochastic optimization is a branch of optimization dealing with problems for which not all factors that might influence the output of decisions are known a-priori. This section gives a short introduction to stochastic optimization and is primarily based on Chapter 1 in *Numerical Techniques for Stochastic Optimization* by Ermoliev and Wets, 1988 [17].

In general, the goal of an optimization problem is to maximize or minimize a quantity such as output, efficiency, or costs. The function $g_0$ describing the aspect one wishes to

optimize is called objective function. As each maximization problem can be transformed into a minimization problem by taking the negative of the objective function, we will focus on minimization problems in the following. The aim is to minimize $g_0 : X \to \mathbb{R}$ over a set of possible multidimensional states $x \in X \subseteq \mathbb{R}^n$, while meeting constraints formalized as inequalities $g_i(x) \leq 0$ with $i \in \{1, \ldots, m\}$:

$$\begin{aligned} &\textit{Minimize} && g_0(x) && \text{for } x \in X \subseteq \mathbb{R}^n \\ &\textit{Subject to} && g_i(x) \leq 0 && \forall\, i \in \{1, \ldots, m\} \end{aligned} \tag{2.1}$$

Equality constraints $g_j(x) = 0$ can be included by two inequalities $g_j(x) \leq 0$ and $-g_j(x) \leq 0$. In the deterministic case, we assume that all model parameters are known a-priori except the state variable $x$, which is to be optimized. However, in many scenarios, the constraints and/or the objective function are subject to uncertainties, such as price fluctuations, available inputs, or environmental conditions. This turns the optimization problem into

$$\begin{aligned} &\textit{Minimize} && g_0(x,\, z) && \text{for } x \in X \subseteq \mathbb{R}^n \\ &\textit{Subject to} && g_i(x,\, z) \leq 0 && \forall\, i \in \{1, \ldots, m\}, \end{aligned} \tag{2.2}$$

where $z \in \Omega$ are realizations of a random variable $Z$ with corresponding probability density $p(z)$. For each fixed realization $z \in \Omega$ of $Z$, the optimization problem is deterministic, which would allow for an optimal solution $x^*$ as a function of the uncertain conditions, i.e. $x^* : \Omega \to X$. However, in many cases, and in particular in decision making problems, a decision needs to be taken before the uncertain event occurs, rendering the parametric approach infeasible.

A simple way of integrating uncertainties into a solution would be to take a weighted average $\tilde{x} = \int_\Omega p(z)\, x^*(z)\, \mathrm{d}z$ of all solutions $x^*(z)$ according to the densities $p(z)$. Another approach would be to solve the deterministic optimization using the expected outcome $\mathbb{E}[Z]$ of the random variable as realization of the external factors, giving $\hat{x} = x^*(\mathbb{E}[Z])$ as solution. While both options might seem reasonable at first sight and could work well in some cases, the solutions resulting from both approaches will in general be far from optimal, and it is not even clear whether they fulfill the constraints of the optimization problem for general realizations of the uncertainties. As an example, a region that over time suffers both from floods and droughts might show unproblematic weather conditions on average, but investments in flood prevention and irrigation systems should still be part of an optimal strategy to minimize losses and are needed to reach goals such as climate resilience.

Instead of relying on a deterministic version of the problem, a new understanding of optimality needs to be adopted in such situations. This notion of optimality will not be global optimality, but optimality accounting for the uncertainty of unknown events and the corresponding risks, to get a solution valid for all possible events. Objectives can be to perform as well as possible in expectation, or to minimize the probability of

a designated outcome. Demanding a solution to satisfy a constraint in every possible realization of the uncertainties can lead to very high costs or even infeasibility, e.g. in the presence of extreme events. Thus, constraints might only have to be met on average or with a given reliability $\alpha \in (0, 1)$ which is chosen exogenously. A stochastic formulation of the general problem given by (2.2) can then be

$$\begin{aligned} &\textit{Minimize} && \mathbb{E}_Z[g_0(x, Z)] && \text{for } x \in X \subseteq \mathbb{R}^n \\ &\textit{Subject to} && \mathbb{P}[g_i(x, Z) \leq 0] \geq \alpha && \forall\, i \in \{1, \ldots, m\} \end{aligned} \tag{2.3}$$

We call the solution to a problem of this form a *robust solution*, as it takes into account all possible realizations of the uncertain parameters. The time-invariant food security model developed in Section 2.2 as well as the time-dependent sustainable food security model developed in Section 3.1 build on a stochastic optimization problem of this type.

## 2.2. Development of a time-invariant food security model

We start with a basic setup of the model. In order to account for the heterogeneous conditions over a large area as West Africa, the region is split into $K$ clusters. Within each cluster, $J$ different crops can be cultivated. The decision variables which we want to optimize are crop allocations within each cluster, hence $x_{j,k}$ for $j \in \{1, \ldots, J\}$ and $k \in \{1, \ldots, K\}$ in hectare. Within each cluster, there will be fixed costs $c_{j,k}$ in [\$/ha] for cultivation of each crop, which include costs of seeds, fertilizer, labor, etc[1]. To simplify subsequent equations, we set $n = J \cdot K$ and redefine our objective variable as $x \in \mathbb{R}^n$ by going through the original matrix row by row, i.e. starting with crop allocations for crop 1 in cluster 1 to $K$, then for crop 2 in all clusters, and so on[2]. Analogous, we define the cost vector $c \in \mathbb{R}^n$. Crop areas have lower a bound as they cannot be negative (Eq. (2.4b)), and an upper bound because the sum over all crop areas within a cluster cannot exceed the total agricultural area $\bar{x}_k$ in the corresponding region (Eq. (2.4c)). Yields for each cluster and crop depend on weather conditions and other uncertainties and therefore are given by a random variable $Y$ with realizations $y = (y_1, \ldots, y_n)^T \in \Omega \subseteq \mathbb{R}^n_{\geq 0}$. The unit for yields is typically [t/ha], but here they are multiplied by their respective energy content per ton to get [kcal/ha], thus making them comparable from the food security point of view. Food security is included in the problem by a demand $A$ in [kcal] which has to be met collaboratively for all of West Africa (Eq. (2.4d)). The overall aim is to minimize the cultivation costs (Eq. (2.4a)) while meeting the above described constraints. This gives the following optimization problem:

$$\textit{Minimize} \qquad f(x) = c^T x \qquad \text{for } x \in \mathbb{R}^n \tag{2.4a}$$

[1] The currency used throughout this thesis is United States Dollar, which we denote by the symbol \$.
[2] Formally, it is $x_i = x_{\lfloor (i-1)/K \rfloor + 1,\ [(i-1) \mod K] + 1}$.

$$
\begin{aligned}
\textit{Subject to} \quad & x_i \geq 0 && \forall\, i \in \{1, \ldots, n\} && \text{(2.4b)}\\
& \sum_{j=0}^{J-1} x_{k+j\cdot K} \leq \bar{x}_k && \forall\, k \in \{1, \ldots, K\} && \text{(2.4c)}\\
& A - y^T x \leq 0 && \forall\, y \in \Omega && \text{(2.4d)}
\end{aligned}
$$

However, requiring to fulfill the food security constraint under all circumstances can lead to extremely high costs, or even be infeasible due to the restricted availability of agricultural area. Therefore, we replace (2.4d) by a probabilistic constraint, requiring that the demand be met with a predefined probability $\alpha \in [0, 1]$:

$$\mathbb{P}[A - Y^T x \leq 0] \geq \alpha \tag{2.4e}$$

Ermoliev and Wets show that a minimization problem with probabilistic constraints as given by (2.4a)–(2.4c) and (2.4e) can be transformed into a problem relying solely on deterministic constraints[17]. This reformulation results in a two stage stochastic optimization problem: Crop allocations $x$ are first stage decisions, which have to be taken before the outcome of the uncertain conditions is known. Later potential shortcomings in cases where the demand could not be met have to be addressed by second stage actions. To this end, a second stage penalty $\rho > 0$ for these shortcomings is introduced, and the expected total costs are minimized:

$$
\begin{aligned}
\textit{Minimize} \quad & f(x) = c^T x + \rho \cdot \mathbb{E}_Y[\max(0, A - Y^T x)] && \text{for } x \in \mathbb{R}^n\\
\textit{Subject to} \quad & x_i \geq 0 && \forall\, i \in \{1, \ldots, n\}\\
& \sum_{j=0}^{J-1} x_{k+j\cdot K} \leq \bar{x}_k && \forall\, k \in \{1, \ldots, K\}
\end{aligned}
\tag{2.5}
$$

This penalty can be interpreted as direct costs of importing food or indirect costs due to socio-economic consequences of food shortages. The higher the penalty, the higher the incentive to invest in first stage actions to prevent shortcomings of the food demand. Therefore, the probability of meeting the food security constraint will increase for higher penalties. For a problem of this structure, Ermoliev and Wets state that there exists a one-to-one mapping between the required reliability $\alpha$ and the penalty $\rho$ as long as the solution is not in conflict with other constraints. Theoretically, if $\rho$ is large enough, the risk of not meeting the food demand will then tend towards zero[17].

## 2.3. Analytical interpretation of the time-invariant food security model

For convex functions, any local minimum is also a global minimum. This is an important property of optimization problems, as it simplifies further analysis and solution strategies. We therefore first proof convexity of the objective function given by (2.5).

**Definition 2.1.** *Let $f : D \to \mathbb{R}$ be a function with $D \subseteq \mathbb{R}^n$. Then $f$ is called* convex *if and only if*

1. *Its domain* $\mathrm{dom}(f) := \{x \in D \mid f(x)$ *is well defined, i.e. finite*$\}$ *is a convex set in* $\mathbb{R}^n$. *The set* $\mathrm{dom}(f)$ *is convex if and only if*

$$x,\ x' \in \mathrm{dom}(f) \Rightarrow \gamma\, x + (1-\gamma)\, x' \in \mathrm{dom}(f) \quad \forall\ \gamma \in (0,\, 1)$$

2. *For all* $x,\ x' \in \mathrm{dom}(f)$ *and* $\gamma \in (0,\, 1)$ *it is*

$$f(\gamma\, x + (1-\gamma)\, x') \leq \gamma\, f(x) + (1-\gamma)\, f(x')$$

**Lemma 2.2.** *Convex functions have the following properties:*

1. *Let* $\{f_i(x)\}_{i \in I}$ *be a set of convex functions on* $D$. *Then the set* $D'$ *of all points for which* $\sup_{i \in I} f_i(x)$ *is finite is a convex set, and* $\sup_{i \in I} f_i(x)$ *is a convex function on this set. In particular it follows: If* $\{f_i(x)\}_{i=1,\,\ldots,\,r}$ *is a finite set of convex functions on* $D$, *then so is* $g(x) = \max(f_1(x),\, \ldots,\, f_r(x))$.
2. *Let* $Z$ *be a random variable and* $f(x, z)$ *a convex function on* $D$ *for all realizations* $z$ *of* $Z$. *If* $\mathbb{E}_Z[f(x,\, Z)]$ *is well defined, it is a convex function on* $D$ *as well.*
3. *Let* $\{f_i(x)\}_{i=1,\,\ldots,\,r}$ *be a finite set of convex functions on* $D$ *and let* $\lambda_i \geq 0$ *be positive constants. Then the function* $\sum_{i=1}^{r} \lambda_i\, f_i(x)$ *is a convex function on* $D$ *as well.*

The corresponding proofs can be found e.g. in *Convex optimization* by Boyd and Vandenberghe, 2004 [55].

**Theorem 2.3.** *The objective function of the time-invariant food security problem*

$$\begin{aligned} f : D_f &\to \mathbb{R} \\ x &\mapsto c^T x + \rho \cdot \mathbb{E}_Y[\max(0,\, A - Y^T x)] \end{aligned}$$

*depending on parameters* $c \in \mathbb{R}^n_{\geq 0},\ A,\ \rho \in \mathbb{R}_{\geq 0}$ *and the random variable* $Y$ *with realizations in* $\Omega \subseteq \mathbb{R}^n_{\geq 0}$ *is a convex function on its domain* $D_f \subseteq \mathbb{R}^n$ *which is defined as follows:*

$$D_f := \Big\{ x \in \mathbb{R}^n \mid x_i \geq 0,\, \forall\, i \in \{1,\, \ldots,\, n\};\ \sum_{j=0}^{J-1} x_{k+j \cdot K} \leq \bar{x}_k,\, \forall\, k \in \{1,\, \ldots,\, K\} \Big\}$$

*Proof of Theorem 2.3.* We first proof convexity of the domain $D_f$. Let $a,\, b \in D_f$ and $\gamma \in (0, 1)$. Then $a_i,\, b_i \geq 0$ and thereby also $\gamma\, a_i + (1-\gamma)\, b_i \geq 0$. Further we have

$$\begin{aligned} \sum_{j=0}^{J-1} (\gamma\, a_{k+j \cdot K} + (1-\gamma)\, b_{k+j \cdot K}) &= \gamma \sum_{j=0}^{J-1} a_{k+j \cdot K} + (1-\gamma) \sum_{j=0}^{J-1} b_{k+j \cdot K} \\ &\leq \gamma\, \bar{x}_k + (1-\gamma)\, \bar{x}_k = \bar{x}_k \end{aligned}$$

Hence, it is $\gamma\, a + (1-\gamma)\, b \in D_f$ which concludes the first part of the proof. Convexity of the objective function now follows by consecutively applying the properties of Lemma 2.2. The term $A - y^T x$ is linear on $D_f$ and thus convex for every realization $y$ of $Y$. Hence, also $\max(0, A - y^T x)$ is convex and subsequently $\mathbb{E}_Y[\max(0, A - Y^T x)]$ as well. Therefore, $f(x)$ is a convex function on $D_f$, being a weighted sum of convex functions with positive weights. □

Given the convexity of the objective function, any local minimum will also be a global minimum. This allows for an analytical consideration of the time-invariant food security optimization problem and an interpretation of the location of potential minima based on the analytical characteristics. The set of points for which the minimum is attained forms a convex subset of the domain [56]. We further know that the random variable $Y$ only attains positive values, as it represents crop yields. We will additionally assume that the probability density function $p_{_Y}$ of $Y$ is a continuously differentiable function. For $n = 1$ and $x > 0$, we can then transform the objective function as follows:

$$\begin{aligned} f(x) &= c\, x + \rho \cdot \mathbb{E}_Y[\max(0, A - Yx)] \\ &= c\, x + \rho \int_0^\infty \max(0, A - xy)\, p_{_Y}(y)\mathrm{d}y \\ &= c\, x + \rho \int_0^{A/x} (A - xy)\, p_{_Y}(y)\mathrm{d}y \end{aligned} \tag{2.6}$$

If there exists an interior point $x \in D_f \backslash \partial D_f$ with vanishing first derivative, it corresponds to the global minimum. According to the *Leibniz integral rule*, the order of derivative and integral can be exchanged and we get

$$\frac{\mathrm{d}}{\mathrm{d}x} f(x) = c - \rho \int_0^{A/x} y\, p_{_Y}(y)\mathrm{d}y \tag{2.7}$$

Let $\Xi_x$ be the event that the food demand $A$ is not met for a given crop allocation $x$. Then

$$\mathbb{E}[Y|_{\Xi_x}] = \int_0^\infty y\, p_{Y|_{\Xi_x}}(y)dy = \int_0^{A/x} y\, \frac{p_{_Y}(y)}{\mathbb{P}[\Xi_x]}\mathrm{d}y \tag{2.8}$$

The derivative in (2.7) can hence be rewritten as

$$\frac{\mathrm{d}}{\mathrm{d}x} f(x) = c - \rho\ \mathbb{P}[\Xi_x]\ \mathbb{E}[Y|_{\Xi_x}] \tag{2.9}$$

The term $\rho\ \mathbb{P}[\Xi_x]\ \mathbb{E}[Y|_{\Xi_x}]$ can be understood as the expected penalty per hectare that can be avoided by increasing the cultivated area. The optimal crop allocation will therefore be where costs per hectare and expected penalty for not using another hectare are balanced, assuming such a point exists within $D_f \backslash \partial D_f$.

Should such a point not exist, the minimum has to lie on the boundary $\partial D_f$ of the domain. In the one dimensional case, the boundary consists of only two points $\partial D_f = \{0, \bar{x}\}$, where $\bar{x}$ is the maximum available area as given in the constraints of the food security problem. The minimum will be attained for $x = 0$, if the cultivation costs are so high

compared to the penalty for not reaching food security, that it is not worth growing anything at all ($c \geq \rho\ \mathbb{E}[Y]$). The solution will be $\bar{x}$, if the relation between costs and expected penalty per unused additional area would still suggest to increase the area, but the maximal available area is reached ($c < \rho\ \mathbb{P}[\Xi_{\bar{x}}]\ \mathbb{E}[Y|_{\Xi_{\bar{x}}}]$).

This approach can be generalized to the multidimensional case. We then get

$$\frac{\partial}{\partial x_i} f(x) = c_i - \rho\ \mathbb{P}[\Xi_x]\ \mathbb{E}[Y_i|_{\Xi_x}] \tag{2.10}$$

In $n$ dimensions the gradient of an interior minimum is zero, i.e. all partial derivatives vanish. If an interior minimum exists, it will therefore be a solution to the equation system $\{c_i - \rho\ \mathbb{P}[\Xi_x]\ \mathbb{E}[Y_i|_{\Xi_x}] = 0\}_{i=1,\dots,n}$. The interpretation is similar as above: a specific crop in a specific cluster will be cultivated on an area that balances the corresponding costs $c_i$ of cultivating additional area with the penalty it would be expected to avoid. However, that penalty now does not only depend on the respective yields $Y_i$ and area $x_i$, but on all other yields $Y_{h\neq i}$ and areas $x_{h\neq i}$ as well. Cultivating a bigger area of a given crop in a specific region might be profitable if it is the only possible choice, but when competing with other crops it might not be chosen at all. Thus, in the multidimensional case, it is not unlikely that the solution lies on the $(n-1)$-dimensional boundary of $D_f$, with crops not used at all in some clusters, while at their maximal available area in others.

The question might arise, whether interior minima can even theoretically exist, or whether the model would always lead to the clusters and crops with the best trade-off being used as much as possible, while the others are not used at all. This is where correlation between yields or rather the lack thereof comes into play: Assume an area $x_i$ is cultivated which is not yet using up the whole agricultural area in the respective cluster. When comparing costs $c_i$ to expected penalties that could be avoided, we do not use the general expected yield, but the conditional expected yield given that the food demand would not be reached by the current crop allocation. Hence, if $x_i$ is already high, corresponding yields and thus the conditional expected value have to be low. This reduces the penalty that can be expected to be avoided and makes an increase in $x_i$ less profitable. To some extent this will translate to any other crop areas $x_{i'}$ with yields $Y_{i'}$ that are correlated to $Y_i$. Meanwhile, another crop with area $x_h$ that in the current situation is used little or not at all, can still have high yields under the condition that the food demand is not reached, if the yields $Y_h$ and $Y_i$ are not correlated. Hence, an increase in $x_h$ could decrease the total expected costs $f(x)$ while an increase in $x_i$ does not, even though the former crop might be much inferior when just looking at costs versus expected yields in general.

As the established dependence structure of yields within West Africa assumes full dependence of the occurrence of extreme yields of different crops within the same cluster (see Section 1.3), this suggests that in general only one crop will be used within each cluster. Yields of different clusters however are independent, hence all crops might be

used in separate clusters by the allocation of arable area that results as solution of the stochastic optimization model.

## 2.4. Solving a stochastic optimization problem

The previous section is an attempt to connect the mathematical behavior of the model to an interpretation based on the meaning of the different parameters. Even though this allows for a more intuitive understanding, it does not give a general method to solve the optimization problem. Even for interior minima, the given system of equations depends on a conditional expected value, which in many cases might not have an analytical solution. In particular, this is the case for normal distributions which we use for the yields. Therefore, the expected value in the objective function is approximated by a sample mean, which gives the following new optimization problem for realizations $y_i^1, \ldots, y_i^N$ of the yields $Y_i$ for each cluster and crop:

$$\begin{aligned} &\textit{Minimize} && f_N(x) = \sum_{i=1}^{n} x_i\, c_i + \rho \cdot \frac{1}{N} \sum_{s=1}^{N} \left[ \max(0, A - \sum_{i=1}^{n} y_i^s\, x_i) \right] && \text{for } x \in \mathbb{R}^n \\ &\textit{Subject to} && x_i \geq 0 && \forall\, i \in \{1, \ldots, n\} \\ & && \sum_{j=0}^{J-1} x_{k+j \cdot K} \leq \bar{x}_k && \forall\, k \in \{1, \ldots, K\} \end{aligned} \tag{2.11}$$

The approximated objective function $f_N(x)$ is piecewise linear and convex according to the properties in Lemma 2.2. The problem can be rewritten as a linear optimization problem by introducing an additional variable $S_s$ for each term that contains a maximum operator. To ensure that these variables assume the right values, two new constraints are included for each: $S_s \geq 0$ and $S_s \geq \frac{\rho}{N}(A - \sum_{i=1}^{n} y_i^s x_i)$. These two constraints alone would only ensure that the additional variables are at least as big as the value they are substituting, but as the total objective function is minimized the solution will have $S_s = \rho \cdot \frac{1}{N} \max(0, A - \sum_{i=1}^{n} y_i^s x_i)$. In its standard form, the linearized problem will then be defined as

$$\begin{aligned} &\textit{Minimize} && \tilde{f}_N(\tilde{x}) = \tilde{c}^T \tilde{x} \\ &\textit{Subject to} && \tilde{A}\tilde{x} \leq \tilde{b} \\ & && \tilde{x} \geq 0, \end{aligned} \tag{2.12}$$

where $\tilde{x} = (x_1, \ldots, x_n, S_1, \ldots, S_N)^T \in \mathbb{R}^{n+N}$ with the first $n$ entries being the crop allocations and the others the newly introduced variables, and

$$\tilde{A} = \begin{pmatrix} -\frac{\rho}{N} y_1^{(1)} & \cdots & -\frac{\rho}{N} y_n^{(1)} & \\ \vdots & \ddots & \vdots & -I_N \\ -\frac{\rho}{N} y_1^{(N)} & \cdots & -\frac{\rho}{N} y_n^{(N)} & \\ & & & \\ I_K & \cdots & I_K & 0 \end{pmatrix} \in \mathbb{R}^{(N+K)\times(n+N)}$$

$$\begin{aligned}\tilde{b} &= \left(-\frac{\rho}{N}A, \ldots, -\frac{\rho}{N}A, \bar{x}_1, \ldots, \bar{x}_K\right)^T && \in \mathbb{R}^{N+K} \\ \tilde{c} &= (c_1, \ldots, c_n, 1, \ldots, 1)^T && \in \mathbb{R}^{n+N}\end{aligned}$$

Two common approaches for solving a linear optimization problem analytically are either to transform it into its *Lagrangian dual problem*, or to apply the *Karush-Kuhn-Tucker conditions*. In general, the maximum of the dual objective function, i.e. the solution of the dual problem, is a lower bound for the minimum of the primal objective function. In the case of a convex objective function with linear constraints, the difference between the solution of the dual problem and the solution of the primal problem, i.e. the *duality gap*, is zero. However, for the given food security optimization problem, the dual problem does not allow for a closed-form solution.

For a linear problem, the Karush-Kuhn-Tucker conditions are necessary but not sufficient conditions for a local minimum of an objective function and are often used in combination with case-distinction in order to handle arising non-linear equations. Due to the high dimensionality of the linearized food security problem, a very high number of cases would have to be considered, making this approach infeasible.

Instead, a numerical method needs to be applied. Python provides several packages for numerical optimization with implementations of different algorithms suited to a wide range of optimization problems. We chose the solving routine `fmin_COBYLA` (Constrained Optimization BY Linear Approximation [57]) of the python package `scipy.optimize` [58], as the given type of constraints can be integrated easily and the method does not rely on gradients, thus avoiding problems arising through non-smoothness of the approximated objective function. The same solver is applied to the time-dependent sustainable food security model discussed in the following chapters. While `fmin_COBYLA` works well in these settings, it does not specifically take into account the convexity of the objective function. This aspect can be leveraged in future work to improve computational performance.

# 3. Time-dependent sustainable food security model

In Chapter 2, a time-invariant food security model was introduced. Due to the relatively low complexity of relations in the model, an in-depth theoretical analysis of the model structure was possible. However, food security is a very complex topic and there are many aspects that can impact food security, in particular over a longer period. Farmers can store surplus output in good years to protect against shortfalls in subsequent years. Investments can be made, e.g. in agricultural area expansion or irrigation systems, which can have high one-time costs but increase yields in coming years. Insurance systems which are established over time can help farmers to attain a higher resilience against extreme events by protecting them from loosing their livelihoods in case of a drought or other disaster.

Including a wide range of time-dependent relations into the food security setting was not feasible within the limits of a master thesis. We therefore focus on implementing a large scale insurance scheme which is funded by taxes that farmers pay on their yearly profits[1]. This makes farmers resilient against extreme events and thereby adds a sustainability dimension to the food security aspect covered by the time-invariant version of the model. It also allows to get more insight on risk transfer mechanisms in this setting. The model output is an optimal crop allocation to meet food demand while ensuring feasibility of the insurance scheme. This information can be used by governments in the process of setting up policies to reach sustainable food security.

In Section 3.1, we develop a time-dependent sustainable food security model and in Section 3.2 the parametrization of the model is described. Section 3.3 introduces a method to quantify the performance of the stochastic optimization model compared to a deterministic approach using the expected yields as model input.

## 3.1. Model development

The time-dependent version of the food security model builds on the model defined in Chapter 2. Relations established for the time-invariant case are still valid, but extended

[1] Note that we assume a non-profit government based insurance system. However, the in the following developed sustainable food security model could be adapted for a setting with a private insurance system.

over a range of years from $T_0$ to $T_{max}$. A given probability of meeting the food demand now has to be attained jointly over all considered years. As long as no other relations are included and no trends in e.g. yields or food demand are used, this model shows the same behavior for each year. However, including a time dimension allows to investigate the influence of developments in agricultural productivity and demography on food security. Furthermore, it allows to introduce a government based insurance system.

The insurance system is implemented by means of a government fund[2], which is built up over the years by farmers paying taxes on their profits. In turn, the government guarantees a given income level $I^{\text{gov}}$ for years with an extreme event and thus lower yields, which in the following are referred to as *catastrophic*. This augments farmers' resilience to extreme events and allows them to sustain their livelihoods even in years with bad harvests. The government decides on the share of low yields for which it will intervene, given by a percentage of covered risk $r$. It could e.g. choose to cover risks posed by 1-in-20-year events, i.e. set $r = 5\%$. The threshold dividing catastrophic from non-catastrophic yields would in this case be the lower 5% quantile of the respective yield distributions. Every time a cluster exhibits catastrophic yields and as a consequence farmers' profits are lower than a fixed guaranteed income level, the government pays the difference between the guaranteed income and profits. However, potential losses of farmers are not covered, i.e. payouts do not exceed the guaranteed income.

Considering multiple independent clusters, a year is called *catastrophic* if at least one of the $K$ clusters exhibits catastrophic yields. The covered risk $r$ still refers to the probability of catastrophic yields in a single cluster, while the resulting probability of a catastrophic year then depends on the underlying covered risk $r$ as well as the number of clusters $K$. Simulations run up to and including the first catastrophic year, or end after the year $T_{max}$ if no catastrophe has happened within the considered time period. The motivation behind this implementation is that in case of insolvency after a catastrophe, the government has to deal with the resulting socio-economic consequences of the failure of the insurance scheme. In order to be in a viable starting position for further catastrophes that might come, it will need to reset funds, potentially relying on foreign aid or loans. Implementing these aspects and their influence on model parameters was beyond the scope of this project, such that we do not cover multiple catastrophes within the model. However, once a catastrophe is overcome, the model could be applied again with updated parameters. In cases where the fund has been sufficiently built up to cover payouts in the catastrophic year, this can be integrated by setting a positive initial fund size $G_{\text{ini}}$.

To formalize the government fund, we first need to formalize the profits of farmers. We write crop area allocations as a three-dimensional matrix $x \in X \subseteq \mathbb{R}^{J\times K\times(T_{\max}-T_0+1)}$, where $J$ is the number of crops, $K$ the number of clusters, and $T_0$ and $T_{\max}$ the first and last considered year respectively. Crop yield values are the source of uncertainty in

[2] As we consider the region of West Africa, this fund is understood as a joint insurance mechanism set up by several governments. For simplicity, we refer to the participating governments as "government".

the model. As we assume full dependence of the occurrence of catastrophic yields for different crops in the same cluster, we first generate a matrix indicating in which years each of the $K$ clusters shows catastrophic yields. For this we use a multi-dimensional random variable $M$ of independent and identically distributed bernoulli variables which take the value 1 with probability $r$, and the value 0 with probability $1 - r$. Hence, 1 describes the presence of catastrophic yields, while 0 refers to non-catastrophic yields. This gives a matrix $m \in \Phi = \{0, 1\}^{K \times (T_{\max} - T_0 + 1)}$. Depending on the presence of catastrophic yields as indicated by a realization $m \in \Phi$ of $M$, yields for each crop, cluster, and year are generated from either the lower $r$-quantile or the remaining upper quantile of the respective distributions. Hence, yield realizations $y \in \Omega \subseteq \mathbb{R}^{J \times K \times (T_{\max} - T_0 + 1)}$ given in [t/ha] are drawn from a random variable $Y|_M$ which describes the yields according to the quantile for catastrophic yields given by the covered risk $r$ and conditional to the presence of catastrophic yields as indicated by $M$. Profits $I$ per cluster are then given by a function $I : X \times \{T_0, \dots, T_{\max}\} \times \Omega \to \mathbb{R}^K$ which is defined by

$$I_k(x,\, t,\, y) = \sum_{j=1}^{J} y_{j,\,k,\,t} \cdot x_{j,\,k,\,t} \cdot p_j - \sum_{j=1}^{J} x_{j,\,k,\,t} \cdot c_j \tag{3.1}$$

Here, $c \in \mathbb{R}^J$ represents the cultivation costs for each crop in [\$/ha] and $p \in \mathbb{R}^J$ are farm-gate prices that farmers earn for each crop in [\$/t], both assumed to be constant over time. From this we can develop a formalization of the payouts $L$ that the government has to make as a function of the crop allocation $x$, the year $t$, and the realizations $m \in \Phi$ and $y \in \Omega$ of $M$ and $Y|_M$, i.e. $L : X \times \{T_0, \dots, T_{\max}\} \times \Phi \times \Omega \to \mathbb{R}$ with

$$L(x,\, t,\, m,\, y) = \sum_{k=1}^{K} m_{k,\,t} \cdot \min\left(I_k^{\text{gov}}(t),\, \max\left(0,\, I_k^{\text{gov}}(t) - I_k(x,\, t,\, y)\right)\right) \tag{3.2}$$

The variable $I^{\text{gov}}(t) \in \mathbb{R}^K$ is the aggregated incomes in [\$] that the government guarantees for each cluster in case of catastrophic yields in year $t$ and will be set as a fixed share $S_{\text{gov}}$ of the expected income for given model settings, which we discuss in more detail in Section 3.2. The matrix $m$ is characterized by

$$m_{k,\,t} = \begin{cases} 1 & \text{cluster } k \text{ has catastrophic yields in year } t \\ 0 & \text{else} \end{cases} \tag{3.3}$$

Hence, there are no payouts for clusters with non-catastrophic yields, even if the guaranteed income level is not reached by the profit farmers make.

We finally get the formalization of the government fund after year $t$ given for a crop allocation $x$ and realizations $m \in \Phi$ and $y \in \Omega$ of $M$ and $Y|_M$ as

$$\begin{aligned} &G : X \times \{T_0, \dots, T_{\max}\} \times \Phi \times \Omega \to \mathbb{R} \\ &G(x,\, t,\, m,\, y) = G_{\text{ini}} + \tau \sum_{t'=T_0}^{t} \sum_{k=1}^{K} \max\left(0,\, I_k(x,\, t',\, y)\right) - \sum_{t'=T_0}^{t} L(x,\, t',\, m,\, y) \end{aligned} \tag{3.4}$$

Here, $G_{\text{ini}}$ is the initial fund size in [\$] and $\tau$ is the tax rate according to which farmers have to pay a share of their profits to the government fund each year.

As mentioned above, the termination year of the simulation depends on the presence of catastrophic yields given by the realization $m \in \Phi$ of $M$, i.e. $T_{\text{fin}} : \Phi \to \{T_0, \ldots, T_{\max}\}$:

$$T_{\text{fin}}(m) = \begin{cases} t_{\text{fin}} & t_{\text{fin}} \text{ is the first year with at least one catastrophic cluster} \\ T_{\max} & \text{no catastrophe happens within } T_{\max} \end{cases} \tag{3.5}$$

We can now include a second objective into the model, in addition to food security: The government's goal is to stay solvent in case of a catastrophic year, i.e. the final fund $G_{\text{fin}}(x, m, y) = G(x, T_{\text{fin}}(m), m, y)$ has to be positive. The time-dependent food security model then takes the form

$$\textit{Minimize} \quad f(x) = \mathbb{E}_M \left[ \sum_{t=T_0}^{T_{\text{fin}}(M)} \sum_{k=1}^{K} \sum_{j=1}^{J} x_{j,k,t} \cdot c_j \right] \qquad \text{for } x \in \mathbb{R}^{J \times K \times (T_{\max} - T_0 + 1)} \tag{3.6a}$$

$$\textit{Subject to} \quad x_{j,k,t} \geq 0 \qquad \forall\, j \in \{1, \ldots, J\},\, k \in \{1, \ldots, K\},\, t \in \{T_0, \ldots, T_{\max}\} \tag{3.6b}$$

$$\sum_{j=1}^{J} x_{j,k,t} \leq \bar{x}_k \qquad \forall\, k \in \{1, \ldots, K\},\, t \in \{T_0, \ldots, T_{\max}\} \tag{3.6c}$$

$$\mathbb{P}\left[ \sum_{k=1}^{K} \sum_{j=1}^{J} (Y|_M)_{j,k,t} \cdot x_{j,k,t} \cdot a_j \geq A_t \,\middle|\, t \in \{T_0, \ldots, T_{\max}\} \right] \geq \alpha_{\text{F}} \tag{3.6d}$$

$$\mathbb{P}\left[ G_{\text{fin}}(x, M, Y|_M) \geq 0 \right] \geq \alpha_{\text{S}} \tag{3.6e}$$

Here, $A \in \mathbb{R}^{(T_{\max} - T_0 + 1)}$ is the aggregated caloric demand for each year in the considered area of West Africa and $a \in \mathbb{R}^J$ is the calorie content of each crop in [kcal/t]. The upper limit for agricultural area in each cluster is given by $\bar{x} \in \mathbb{R}^K$ and is assumed to be constant over time. As in Section 2.2, we do not require the food security constraint in (3.6d) and the government solvency constraint in (3.6e) to be met for every realization $m \in \Phi$ and $y \in \Omega$ of $M$ and $Y|_M$. Instead we introduce probabilities $\alpha_{\text{F}}$ and $\alpha_{\text{S}}$ according to which the food security and solvency constraints have to be met. Note, that high probabilities might not always be feasible. While high $\alpha_F$ could be reached in theory if crop areas $x$ were not constrained, it is possible that the probability $\alpha_S$ is limited even for $x \to \infty$. Assume a catastrophe in the first year, such that the fund could not yet be built up. The government can only stay solvent if farmers' profits are at least as high as the guaranteed income despite the catastrophe, and thus no payouts need to take place. However, with very low yields farmers might have losses per cultivated hectare, if cultivation costs exceed possible income. Increasing the crop allocation $x$ will not make solvency possible in this case.

We now introduce penalties $\rho_{\mathrm{F}}$ and $\rho_{\mathrm{S}}$ describing the negative consequences of a shortcoming of the objectives. We thereby can turn the problem given by (3.6a)–(3.6e) with probabilistic constraints into a two-stage stochastic optimization problem, as done for the time-invariant food security model in Section 2.2. After the reformulation, the problem still depends on uncertain parameters in the objective function but includes only deterministic constraints. Taking the expected value of the objective function over the above defined random variables, while keeping in mind their dependencies, then gives the following final formulation of the time-dependent food security model:

$$
\begin{aligned}
\textit{Minimize} \quad & f(x) = \mathbb{E}\Bigg[ \sum_{t=T_0}^{T_{\mathrm{fin}}(M)} \sum_{k=1}^{K} \sum_{j=1}^{J} x_{j,k,t} \cdot c_j \\
& \quad + \sum_{t=T_0}^{T_{\mathrm{fin}}(M)} \Big( \rho_{\mathrm{F}} \cdot \max\Big(0,\, A_t - \sum_{k=1}^{K} \sum_{j=1}^{J} (Y|_M)_{j,k,t} \cdot x_{j,k,t} \cdot a_j \Big) \Big) \qquad (3.7a) \\
& \quad + \rho_{\mathrm{S}} \cdot \max\Big(0,\, -G_{\mathrm{fin}}(x,\, M,\, Y|_M)\Big) \Bigg] \qquad \text{for } x \in \mathbb{R}^{J \times K \times (T_{\max} - T_0 + 1)} \\
\textit{Subject to} \quad & x_{j,k,t} \geq 0 \qquad \forall\, j \in \{1, \ldots, J\},\, k \in \{1, \ldots, K\},\, t \in \{T_0, \ldots, T_{\max}\} \qquad (3.7b) \\
& \sum_{j=1}^{J} x_{j,k,t} \leq \bar{x}_k \qquad \forall\, k \in \{1, \ldots, K\},\, t \in \{T_0, \ldots, T_{\max}\} \qquad (3.7c)
\end{aligned}
$$

The model output will be crop allocations $x$ that minimize the cost function $f(x)$ and thereby are the optimal choice regarding food security and solvency of the insurance scheme for the given settings.

As the expected value cannot be calculated analytically, we again approximate the objective function by the sample mean of $N$ realizations $m^s \in \Phi$ of the presence of catastrophic yields given by $M$ and $y^s \in \Omega$ of the corresponding yields $Y|_M$:

$$
\begin{aligned}
f_N(x) = \frac{1}{N} \sum_{s=1}^{N} \Bigg[ & \sum_{t=T_0}^{T_{\mathrm{fin}}(m^s)} \sum_{k=1}^{K} \sum_{j=1}^{J} x_{j,k,t} \cdot c_j \\
& + \sum_{t=T_0}^{T_{\mathrm{fin}}(m^s)} \Big( \rho_{\mathrm{F}} \cdot \max\Big(0,\, A_t - \sum_{k=1}^{K} \sum_{j=1}^{J} y^s_{j,k,t} \cdot x_{j,k,t} \cdot a_j \Big) \Big) \\
& + \rho_{\mathrm{S}} \cdot \max\Big(0,\, -G_{\mathrm{fin}}\Big(x,\, m^s,\, y^s\Big)\Big) \Bigg] \qquad (3.8)
\end{aligned}
$$

Thus, analogous to the time-invariant model in Chapter 2, the time-dependent sustainable food security model is an approximation of a two stage stochastic optimization model with a convex, piecewise linear objective function.

## 3.2. Model parametrization

The time-dependent sustainable food security model depends on a group of parameters, in part exogenous parameters that we set according to literature while others are model specific parameters which can be varied to analyze different scenarios.

Farm-gate prices in [$/t], i.e. prices received by farmers at the point of initial sale, can be found in a data set of the Food and Agriculture Organization Corporate Statistical Database (FAOSTAT)[59]. Prices are given as yearly timeseries on country level for over 200 crop and livestock commodities, including rice and maize which are the crops considered in this project. The timeseries cover the period from 1901 to 2018 but are incomplete for some countries. As we work with fixed prices valid for all simulation years, we take the average of all available years per country for both crops respectively. We then calculate a weighted average of the resulting country values using the shares each country has of the considered area in West Africa to get a single price value $p_j$ per crop. Resulting prices are 343.82 $/t for rice and 266.30 $/t for maize.

FAOSTAT does not report crop cultivation costs, and to our knowledge other sources for consistent values on national level are not available either. Hence, we rely on regional case studies and average obtained cost values to get a single value $c_j$ per crop valid for the total area and all years. Case studies for rice cultivation costs cover Liberia[60], Kaduna State in Nigeria[61], and Benin, Burkina-Faso, Mali and Senegal[62]. For maize, case studies cover Niger State in Nigeria[63] and Benin, Ghana and Cote D'Ivoire[64]. The resulting cultivation costs are 643.44 $/ha for rice and 281.22 $/ha for maize.

A naive upper bound $\bar{x}_k$ for crop areas per cluster is the actual area of the cluster, but the arable area is much smaller. Cotillon and Tappan, 2016, declare that by 2013 a share of 22.4% of the land surface of West Africa was cultivated[65]. They also state, that trends in West Africa show an expansion of agricultural area over time[65]. However, as expansion of agricultural area comes with investment costs and this aspect is not included in the model, we work with a fixed agricultural area share of 22.4%. We further assume that agricultural area is spread evenly over West Africa, such that we can calculate available agricultural area in each cluster by scaling the total cluster area.

Food demand in the model is calculated based on population data and a fixed average daily energy intake of 2360kcal per person. This is an estimated value for Sub-Saharan Africa excluding South Africa in 2015, as reported by FAO in a report on *World Agriculture*[66]. The United Nations World Population Prospects (UN WPP) provides a data set of yearly historic population data on national and sub-continental level since 1950 as well as nine population scenarios projected up to the year 2100 on the same scale[67,68]. As the area considered in the food security model does not follow political boundaries, the UN WPP data set can not be directly applied. The National Aeronautics and Space

Administration (NASA) Socioeconomic Data and Applications Center (SEDAC) provides a gridded population count at $0.5° \times 0.5°$ spatial resolution, adjusted to match the 2015 revision of UN WPP country totals[69,70]. We use this data set to calculate the share of the population reported by UN WPP for Western Africa in 2015 that lives in the area considered by the sustainable food security model. This share is then used to calculate population scenarios for the model by scaling UN WPP projections. We use the low, medium and high fertility scenarios for comparison of possible demographic developments[68]. For the analysis of other model parameters, we keep the population size fixed over time.

The main input to the time-dependent food security model are yield distributions for each cluster, crop, and year. Yield distributions are based on the GDHY data set (see Section 1.3.2), and two scenarios are used in the model analysis. One uses trends in mean yield values as calculated in Section 1.3.2, while the other uses fixed yield distributions for all years, as this simplifies the interpretation of other parameters' influence on the model output. As 2016 is the last year covered by the GDHY data set, we set $T_0 = 2017$ as the first year covered by the model, using the yield distribution estimated for 2016 in the fixed yield distributions scenario, and the distributions projected by the trend for the respective years otherwise. We set $T_{\max} = 2041$, as a time window of 25 years roughly corresponds to one generation which is assumed to be a reasonable planning horizon.

Exact parameters of the crop yield distributions depend on the year and cluster, but distributions for rice yields generally have a higher mean but also slightly higher standard deviation than maize yield distributions. The number of crops is set as $J = 2$, with the crops being maize and rice, due to the availability of data as explained in Section 1.2. In 2016, for $K = 1$ the mean of the estimated distribution for cluster average rice yields is 2.54 t/ha with a standard deviation of 0.14, while for maize it is 1.62 t/ha with a standard deviation of 0.08. For $K = 7$ the means of rice yield distributions are in the range of 1.63 t/ha to 4.45 t/ha with standard deviations between 0.13 and 0.31, while maize yields are more stable over West Africa with means ranging from 1.03 t/ha to 2.37 t/ha for the different clusters and standard deviations between 0.09 and 0.24. Yield realizations are transformed from [t/ha] to [kcal/ha] using values $a_j$ on their energy content taken from the Standard Reference on nutrient values of the U.S. Department of Agriculture (USDA)[71], being 360 kcal/100g for rice and 365 kcal/100g for maize.

The number of clusters $K$ can be varied in order to analyze the influence of changing regional subdivision on the model output. Including a higher number of clusters leads to reduced spatial correlation of yields within the model, as different clusters are assumed to be independent. In particular, when some clusters exhibit catastrophic yields, others might still be non-catastrophic. We use $K = 1$ to analyze the influence of other parameters on the model's behavior, and compare runs between $K = 1$ and $K = 2$ to understand the direct effect of reduced correlation. Furthermore, $K = 7$ is included

as this is the optimal number of clusters obtained by the cluster analysis (see Section 1.2).

The risk level $r$ covered by the government is given as the quantile separating catastrophic from non-catastrophic yields and thus defines also the probability for a single cluster to exhibit catastrophic yields, in which case farmers of that cluster will be guaranteed a designated aggregated income. A common threshold to define extreme events is the 5%-quantile[72], therefore we use $r = 5\%$ as default, but also run scenarios with $r = 2.5\%$ and $r = 10\%$ for comparison.

The fund from which farmers are payed in catastrophic years has an initial size $G_{\text{ini}}$ and is then built up by farmers paying taxes on their profits. We set the default tax rate to $\tau = 3\%$, with $\tau = 1\%$ and $\tau = 5\%$ used to analyze the influence of the tax rate on model results. The initial fund is $G_{\text{ini}} = 0$ in all runs, but can be set to a positive value either to represent the government's contribution to the fund, or in case the model is reapplied after a catastrophic year. For the latter, the initial fund of the new model run will be positive if the government stayed solvent despite the payouts.

The guaranteed income per cluster is calculated from the aggregated expected income of farmers in that cluster without government involvement. As policy decisions in general are often based on present-day values, we are interested in the expected income in the year previous to the simulation start $T_0$, i.e. in 2016. We assume that independent of the probabilities of reaching food demand and government solvency for which the model is to be run, farmers aim for a high probability of food security $\alpha_F = 95\%$ before the government is involved, as they are missing the security to take higher risks. By running the model for the single year 2016 with $\alpha_F = 95\%$ and a solvency penalty of $\rho_S = 0\,\$/\$$ to exclude government intervention, we can determine the expected income under these circumstances. We then set the guaranteed income per cluster to a given share $S_{\text{gov}}$ of the respective expected income, with a default value of $S_{\text{gov}} = 75\%$. With increasing population, we assume the absolute number of farmers to increase proportionally. To keep the guaranteed income per farmer constant, we therefore scale the resulting aggregated guaranteed income per cluster by the ratio between population size in year $t$ and 2016 to get the guaranteed income in year $t$.

However, the food security and solvency probabilities $\alpha_F$ and $\alpha_S$ are not a direct input to the developed optimization model, which instead relies on penalties $\rho_F$ and $\rho_S$ describing second stage costs in case of violation of the constraints. We therefore need to numerically find the proper penalty values to meet the constraints with the given probabilities. As both constraints are not independent of each other, a pair of penalties $\rho_F$ and $\rho_S$ leading to probabilities $\alpha_F$ and $\alpha_S$ is not necessarily unique. Furthermore, a simultaneous search for both penalties would lead to high computational costs. Instead, we determine each of the penalties $\rho_F(\alpha_F)$ and $\rho_S(\alpha_S)$ separately while the respective other penalty is set to 0, requiring an accuracy of 1% in the probabilities. As both penalties lead to investment in the same first stage action, i.e. crop cultivation, there is

no trade-off but rather a synergy between the penalties when both are included in the model. Hence, resulting probabilities $\alpha'_F$ and $\alpha'_S$ of the run with penalties $\rho_F = \rho_F(\alpha_F)$ and $\rho_S = \rho_S(\alpha_S)$ can be higher than demanded due to the interaction of the penalties, but they are always at least as high as the probabilities given as input to the model, i.e. $\alpha'_F \geq \alpha_F$ and $\alpha'_S \geq \alpha_S$.

## 3.3. Benchmarking model performance against a deterministic alternative

In Section 2.1, several approaches of integrating uncertainty into a model were discussed and the usage of stochastic optimization models in order to get a robust solution was motivated. We now introduce a method to quantify potential benefits of setting up a stochastic optimization model over a deterministic approach based on expected values of uncertainties.

Let $g_0 : X \times \Omega \to \mathbb{R}$ be a real-valued objective function defined on a set of feasible decisions $x \in X$ and realizations $z \in \Omega$ of a random variable $Z$. As introduced in Section 2.1, the optimal solution using the expected value as realization of $Z$ can be calculated as $EV = \min_{x \in X} g_0(x, \mathbb{E}[Z])$ attained at $\hat{x}$, while the robust solution is given by $RS = \min_{x \in X} \mathbb{E}_Z[g_0(x, Z)]$. *Jensen's inequality* states that for any convex function $\varphi$ and a random variable $Z$ the following relation holds[73]:

$$\varphi(\mathbb{E}[Z]) \leq \mathbb{E}_Z[\varphi(Z)] \tag{3.9}$$

If we assume convexity of the objective function $g_0$, we can deduce

$$EV = \min_{x \in X} g_0(x, \mathbb{E}[Z]) \leq \min_{x \in X} \mathbb{E}_Z[g_0(x, Z)] = RS \tag{3.10}$$

However, note that the value $EV$ only reflects real costs if the actual realization of $Z$ is its expected value. In any other case, the solution $\hat{x}$ given by minimizing $g_0(x, \mathbb{E}[Z])$ might lead to additional costs after the uncertain events have happened. In particular it is

$$\min_{x \in X} \mathbb{E}_Z[g_0(x, Z)] \leq \mathbb{E}_Z[g_0(\hat{x}, Z)] \tag{3.11}$$

Hence, the expected costs using the solution $\hat{x}$ are always at least as high as the expected costs of the robust solution. This motivates the following definition[74]:

**Definition 3.1.** *Let $g_0 : X \times \Omega \to \mathbb{R}$ be a function defined on a set of feasible decisions $x \in X$ and realizations $z \in \Omega$ of a random variable $Z$. Let $\hat{x} = \operatorname{argmin}_{x \in X} g_0(x, \mathbb{E}[Z])$. Then the* value of stochastic solution *is defined as*

$$VSS = \mathbb{E}_Z[g_0(\hat{x}, Z)] - \min_{x \in X} \mathbb{E}_Z[g_0(x, Z)] \tag{3.12}$$

The value of stochastic solution thereby reflects the costs that can be saved on average by using the robust solution instead of the solution based on the expected value of the uncertainties.

In the case of the time-dependent sustainable food security model, the objective function does not only rely on one random variable, but on a random variable $M$ describing the presence of catastrophic yields and random yields $Y|_M$ conditional on $M$. However, the expected yields $\bar{y} = \mathbb{E}_M[\mathbb{E}_{Y|_M}[Y|_M]]$ in general are non-catastrophic. Therefore, we assume no occurrences of catastrophes as realization of $M$ and average yields $\bar{y}$ as realization of $Y|_M$ for the deterministic approach when analyzing the performance of the sustainable food security model in Section 4.7.

# 4. Results on the sustainable food security model

In Section 3.2, the parametrization of the time-dependent sustainable food security model is discussed. Some parameters as the available agricultural area or farm-gate prices are set according to literature research and are assumed to be constant. To understand how different aspects influence the performance and feasibility of a large scale insurance system as strategy towards sustainable food security, the influence of the other parameters on the model output is analyzed in this chapter.

We start with more technical points as basis, first analyzing how the accuracy of the model is affected by the sample size $N$ used to approximate the objective function. This analysis allows to determine a reasonable sample size for subsequent runs, considering the trade-off between accuracy and computational costs (Section 4.1). In the following section, the relation between probabilities $\alpha_F$ and $\alpha_S$ for meeting the food demand in West Africa and solvency of the government fund and the corresponding penalties $\rho_F$ and $\rho_S$ is examined. We also analyze the interaction between both penalties and their influence on model dynamics, and based on this choose viable probabilities for further runs (Section 4.2).

Building on these results, we then consider the aspects relating to more substantive questions. In Section 4.3, the functionality of government instruments to influence the financing mechanism and assure fund solvency are analyzed. As such instruments the model includes the covered risk level $r$, the tax rate $\tau$ that farmers have to pay on their profits, and the share $S_{\text{gov}}$ of expected income that is guaranteed by the government in case of extreme events. In a next step, different combinations of population and yield scenarios are compared to understand the influence of future developments on sustainable food security (Section 4.4). In Section 4.5, we investigate the effect of subdividing the considered region into several clusters to address the issue of spatial dependence between crop yields in different areas of West Africa.

When analyzing specific parameters or scenarios, the other settings are kept at their default to simplify the interpretation of changes in the model output. Default settings are the scenarios with fixed yield distributions and population numbers and thereby also constant food demand over time, a covered risk of $r = 5\%$, a tax rate of $\tau = 3\%$, a guaranteed income of $S_{\text{gov}} = 75\%$ of the expected income, and a single cluster, i.e. $K = 1$. As probabilities, we generally consider a set of combinations of different values

for $\alpha_F$ and $\alpha_S$, focusing on $\alpha_F = 95\%$ and $\alpha_S = 85\%$ for visualizations in Section 4.3 and later.

After analyzing all settings separately, we combine the optimal cluster number $K = 7$ with a possible scenario of future developments for a realistic run of the model, using yield distributions that include the trend in mean yields and population predictions based on a medium fertility assumption (Section 4.6). We conclude this chapter with a section assessing the hypothesis that a stochastic food security model substantially outperforms a deterministic modeling approach that relies on expected yield values (Section 4.7).

The implementation of the sustainable food security model as well as the analysis of all results was done using Python 3.7 [18] (see Appendix A).

## 4.1. Accuracy of the sustainable food security model depending on the sample size

The exact objective function $f(x)$ of the sustainable food security optimization problem, given by (3.7a), depends on an expected value and is therefore approximated by an objective function $f_N(x)$ using the sample mean over $N$ realizations of the uncertainties (see Eq. (3.8)). The accuracy of the model relies on the number of realizations included in this approximation.

In order to compare the model output for different runs with only the sample size $N$ changing, the guaranteed income $I^{\text{gov}}$ and the penalties $\rho_F$ and $\rho_S$ have to be predetermined, as these are otherwise calculated depending on the given settings including $N$. We used default settings with $N = 10\,000$ and probabilities $\alpha_F = 80\%$ and $\alpha_S = 90\%$ to calculate these parameters[1]. The model was then run for different $N$ in the range from 10 to 30 000. For each sample size the run was repeated 50 times using different seeds for the realizations of uncertainties. To quantify the accuracy of the model, for each sample size we consider the Relative Standard Deviation (RSD) for two different aspects of the model results obtained from the different runs. First, we calculate the RSD of the minima of the objective functions $f_N(x)$, i.e. the minimized total costs, then we determine the average RSD of the corresponding crop areas for which the minima are attained. Results for both metrics are visualized in Figure 4.1.

For the minimized total costs, all runs with a sample size of at least 3 500 had a RSD of less than 1% (see Figure 4.1 (A)). We chose 1% as threshold, as it approximately marks the point where only marginal improvements in RSD can be achieved by further increasing the sample size. For the crop areas, already a sample size of 750 was sufficient to get an average RSD value below 1%, however the sample size above which only

[1] The probabilities used for analysis of the sample size differ from the typical probabilities used throughout the following result sections. This is because accuracy tests had to be run before analyzing the relation between penalties and probabilities on which the decision of default probabilities was based.

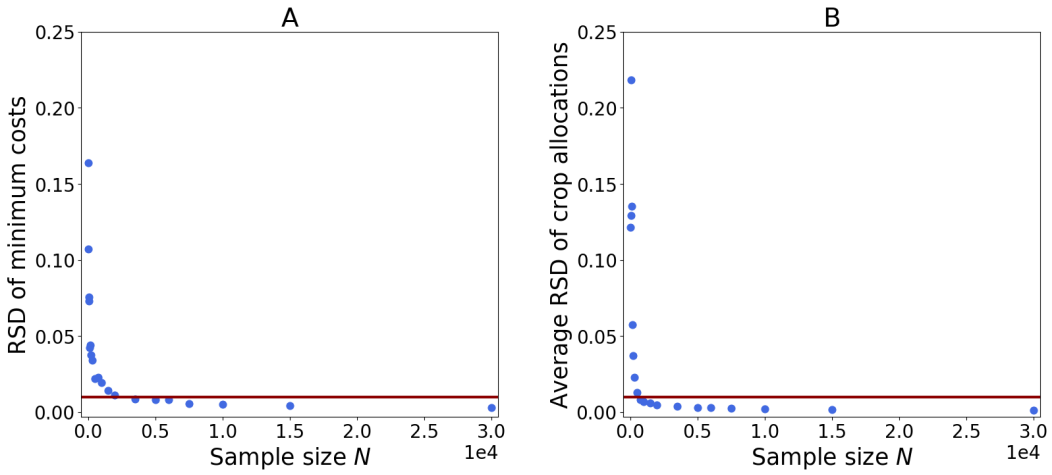



Figure 4.1.: For $N$ in the range of 10 to 30 000, (A) shows the Relative Standard Deviation (RSD) of min $f_N(x)$ over feasible crop area allocations $x$, while (B) shows the average RSD of the corresponding crop areas for which the minimum is attained. The red line indicates a RSD of 1%, which is used as threshold for the required accuracy.

marginal improvements can be realized is again around $N = 3\,500$ (see Figure 4.1 (B)). Therefore, $N = 3\,500$ was chosen as default sample size.

However, the model accuracy depends on other settings as well. For higher K, the dimension of the random yield vector increases as yields are given for each cluster, and thus the sample size needs to be increased to reliably cover the joint yield distribution. In general, results get less stable for very high probabilities $\alpha_F$ or $\alpha_S$, as in these cases the respective constraint has to be satisfied also in the worst cases. This increases the influence of a small share of the realizations. Instabilities also increase in the later years of the considered time period: Each year, a percentage of realizations according to the covered risk $r$ is defined as catastrophic. Those realizations are excluded in the later years as the simulation of each realization terminates after the first catastrophe, such that the sample size decreases over the years. This effect is intensified in the case where only a penalty for government solvency is included, i.e. where $\rho_F = 0\,\$/\text{kcal}$, as in this case crop area allocations in the later years are mainly influenced by the years in which a catastrophe and thus payouts take place. Therefore, crop area allocations for $\rho_F = 0\,\$/\text{kcal}$ can show an unstable behavior in the last years even if the sample size is sufficient for other penalty levels, especially if combined with very high or low $\rho_S$ (see Figure 4.2 (B) for $\rho_S = 50\,\$/\$$).

Hence, while most runs were conducted for $N = 3\,500$, for some settings runs are repeated with a higher sample size, reaching up to $N = 50\,000$ for cases considering more than one cluster, to obtain a more steady behavior of crop area allocations and other resulting variables as probabilities over time.

## 4.2. Interaction of probabilities and penalties and their influence on model dynamics

The penalties $\rho_F$ and $\rho_S$ are second stage costs which arise in the case of violations of the constraints, and are multiplied by the strength of the violation, i.e. the shortcoming of the food demand or the amount of payouts in a catastrophic year that cannot be covered by the fund. They are included in the model as a tool to ensure that the constraints are met with given probabilities $\alpha_F$ and $\alpha_S$ for food security and solvency of the government fund. We can interpret the penalties as an abstract measure of the negative consequences of not meeting constraints. Hence, for higher penalties, the probability of meeting the constraints will increase, as the consequences are worse and it thus makes sense to invest more money in preventing these consequences from happening. Note that his argumentation can be turned around: Instead of aiming to reach certain probabilities of solvency or food security, the government could analyze the severity of possible negative consequences when not meeting the constraints and quantify direct and indirect monetary and socio-economic costs. These costs would then correspond to externally given penalties in the model. The model informs on optimal crop allocations to minimize the aggregated costs of first stage actions and second stage penalties, whereby the corresponding probabilities to meet the constraints would then be a result of the model. While a quantification of second stage costs can be feasible if only direct costs are included, the quantification of socio-economic consequences in general is not possible. Therefore, we use the former interpretation and understand the probabilities $\alpha_F$ and $\alpha_S$ for food security and solvency as input to the model.

Penalties corresponding to given probabilities are found numerically and independent from one another, by keeping the other penalty at zero in the process (for details see Section 3.2). However, each penalty will not only influence the associated probability, but also the probability of meeting the other ojective, as in both cases a higher penalty leads to larger crop area allocations, thus increasing the probability to meet either constraint. Figure 4.2 shows both probabilities for $\rho_F = 0\,\$/\text{kcal}$ and varying $\rho_S$ in (A) and for $\rho_S = 0\,\$/\$$ and varying $\rho_F$ in (B). To result in second stage costs of same order of magnitude for both penalties, ranges of the penalties $\rho_F$ and $\rho_S$ need to differ, as the shortcomings with respect to the corresponding objectives have different orders of magnitude. Considering e.g. the run with $\rho_F = 2 \cdot 10^{-4}\,\$/\text{kcal}$ and $\rho_S = 500\,\$/\$$, which roughly corresponds to respective probabilities of $\alpha_F = \alpha_S = 95\%$, the average penalty paid for insolvency is $5.4 \cdot 10^9\$$ and the average food demand penalty aggregated over all years is $8.2 \cdot 10^9\$$. These penalties account for less than 5% of the total expected cost each.

For both cases visualized in Figure 4.2, the corresponding probability, i.e. the solvency probability $\alpha_S$ in (A) and the food security probability $\alpha_F$ in (B), tends towards one for high penalties. A high food security penalty $\rho_F$ without a penalty $\rho_S$ also leads to a high solvency probability $\alpha_S$, although not tending towards one. This effect is a

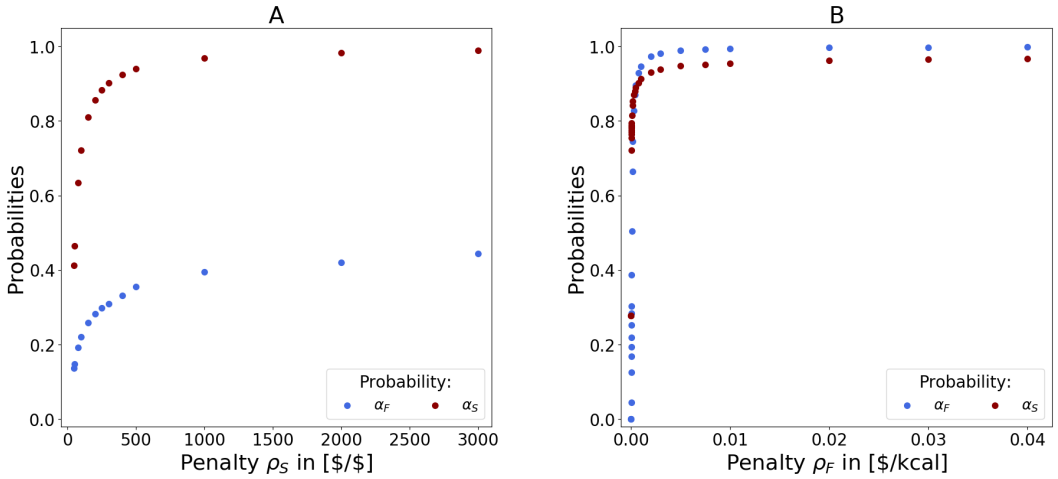



Figure 4.2.: Probabilities $\alpha_F$ and $\alpha_S$ corresponding to different penalty values while all other settings are kept at default values. In (A) $\rho_S$ is varied while $\rho_F = 0\,\$/\text{kcal}$, in (B) $\rho_F$ is varied while $\rho_S = 0\,\$/\$$.

consequence of the initial fund size $G_{\text{ini}}$ being equal to zero and the resulting difficulty to stay solvent in case of a catastrophe in the first years after initiating the insurance scheme. To reach a high probability $\alpha_S$, the government fund has to cover such cases as well. This can only be achieved by cultivating large crop areas which will lead to overproduction in non-catastrophic years (see Figure 4.3 (B)). These large areas effect the fund in two ways: On the one hand, the fund will be build up quicker due to higher tax payments. On the other hand, farmers' aggregated profits in the case of catastrophe will not be as low as for smaller cultivated areas, such that the amount of payments the

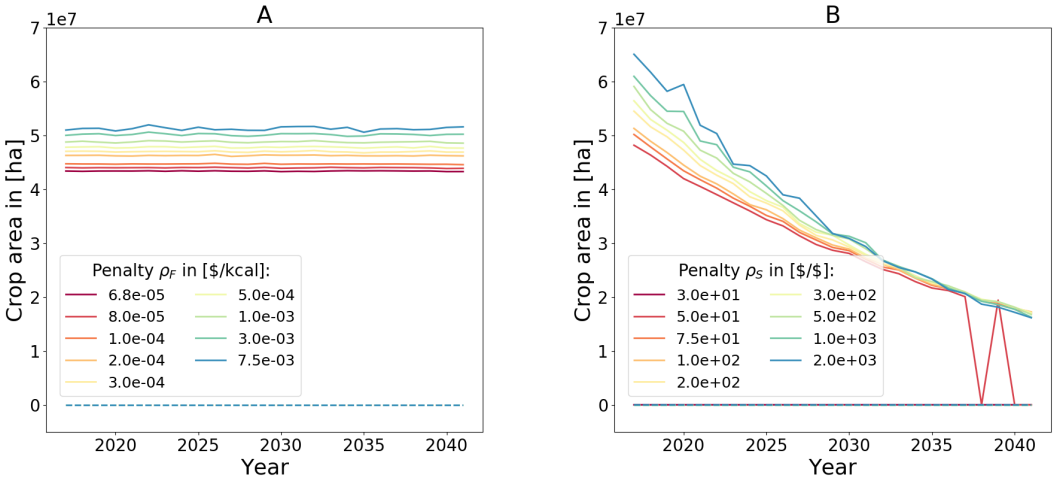



Figure 4.3.: Yearly crop area allocations for maize (solid line) and rice (dashed line) corresponding to different values of each penalty while all other settings are kept at default values. In (A) different lines correspond to different values of the food security penalty $\rho_F$ while $\rho_S = 0\,\$/\$$, in (B) different lines correspond to different values of the solvency penalty $\rho_S$ while $\rho_F = 0\,\$/\text{kcal}$. Note that rice allocations in both cases are constant at 0 ha, such that the dashed lines coincide. Unstable behavior of maize allocations for $\rho_S = 50\,\$/\$$ in (B) is due to numerical instability.

government has to make is reduced, as they only cover the difference between profits and guaranteed income. Meanwhile, crop allocations resulting from only using a food security penalty $\rho_F$ with $\rho_S = 0\,\$/\$$ under default settings, i.e. without trends in yield or population size, will be constant over the entire considered period (see Figure 4.3 (A)). In particular, the level of crop allocation to guarantee food security is lower than the crop areas in the first years resulting for high values of the solvency probability $\alpha_S$. Therefore, the solvency probability $\alpha_S$ will tend towards a value slightly smaller than one when just including a food security penalty $\rho_F$. On the other hand, a high solvency penalty $\rho_S$ with $\rho_F = 0\,\$/\text{kcal}$ only tends towards a food security probability of slightly over $\alpha_F = 40\%$. This happens, as once the fund has been established, the crop areas can be reduced drastically while still guaranteeing solvency in most cases (see Figure 4.3 (B)). The crop allocation levels drop below the essential area to meet the food demand, such that averaged over all years a certain probability $\alpha_F$ cannot be surpassed.

The dynamics of crop area allocation thereby also affect the yearly probability of meeting the food demand. A specific overall food security probability $\alpha_F$, when resulting from a high solvency penalty $\rho_S$ as shown in Figure 4.2 (A), will not be reflected evenly by the probabilities in each year, as can be seen in Figure 4.4 (B). Instead, in the first years the food security probability is very high, but then drops to zero once the fund has sufficiently been built up. If both penalties are included, the probability will not drop to zero but to a level corresponding the food security penalty $\rho_F$. This is, apart from computational restrictions, the reason for determining the penalties independently, as the food demand should be met with a certain probability in every year and not just on average. In Figure 4.4 (A), we see that the food security penalty $\rho_F$ without including a penalty $\rho_S$ achieves a given probability $\alpha_F$ in every year. However, this is only valid as long as no trends in yields or population size are included, as e.g. positive yield

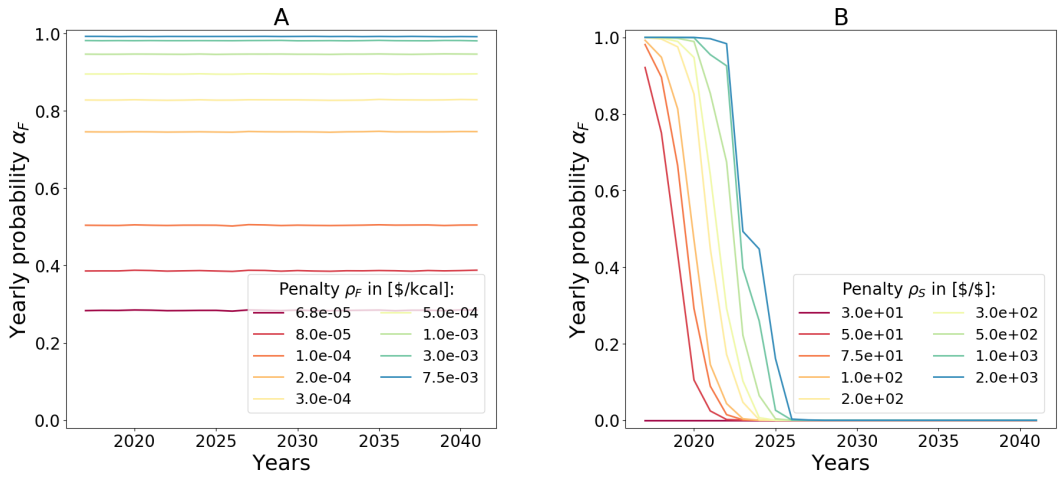



Figure 4.4.: Yearly probability for meeting the food demand corresponding to different penalty values while all other settings are kept at default values. In (A) different lines correspond to different values of $\rho_F$ while $\rho_S = 0\,\$/\$$, in (B) different lines correspond to different values of $\rho_S$ while $\rho_F = 0\,\$/\text{kcal}$.

trends would simplify meeting the food demand in later years, such that the probability for food security would increase over time. To obtain a specific probability of meeting the food demand in every year independent of the scenario, food security in each year would need to be a separate constraint, leading to time dependent penalties $\rho_F(t)$ and increasing the complexity and computational requirements of the model.

In both cases visualized by Figure 4.3, the model suggests to only cultivate maize. This is due to the fact that according to the data used in the model maize is superior to rice when using only one cluster, both regarding the food security constraint and the solvency constraint. Considering the solvency constraint, high profits for a given investment are needed to build up the fund through the corresponding high tax payments. In contrast, for the food security constraint high production in terms of calories for a given investment is desired. We can quantify the performance of each crop with reference to the constraints by calculating the expected net profit per dollar invested in crop cultivation and the expected produced calories per dollar invested in crop cultivation[2]. Resulting values when considering West Africa as a single cluster are given in Table 4.1.

| Crop | profit per invested \$ (in [\$]) | production per invested \$ (in [kcal]) |
|---|---|---|
| Rice | 0.36 | 14213.37 |
| Maize | 0.54 | 21056.81 |

Table 4.1.: Performance of crops for $K = 1$, quantified by expected net profit per investment in [\$/\$] regarding the solvency constraint and by expected production per investment in [kcal/\$] regarding the food security constraint.

We see that here maize is superior with regard to both constraints. This however is not the case in general. Assuming that one crop has a better trade-off in food production while the other is more profitable, the latter crop will be the main crop in the first years to quickly build up the fund, but once the fund is sufficiently built up and the food security penalty $\rho_F$ dominates the dynamics, the share of the first crop will increase. We can observe this effect in some clusters when considering $K > 1$ in Section 4.5. Furthermore, the performance of individual crops can change over time when including trends in mean yields, which could also lead to a substitution of crop area allocation between different crops.

The effect of changing dominance of the penalties can also be noticed when looking at crop area allocations resulting from runs with both penalties included in the model. Figure 4.5 shows crop allocations over the considered time period for different food security penalties $\rho_F$ in (A) and for different solvency penalties $\rho_S$ in (B). In contrast to Figure 4.3, instead of keeping the other penalty at zero, it is $\rho_S = 200\,\$/\$$ in (A) and $\rho_F = 10^{-4}\,\$/\text{kcal}$ in (B). In (A) crop allocations behave very similar in the initial period and decrease over the years, until they flatten out to different levels corresponding to

[2] Specifically, we calculate $(\bar{y}_{j,k} \cdot p_j - c_j)/c_j$ and $(\bar{y}_{j,k} \cdot a_j)/c_j$, where $\bar{y}_{j,k}$ is the expected yield for the respective cluster and crop in [t/ha], $p_j$ is the farm-gate price for the respective crop in [\$/t], $c_j$ are the cultivation costs for the respective crop in [\$/ha], and $a_j$ is the energy content of the respective crop in [kcal/t]. Time dependence of yields is excluded as we consider the scenario using fixed yield distributions.

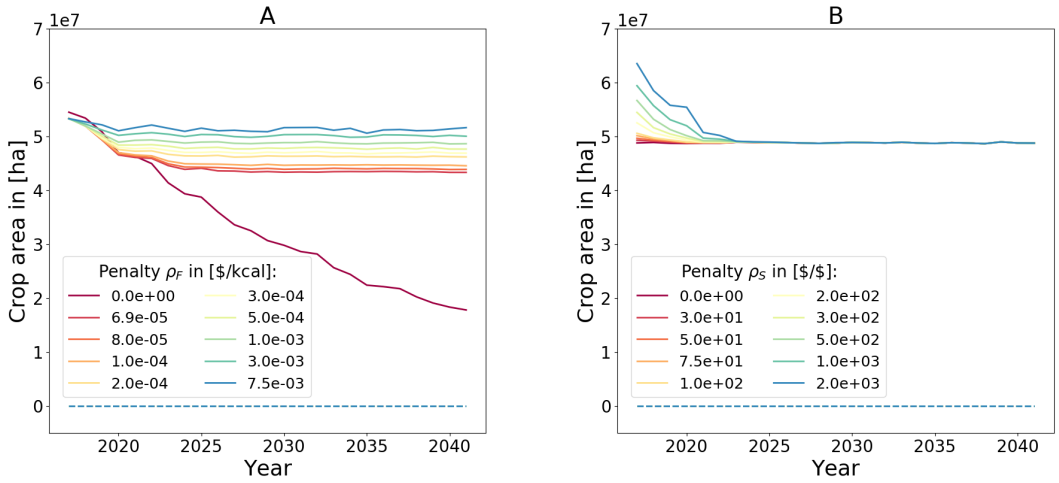



Figure 4.5.: Yearly crop area allocations for maize (solid line) and rice (dashed line) corresponding to different combinations of penalties while all other parameters are kept at default values. In (A) different lines correspond to different values of $\rho_F$ while $\rho_S = 200\,\$/\$$, in (B) different lines correspond to different values of $\rho_S$ while $\rho_F = 10^{-4}\,\$/\text{kcal}$. Note that rice allocations in both cases are constant at 0 ha, such that the dashed lines coincide.

the value of $\rho_F$ once the food security penalty dominates. Crop allocations in the first years are not exactly the same, as the higher crop allocation in later years for higher $\rho_F$ is counteracted by a slightly lower crop allocation in the first years. The interaction of both penalties does not only influence the level of crop allocations in the first and last years, but also the position and duration of the transition phase between the dominance of either penalty. This can be seen by the different years in which crop area allocations change from decreasing to an approximately constant level. In (B) the food security penalty $\rho_F$ is kept constant, thus the crop allocation in the final years is very similar for all shown cases. Instead, the initial crop allocations vary according to the different solvency penalties $\rho_S$.

We observed that without an initial fund $G_{\text{ini}}$ it is difficult for the government to stay solvent if catastrophes occur in the first years, which leads to high crop area allocations in the first years. As we assume the resulting uneven production over time to be an undesired effect and very high crop allocation levels are potentially infeasible due to limited resources of farmers, we use a probability of $\alpha_S = 85\%$ for visualization of model results in upcoming sections, while demanding a high level of food security by setting $\alpha_F = 95\%$.

## 4.3. Policy interventions for financing farmers' resilience against extreme events

The government can influence the solvency of the financing mechanism by three kinds of intervention according to the type of insurance it wants to provide. These parameters are

the risk level $r$ that is covered, the share $S_{\text{gov}}$ of the expected income that is guaranteed in case of catastrophic yields, and the tax rate $\tau$ determining the share of profits farmers have to pay into the government fund. For each we compare the default value to a lower and a higher value, while all other settings are kept at default. Resulting crop area allocations for probabilities $\alpha_F = 95\%$ and $\alpha_S = 85\%$ are shown in Figure 4.6.

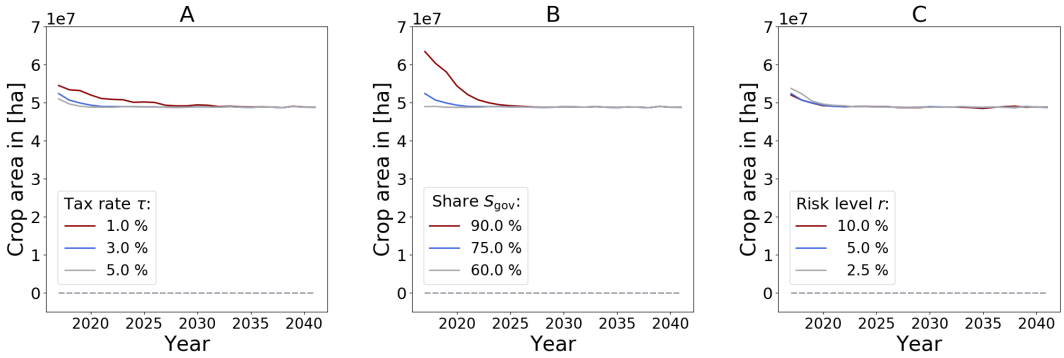



Figure 4.6.: Yearly crop area allocations for maize (solid line) and rice (dashed line) for $\alpha_F = 95\%$ and $\alpha_S = 85\%$. In (A) the tax level $\tau$ is varied, in (B) the share $S_{\text{gov}}$ of expected income that will be guaranteed by the government is varied, and in (C) the covered risk $r$ is varied. In each case, all other settings are kept at default values. Note that rice allocations in all cases are constant at 0 ha, such that the dashed lines coincide.

In all cases, crop allocations start with higher values in the first years and then decrease until they reach a constant level. This level is the same for all parameter choices, as in these years the model behavior is dominated by the food security penalty $\rho_F$ and is therefore not influenced by the government related parameters $\tau$, $S_{\text{gov}}$ and $r$.

Using a different tax rate $\tau$ directly affects the profits needed to build up the government fund. Therefore, a lower tax rate $\tau$ leads both to a higher initial crop allocation and to the decrease in crop allocations being spread over a longer time period, thus also shifting the point at which the dominance of the penalties switches (see Figure 4.6 (A)). While there is a significant difference between $\tau = 1\%$ and $\tau = 3\%$, increasing the tax rate further to $\tau = 5\%$ only shows a small reduction in overall crop area allocations. Thus, a tax rate this high seems to put unnecessary pressure on farmers.

The share $S_{\text{gov}}$ of the expected income that is guaranteed by the government in catastrophic years has direct influence on the payouts carried out by the government. To reach the same solvency probability $\alpha_S$ for a larger share $S_{\text{gov}}$, the crop areas thus have to be bigger to build a larger fund. Bigger crop areas will also reduce the effect of increased payouts, as farmers' profits in catastrophic years in general will be higher if a bigger area is cultivated and the government thus has to cover a smaller part of the guaranteed income. The effect of $S_{\text{gov}}$ on the crop allocations is visualized in Figure 4.6 (B). For $S_{\text{gov}} = 60\%$, payouts are so low that the food security penalty $\rho_F$ dominates immediately, leading to a constant level of crop areas over the full period. A guaranteed share of $S_{\text{gov}} = 90\%$ on the other hand leads to a substantially bigger crop area, which is

accompanied by high overproduction in the case of non-catastrophic years. This suggest that such a high share $S_{\text{gov}}$ is infeasible, at least when setting up the insurance scheme without an initial fund $G_{\text{ini}}$.

For a higher covered risk r, a bigger share of yield realizations is labeled as catastrophic and leads to government payouts. As the probability of a year to be catastrophic is thus higher, the expected year of the first catastrophe is earlier. Hence, the government fund needs to be built up faster, which explains the slightly higher crop areas shown in Figure 4.6 (C). This effect however is counteracted by reduced payouts, as the threshold for catastrophic yields is higher and thus farmers on average earn more money in a year labeled catastrophic than they would for a lower covered risk $r$. Average payouts for a covered risk of $r = 2.5\%$ are around $7 \cdot 10^8$\$, while they are only $4.7 \cdot 10^8$\$ and $4.6 \cdot 10^8$\$ for $r = 5\%$ and $r = 10\%$ respectively. Therefore, there is only little difference in crop allocations for changing covered risk $r$, and $r = 5\%$ and $r = 10\%$ lead to almost identical graphs as shown in Figure 4.6 (C). The similar effects arising for $r = 5\%$ and $r = 10\%$ can be seen as an argument for covering extreme events corresponding to the lowest 10% of yields to give farmers a higher resilience against bad harvests and thus further increase sustainability. However, an effect that is not covered by the food security model is that with a higher covered risk level more catastrophes can happen within a fixed time period, leading to multiple payouts which put additional pressure on the government fund. To cover this aspect, the structure of the model would need to be adapted to allow for multiple catastrophes. This however needs further information on consequences of a catastrophe and resulting government policies, to understand how other model parameters are affected. As this was beyond the scope of this project, we instead analyze the distribution of the final fund size in the different settings, as this is the main aspect within the model that influences post-catastrophe settings. The distributions are shown in Figure 4.7.

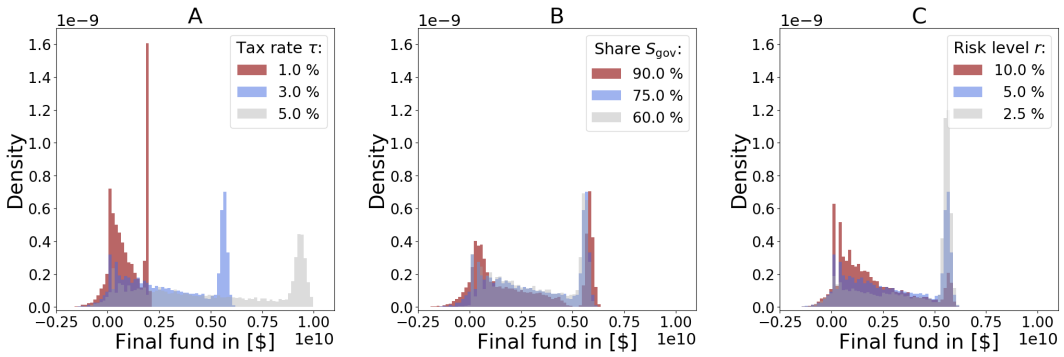



Figure 4.7.: Distribution of final fund for the different yield realizations in each run for $\alpha_F = 95\%$ and $\alpha_S = 85\%$. In (A) the tax level $\tau$ is varied, in (B) the share $S_{\text{gov}}$ of expected income that will be guaranteed by the government is varied, and in (C) the covered risk $r$ is varied. In each case all other settings are kept at default values.

In general, the distributions are bimodal with one mode corresponding to the final fund following a catastrophe and the other to the final fund after the last year $T_{\max}$ if no catastrophe took place. The share of negative values of the final fund corresponds to

the solvency probability $\alpha_S$. Even though all histograms show runs with $\alpha_S = 85\%$, this share does vary. This is due to the changing influence of the food security penalty $\rho_F$, as the penalties are determined independently.

A lower tax rate $\tau$ mainly compresses the distribution of the final fund towards the lower end, which can be seen in Figure 4.7 (A). As in most years the crop allocation is dominated by $\rho_F$ and thereby identical for all tax rates, after the initial years the fund is built up more slowly for lower tax rates, thus leading also to a shifted second mode. The relation between the value of the higher peaks in the final fund roughly corresponds to the relation between the different tax shares, as this peak is given by realizations for which the simulation runs over the full 25 years without catastrophe. For $\tau = 1\%$ we see that the share of negative final funds is slightly higher, and thus the solvency probability $\alpha_S$ slightly lower than for higher tax rates. In contrast, $\tau = 3\%$ and $\tau = 5\%$ show very similar behavior at the lower end of the distribution. However, there is a much lower accumulation of fund over time for $\tau = 3\%$, meaning that the solvency constraint is met more efficiently for this tax rate. Similar to the resulting crop allocations, this suggests that $\tau = 5\%$ puts unnecessary pressure on farmers and that $\tau = 3\%$ is a better choice to ensure feasibility of the financing mechanism while providing resilience against extreme events to farmers.

The different shares $S_{\text{gov}}$ of expected income that are guaranteed by the government in case of catastrophic yields show the biggest influence on the share of negative values in the distribution of the final fund, as shown in Figure 4.7 (B). Hence, this government instrument has the strongest connection to the extent to which the food security penalty $\rho_F$ influences the solvency probability $\alpha_S$. The dominance of the food security penalty over all years combined with low payouts for $S_{\text{gov}} = 60\%$ results in a solvency probability of $\alpha_S = 99.6\%$, even though the solvency penalty $\rho_S$ is calibrated to $\alpha_S = 85\%$. Apart from this, the upper peak of the distribution is slightly shifted for different values of $S_{\text{gov}}$, as for a lower guaranteed share and thus lower payouts the crop allocations in the first years are also lower. This is reflected by lower profits in the first years and the fund is built up to a slightly lower level.

The level $r$ of risk covered by the government does not have a significant influence on the dynamics of payouts, as explained above. Therefore, the negative part of the distribution shown in Figure 4.7 (C) is very similar in all cases. It does however strongly influence the number of realizations that do not exhibit a catastrophe within the considered time period. Hence, for a lower covered risk $r$ and thus lower probability of a year to be catastrophic, the upper mode, which corresponds to realizations without a catastrophe, has a higher density. Contrary to this, for a high covered risk level the probability of positive but very low final fund is higher. Even though it is desirable to reduce unnecessary accumulation of funds and high final funds should not be the aim of the insurance scheme, a very low final fund has negative effects as well: Crop area allocations in the years after the catastrophe will need to be high to rebuild the fund, and thereby

lead to additional cultivation costs. This is especially relevant when considering high covered risk levels and thus more frequent catastrophes.

## 4.4. Impact of yield and population projections on sustainable food security

So far, the model was analyzed assuming fixed yield distributions over time, a fixed population size, and thereby a constant food demand. However, sustainable food security over the years will be affected by changes in yields which are caused by factors such as agricultural developments or climate change, and can face an increasing food demand due to a growing population as predicted for West Africa. In addition to the fixed yield and population scenarios, we therefore consider a yield scenario including trends in mean yields given by historic data, and three different population scenarios, i.e. a low, medium, and high fertility projection of population numbers. All combinations of yield and population scenarios were run for the default settings and a set of different food security and solvency probabilities $\alpha_F$ and $\alpha_S$. Resulting crop allocations are visualized in Figure 4.8 for $\alpha_F = 95\%$ and $\alpha_S = 85\%$. In Figure 4.8 (A), the results of different population scenarios for fixed yield distributions are shown, while for Figure 4.8 (B) trends in yield distributions are included.

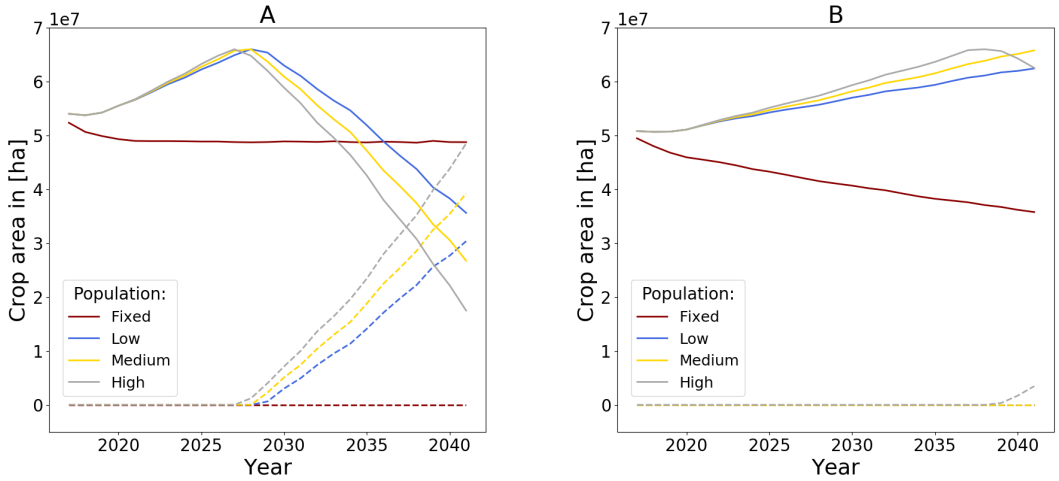



Figure 4.8.: Yearly crop area allocations for maize (solid line) and rice (dashed line) for different combinations of yield and population scenarios with $\alpha_F = 95\%$, $\alpha_S = 85\%$, and all other settings at default values. In (A) fixed yield distributions are used and in (B) yield trends are included. Different line colors correspond to different population scenarios. Note that rice allocations in (B) for the fixed, low and medium population scenarios are constant at 0 ha, such that these dashed lines coincide.

For fixed yield distributions and fixed population size, the crop allocation shows the behavior described in the Section 4.2, starting at a higher value and then decreasing to a constant level (dark red line in Figure 4.8 (A)). If the population is kept constant but yield trends are included, the crop area over time linearly decreases in the later years, as

yields are expected to increase over time (dark red line in Figure 4.8 (B)). The different population predictions all use estimates of the actual population size until 2020 and only differ starting from 2021, leading to these scenarios showing very similar behavior in the first years. The fixed population scenario shows a slightly different crop allocation even in the first year, as fixed population size and yield distributions are taken from 2016 because this is the last year covered by the historic yield data set, while projections start in 2017 which is the first year covered by the model.

Starting from 2021, the crop allocations for different population predictions start to diverge, with crop allocation for the high fertility scenario increasing the fastest. After some years, the limit of arable land is reached. While maize is cheaper in terms of production in kcal/$, it also has lower yields per area. For rice the expected yields are $9.15 \cdot 10^6$ kcal/ha and for maize only $5.92 \cdot 10^6$ kcal/ha. Hence, once the area of maize can no longer be increased to match the increasing food demand, maize starts to be substituted by rice even though rice is more expensive to cultivate. Increasing yields counteract the influence of population growth. Therefore, the effects are reduced in Figure 4.8 (B) compared to Figure 4.8 (A), they are however still present. This shows that current trends in crop yields in West Africa cannot keep up with the predicted increase in population.

## 4.5. Impact of reduced spatial correlation on sustainable food security

By using just one cluster for the whole region, we assume full spatial correlation of crop yields over the whole area. This was used to understand the influence of model parameters such as the government instruments to regulate the insurance scheme, but realistically yields will not be correlated over such long distances. The influence of spatial correlation is one of the main points of interest in this project, as pooling of uncorrelated risks can reduce the overall severity of local extreme events and improve the performance of the insurance scheme.

We first compare the default run for $K = 1$ with the model outcome for $K = 2$ while keeping all other settings constant. Figure 4.9 shows the respective crop area allocations, including the total cultivated area and maximum available arable land per cluster. The most apparent change in crop dynamics when considering two clusters is that area is allocated to both crops in the first cluster, as shown in Figure 4.9 (B). This is not a direct effect of the reduced spatial correlation, but due to the different crop yield distributions. In Section 4.2, it was already indicated that this is possible since both crops can show different performance regarding the generation of profits to build up the fund, and the production of the necessary calories to meet the food demand. We saw that for $K = 1$ maize is superior to rice with respect to both food security and solvency of the government fund. Analogous to calculations in Section 4.2, the performance of

each crop in each cluster for $K = 2$ can be quantified by the expected values given in Table 4.2.

| | Crop | profit per invested \$ (in [\$]) | production per invested \$ (in [kcal]) |
|---|---|---|---|
| Cluster 1 | Rice | 0.51 | 15 820.99 |
| | Maize | 0.19 | 16 348.43 |
| Cluster 2 | Rice | 0.20 | 12 557.05 |
| | Maize | 0.89 | 25 864.18 |

Table 4.2.: Performance of each crop in each cluster for $K = 2$, quantified by expected net profit per investment in [\$/\$] regarding the solvency constraint and by expected production per investment in [kcal/\$] regarding the food security constraint.

Here, maize in cluster 2 is the superior choice over all crops in all clusters regarding both objectives, and is therefore used for the full available area in cluster 2, as shown in Figure 4.9 (B). In cluster 1 however, rice is superior to maize in building up the fund to meet the solvency constraint, while maize has a slightly better result regarding calorie production to meet the food security constraint. This explains the use of both crops within cluster 1 and the change of dominating crop according to the shifting dominance of the two constraints.

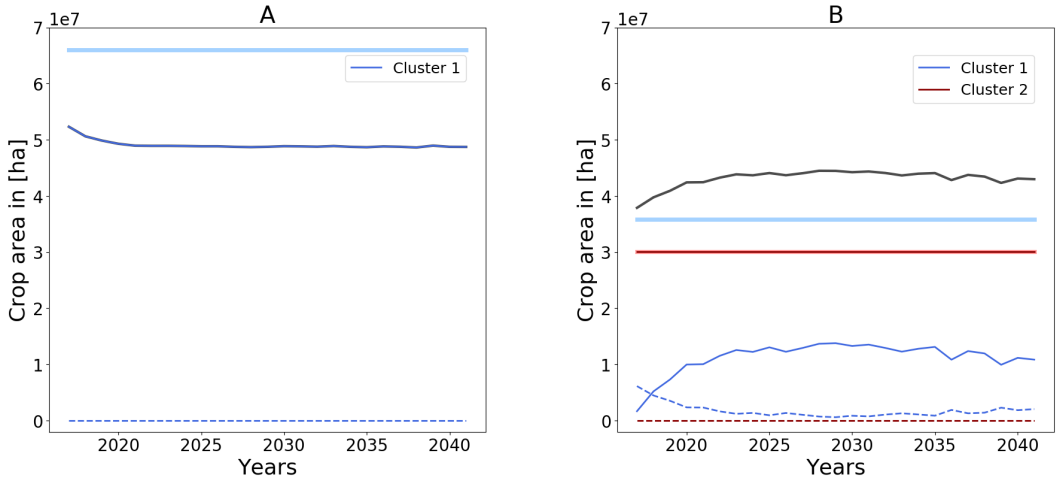



Figure 4.9.: Yearly crop area allocations per cluster (dark blue and dark red) for maize (solid lines) and rice (dashed lines). The maximum arable land per cluster is included as thicker lines in light blue and red respectively. The black line is the aggregated cultivated land for all clusters and crops. In (A), default settings with $K = 1$ and probabilities $\alpha_F = 95\%$ and $\alpha_S = 85\%$ are used. (B) uses $K = 2$ while keeping the other settings the same. Note, that in (A) the black and the solid dark blue lines coincide, while in (B) the bright red and solid dark red lines coincide.

Following this argumentation, we could expect that only maize will be used once the security penalty is dominating the dynamics. However, the area used for cultivation of rice in cluster 1 levels at a very low but non-zero value. This can be explained by the characteristics of the yield distributions as well: In the case of a catastrophe, yields are in the lower 5%-quantile according to the covered risk $r$. The standard deviation relative to the mean of the distribution in cluster 1 is lower for rice yields than for maize yields, hence the relative impact of catastrophic yields is lower for rice. The resulting

expected production in case of a catastrophe is 13 648.72 kcal/$ for rice and 13 801.34 kcal/$ for maize. Hence, maize is still slightly superior in expectation, but there will also be some realizations where rice is the better choice regarding the food security constraint, which explains the small share of rice in cluster 1 in the later years. It does however not explain why this share increases again in the last years. This is assumed to be a numerical effect due to accuracy limits of the model, as this feature is more apparent when using a smaller sample size $N$.

The reduction in spatial correlation of crop yields leads to a decrease in total cultivated area, also shown in Figure 4.9. Considering two clusters instead of one, when one cluster exhibits catastrophic yields, the other can still be non-catastrophic, thus lessening the overall impact of a catastrophe. This allows to reach the same probabilities $\alpha_F$ and $\alpha_S$ with a smaller cultivated area. Therefore, the total cultivation costs over the full time period for $K = 1$ are $3.45 \cdot 10^{11}$\$, while they are only $3.19 \cdot 10^{11}$\$ for $K = 2$. Consequently, a collaborative approach to food security aggregating demand and production over areas with uncorrelated yields can reduce the joint requirement for resources such as agricultural area.

Increasing the number of clusters also changes the dynamics of total cultivated area over time shown in Figure 4.9. This is another effect of the substitution between rice and maize in cluster 1 for $K = 2$. Considering only one cluster, in order to have higher profits to build up the fund, a larger area needs to be cultivated in the beginning. Using two clusters, the necessary profit can be gained by cultivating rice instead of maize. As rice is not only more profitable, but also has higher yields per hectare, the total cultivated area is actually smaller in the first years. Despite the higher yields, production in terms of calories in cluster 1 is more expensive using rice due to higher cultivation costs. Therefore, with the fund building up, the share of maize is increased, even though this increases the total cultivated area.

The behavior regarding the constraints differs as well when considering two clusters instead of one. Figure 4.10 (A) visualizes the respective yearly average shortcomings of the food demand with fluctuations resulting from computational restrictions. For $K = 1$, the solvency constraint leads to higher production in the first years to build up the fund as described in Section 4.2, resulting in much lower average shortcomings of the food demand in the beginning, that then flatten out to a level of about $3.31 \cdot 10^{11}$ kcal. Since for $K = 2$ the fund is built up by increasing the share of rice instead of a higher production in general, the average shortcomings fluctuate around a constant level of $2.86 \cdot 10^{11}$ kcal for the whole time period. This is a positive effect of reduced spatial correlation, as a stable approach to food security over time is beneficial. The lower level of shortcomings for $K = 2$ relates to a higher food security penalty $\rho_F = 1.33 \cdot 10^{-3}$ \$/kcal compared to $\rho_F = 1.0 \cdot 10^{-3}$ \$/kcal for $K = 1$, as in general if the penalty is higher, only a smaller violation of the constraint is acceptable. The higher food security penalty for $K = 2$ is a result of the higher probability for catastrophic years. For a covered risk of $r = 5\%$, the probability of a catastrophe considering just one cluster is 5%, while for

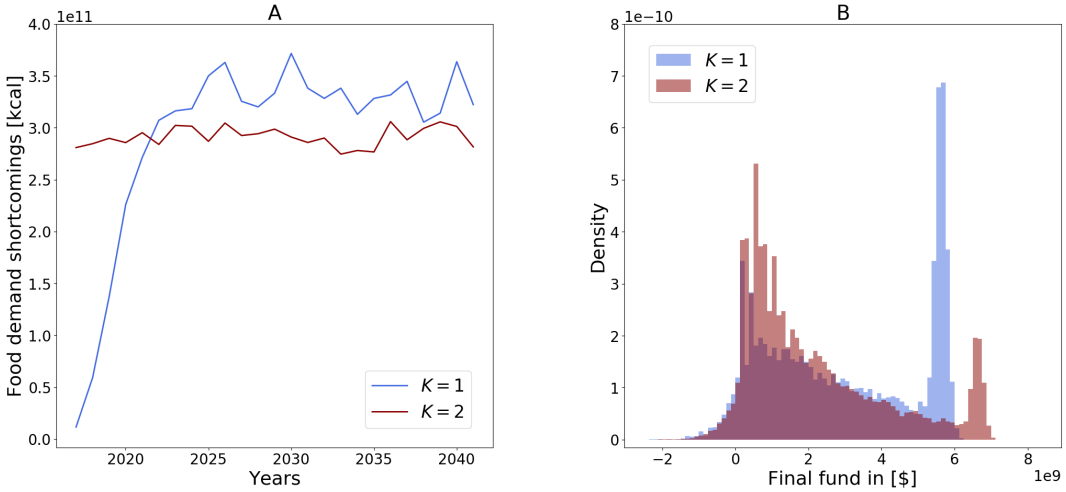



Figure 4.10.: In (A) the yearly average shortcomings of the food demand are shown, while (B) visualizes the distribution of the final fund after catastrophe given by the different yield realizations. Dark blue and dark red represent the respective model results for $K = 1$ and $K = 2$ respectively.

$K = 2$ it is 9.75%, as either cluster (or both) can be catastrophic. Therefore, the food demand needs to be met for a larger share of realizations with catastrophic years when considering two clusters. For this, higher investments in crop production are needed, which are only realized if the penalty $\rho_F$ for violation of the food security constraint and thereby the incentive to invest in the prevention of this case is high. This effect is to some extent counteracted by the fact that it is easier to meet the food demand despite a catastrophe for $K = 2$ than for $K = 1$, as for $K = 2$ in most cases only a part of West Africa exhibits a catastrophe, i.e. only one of the two clusters.

Figure 4.10 (B) shows the distribution of the final fund after a catastrophe for $K = 1$ and $K = 2$ respectively. The main difference between the two distributions is the shift of density between the two modes. Since for $K = 1$ the probability of a catastrophic year is 5%, a much higher share of the realizations does not exhibit a catastrophe within the 25 considered years than for $K = 2$, which has a probability of 9.75% for a year to be catastrophic. Realizations without a catastrophe correspond to the upper mode in the distribution, which therefore has a lower density for $K = 2$. This is another positive effect of a joint insurance scheme over uncorrelated areas, as there is less accumulation of unused taxes in the fund, thereby putting less unnecessary pressure on farmers and meeting the solvency constraint more efficiently. Analogous to Section 4.3, the share of negative values in the distribution describes the probability of insolvency after a catastrophe. Even though both versions were run for $\alpha_S = 85\%$, this share differs. This is due to the separate calculation of food security penalty $\rho_F$ and solvency penalty $\rho_S$ and the interaction of the constraints when including both. For $K = 1$ the resulting probability for solvency is $\alpha_S = 95.81\%$, for $K = 2$ it is $\alpha_S = 96.79\%$. In both cases, the probability of solvency is increased a lot by the introduction of the food security penalty, with the effect being slightly higher for $K = 2$, as the food security penalty

for $K = 2$ is higher. The solvency penalty however is lower for $K = 2$, being only $\rho_S = 137.5\,\$/\$$ for $K = 2$ versus $\rho_S = 194\,\$/\$$ for $K = 1$. This reflects that for reduced spatial correlation the government can more easily stay solvent, as they only need to pay the share of the total guaranteed income corresponding to the catastrophic cluster, and thus a lower investment in crop cultivation is already sufficient to assure a certain probability of solvency $\alpha_S$.

Increasing the number of considered clusters further intensifies the effects discussed for the case $K = 2$. The model was run for default settings with $K = 7$, as this is the optimal number of clusters according to the cluster analysis in Section 1.2.2. Assuming that the insurance scheme covers seven uncorrelated areas, government payouts in case of catastrophe are even lower, as in most cases only one or two of the clusters exhibit catastrophic yields simultaneously. In addition, the fund is not only built up in the years before a catastrophe, but non-catastrophic clusters still pay taxes in the catastrophic year. This makes it much easier for the government to stay solvent, hence a lower solvency penalty of $\rho_S = 75\,\$/\$$ is already enough incentive to reach the requested probability of $\alpha_S = 85\%$.

For the food security penalty we discussed two counteracting effects in the case of $K = 2$. On the one hand, the consequences of a catastrophe are reduced for higher $K$, as only a share of the total region of West Africa exhibits catastrophic yields, which makes it easier to reach the food demand despite a catastrophe. On the other hand, the probability of some clusters exhibiting catastrophic yields is higher when considering a higher number of cluster $K$, that each have an independent probability of catastrophic yields of $r = 5\%$. For $K = 2$, the latter of the two arguments dominated, such that the food demand penalty $\rho_F$ was slightly higher for $K = 2$ than for $K = 1$. However, for $K = 7$, the reduced consequences of a catastrophe have a stronger influence, such that the food security penalty for $K = 7$ is $\rho_F = 8.6 \cdot 10^{-4}\,\$/\text{kcal}$, actually being lower than for $K = 1$ which had a penalty of $\rho_F = 1 \cdot 10^{-3}\,\$/\text{kcal}$.

Figure 4.11 (A) shows the crop allocations for $K = 7$ clusters, including the maximum available arable area per cluster and the aggregated cultivated area for all clusters and crops. In (B), crop allocations are shown in separate plots for each cluster to allow the interpretation of individual dynamics. Again, the dynamics of crop areas over time can be explained by analyzing the expected performance of each crop in each cluster regarding profit generation and food supply. Maize in cluster 1 and 5 and rice in cluster 4 are the most profitable crops, with net profits of over 1$ per invested dollar, followed by maize in cluster 2 with 0.71$ net profit per invested dollar. Therefore, these are the crops used in the earlier years to build up the fund. As a side effect, some clusters are not used at all in the first years. This is possible due to a combination of different model properties: The food security constraint is applied to the whole area of West Africa by using an aggregated food demand rather than regional values, such that there is no direct incentive to spread production over many clusters from the food security point of view. Regarding the solvency constraint, two aspects influence the crop allocation.

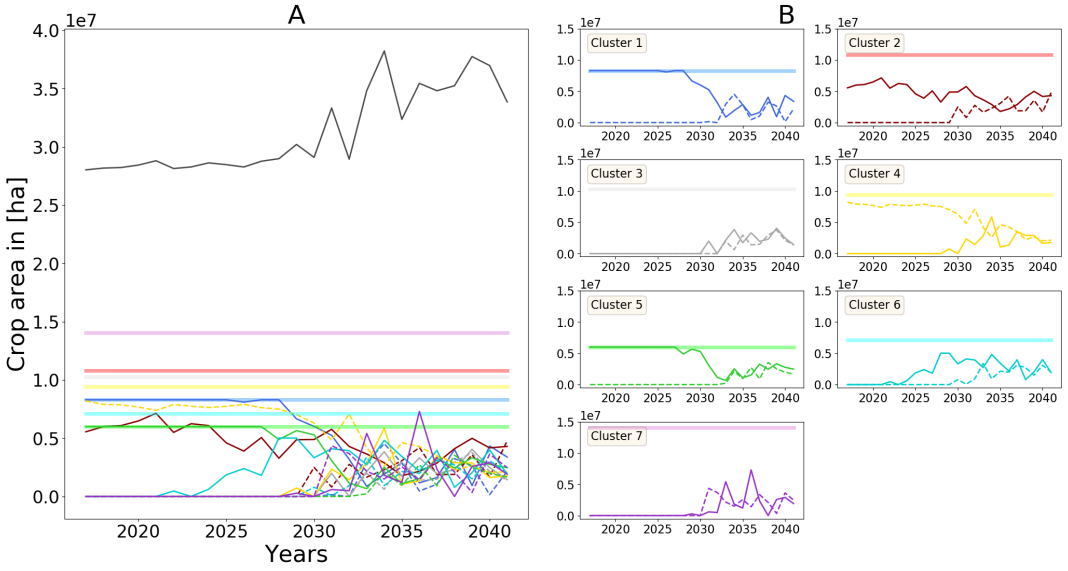



Figure 4.11.: Results for $K = 7$ with other parameters kept at default values. In (A), yearly crop area allocations per cluster for maize (solid lines) and rice (dashed lines) are shown. Different colors correspond to different clusters. The maximum arable land per cluster is included as thicker lines in a lighter shade of the respective color. The black line is the aggregated cultivated land for all clusters and crops. In (B), the same crop area allocations are shown, but in separate plots for easier interpretation. Here, labels of the axes are omitted due to limited available space, but as in (A) the x-axis shows the years, while the y-axis shows crop areas in [ha].

Mainly, the fund needs to be built up, which leads to the use of the most profitable crops in the beginning. However, payouts can also be reduced by assuring that some profit is earned even in the case of a catastrophe, as the government only needs to pay the difference between profits and the guaranteed income. Some clusters have no or only very little guaranteed income, as this is calculated as a share of the expected income in a scenario including only the food security objective which does not distinguish between the different clusters. Therefore, the last argument of crop allocation to reduce payouts in case of a catastrophe does not apply to these clusters. The interpretation of some clusters having very low guaranteed income is that the insurance scheme focuses on the areas with better performance. Aggregated over the whole area this will lead to a better trade-off between costs and benefits.

While for $K = 1$ and $K = 2$ the transition between dominating solvency penalty in the beginning and dominating food security penalty once the fund is built up happens quite early, considering seven clusters there is a longer period in which mainly the afore mentioned more profitable crops are cultivated. This might be, since if only one cluster exhibits catastrophic yields, government payouts are quite low and a small fund size is sufficient to stay solvent. Most catastrophes are covered by this case, therefore the share of realizations at risk of insolvency in the first years is lower than for smaller $K$. As only a probability of $\alpha_S = 85\%$ for solvency is demanded, the remaining 15% can thus be spread over a longer period of time, such that the fund does not need to be built up as fast as for smaller $K$. However, it is possible that multiple clusters are catastrophic

simultaneously, such that over time the fund will still be built up to a higher level to cover these cases as well, and therefore the solvency penalty $\rho_S$ is dominating over a longer period of time.

After the profit-oriented phase, the remaining clusters and crops start to be used as well. Similar to the case $K = 2$, this is linked to an increase in total cultivated area, as the crops that are cheaper for production in terms of calories generally also have lower yields per hectare. A detailed analysis of the dynamics of crop area allocation in the last years is not possible, as the sample size for those years is too low. For $K = 7$, the probability of a year being catastrophic is about 30%, such that most realizations already exhibit a catastrophe, and thus terminate, in the early years and only about 0.5% of the realizations cover more than the first 15 years. Hence, even for a high sample size as $N = 50\,000$, less than 250 realizations remain after the year 2031 when considering the default time period, and only 10 are expected to reach the last year. A higher sample size was not possible due to computational restrictions.

## 4.6. Mitigating the impact of yield and population projections by reduced spatial correlation

As last model run, we combine the optimal number of clusters $K = 7$ given by the cluster analysis in Section 1.2.2 with yield distributions following trends in mean yields and the medium fertility population projection and thereby increasing food demand. This allows to analyze one of the most realistic scenarios possible in the scope of the given stochastic optimization model. We compare this run to the results for $K = 7$ with default settings, to understand how changing conditions over time can influence the approach to sustainable food security implemented by the model.

Similar to Figure 4.11, in Figure 4.12 (A) crop area allocations for each of the seven clusters are visualized, including the maximum available arable area per cluster and the aggregated cultivated area for all clusters and crops. In (B), crop allocations are shown in separate plots for each cluster to allow the interpretation of individual dynamics. Again, the exact behavior in the final years of the considered time period has to be seen with caution due to a reduced sample size. However, some general trends resulting from changing yield distributions and population numbers over time can be deduced.

The more profitable crops, i.e. rice in cluster 4 and maize in cluster 1, 2, and 5, are dominant over a longer period of time and are than reduced to a lower level more slowly than in the default run for $K = 7$, as the guaranteed income level increases over time and thus the fund needs to be built up higher to reach the same probability of solvency $\alpha_S$. As the dominance is shifted from the solvency penalty to the food security penalty, crop area allocations assume higher levels as for the default run and seem to increase over the last years to keep up with the increasing food demand. This can also be seen when considering the total cultivated area shown in Figure 4.12 (A) and comparing it to the

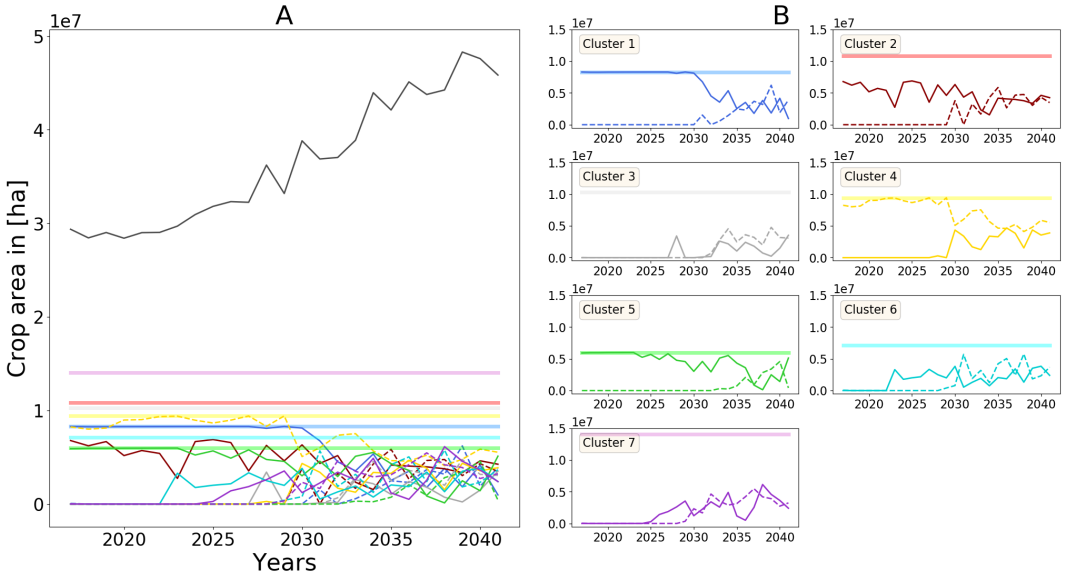



Figure 4.12.: Results for $K = 7$, including yield trends and using the medium fertility population scenario, with other settings kept at default values. In (A), yearly crop area allocations per cluster for maize (solid lines) and rice (dashed lines) are shown. Different colors correspond to different clusters. The maximum arable land per cluster is included as thicker lines in a lighter shade of the respective color. The black line is the aggregated cultivated land for all clusters and crops. In (B), the same crop area allocations are shown, but in separate plots for easier interpretation. Here, labels of the axes are omitted due to limited available space, but as in (A) the x-axis shows the years, while the y-axis shows crop areas in [ha].

respective values of the default run for $K = 7$ shown in Figure 4.11 (A). While in both cases the total cultivated area is at a similar level of slightly less than $3 \cdot 10^7$ ha in the beginning, for the case including yield trends and population projections this value soon starts to increase over time. For the default case however it keeps fluctuating around a constant level for ten years, before starting to increase. It then increases during the transition phase between the dominance of the solvency penalty and the food security penalty, and levels again once the food security penalty dominates the dynamics. When including yield trends and population projections, we do not see a leveling of total cultivated area in the final years, as the increasing food demand resulting from an increasing population needs to be met with higher agricultural production. Thereby, the total area cultivated in the final years in this case is much higher than for the default case which excludes yield or population changes. In particular, this also means that the current trend in yields cannot keep up with the speed of population growth, as otherwise the same or even a smaller area should be sufficient to reach food security in the final years of the considered period. However, in the case of reduced spatial correlation, the total allocated crop areas do not reach the limit of available arable land within the considered time period. In contrast, runs in Section 4.4 using yield and population projections for $K = 1$ lead to a substitution in crops as the total available area was not sufficient relying only on the crop showing the better performance in terms of calories per invested money.

## 4.7. Performance of the stochastic model compared to a deterministic alternative

In Section 2.1, the usage of a stochastic optimization model in the context of sustainable food security was motivated. It was explained why deterministic approaches using scenario analysis will give suboptimal results when applied to real situations, as they omit variability in uncertain parameters. In Section 3.3, a method to quantify the benefit of using a stochastic optimization model over a deterministic model relying on the expected value of uncertain parameters as model input is explained. We now apply this method and calculate the value of stochastic solution for the sustainable food security model.

In the setting of sustainable food security which is subject to uncertain crop yields, the deterministic approach uses average yields to determine optimal crop area allocations. Resulting values in comparison to results of the stochastic model are visualized in Figure 4.13 for $K = 7$, using yield distributions including trends in mean yields and the medium fertility population projections. As the average yields in general are non-catastrophic, the deterministic approach assumes that no government payouts need to be carried out and thereby neglects the potential of a large scale insurance system. Applying crop area allocations from the deterministic model to crop yield realizations drawn from the full yield distributions will lead to payouts and, with the fund not being built up accordingly, result in potentially high solvency penalties. As crop areas are allocated to exactly match the food demand assuming average yields, for yield realizations drawn from the full distributions the food demand will in many cases not be met, thereby leading to high food security penalties as well.

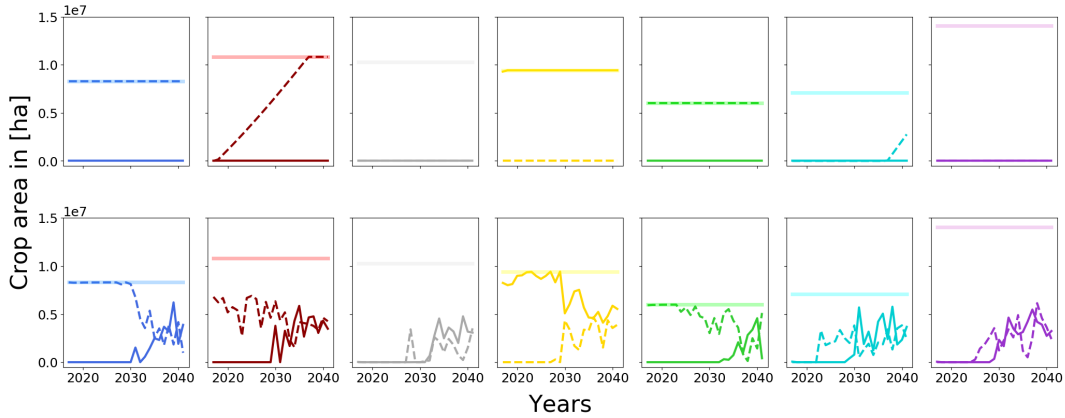



Figure 4.13.: Results using $K = 7$, including trends in mean yields and medium fertility population projections, while keeping other settings at default value. Yearly crop area allocations per cluster for maize (solid lines) and rice (dashed lines) are shown. The maximum arable land per cluster is included as thicker lines in a lighter shade of the respective color. The upper row shows results of the deterministic model relying on mean yield values as input, while the lower row shows results of the stochastic model.

Considering seven clusters and using yield distributions including trends in mean yields and the medium fertility population projections, the expected total costs including cultivation costs, food security penalties, and solvency penalties is $3.88 \cdot 10^{10}$ \$ when applying crop area allocations resulting from the stochastic optimization model. When using crop area allocations resulting from the deterministic model instead, the total expected costs are $4.44 \cdot 10^{10}$ \$. Hence, the value of stochastic solution is $5.6 \cdot 10^{9}$ \$, being 14.3% of the expected costs of the robust solution. This percentage increases when considering a smaller number of clusters, being 33.4% for the default run using just one cluster, because costs increase more using the deterministic solution than using the robust solution. This underlines the potential of risk pooling through covering large areas with uncorrelated yield distributions. It allows to cultivate a smaller surplus area compared to the area needed under expected yields, while still meeting the constraints with certain probabilities.

# 5. Discussion

There are many models that cover the allocation of crop areas, agricultural production, and related economic, environmental or social aspects, such as the global land-use allocation model MAgPIE (Model of Agricultural Production and its Impact on the Environment) [75], the global agricultural sector model CAPRI (Common Agricultural Policy Regionalised Impact Modelling System) [76], or the global economic land-use model GLOBIOM (Global Biosphere Management Model) [77]. These are deterministic models based on scenario analysis. While they can give detailed information for different scenarios, they do not include variability in input parameters. This however is a central aspect when analyzing risk management, and in the context of food security, the uncertainty of crop yields can have a significant impact on strategies followed by farmers and the government. This is e.g. acknowledged by Fuss et al., 2011 [78], who apply a stochastic version of GLOBIOM [79] to investigate global food security under crop yield uncertainty. However, risks related to agricultural production also affect the sustainability of food security. Especially in developing regions as West Africa, farmers often lack the resources to get through an extreme event without losing their livelihoods [8]. Hence, resilience against extreme events is necessary to prevent farmers from migrating or changing to activities different from farming, further jeopardizing food security [8].

This project addresses this aspect by developing a stylized stochastic model for sustainable food security in West Africa, which is described in Chapter 3. The availability dimension of food security is implemented in the model through a certain food demand that needs to be met. This food demand is given in terms of calories aggregated for the whole considered area. The sustainability dimension of food security is included by a large scale insurance scheme financed through taxes on agricultural profits to make farmers resilient against extreme events. Shortcomings of the food demand as well as insolvency of the catastrophe fund lead to second stage costs. The model informs on optimal crop area allocations to minimize the sum of first stage crop cultivation costs and second stage costs. To this end, it assumes the government's point of view, without taking into account the interests of individual farmers per se. The optimal crop allocation resulting from the model is prescriptive, and can be used by the government to set up policy schemes in order to work towards sustainable food security.

The central question is to what extent implementing a large scale insurance scheme to achieve sustainable food security in West Africa is feasible under the constraints of available arable area, agricultural production and financial resources. Related aspects that need to be analyzed are the influence of the design of the insurance scheme and

the trade-off between high probabilities of meeting the food demand and maintaining solvency of the catastrophe fund, and necessary first stage investments to achieve these probabilities. To improve the performance of the insurance scheme, risk pooling can be applied, which we included by subdividing West Africa into subregions, referred to as clusters, that feature low risk correlation between each other. These clusters are determined based on the drought index SPEI, and we assume the clusters to have an independent risk of catastrophic yields. In this context, it is interesting to what extent risk pooling can increase the feasibility of achieving sustainable food security under the model assumptions. Over time, food security will also be affected by environmental, technological, and demographic developments. We therefore included different yield and population projections to investigate the influence of these developments on sustainable food security.

The developed model can be considered as a proof-of-concept for a sustainable food security system. It offers a stylized view on the complex topic of food security and its sustainability, excluding aspects such as investments in infrastructure or storage of surplus food production. Partially, this project was focused on theoretical analysis, model development, and understanding the influence of parameter variations on the model output. At the same time, the results described in Chapter 4 allow a broad range of observations to discuss the above mentioned research topics. In Section 5.1 we address feasibility of the implemented approach to sustainable food security by discussing the necessary agricultural area and arising first and second stage costs, as well as indicating a possible usage of model outputs by the government. In Section 5.2 we review the trade-off between solvency of the insurance scheme and usefulness for farmers arising through policy interventions in the financing mechanism. We then discuss the potential of risk pooling in the setting of sustainable food security in Section 5.3, and finally consider the influence of developments in population growth and agricultural productivity in Section 5.4.

## 5.1. Feasibility of sustainable food security and possible use of model results

In Section 4.2, we showed that even if the risk of poor harvests were correlated over the whole of West Africa, which in our model is implemented as a single cluster with fully dependent yields, the available agricultural area is sufficient to reach food security and solvency of the government fund with high probabilities. While this is a requirement for feasibility of the implemented approach to sustainable food security, the costs arising both from crop cultivation and from dealing with negative consequences when the objectives cannot be met are a second aspect to consider. In general, there is a trade-off between risk prevention by higher first stage investments and post-disaster payments in case the objectives are not met [16]. If the first stage costs as well as penalties representing the negative consequences of shortcomings of the objectives are known, the model

can inform on the best balance between first stage strategic decisions and second stage adaptive actions, resulting in a certain probability of meeting the constraints. However, post-disaster actions can comprise direct costs such as interests on loans or food import, but also indirect costs due to socio-economic consequences of food shortages or loss of resilience if the guaranteed income can no longer be provided by the government. As the exact consequences of violation of constraints cannot be quantified, assessment of the trade-off between high first stage costs and potential second stage penalties and thereby the feasibility of sustainable food security under the model assumptions is only possible on a conceptual level.

Regarding food security, aiming for a higher probability to meet the corresponding constraint increases the required crop area allocations in a consistent manner over time to account for the possibility of low yields. This means that in each year the first stage investments are higher by a certain amount, balancing higher second stage penalties. High solvency probabilities require the fund to build up quickly, in order to ensure solvency also in the case of an extreme event in the first years after setting up the insurance scheme. This needs high overproduction in the first years compared to the level necessary for food security. The overproduction implies high first stage investments in crop cultivation, which within the model are only justified for severe negative consequences of insolvency, represented by high penalties in case of violation of the solvency constraint. A direct way to decrease overproduction in the model without compromising the probability of solvency is to initialize the insurance scheme with a positive fund, such that the building phase of the fund falls away. This would however be connected to additional costs for the government when setting up the insurance scheme as well as after each catastrophe that leaves the fund with with less money than the minimal amount needed to avoid overproduction. As we did not cover changing initial fund sizes in the model analysis, we use a default probability for solvency of 85% to limit overproduction in the first years. This can be feasible if the negative impact of insolvency is not too severe, e.g. if the government can take a loan and the only resulting additional costs are interests that have to be paid. However, if the actual negative consequences of insolvency are higher than described by the penalty corresponding to a probability of 85%, e.g. due to socio-economic consequences when the government is unable to provide the guaranteed income, this will lead to even higher and potentially infeasible second stage costs.

While within the model the overproduction leads to the fund being built up through higher profits and corresponding higher tax revenues, the actual implementation of this might not be realizable. On the one hand, farm-gate prices could decrease in case of overproduction, reducing the effect of higher profits and taxes, and on the other hand, starting with high crop areas but then quickly diminishing these areas might not be in the personal interests of farmers. Resulting crop area allocations can be understood as reference values to reach sustainable food security, which can be used by the government in the process of setting up policies, but these crop areas in general would not follow from individual farmer decisions. A possible government approach that reflects the

model structure could be to each year announce the amount of area for each crop and region that will be covered by the catastrophe fund, and only include farmers in the insurance scheme which participate in cultivating these areas. In that case, the incentive for farmers to conform with the crop area allocations resulting from the optimization model would be to increase their resilience against bad harvests. Of course, this is only a basic concept to give an example of possible policies, and leaves many questions open: How can farmers be compensated if they took part in the insurance scheme for several years, but then are left out due to reduction of crop area allocation in their area without ever directly profiting from the insurance scheme? How would the necessary crop allocation be distributed in areas with more farmers interested in the insurance scheme than needed to realize the output of the optimization model? How would crop allocations be filled up in areas with less participating farmers than needed? In case of policies including additional costs such as subsidies, these could have feedbacks on the optimal solution and would also need to be taken into account in the decision making progress.

## 5.2. Trade-off between fund solvency and farmers' resilience: policy options of the government

The government has different possibilities to influence the feasibility of the insurance scheme. In the model, three options are included: the tax rate according to which farmers have to pay a share of their profits to the government fund, the risk level that is covered in order to increase farmers' resilience against extreme events, and the level of guaranteed income in case of bad harvests covered by the insurance. As the model takes the government's point of view, an increase in tax rate or decrease in guaranteed income will obviously improve feasibility of the insurance scheme, as shown in Section 4.3. However the higher the tax rates and the lower the guaranteed income, the lower also the positive effect for farmers. If tax rates are too high, they will not be able to afford this type of insurance, while low guaranteed income will not protect their livelihoods in case of an extreme event. Therefore, a balance between feasibility of the insurance scheme and usefulness for farmers needs to be found.

We observed that when considering West Africa as a single cluster, there was a significant positive effect for solvency when increasing the tax rate from 1% to 3%. However, the additional improvements when increasing the tax rate further to 5% were modest, with a large amount of taxes accumulated in the fund but not needed for payouts and thereby putting unnecessary pressure on farmers. The same kind of trade-off between feasibility for the government and resilience provided to farmers can also be expected when changing the risk level covered by the insurance scheme. With a higher percentage of bad harvests covered, total payouts from the government fund to farmers are higher. This is not directly visible in the model results, as the expected payouts for a single extreme event are lower due to a reduced average severity, and the model only considers

the years up to, and including, the first catastrophe. However, the frequency of payouts increases with higher risk levels covered by the insurance scheme.

## 5.3. The concept of risk pooling and its potential for sustainable food security

Government interventions to the financing mechanism can only be used within limits to increase feasibility of the insurance scheme without degrading the original purpose of adding sustainability to food security by offering farmers resilience against extreme events. However, risk pooling is a way to improve performance both regarding food security and the solvency of the insurance scheme from the government's point of view without compromising farmers' benefits. The general concept of risk pooling is to reduce the relative severity of extreme events by insuring larger areas or groups of individuals with uncorrelated risks in a centralized form. In analogy to an example given by Ermolieva et al., 2016 [80], imagine 100 individuals incurring a risk of 10% of an event that will cost them 10$. As it is not known when these events will happen, an insurance company that covers this risk at all times needs to have 10$ of capital saved for each individual. Aggregated, this would lead to 1000$ of savings. However, if we assume that the risk for each individual is independent of the others and then look at the joined risk, the probability of needing 1000$ at once is $0.1^{100}$, which is extremely low. Already the probability of 20 or more individuals facing the extreme event at the same time is only 0.2%. Hence, savings of only 200$ would be enough to cover potential consequences for all individuals in almost every case.

This can be applied to the setting of sustainable food security. If the approach covers a large area with sub-regions of different agro-climatic conditions and therefore low risk correlation, a centralized approach will not need to prepare for catastrophic yields happening in all regions simultaneously. This however would amount to each region preparing separately. Thereby, risk pooling allows for a high probability of solvency with a lower fund size than the sum of separate funds would need to be. In particular, very high probabilities of solvency of e.g. 99%, which might be required by the provider of the insurance, will only be feasible by applying a centralized supranational approach. Risk pooling also positively affects the food security objective of the sustainable food security model. By using an aggregated food demand, if one cluster exhibits poor yields, the others can support it by supplying potential surplus food production.

The developed model does not consider different sizes of covered area or independent regions covered by separate insurance schemes for sustainable food security. The implemented spatial structure based on a variable number of independent clusters however still allows to analyze the potential effect of risk pooling. Full spatial correlation of yields in West Africa as given by considering only a single cluster is obviously a wrong

assumption, but if we consider an average subregion for which the assumption of correlated yields is reasonable, this can be expected to behave similar to West Africa understood as a single region. Of course, each subregion has different characteristics and yield distributions which will therefore differ from the average values. Nonetheless, results of running the stochastic optimization model for West Africa considering only one cluster can be seen as a proxy for the case in which each subregion implements its own insurance scheme for sustainable food security. Therefore, we compare the model output considering multiple clusters and the model output considering a single cluster to understand the potential of a centralized approach with risk pooling compared to each cluster implementing its own insurance scheme for sustainable food security.

In Section 4.5, results of the centralized approach including risk pooling are analyzed. Regarding the financing scheme, the direct effect is that the fund does not need to be built up as high, because expected payouts are lower. In the considered settings, this resulted in lower crop area allocations, but the positive effect could also be used to instead reduce tax rates or increase the guaranteed income level, in order to further improve farmers' resilience against extreme events. Also regarding the food security objective, the required crop area allocations are reduced when including risk pooling. This increases the feasibility of food security over a large area by allowing the same level of security with lower investments. However, the assumption of a joint food demand for all of West Africa has some implications that should be considered. Allowing shortages in one cluster to be directly counterbalanced by higher production in other clusters neglects potential transportation costs that would arise, and assumes free trade between the different areas. International trade can be one adaption method to climate change induced increase in hunger around the world, but trade barriers and tariffs diminish its potential[81]. Within the model, the joint food demand can even lead to some clusters not being used at all, especially in the first years when the focus is on the more profitable crops to build up the fund. While this is a way to reduce overall costs within the model, realistically this makes only sense to some extent. Even though some areas e.g. close to the desert might actually not be used much for agriculture, in areas with low expected yields there still might be farmers whose personal interest will be to cultivate land, even if seen over the whole region this is not the best choice.

When considering a higher number of clusters, we also saw a substitution of the two crops in the first years after initialization of the financing scheme. This was however not a direct effect of risk pooling, but due to the more diverse yield distributions considering more clusters. The effect would therefore also be seen in some clusters when assuming that each cluster implements its own insurance scheme for sustainable food security. The general potential of crop diversity is covered in a limited way in the model due to focusing on only two crops. Furthermore, by including food security only as a caloric demand, the model does not address the need for diversity to allow for healthy and nutritious diets[82].

## 5.4. Sustainable food security under environmental, technological, and demographic developments

While improvements in agricultural technology and management practices can increase crop yields over time, this effect is counteracted by negative effects due to climate change. Already now, losses in crop production due to climate change can be identified[83]. In addition, West Africa faces the fastest growing population in the world, further increasing difficulties of achieving food security[7]. These effects were included in our analysis by a yield scenario with distributions following a historic trend in average yields and different population scenarios for West Africa given by the UN. These scenarios were considered both using just one cluster for the full region of West Africa and including risk pooling. In both cases we see an increase in total cultivated area over time, showing that even for the low fertility population predictions the current trends in average yields cannot keep up with the increase in food demand. Without risk pooling, the rising crop area allocations are predicted to reach the limit of available area around 2040 depending on the population scenario. Food production within the model then has to rely increasingly on the more expensive crop which has a higher yield per area. Another possibility could be to expand agricultural area or invest in increased productivity, however these options are not included in the model. Agricultural area in West Africa is still predominantly rainfed[8] and fertilizer usage is low[84,85], such that investments have high potential in increasing crop yields as shown e.g. in a Mali case study for climate-smart agriculture[86].

When including risk pooling, the increasing crop area allocations do not reach the limit of available agricultural area within the considered time period, again showing the positive effect of risk pooling on the feasibility of this approach to sustainable food security. However, the medium fertility population predictions which we considered in Section 4.6 do result in a trend towards larger crop areas. Extrapolated, this trend implies that the limit of available area will be reached around 2050. Using the high fertility population prediction would shift this to an earlier year.

Overall, it is apparent that population growth is a critical factor for sustainable food security in West Africa, and that current trends in average crop yields are not sufficient to counteract increasing food demand. Increasing variability of yields due to climate change, which is not covered by the trend in yield averages, can put additional pressure on food production[8]. While a large scale insurance scheme to increase farmers resilience does show potential, in order to reach the SDG *Zero Hunger*, it needs to be coupled with investments in closing the yield gap through sustainable intensification of agricultural production[7] as aimed at by the *African Green revolution*[87], long-term investments in disaster preparedness e.g. through infrastructure such as irrigation or management of dams[8,13], and investment in farmer advisory systems to promote the application of best management practices[88]. The fourth SDG, *Quality Education*, is also expected to

positively impact the road towards *Zero Hunger*, by slowing down the rapid population growth in developing countries[89].

# 6. Summary and outlook

Achieving global sustainable food security is a key challenge of current development, and as such is addressed by the UN as second Sustainable Development Goal (SDG) for 2030, generally referred to as *Zero Hunger* [5]. Currently, Africa is not on track to meet this goal [8]. On the contrary, prevalence of undernourishment is rising, with the main increase occurring in West Africa [8]. Impacts of climate change on African ecosystems are expected to be high [10], with Sub-Saharan Africa being especially vulnerable, as it strongly relies on rainfed agriculture, struggles with high poverty rates and poor infrastructure, and faces the fastest population growth world wide [7,8].

In this thesis, we applied a stochastic modeling approach to risk management for sustainable food security in West Africa. The project consisted of three main steps. First, extended data analysis was conducted to build an understanding of the situation at hand, to develop the structure and relations within the resulting model, and to generate required input data. A spatial dependence structure for crop yields was established based on cluster analysis using the drought index SPEI. For each crop, yields within a cluster are assumed to be fully dependent, allowing to work with average yields per cluster. Furthermore, the occurrence of catastrophic versus non-catastrophic yields for different crops within the same cluster are assumed to be fully dependent. In contrast, crop yields and hence the occurrence of catastrophic yields between different clusters are assumed to be fully independent. According to this dependence structure, yield distributions for maize and rice were estimated from the historic crop yield dataset GDHY, including the observed trends in mean yields in the projection of distributions for future years. These are the main input to our sustainable food security model.

In a second step, a stochastic optimization framework was set up. We considered the underlying theory of stochastic optimization as a background and a time-invariant version of the model in the interest of an analytical interpretation of the model behavior. Then, a time-dependent sustainable food security model and its parametrization was developed. The model includes food security in terms of a certain food demand that needs to be met. The sustainability dimension is implemented by increasing farmers resilience to extreme events through a large scale insurance scheme financed by taxes on agricultural profits. Apart from food security, the model includes solvency of the insurance scheme after payouts as second objective to ensure a successful operation over time. As such objectives in general cannot be met for every realization of the uncertain parameters, i.e. in our case for all possible yield realizations, they are included as probabilistic constraints that only need to be satisfied with certain probabilities. To ensure that the

required probabilities are met, second stage penalties for violation of the constraints are included. The model output informs on optimal crop area allocations for each crop in each cluster and year to minimize the sum of first stage crop cultivation costs and second stage penalties that arise if objectives are not met, respecting limits given by available arable area in each cluster. It thereby decides on the best balance between risk prevention through first stage investments and post-disaster payments according to second stage penalties.

At last, a theoretical analysis of the model behavior was conducted and the results were interpreted. The central question was to what extent implementing a large scale insurance scheme to achieve sustainable food security in West Africa is feasible under the constraints of available arable area, agricultural production, and financial resources. Furthermore, it was discussed how this feasibility is impacted by policy interventions, risk pooling, or environmental, technological, and demographic developments. In general, the available arable area in West Africa was sufficient to meet the food demand with a high probability. However, high probabilities of fund solvency lead to overproduction in the first years after implementing such an insurance scheme. We assumed this unbalanced production over time to be an undesired effect with very high levels of crop allocation potentially being infeasible due to limited resources of farmers. Policy interventions, such as increasing the tax rate according to which farmers pay a share of their profits to the catastrophe fund, reducing the share of risk that is covered by the insurance scheme, or reducing the guaranteed income that is payed in case of a catastrophe, improved the performance of the insurance scheme regarding solvency. However, these actions also reduced the utility of the insurance for farmers. If farmers' resilience against extreme events is no longer provided, the sustainability dimension of food security is also no longer given.

We then showed that risk pooling is an effective way to improve performance both regarding food security and the solvency of the insurance scheme without compromising farmers' benefits. The general concept is to reduce the relative severity of extreme events by insuring larger areas or groups of individuals with uncorrelated risks in a centralized form. In our case, this is realized by a joint catastrophe fund covering areas of different agro-climatic conditions and therefore low risk correlation, for which we assume uncorrelated yields in the model. To allow for very high probabilities of solvency of the catastrophe fund, which might be required by the insurance provider, such a centralized supranational approach is essential. Finally, we discussed the impact of future developments on sustainable food security. The main result was that population growth is a critical factor for sustainable food security in West Africa and that current positive trends in average crop yields are not sufficient to counteract increasing food demand. Therefore, any approach towards sustainable food security should include diverse actions such as investments in closing the yield gap through sustainable intensification of agricultural production, long-term investments in disaster preparedness, or investment in farmer advisory systems to enhance the application of best management practices.

Nonetheless, we understand a large scale insurance scheme to increase farmers resilience to be a key measure on the road to reach the SDG *Zero Hunger*.

Throughout the thesis, several simplifications and technical restrictions due to data availability and constraints of computational resources were addressed, and throughout the discussion in Chapter 5 different caveats in the interpretation of model results were noted. Many of these can be dealt with by an extension of the model structure or intensified data analysis. The following paragraphs give an overview of such possible next steps to improve the current model.

Applying more advanced statistical methods such as copulas can both improve the dependence structure used as basis for the modeling approach and allow to include trends in yield variability in the projection of yield distributions over time. Extended literature research or data collection can result in spatially differentiated values for cultivation costs and farm-gate prices that reflect local conditions, and also available agricultural area could be estimated per cluster instead of assuming equal shares throughout West Africa. A more differentiated approach to population numbers also allows to include multiple local food security objectives in addition to a joint food demand. The aggregated food security objective allows to analyze the potential of free trade, permitting subregions with high yields to support other subregions that are facing difficulties. However, additional local objectives can ensure that no subregions are left out in the calculation of crop area allocations and that food availability is spread more evenly across the whole territory. Risk pooling would still apply to the insurance scheme by a centralized fund. In order to analyze the effects of risk pooling in more depth, the model can be applied to the different subregions separately instead of using the model results from considering West Africa as a single region as proxy. A further point addressed in the discussion is the diversity of crops, pointing out that additional crops can be included to allow for more substitution between crops with different performances. Furthermore, the food security objective could be extended to include the necessity of a diverse and nutritious diet.

A more fundamental adaption of the model would be a change in the structure of the solvency constraint to not only cover the first catastrophe but a fixed number of years for all realizations. This eliminates the technical problem of reduced sample size in the later years of the considered time period. At the same time it allows for a better analysis of the effect of changing the risk level covered by the insurance scheme, as the resulting change in frequency of catastrophes would be represented in the model. An arising difficulty is how to quantify the impact of a catastrophe on model parameters such as the fund size. Therefore, it will be important to get a more realistic view on possible strategies to deal with consequences of not maintaining solvency. In the current model set up, the solvency constraint also leads to overproduction in the first years after the insurance scheme is launched. This effect can be reduced by including a positive initial fund size, which is already implemented in the sustainable food security model, but was not analyzed in the scope of this project. A second aspect mentioned

in relation to overproduction was the possibility of falling prices as a result. In general, farm-gate prices are subject to fluctuations, but are assumed constant over time in the model. Variability in prices can be included as a second source of uncertainty in the optimization framework.

Finally, different investment options can be included in the model. These can range from short-term investments such as fertilization and high quality seeds to long-term investments such as expansion of agricultural area, improved agricultural technology, or irrigation systems. Furthermore, food storage can be implemented to allow good years with a higher production to support the following years. All of these actions would be additional first stage actions within the model. The extended sustainable food security model can then determine the optimal balance between crop cultivation, investments, and post-catastrophe actions.

The stylized sustainable food security model allowed for an analysis of a wide range of fundamental questions regarding a potential path towards sustainable food security. In particular, it highlighted the need for a supranational approach to deal with risks and challenges affecting sustainable food security. The above discussed points show the potential of future work to further assess the approach using more detailed simulation models, and hopefully result in a model able to impact the behavior of policy makers regarding sustainable food security.

# Bibliography


[1] FAO. Rome Declaration on World Food Security. *Food and Agriculture Organization*, 1996. URL http://www.fao.org/3/w3613e/w3613e00.htm. Date of Access: 27.12.2019.

[2] FAO. Declaration of the World Summit on Food Security. *Food and Agriculture Organization*, 2009.

[3] HLPE. Food security and nutrition: building a global narrative towards 2030. *High Level Panel of Experts on Food Security and Nutrition*, 2020.

[4] UN. Transforming our World: the 2030 Agenda for Sustainable Development. *United Nations*, 2015.

[5] UN. Sustainable Development Goals. *United Nations*, 2015. URL https://sustainabledevelopment.un.org/?menu=1300. Date of Access: 27.12.2019.

[6] S. Fujimori, T. Hasegawa, V. Krey, K. Riahi, C. Bertram, et al. A multi-model assessment of food security implications of climate change mitigation. *Nature Sustainability*, 2(5):386–396, 2019. doi: 10.1038/s41893-019-0286-2.

[7] C. Hall, T.P. Dawson, J.I. Macdiarmid, R.B. Matthews, and P. Smith. The impact of population growth and climate change on food security in Africa: looking ahead to 2050. *International Journal of Agricultural Sustainability*, 15(2):124–135, 2017. doi: 10.1080/14735903.2017.1293929.

[8] FAO and ECA. Regional Overview of Food Security and Nutrition. *Food and Agriculture Organization, United Nations Economic Commission for Africa*, page 116, 2018.

[9] O.A. Otekunrin, O.A. Otekunrin, S. Momoh, and I.A. Ayinde. How far has Africa gone in achieving the Zero Hunger Target? Evidence from Nigeria. *Global Food Security*, 22:1–12, 2019.

[10] I. Niang, O.C. Ruppel, M.A. Abdrabo, A. Essel, C. Lennard, et al. Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. pages 1199–1265, 2014.

[11] J.R. Porter and M.A. Semenov. Crop responses to climatic variation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1463):2021–2035, 2005. doi: 10.1098/rstb.2005.1752.

[12] FAO. Executive Brief: The Sahel crisis 2012. *Food and Agriculture Organization of the United Nations*, 2012. URL http://www.fao.org/fileadmin/user_upload/emergencies/docs/EXECUTIVE_BRIEF_TCE2-_Sahel_Crisis_14_February.pdf. Date of Access: 12.07.2020.

[13] S.M. Vicente-Serrano, S. Beguería, L. Gimeno, L. Eklundh, G. Giuliani, et al. Challenges for drought mitigation in Africa: The potential use of geospatial data and drought information systems. *Applied Geography*, 34:471–486, 2012. doi: 10.1016/j.apgeog.2012.02.001.

[14] M.B. Burke, E. Miguel, S. Satyanath, J.A. Dykema, and D.B. Lobell. Warming increases the risk of civil war in Africa. *Proceedings of the national Academy of sciences*, 106(49):20670–20674, 2009.

[15] D. Gautier, D. Denis, and B. Locatelli. Impacts of drought and responses of rural populations in West Africa: a systematic review. *Wiley Interdisciplinary Reviews: Climate Change*, 7(5):666–681, 2016. doi: 10.1002/wcc.411.

[16] Y.M. Ermoliev, S.M. Robinson, E. Rovenskaya, and T. Ermolieva. Integrated Catastrophic Risk Management: Robust Balance between Ex-ante and Ex-post Measures. *SIAM News*, 51(6):4, 2018.

[17] Y.M. Ermoliev and R.B. Wets. *Numerical Techniques for Stochastic Optimization*. Springer Verlag, 1988. ISBN 0-387-18677-8.

[18] G. van Rossum and F.L. Drake. *Python 3 Reference Manual*. CreateSpace, 2009. ISBN 1-4414-1269-7.

[19] S.M. Vicente-Serrano, S. Beguería, and J.I. López-Moreno. A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. *Journal of Climate*, 23(7):1696–1718, 2010. doi: 10.1175/2009jcli2909.1.

[20] W.C. Palmer. Meteorological drought. *Res. Pap*, 45:58, 1965.

[21] N. Wells, S. Goddard, and M.J. Hayes. A Self-Calibrating Palmer Drought Severity Index. *Journal of Climate*, 17(12):2335–2351, 2004. doi: 10.1175/1520-0442(2004)017⟨2335:aspdsi⟩2.0.co;2.

[22] U. Diasso and B.J. Abiodun. Drought modes in West Africa and how well CORDEX RCMs simulate them. *Theoretical and Applied Climatology*, 128(1–2):223–240, 2015. doi: 10.1007/s00704-015-1705-6.

[23] W.M. Alley. The Palmer Drought Severity index: Limitations and Assumptions. *Journal of Climate and Applied Meteorology*, 23(7):1100–1109, 1984. doi: 10.1175/1520-0450(1984)023⟨1100:tpdsil⟩2.0.co;2.

[24] S. Beguería, S.M. Vicente-Serrano, and M. Angulo-Martínez. A Multiscalar Global Drought Dataset: The SPEIbase: A New Gridded Product for the Analysis of Drought Variability and impacts. *Bulletin of the American Meteorological Society*, 91(10):1351–1356, 2010. doi: 10.1175/2010bams2988.1.

[25] D.A. Wilhite and M.H. Glantz. Understanding: the drought phenomenon: the role of definitions. *Water international*, 10(3):111–120, 1985.

[26] T.B. McKee, N.J. Doesken, and J. Kleist. The relationship of drought frequency and duration to time scales. In *Proceedings of the 8th Conference on Applied Climatology*, volume 17, pages 179–183. American Meteorological Society Boston, MA, 1993.

[27] M.G. Miah, H.M. Abdullah, and C. Jeong. Exploring standardized precipitation evapotranspiration index for drought assessment in Bangladesh. *Environmental Monitoring and Assessment*, 189(11), 2017. doi: 10.1007/s10661-017-6235-5.

[28] B.Y. Tam, K. Szeto, B. Bonsal, G. Flato, A.J. Cannon, et al. CMIP5 drought projections in Canada based on the Standardized Precipitation Evapotranspiration Index. *Canadian Water Resources Journal / Revue canadienne des ressources hydriques*, 44(1):90–107, 2018. doi: 10.1080/07011784.2018.1537812.

[29] C.W. Thornthwaite. An approach toward a rational classification of climate. *Geographical review*, 38(1):55–94, 1948.

[30] T. Mavromatis. Drought index evaluation for assessing future wheat production in Greece. *International Journal of Climatology*, 27(7):911–924, 2007. doi: 10.1002/joc.1444.

[31] I. Harris, T.J. Osborn, P. Jones, and D. Lister. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific Data*, 7(1), 2020. doi: 10.1038/s41597-020-0453-3.

[32] AgMIP. AgMIP Charter. 2020. URL https://agmip.org/agmipcharter2/#. Date of Access: 06.06.2020.

[33] C. Rosenzweig, J.W. Jones, J.L. Hatfield, A.C. Ruane, K.J. Boote, et al. The Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and pilot studies. *Agricultural and Forest Meteorology*, 170:166–182, 2013. doi: https://doi.org/10.1016/j.agrformet.2012.09.011.

[34] C. Müller, J. Elliott, D. Kelly, A. Arneth, J. Balkovic, et al. The Global Gridded Crop Model Intercomparison phase 1 simulation dataset. *Scientific data*, 6(1):1–22, 2019. doi: https://doi.org/10.1038/s41597-019-0023-8.

[35] T. Iizumi, M. Yokozawa, G. Sakurai, M.I. Travasso, V. Romanenkov, et al. Historical changes in global yields: major cereal and legume crops from 1982 to 2006. *Global ecology and biogeography*, 23(3):346–357, 2014.

[36] FAOSTAT. Crops. License: CC BY-NC-SA 3.0 IGO. 2020. URL: http://www.fao.org/faostat/en/#data/QC. Date of Access: 28.06.2020.

[37] C. Monfreda, N. Ramankutty, and J.A. Foley. Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. *Global Biogeochemical Cycles*, 22(1), 2008. doi: 10.1029/2007gb002947.

[38] T. Iizumi and T. Sakai. The global dataset of historical yields for major crops 1981–2016. *Scientific Data*, 7(1), 2020. doi: 10.1038/s41597-020-0433-7.

[39] A. Sklar. Fonctions de reprtition an dimensions et leursmarges. *Université Paris 8*, 8:229–231, 1959.

[40] F. Gaupp, G. Pflug, S. Hochrainer-Stigler, J. Hall, and S. Dadson. Dependency of Crop Production between Global Breadbaskets: A Copula Approach for the Assessment of Global and Regional Risk Pools. *Risk Analysis*, 37(11):2212–2228, 2016. doi: 10.1111/risa.12761.

[41] P. Kumar. Statistical Dependence: Copula Functions and Mutual Information Based Measures. *Journal of Statistics Applications & Probability*, 1(1), 2012.

[42] B.S. Everitt, S. Landau, M. Leese, and D. Stahl. *Cluster analysis*. 2011. ISBN 978-0-470-74991-3.

[43] T.S. Madhulatha. An Overview on Clustering Methods. *IOSR Journal of Engineering*, 02(04):719–725, 2012. doi: 10.9790/3021-0204719725.

[44] H.-P. Kriegel, P. Kröger, J. Sander, and A. Zimek. Density-based clustering. *WIREs Data Mining and Knowledge Discovery*, 1(3):231–240, 2011. doi: 10.1002/widm.30.

[45] P. Foggia, G. Percannella, C. Sansone, and M. Vento. Benchmarking graph-based clustering algorithms. *Image and Vision Computing*, 27(7):979–988, 2009. doi: 10.1016/j.imavis.2008.05.002.

[46] M. Meilã and D. Heckerman. An experimental comparison of model-based clustering methods. *Machine learning*, 42(1–2):9–29, 2001.

[47] L. Kaufmann and P. Rousseeuw. Clustering by Means of Medoids. *Data Analysis based on the L1-Norm and Related Methods*, pages 405–416, 1987.

[48] E. Schubert and P.J. Rousseeuw. Faster k-Medoids Clustering: Improving the PAM, CLARA, and CLARANS Algorithms. In *Similarity Search and Applications*, pages 171–187. Springer International Publishing, 2019. ISBN 978-3-030-32047-8.

[49] S. Mannor, X. Jin, J. Han, and X. Zhang. K-Medoids Clustering. In *Encyclopedia of Machine Learning*, pages 564–565. Springer US, 2011. doi: 10.1007/978-0-387-30164-8_426.

[50] S. van Dongen and A.J. Enright. Metric distances derived from cosine similarity and Pearson and Spearman correlations. *arXiv preprint arXiv:1208.3145*, 2012.

[51] W. Shi and F. Tao. Vulnerability of African maize yield to climate change and variability during 1961–2010. *Food Security*, 6(4):471–481, 2014. doi: 10.1007/s12571-014-0370-4.

[52] W.J. Sacks, D. Deryng, J.A. Foley, and N. Ramankutty. Crop planting dates: an analysis of global patterns. *Global Ecology and Biogeography*, 2010. doi: 10.1111/j.1466-8238.2010.00551.x.

[53] J. Liu, J.R. Williams, A.J.B. Zehnder, and H. Yang. GEPIC – modelling wheat yield and crop water productivity with high resolution on a global scale. *Agricultural Systems*, 94(2):478–493, 2007. doi: 10.1016/j.agsy.2006.11.019.

[54] C. Folberth, A. Baklanov, J. Balkovič, R. Skalskỳ, N. Khabarov, et al. Spatio-temporal downscaling of gridded crop model yield estimates based on machine learning. *Agricultural and forest meteorology*, 264:1–15, 2019.

[55] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004. ISBN 78-0-521-83378-3.

[56] W. Fenchel and D.W. Blackett. *Convex cones, sets, and functions*. Princeton University, Department of Mathematics, Logistics Research Project, 1953.

[57] M.J.D. Powell. A Direct Search Optimization Method That Models the Objective and Constraint Functions by Linear Interpolation. In *Advances in Optimization and Numerical Analysis*, pages 51–67. Springer Netherlands, 1994. doi: 10.1007/978-94-015-8330-5_4.

[58] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272, 2020. doi: 10.1038/s41592-019-0686-2.

[59] FAOSTAT. Producer Prices. License: CC BY-NC-SA 3.0 IGO. 2020. Extracted from: http://www.fao.org/faostat/en/#data/PP. Date of Access: 28.06.2020.

[60] Ministry of Agriculture (Liberia). Liberia: Invest Agriculture, 2019. URL https://ekmsliberia.info/document/liberia-invest-agriculture/. Date of Access: 28.06.2020.

[61] G.N. Ben-Chendo, N. Lawal, and M.N. Osuji. Cost and returns of paddy rice production in Kaduna State. *European Journal of Agriculture and forestry Research*, 5(3):41–48, 2017.

[62] A.A. Fall. Synthèse des études sur l'état des lieux chaine de valeur riz en Afrique de l'ouest: Bénin, Burkina Faso, Mali, Niger et Sénégal. *Rapport final, Réseau des Organisations Paysannes et de Producteurs de l'Afrique de l'Ouest*, pages 57–59, 2016.

[63] M.S. Sadiq, M.T. Yakasai, M.W. Ahmad, T.Y. Lapkene, and M. Abubakar. Profitability and production efficiency of small-scale maize production in Niger State, Nigeria. *IOSR Journal of Applied Physics*, 3(4):19–23, 2013.

[64] M.N. Ba. Competiveness of Maize Value Chains for Smallholders in West Africa: Case of Benin, Ghana and Cote D'Ivoire. *Agricultural Sciences*, 08(12):1372–1401, 2017. doi: 10.4236/as.2017.812099.

[65] S.E. Cotillon and G.G. Tappan. Landscapes of West Africa – A Window on a Changing World. *U.S. Geological Survey*, 2016. doi: 10.5066/F7N014QZ.

[66] FAO. World Agriculture: towards 2015/2030: Summary Report. *Food and Agriculture Organization of the United Nations*, 2002.

[67] UN. Probabilistic Population Projections Rev. 1 based on the World Population Prospects 2019 Rev. 1. *United Nations, Department of Economic and Social Affairs, Population Division*, 2019. URL https://population.un.org/wpp/Download/Probabilistic/Population/. Date of Access: 12.04.2020.

[68] UN. World Population Prospects 2019: Methodology of the United Nations population estimates and projections. *United Nations, Department of Economic and Social Affairs, Population Division*, 2019.

[69] Center For International Earth Science Information Network – CIESIN – Columbia University. Gridded Population of the World, Version 4 (GPWv4): Population Count Adjusted to Match 2015 Revision of UN WPP Country Totals, Revision 11. *NASA Socioeconomic Data and Applications Center (SEDAC)*, 2018. doi: 10.7927/H4PN93PB. Date of Access: 12.04.2020.

[70] Center For International Earth Science Information Network – CIESIN – Columbia University. Documentation for the Gridded Population of the World, Version 4 (GPWv4), Revision 11 Data Sets. *NASA Socioeconomic Data and Applications Center (SEDAC)*, 2018. doi: 10.7927/H45Q4T5F. Date of Access: 29.06.2020.

[71] USDA. Standard Reference on Nutrient Values. *U.S. Department of Agriculture*, 2016. URL https://www.ars.usda.gov/northeast-area/beltsville-md-bhnrc/beltsville-human-nutrition-research-center/methods-and-application-of-food-composition-laboratory/mafcl-site-pages/sr11-sr28/. Date of Access: 11.06.2020.

[72] N.D. Pearson. *Risk budgeting: portfolio problem solving with value-at-risk*. John Wiley & Sons, 2011.

[73] J.L.W.V. Jensen. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica*, 30(1):175–193, 1906. doi: 10.1007/bf02418571.

[74] J.R. Birge. The value of the stochastic solution in stochastic linear programs with fixed recourse. *Mathematical programming*, 24(1):314–325, 1982.

[75] H. Lotze-Campen, C. Müller, A. Bondeau, S. Rost, A. Popp, et al. Global food demand, productivity growth, and the scarcity of land and water resources: a spatially explicit mathematical programming approach. *Agricultural Economics*, 2008. doi: 10.1111/j.1574-0862.2008.00336.x.

[76] W. Britz, P. Witzke, et al. Capri model documentation 2014. 2014.

[77] P. Havlík, U.A. Schneider, E. Schmid, H. Böttcher, S. Fritz, et al. Global land-use implications of first and second generation biofuel targets. *Energy Policy*, 39(10): 5690–5702, 2011. doi: 10.1016/j.enpol.2010.03.030.

[78] S. Fuss, P. Havlík, J. Szolgayová, E. Schmid, and M. Obersteiner. Large-Scale Modelling of Global Food Security and Adaptation under Crop Yield Uncertainty. *EAAE 2011 Congress Change and Uncertainty*, 2011.

[79] T. Ermolieva, P. Havlík, Y. Ermoliev, A. Mosnier, M. Obersteiner, et al. Integrated Management of Land Use Systems under Systemic Risks and Security Targets: A Stochastic Global Biosphere Management Model. *Journal of Agricultural Economics*, 67(3):584–601, 2016. doi: 10.1111/1477-9552.12173.

[80] T. Ermolieva, T. Filatova, Y. Ermoliev, M. Obersteiner, K.M. de Bruijn, et al. Flood Catastrophe Model for Designing Optimal Flood Insurance Program: Estimating Location-Specific Premiums in the Netherlands. *Risk Analysis*, 37(1):82–98, 2016. doi: 10.1111/risa.12589.

[81] C. Janssens, P. Havlík, T. Krisztin, J. Baker, S. Frank, et al. Global hunger and climate change adaptation through international trade. *Nature Climate Change*, 2020. doi: 10.1038/s41558-020-0847-4.

[82] W. Willett, J. Rockström, B. Loken, M. Springmann, T. Lang, et al. Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems. *The Lancet*, 393(10170):447–492, 2019. doi: 10.1016/s0140-6736(18) 31788-4.

[83] B. Sultan, D. Defrance, and T. Iizumi. Evidence of crop production losses in West Africa due to historical global warming in two crop models. *Scientific Reports*, 9 (1), 2019. doi: 10.1038/s41598-019-49167-0.

[84] B. Vanlauwe, A. Bationo, J. Chianu, K.E. Giller, R. Merckx, et al. Integrated Soil Fertility Management: Operational definition and consequences for implementation and dissemination. *Outlook on Agriculture*, 39(1):17–24, 2010. doi: 10.5367/000000010791169998.

[85] Y. Luan, W. Zhu, X. Cui, G. Fischer, T.P. Dawson, et al. Cropland yield divergence over Africa and its implication for mitigating food insecurity. *Mitigation and Adaptation Strategies for Global Change*, 24(5):707–734, 2018. doi: 10.1007/s11027-018-9827-7.

[86] N. Andrieu, B. Sogoba, R. Zougmore, F. Howland, O. Samake, et al. Prioritizing investments for climate-smart agriculture: Lessons learned from Mali. *Agricultural Systems*, 154:13–24, 2017. doi: 10.1016/j.agsy.2017.02.008.

[87] MDG Center of East and Southern Africa. Africa's Green Revolution: A Call to Action: Innovative approaches to meet the hunger Millennium Development Goal in Africa. *Millennium Development Goals Technical Support Centre, Nairobi*, 2004.

[88] W. Settle and M.H. Garba. Sustainable crop production intensification in the Senegal and Niger River basins of francophone West Africa. *International Journal of Agricultural Sustainability*, 9(1):171–185, 2011. doi: 10.3763/ijas.2010.0559.

[89] G.J. Abel, B. Barakat, S. Kc, and W. Lutz. Meeting the Sustainable Development Goals leads to lower world population growth. *Proceedings of the National Academy of Sciences*, 113(50):14294–14299, 2016. doi: 10.1073/pnas.1611386113.

# Acronyms

| | |
|---|---|
| **AgMIP** | Agricultural Model Intercomparison and Improvement Project. |
| **CRU** | Climatic Research Unit. |
| **ECA** | United Nations Economic Commission for Africa. |
| **EPIC** | Environmental Policy Integrated Climate Model. |
| **FAO** | Food and Agriculture Organization of the United Nations. |
| **FAOSTAT** | Food and Agriculture Organization Corporate Statistical Database. |
| **GDHY** | Global Dataset of Historic Yields. |
| **GEPIC** | GIS-based Environmental Policy Integrated Climate Model. |
| **GGCM** | Global Gridded Crop Model. |
| **GGCMI** | Global Gridded Crop Model Intercomparison. |
| **HLPE** | High Level Panel of Experts on Food Security and Nutrition. |
| **IIASA** | International Institute for Applied System Analysis. |
| **NASA** | National Aeronautics and Space Administration. |
| **NPP** | Net Primary Production. |
| **PCA** | Principal Component Analysis. |
| **PDSI** | Palmer Drought Severity Index. |
| **PET** | Potential Evapotranspiration. |
| **PHU** | Potential Heat Units. |
| **RSD** | Relative Standard Deviation. |
| **sc-PDSI** | self-calibrating Palmer Drought Severity Index. |
| **SDG** | Sustainable Development Goal. |
| **SEDAC** | Socioeconomic Data and Applications Center. |

**SPEI** Standardized Precipitation-Evapotranspiration Index.

**SPI** Standardized Precipitation Index.

**UN** United Nations.

**UN WPP** United Nations World Population Prospects.

**USDA** U.S. Department of Agriculture.

# Appendices

# A. Python Code

The full python code used in this project is openly accessible through a GitHub repository and can be found at https://github.com/deleip/FoodSecurityWestAfrica/tree/master/Thesis. The functionality is described in the README.md, which is also included in the following.

## A.1. Readme

### Aim

The aim of this project is a stylized two-stage stochastic optimization model for food security in West Africa. The region is spatially subdivided in a modifiable number of clusters and uncertainty is included by yield distributions per cluster, year, and crop. The model covers a given time period (default is 2017–2041) and includes a government fund which is built up by farmers paying taxes on their profits and is used for payouts to guarantee a certain income for farmers in catastrophic years. The model output is an allocation of arable land in each year and cluster to the different crops (i.e. maize or rice). The objective is to minimize costs while assuring government solvency (i.e. the final government fund should be positive) and food security (i.e. producing a certain amount of calories every year) with given probabilities each.

### Content

- `Data.py`
- `Functions_Data.py`
- `Analysis.py`
- `Functions_Analysis.py`
- `StochasticOptimization.py`
- `Functions_StochasticOptimization.py`
- `IntermediateResults`

### Functionality

The first step is the preparation of different data sets (drought indices, AgMIP crop yield data, GDHY crop yield data, crop calendar of 2000, CRU data on precipitation and PET, UN world population scenarios, SEDAC gridded world population in 2015, farm-gate prices) for further usage. This is implemented in `Data.py` and depends on routines defined in `Functions_Data.py`. Next, drought indices are used to subdivide West Africa into clusters and the optimal cluster number is determined; then a yield model is set up through regression analysis to estimate yield distributions for the model input. Both steps are implemented in `Analysis.py`, depending on routines defined in `Functions_Analysis.py`. As the underlying datasets have an aggregated size of over 16GB, the files `Data.py` and `Analysis.py` cannot be run based on the content provided on GitHub. However, URLs to the respective datasets are documented in `Data.py`.

The core of the project is the stochastic optimization model. This is implemented in `Functions_StochasticOptimization.py`, and is run by `StochasticOptimization.py` to analyze the behavior of the model and the influence of the parameters that can be varied. `StochasticOptimization.py` also includes visualization of this analysis. The GitHub repository provides all data necessary to run the model for numbers of clusters $k = 1$, 2, 3, or 7 in the folder `IntermediateResults`.

Modifiable model settings with their default values are:

| parameter description | variable name | default value |
|---|---|---|
| number of clusters | $k$ | 1 |
| number of clusters that have to be catastrophic for a catastrophic year[1] | $num\_cl\_cat$ | 1 |
| yield scenario (“fixed” or “trend”) | $yield\_projection$ | “$fixed$” |
| first year of the simulation | $yield\_year$ | 2017 |
| should stylized values be used ($True$ or $False$, leftover from test phase) | $stilised$ | $False$ |
| UN population scenario (“Medium”, “High”, “Low”, “ConstantFertility”, “InstantReplacement”, “ZeroMigration”, “ConstantMortality”, “NoChange”, “Momentum”) or fixed yield distributions over time (“fixed”) | $pop\_scenario$ | “$fixed$” |
| risk level (given as frequency) | $risk$ | 20 |
| sample size | $N\_c$ | 3500 |
| number of covered years | $T\_max$ | 25 |
| seed for reproducible yield realizations | $seed$ | 150620 |
| tax rate | $tax$ | 0.03 |

[1] This was introduced in the course of the project, but is not included in the final interpretation of the model

| parameter description | variable name | default value |
|---|---|---|
| percentage of expected income that will be covered by government in case of catastrophes | *perc_guaranteed* | 0.75 |

Table A.1.: Overview of model settings and their notation within the implementation, including default values

The model can be called by the function

$$crop\_alloc,\ meta\_sol, rhoF,\ rhoS,\ settings,\ args = \texttt{OptimizeFoodSecurityProblem}(probF,\ probS,\ rhoFini,\ rhoSini,\ {}^{**}kwargs)$$

Here, $^{**}kwargs$ is a placeholder for the above listed settings. Settings that are kept at their default do not need to be included. The parameter *probF* is the probability with which the food constraint needs to be met, *probS* is the probability with which the government needs to stay solvent. Note, that for some settings very high probabilities might be infeasible. The function will first call

$$settings = \texttt{DefaultSettingsExcept}({}^{**}kwargs)$$

to get a dictionary of all settings, including the expected income which is calculated depending on the other settings. Then

$$rhoF,\ rhoS = \texttt{GetPenalties}(settings,\ probF,\ probS,\ rhoFini,\ rhoSini)$$

will determine the correct penalties *rhoF* and *rhoS* for the given probabilities, starting with *rhoFini* and *rhoSini* as first guesses and using an algorithm based on the bisection method. Next

$$x\_ini,\ const,\ args,\ meta\_cobyla,\ other = \texttt{SetParameters}(settings)$$

will prepare all other model inputs depending on the settings. The outputs of the function are an array *x_ini* as an initial guess for the crop allocation, a list *const* defining the model constraints (i.e. crop areas need to be positive and respect the available arable area), all arguments that need to be passed to the objective function by the optimizer except the crop areas as a dictionary *args* (i.e. yield realizations, food demand, terminal years, cultivation costs etc.), technical information for the solver as a dictionary *meta_cobyla* and some additional information on the parameters for potential analysis in the dictionary *other*. Finally, the solver `scipy.optimize.fmin_cobyla` is called within the function

$$crop\_alloc,\ meta\_sol,\ duration = \texttt{OptimizeMultipleYears}(x\_ini,\ const,\ args,\ meta\_cobyla,\ rhoF,\ rhoS)$$

This returns the optimal crop allocation *crop_alloc*, meta information *meta_sol* about the solution (e.g. the minimized value of the objective function, the final fund for all realizations, or the yearly shortcomings from the food demand for each realization), and the time the solver took to find the solution as *duration*.

The main function `OptimizeFoodSecurityProblem` finally returns the optimal crop allocation *crop_alloc*, the meta information *meta_sol*, the penalties *rhoF* and *rhoS* corresponding to the input probabilities *probF* and *probS*, the dictionary *settings* of all settings, and the dictionary *args* of all additional arguments that are passed to the objective function.

The model can be called for specific penalties *rhoF* and *rhoS* instead of given probabilities by the following combination:

$$\begin{aligned} settings &= \texttt{DefaultSettingsExcept}(^{**}kwargs) \\ x\_ini,\ const,\ args,\ meta\_cobyla,\ other &= \texttt{SetParameters}(settings) \\ crop\_alloc,\ meta\_sol,\ duration &= \texttt{OptimizeMultipleYears}(x\_ini,\ const,\ args, \\ &\qquad meta\_cobyla,\ rhoF,\ rhoS) \end{aligned}$$

# Eigenständigkeitserklärung

Hiermit versichere ich, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Außerdem versichere ich, dass ich die allgemeinen Prinzipien wissenschaftlicher Arbeit und Veröffentlichung, wie sie in den Leitlinien guter wissenschaftlicher Praxis der Carl von Ossietzky Universität Oldenburg festgelegt sind, befolgt habe.

Debbora Leip
Matrikelnummer: 4587207
Ortenberg, den 14.8.2020