A general framework for the generation of probabilistic socioeconomic scenarios and risk quantification concerning food security with application in the Upper Nile river basin.

Koundouri, Phoebe and Papayiannis, Georgios and Vassilopoulos, Achilleas and Yannacopoulos, Athanasios

January 2022

A general framework for the generation of probabilistic socioeconomic scenarios and risk quantification concerning food security with application in the Upper Nile river basin

P. Koundouri[a], G. I. Papayiannis[b,c], A. Vassilopoulos[d], and A. N. Yannacopoulos[b]

[a]Athens University of Economics & Business, Department of International and European Economic Studies, Research Laboratory on Socio-Economic and Environmental Sustainability, Athens, GR
[b]Athens University of Economics & Business, Department of Statistics, Stochastic Modeling and Applications Laboratory, Athens, GR
[c]Hellenic Naval Academy, Department of Naval Sciences, Section of Mathematics, Mathematical Modeling and Applications Laboratory, Piraeus, GR
[d]University of Ioannina, Department of Economics, Ioannina, GR

Abstract

Food security is a key issue in sustainability studies. In this paper we propose a general framework for providing detailed probabilistic socioeconomic scenarios as well as predictions across scenarios, concerning food security. Our methodology is based on the Bayesian probabilistic prediction model of world population (Raftery et al [10]) and on data driven prediction models for food demand and supply and its dependence on key drivers such as population and other socioeconomic and climate indicators(e.g. GDP, temperature, etc). For the purpose of risk quantification, concerning food security, we integrate the use of recently developed convex risk measures involving model uncertainty (Papayiannis et al [8], [9]) and propose a methodology for providing estimates and predictions across scenarios, i.e. when there is uncertainty as to which scenario is to be realized. Our methodology is illustrated by studying food security for the 2020-2050 horizon in the context of the SSP-RCP scenarios, for Egypt and Ethiopia.

Keywords: food security; probabilistic projections; risk quantification; shared socioeconomic pathways scenarios;

1 Introduction

Food systems represent the food value chain – from input supply and production of crops, livestock, fish, and other agricultural commodities to transportation, processing, retailing, wholesaling, and preparation of foods to consumption and disposal. A sustainable food system (SFS) is a food system that delivers food security and nutrition for all in such a way that the economic, social and environmental bases to generate food security and nutrition for future generations are not compromised. This means that: (a) it is profitable throughout (economic sustainability), (b) it has broad-based benefits for society (social sustainability), and (c) it has a positive or neutral impact on the natural environment (environmental sustainability). The transition to sustainable food systems lies at the heart of the United Nations' Agenda 2030 and the 17 Sustainable Development Goals[1] (SDGs), which call for major transformations in agriculture and food systems in order to end hunger, achieve food security and improve nutrition by 2030.

Food sufficiency is a key issue in sustainability studies. One of the key factors in the study of food sufficiency is evidently population, since food demand sufficient for subsistence clearly

[1] https://sdgs.un.org/goals

depends in an inelastic way on the population as well as on its detailed age structure. Not much can be made from a policy perspective on that, except perhaps on campaigns concerning people awareness on population control. On the other hand, another important factor in food sufficiency is food supply. This is a factor in which, apart from population, the economy plays an important role as well (e.g. through proper planning of the economic sectors involved in food production and/or distribution) and where science can greatly assist (e.g. by adopting modern and more efficient modes of agriculture or breeding etc).

Trustworthy predictions of food demand and food supply in a sufficiently long time horizon are very important in determining future food balance and will be of great help to policy makers. Predicting a potential food shortage on a sufficiently long horizon, provides policy makers the luxury of adopting long term measures combining a portfolio of production restructuring policies, international trade treaties, adoption of scientific measures or modern technologies etc, which may efficiently alleviate the risk of future food shortage. Such trustworthy predictions will inevitably rely upon predictions of the future population trends, and these predictions are expected to be the key drivers of food demand and supply. Such predictions must be probabilistic in nature (i.e. provide the distribution of the relevant random variable - e.g. food demand or supply - which gives full information of the trends along with their validity, rather than point estimates which carry less information concerning the predictions) and should also take into account model uncertainty, which is endemic especially when long term predictions are involved, for which the stochastic factors that affect them may be partially known.

It is the aim of this paper to provide probabilistic predictions of food demand and supply using as a starting point the detailed Bayesian probabilistic prediction model of the world population proposed by Raftery and his coworkers (see e.g., [10]). This model provides detailed information on the future world population and its age structure. Building on that, and using the detailed predictions concerning the future age structure for the population as well as the minimal required calories consumption per age group, we propose a model for the food consumption required for subsistence as well as a Cobb-Douglas type model for the food supply, involving population, GDP and environmental factors. Calibrating these models on past data from the period 1990-2019 and combining them with the Bayesian population model we provide probabilistic projections for food consumption and supply in the horizon 2030-2050 compatible with the standard SSP and RCP scenarios. To properly take into account the effects of model uncertainty we further propose a convex risk measure approach to the estimation of important quantities indicating possible food shortage such as for instance the difference between supply and consumption. Finally, the problem concerning the sensitivity of the projections on the SSP-RCP scenario chosen is addressed, by proposing a methodology for providing projections which are robust with respect to the scenario that materializes. Such considerations are important especially for long term effects and measures that have to be predicted and implemented long before the actual scenario that materializes has been fully clarified. Such prior estimates are based on the concept of Fréchet utilities or risk measures ([8], [9]) and allows one to obtain a robust estimation of the future values of the quantities of interest across scenarios.

The proposed methodology is applied in two major countries from the upper Nile river basin, Egypt and Ethiopia, constructing probabilistic socio-economic scenarios and then deriving food security risk valuations for each SSP-RCP scenario and across.

2 A general probabilistic socio-economic modelling framework for food security

2.1 Probabilistic population modelling

Population growth and evolution is a key factor driving many socioeconomic indices, including economic growth, production, environmental issues, food and water demand etc. In this respect, demographics must be the starting point for any socioeconomic modeling study. In this section we present some key results concerning scenario development for future population growth.

2.1.1 The main probabilistic population model

The state of the art model concerning world population is the probabilistic population model of Raftery and coworkers (see e.g., [10]). This model takes into account the inherent uncertainty of the phenomenon and its effects on future population projections using a Bayesian hierarchical model. The model is based on the natural evolution of the population phenomenon as characterized by the standard population model employed by the United Nations (UN),

$$P_c(t) = P_c(t-1) + B_c(t) - D_c(t) + M_c(t) \tag{1}$$

where $P_c(t)$ denotes the population of country c at time t (corresponding either to a single year or a 5-year period), $B_c(t)$ stands for the number of births (which depends on the total fertility rate), $D_c(t)$ denotes the number of deaths (which depends on the life expectancy) and $M_c(t)$ measures the net international migration. Uncertainty is introduced into the model by assuming that various key demographic processes such as fecundity, mortality etc are subject to random factors which subsequently through an appropriate hierarchical procedure can be introduced in (1), leading to the introduction of uncertainty in the fundamental quantities of interest such as $P_c(t)$, or its breakdown into age groups and sex for various future times t. Based on an extensive database of past world population data, the fundamental law (1), and the principles of Bayesian statistics, the probabilistic features of the uncertainty factors driving the population fluctuations are recovered. Then, using this information, the fundamental law (1) is iterated forward and used to obtain estimates for the future evolution of the quantities of interest. The estimates incorporate in a dynamically consistent fashion the effects of uncertainty as documented at least from the past data, and thus provide uncertainty consistent predictions for the future.

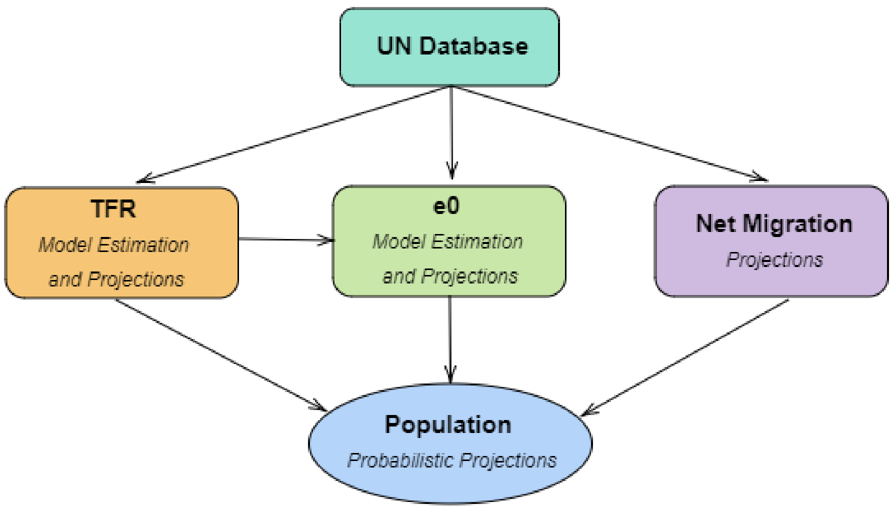

Figure 1: The probabilistic population modeling procedure

One of the key features of the model is this: it allows for quantities related to population projections to be random variables characterized by a probability distribution rather than point estimates. In particular, rather than producing a point estimate for a population related quantity $P(t)$ at time t (P can represent for example population for a particular age group or sex, or quantities such as fertility etc) the model, based on possible realizations of the uncertainty factors driving the phenomenon, treats $P(t)$ as a random variable, and produces (dynamically) a set of possible realizations $\{P^{(j)}(t) \ : \ j = 1, \dots, n\}$, which allow us to obtain approximations for the probability distribution of the random variable $P(t)$. Then, using this probability distribution one characterizes the quantity $P(t)$ with quantities carrying more information about it rather than just a point estimate, such as for example its percentiles at certain confidence levels or conditional means. These different realizations

$$\Pi := \{P^j(t) \ : \ t = T_0, \dots, T, \ j = 1, \dots, n\},$$

where $T_0 < T$ are two selected time horizons, will be referred to as trajectories, with $\{P^j(t) \ : \ t = T_0, \dots, T\}$ for fixed j representing a particular realization (i.e. a particular possible path) for the evolution of population in the time interval $[T_0, T]$ in the future. Clearly, only one of the above paths in Π, if any, will materialize. However, the set of paths Π provides us with important information concerning the probability of occurrence of paths with certain characteristics and allows for prediction of future population trends as well as the formulation of scenarios concerning these trends.

The following information on the structure of the probabilistic population model must be introduced here in order to make the SSP scenario generation procedure described in Section 2.1.2 more clear. In particular, the model of Raftery et al [11], relies on model (1), however treats separately the components $B_c(t)$ and $D_c(t)$ according to the probabilistic modeling approach mentioned above, using estimates and projections for the net migration from the UN base (or other databases) and then combines these approaches in order to construct projections per country or regionally by simulating trajectories. First, a hierarchical model is constructed for the Total Fertility Rate (TFR) component which provides projections for the fertility rates distribution at the country level and then the number of births distribution according to the approach presented in [1]. Then, this information is used to feed and build an hierarchical model for the Life Expectancy (e0) component according to the approach presented in [11], which is used to provide projections for life expectancy distributions of females and males per country at the various age-groups as well as to provide the mortality rates distribution on each age-group by gender. Next, available projections for migration (MIG), and in particular net migration per country, are collected by UN database and other data providers (like Wittgenstein Center[2] and are used as input to complete the required components for the population model (1). Note that there are similar hierarchical modeling approaches in the literature (e.g. [2]) however the lack or insufficiency of migration data for all the countries of interest makes the implementation of this model yet infeasible. At the final step, all the above components are combined and aggregated by the general population model (1) to provide the future population projections in terms of trajectories (possible scenarios) or distributions if conditioned to certain time instants. Note the above modeling task is implemented to the statistical software R[3] through the related package **bayesPop** described in detail in [12]. The whole modeling task is illustrated in Figure (1).

2.1.2 Socio-economics scenarios building and population projections

The concept of scenario making concerning future events has infiltrated environmental economics and has become a fundamental tool in the analysis. An important set of scenarios used frequently in analyses are the Sustainable Socio-economic Pathways scenarios (SSPs), which set certain plausible assumptions for key quantities (such as growth or fertility) in the future for certain parts of the world. One characteristic of these scenarios is that they are phrased in a qualitative fashion (see Table 1 and 2) so that they have to be transcribed to quantitative counterparts if they are to be used in concrete models.

[2] http://www.wittgensteincentre.org
[3] https://www.r-project.org/

	Country groupings	Fertility	Life expectancy	Migration
SSP1: rapid development	HiFert	Low	High	Medium
	LoFert	Low	High	Medium
	Rich-OECD	Medium	High	Medium
SSP2: medium	HiFert	Medium	Medium	Medium
	LoFert	Medium	Medium	Medium
	Rich-OECD	Medium	Medium	Medium
SSP3: stalled development	HiFert	High	Low	Low
	LoFert	High	Low	Low
	Rich-OECD	Low	Low	Low
SSP4: inequality	HiFert	High	Low	Medium
	LoFert	Low	Medium	Medium
	Rich-OECD	Low	Medium	Medium
SSP5: conventional development	HiFert	Low	High	High
	LoFert	Low	High	High
	Rich-OECD	High	High	High

Table 1: Shared Socio-economic Pathways (SSP) definitions

An important step in our approach is the transcription of the various Sustainable Socio-economic Pathways scenarios (SSPs) in the probabilistic setting in order to provide the population projections under each one of these directions. This approach divides the possible states of the world into five qualitative scenarios (rapid development, medium development, stalled development, inequality and conventional development) according to the levels of specific demographic characteristics and specifically fertility, life expectancy (or mortality), migration and education. Since the population projection method we follow does not take into account the education levels we omit this factor for the purposes of this work since does not directly affect the population evolution from the Bayesian model's perspective. Each country's SSP scenario may differ depending on its grouping as (a) high fertility country (HiFert), (b) low fertility country (LoFert) or (c) Rich-OECD country[4]. Specifications of the various SSPs with respect to the country groupings, as described in [6], are illustrated in Table 1.

One serious drawback of defining the scenarios in the above fashion is the inability to specify concretely what exactly is meant by low, medium and high. This lack of quantitative definition of the scenarios makes them difficult to apply in environmental modeling. However, as we show in this section, the probabilistic approach to modeling (such as for example the population model described in Section 2.1.1) can be very well blended with SSP qualitative scenarios, providing a concrete and realistic framework for scenario building, in which the various categories Low, Medium and High are endogenously and consistently selected by the dynamics and the evolution of the system under study.

More concretely, using the population modeling approach discussed in the previous section, the fertility and life expectancy components for each country of interest are provided in terms of the samples of trajectories created under the underlying probabilistic models. Therefore, the definition of the thresholds that separate *Low*, *Medium* and *High* scenarios, as mentioned in Table 1 specifications, for each one of these quantities should be done according to the observed (from the simulated sample, i.e. the components trajectories) variation. If the year 2100 is the time horizon we set, then the empirical distribution of the quantity of interest, as obtained from the simulated trajectories, can be used to define and quantify the various intensity levels corresponding to each scenario.

In particular, if we need to determine three different intensity levels, low, medium and high, then the 33% and 66% quantiles for fertility rate and life expectancy of the whole sample's

[4] http://www.oecd.org/about/members-and-partners/

distribution at the terminal time 2100 can serve as the discrimination thresholds. According to this rule, a trajectory is assigned to the high scenario if at the terminal year 2100 the observation for the corresponding quantity lies on the top 66% of the empirical distribution. Similarly, for the other levels. One advantage of this methodology is that the corresponding levels for the scenarios are not preassigned but are determined endogenously by the history and dynamics of the data from the population model. Repeating the above procedure for all trajectories, we end up with three sub-samples each corresponding to different possible realizations of the low, medium and high scenario. These sub-samples can be used to provide statistical information, such as moments, variability, etc. within scenarios. A possible question to this procedure could be about the validity of the initial sample and its ability to represent all possible future states of the world. Since the Bayesian model relies on the observed data from previous periods and takes into account possible relations of the country or region under study with all other countries and regions of the world, then any reasonable scenario (with respect to the data that have been collected up to the time the projection task is executed) should be amenable to simulation. Then, using a sufficiently large number of simulated trajectories, e.g. 100,000 trajectories, should guarantee that the results are reliable.

Let us describe the filtering rule that is incorporated for the classification of the initial sample of trajectories into different subgroups for each one of fertility rate and life expectancy simulated paths. In particular, from the induced distribution of the simulated TFR values for a country at a terminal time horizon T the various level scenarios are defined according to the rule:

- **Low TFR scenario**: all trajectories with TFR value lower than the 33% quantile value at time $t = T$
- **Medium TFR scenario**: all trajectories with TFR value larger than the 33% quantile value and lower than the 66% quantile value at time $t = T$
- **High TFR scenario**: all trajectories with TFR value larger than the 66% quantile value at time $t = T$

Similarly, from the simulated life expectancy paths for a country at time T the scenarios are defined according to the rule:

- **Low e0 scenario**: all trajectories with e0 value lower than the 33% quantile value at time $t = T$
- **Medium e0 scenario**: all trajectories with e0 value larger than the 33% quantile value and lower than the 66% quantile value at time $t = T$
- **High e0 scenario**: all trajectories with e0 value larger than the 66% quantile value at time $t = T$

Migration levels could be defined in the same manner if a probabilistic approach had been used, however in our case we use the deterministic net migration projections under each SSP scenario as provided by Wittgenstein Center database which are already discriminated to three intensity levels low, medium and high for each country.

In this manner, assigning each one of the countries of interest in the appropriate Country Grouping, country's specifications regarding fertility, life expectancy and migration for each one of the SSPs scenarios are set. Then, in order to construct a sufficient database for each SSP scenario, simulation of population trajectories for all countries of interest are drawn from population model (1) under each SSP's specifications defined by the aforementioned rules. For example, for a country belonging to the HiFert grouping, as mentioned in Table 1, the SSP2 scenario consist of the TFR trajectories that belong to the Medium TFR scenario, the e0

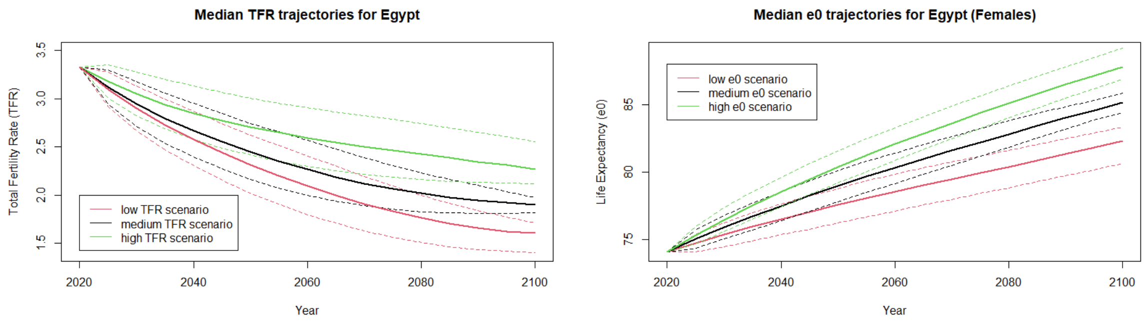

Figure 2: Median, 10% and 90% quantile fertility rate and life expectancy trajectories under low, medium and high level scenarios

trajectories that belong to the Medium e0 scenario and the net migration projection under the medium scenario. These three components, integrated by the population model (1) will provide the population trajectories that consist the SSP2 scenario for this country in a probabilistic manner, in the sense that we are able to compute quantiles, moments, etc. for the certain population dimensions for this country.

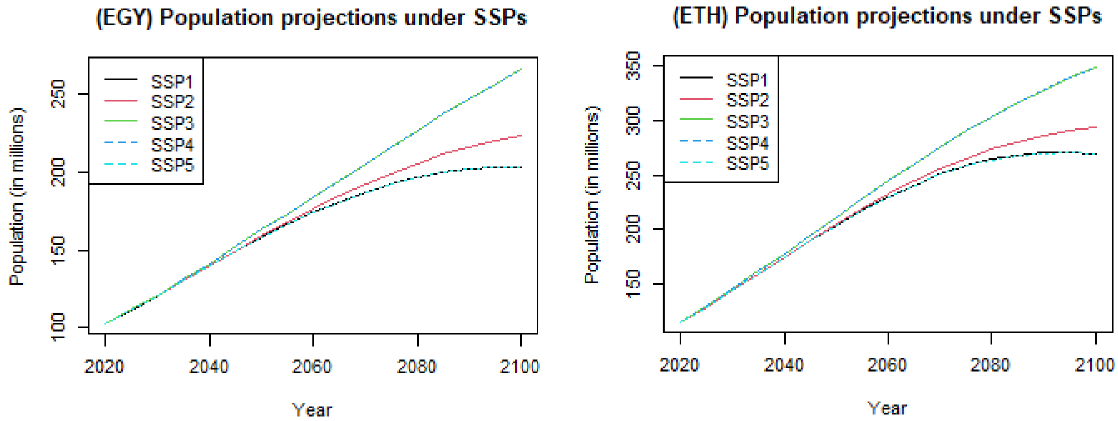

Figure 3: Median probabilistic population projections under each SSP scenario for Egypt and Ethiopia.

2.1.3 Generation of SSP compatible socio-economic and environmental parameters

Population is clearly the key driver for an economy's evolution. Building scenarios representing different pathways for the population evolution, offers a vehicle for the estimation of socio-economic quantities that depend mostly on population like labour force, gross domestic product (GDP) and others. There are several macroeconomic models that can provide such predictions based on population projections and provide projections for various socio-economic indicators of interest under the SSP scenarios. In this paper, the MaGE model [4] is employed for this purpose which based on UN database and IIASA[5] estimates provides projections up to year 2100 for the socio-economic activities of all countries of the world (being a global model) under each one of the SSP scenarios. Based on the strategy described in Section 2.1.2, MaGE model is employed by substituting its population inputs with the conditional means for population under each socio-economic scenario in order to provide estimates for GDP taking into account simultaneously all

[5] https://iiasa.ac.at/

economies of the world (under the perspective of MaGE model). This task could be done per population trajectory from the population scenario database, however such a task is extremely expensive in computational time and it would be required a new macroeconomic model to be built from scratch. Therefore, for the purposes of this work we use the projections (non-probabilistic ones) for GDP provided by MaGE for each one of the scenarios having in mind the possible limitations and drawbacks from this approach.

Scenario	Emissions Level	Temperature Change	Mitigation Measures
RCP1.9	Best-case	between $1 - 1.5^oC$	Extremely stringent
RCP2.6	Low	between $1.5 - 2^oC$	Very stringent
RCP4.5	Medium - Low	between $2.5 - 3^oC$	Less stringent
RCP6.0	Medium - High	between $3 - 3.5^oC$	Very loose
RCP7.0	High	up to 4^oC until 2100	Extremely loose
RCP8.5	Worst-case	up to 5^oC until 2100	No mitigation

Table 2: Representative Concentration Pathways (RCP) definitions

Since we are investigating food security issues, we should take into account environmental quantities like temperature, precipitation, etc since they directly affect water stress of countries and agricultural operations which is depicted in the the respective domestic production figures. The so called Representative Concentration Pathways (RCPs) determine scenarios regarding the increase on the mean temperature on the planet by the horizon 2100 taking into account the environmental policies that are to be or not to be adopted. These scenarios are divided by the levels of the increase concerning the mean temperature in the planet and the intensity of the mitigation measures that have to be adopted per scenario and some standard RCP scenarios are illustrated in Table 2.

Scenario Name	SSP scenario	RCP scenario
SSP1-1.9	SSP1	RCP 1.9
SSP1-2.6	SSP1	RCP 2.6
SSP2-4.5	SSP2	RCP 4.5
SSP3-7.0	SSP3	RCP 7.0
SSP4-6.0	SSP4	RCP 6.0
SSP5-8.5	SSP5	RCP 8.5

Table 3: The list of the SSP-RCP scenarios investigated in this work

The cutting-edge approach in creating realistic scenarios for the future states of the world, is to blend the concept of SSPs with that of RCPs. With a first glance, one would naively provide 30 different SSP-RCP scenarios, however this is not exactly the case since both types of scenarios, although referring to different target quantities, are not independent. The Coupled Model Intercomparison Project[6] (CMIP) studies these climate scenarios by integrating different environmental models and using very dense databases. Some mixed-type scenarios (SSP-RCP) that are up to now well tested and available are illustrated in Table 3. In order to provide food security evaluations in a framework which is compatible with both the socio-economic and climate pathways we perform estimations according to the six scenarios illustrated in Table 3.

2.2 Probabilistic projections on food demand and supply

In this section we propose a statistical model that may provide probabilistic projections on food demand and supply and hence also food security. As mentioned above, the key driver in this model will be population growth, which will be treated in a full probabilistic fashion using the detailed probabilistic model of Raftery et al [11]. The food security model also requires other

[6] https://www.wcrp-climate.org/wgcm-cmip/wgcm-cmip6

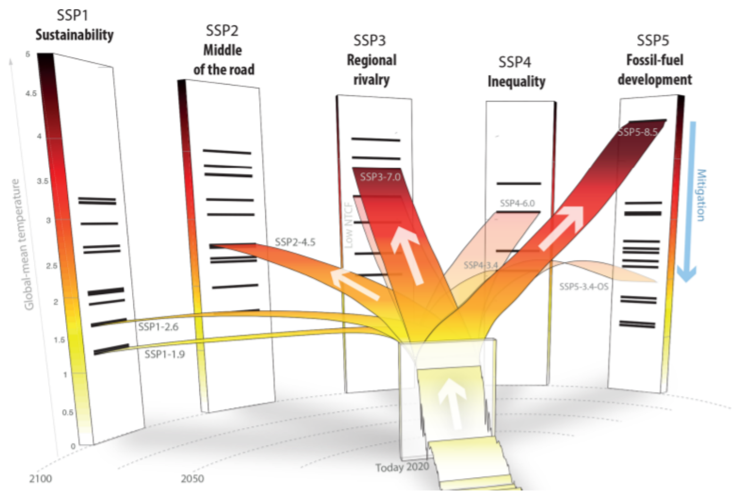

Figure 4: Graphical illustration of the SSP-RCP scenarios combination ([5])

important quantities, such as for example economic indices (GDP) or natural resources (land use or water, temperature, precipitation, etc) which will not be treated in a full probabilistic fashion for lack of sufficient computational resources.

2.2.1 Estimating the need for food

The need for food is inelastic, in the sense that humans need a minimum and a maximum daily intake of calories for subsistence. The calories intake requirements vary per age group, sex, and lifestyle (e.g. level of activity) varying in a range of 1000 to 3200 calories daily depending on the above mentioned categories. In Tables 4 and 5 these requirements are shown, as proposed by the HHS/USDA for the male and female population according to age group and activity level.

Age	Not Active	Somewhat Active	Very Active
2–3 years	1,000–1,200 calories	1,000–1,400 calories	1,000–1,400 calories
4–8 years	1,200–1,400 calories	1,400–1,600 calories	1,600–2,000 calories
9–13 years	1,600–2,000 calories	1,800–2,200 calories	2,000–2,600 calories
14–18 years	2,000–2,400 calories	2,400–2,800 calories	2,800–3,200 calories
19–30 years	2,400–2,600 calories	2,600–2,800 calories	3,000 calories
31–50 years	2,200–2,400 calories	2,400–2,600 calories	2,800–3,000 calories
51 years and older	2,000–2,200 calories	2,200–2,400 calories	2,400–2,800 calories

Table 4: Calories Needed Each Day for Boys and Men (Source: HHS/USDA Dietary Guidelines for Americans, 2010)

Given the age structure of the male and female population for the various countries we may then obtain detailed estimates for the food requirement in terms of total daily calories intake in terms of the quantity

$$C_c^R(t) = \sum_a R_a^f P_{a,c}^f(t) + \sum_a R_a^m P_{a,c}^m(t) \tag{2}$$

where $C_c^R(t)$ is the total daily recommended calories intake, c and t corresponds to country and t respectively, a corresponds to the age groups mentioned in Tables 4 and 5, $P_{a,c}^m(t), P_{a,c}^f(t)$ the total male and female population for the respective age groups and R_a^m, R_a^f are the requirements fiven in Tables 4 and 5. This estimate varies, depending on the activity level distribution of the population, however, one may obtain a lower bound for this quantity using the values for R_a^m, R_a^f

Age	Not Active	Somewhat Active	Very Active
2–3 years	1,000 calories	1,000–1,200 calories	1,000–1,400 calories
4–8 years	1,200–1,400 calories	1,400–1,600 calories	1,400–1,800 calories
9–13 years	1,400–1,600 calories	1,600–2,000 calories	1,800–2,200 calories
14–18 years	1,800 calories	2,000 calories	2,400 calories
19–30 years	1,800–2,000 calories	2,000–2,200 calories	2,400 calories
31–50 years	1,800 calories	2,000 calories	2,200 calories
51 years and older	1,600 calories	1,800 calories	2,000–2,200 calories

Table 5: Calories Needed Each Day for Girls and Women (Source: HHS/USDA Dietary Guidelines for Americans, 2010)

for non active individuals, an upper bound using these values for the very active individuals and a mean estimate using the average for the values of R.

The quantity $C_c^R(t)$ can be used as a proxy for the food demand (either a lower estimate, or an upper estimate or an average estimate). Clearly, in certain countries this may deviate from the actual total food demand (again measured in in total calories per day) on account of malnutrition issues related to poverty or unequal income distribution, etc. However, as actual data for total food demand are not easy to find, we use $C_c^R(t)$ as a reasonable proxy for food demand.

The probabilistic population model (see Section 2.1.1) provides accurate probabilistic predictions for the population pyramid, i.e., for the quantities $P_{a,f,c}(t), P_{a,c,m}(t)$. In particular, using the hierarchical Bayesian model of Raftery et al [10, 11]) we may obtain M different possible realizations of the population pyramid trajectories

$$\left\{P^{(j)}_{a,c,f}(t),\ t = T_0, \ldots, T\right\},\ \left\{P^{(j)}_{a,c,m}(t),\ t = T_0, \ldots, T\right\},\ j = 1, \ldots, M,$$

for the evolution of the female and male population per age group over the time period $[T_0, T]$. As already stated the uncertainty effects are properly accounted for in these trajectories and in accordance to past data. Taking a slice of, e..g., $\{P^{(j)}_{a,c,f}(t),\ t = T_0, \ldots, T\}$, $1, \ldots, M$ at a fixed $t' \in [T_0, T]$ will provide a sample $\{P^{(j)}_{a,c,f}(t'),\ j = 1, \ldots, M\}$ of M possible observations of the random variable $P_{a,c,f}(t')$ which can be used to obtain useful information concerning its distribution (i.e. moments, quantiles, etc). In fact, the general trajectories can be classified according to various criteria that characterize the SSP scenarios (see Section 2.1.2) so as to obtain subsets of the trajectories which are compatible with the various SSP scenarios, and hence obtain trajectories per scenario. Using the trajectories for each scenario we may obtain conditional means or quantiles for the conditional distribution of the quantities $P_{a,c,f}(t'), P_{a,c,m}(t')$ per SSP scenario. This procedure allows us to have a detailed probabilistic scenario based description of the possible evolution of future population related quantities per SSP scenario.

Having obtained the detailed trajectories and probabilistic scenarios for the population, using the estimate (2) for the total food demand, we may generate similar probabilistic scenarios for $C_c^R(t)$ and generate similar probabilistic projections for future food demand, based on the detailed modelling of the population structure. To this aim we have to use the trajectories $\{P^{(j)}_{a,c,f}(t),\ t = T_0, \ldots, T\}$, $\{P^{(j)}_{a,c,m}(t),\ t = T_0, \ldots, T\}$, $j = 1, \ldots, M$, for the evolution of the female and male population per age group over the time period $[T_0, T]$ already obtained, to generate similar trajectories $\{C_c^{R,(j)}(t),\ t = T_0, \ldots, T\}$, $j = 1, \ldots, M$, which will subsequently be used to generate samples for projections of C_c^R on various future dates $t' \in [T_0, T]$ and from those as described above generate probabilistic information on this important quantity. Clearly, when this quantity is needed in the context of SSP scenarios the relevant trajectories for the population quantities corresponding to these scenarios must be used in the generation of trajectories in (2) for the total food demand.

2.2.2 Total food supply modeling

Modeling and predicting the total food supply (in terms of calories per day) is also an important issue. For instance if the total food supply cannot cover $C_c^R(t)$ for a particular country at a particular time, then major issues related to food sufficiency arise. Predicting possible future food shortages is of paramount importance to policy makers as it allows for proactive measures to be taken for example in restructuring food production, land use, etc or planning for sufficient imports.

Socioeconomic conditions and natural resources are expected to play a crucial role in total food supply modeling. To better incorporate these features in our model, we break the total food supply into two distinct contributions, domestic produced supply S_c^D and imported supply S_c^I. Domestic supply corresponds to the total quantity of food (in calories) produced within country c, which crucially depends on natural resources in country c, such as land and water, as well as labour in the food production sector (agriculture or livestock). On the other hand, imported food supply S_c^I for country c depends more on socioeconomic factors such as for example GDP or the population level. The total food supply is the sum of these two contributions.

Historical data for total food supply can be found in FAO database[7]. Our working hypothesis is that total food supply must depend on the population structure (which essentially forms food demand) as well as on economic quantities. Relevant economic quantities can be land use data, the percentage of labor force employed in the food production sector, etc. However, often such detailed data may not be available for all countries of interest, and particularly for those for those in the developing world (non-OECD countries). Moreover, even if such data were available, predicting future values of these data in various scenarios and in a form compatible with the probabilistic population projections, may be difficult or computationally impossible at least with the current state of the art of economic modeling and computational resources respectively.

To this end we opt for a minimal model, for the prediction of the total food supply per country, based on its major driver which is population, and for which we have a detailed probabilistic model which can be used for probabilistic predictions on future dates, and an aggregated economic quantity, which we choose to be the per capita GDP, I_c. Our modeling assumption is that total food supply S_c can be expressed as

$$
(3) \qquad \begin{aligned}
&(3a)\;\; S_c(t) = \gamma_0 (S_c^D)^{\gamma_1}(t) I_c^{\gamma_2}(t), \;\; \text{(total food supply)} \\
&(3b)\;\; S_c^D(t) = \beta_0\, e^{\beta_1 t} A_c^{\beta_2}(t) L_c^{\beta_3}(t) W_c^{\beta_4}(t) T_c^{\beta_5}(t), \;\; \text{(domestic production)} \\
&(3c)\;\; W_c(t) = \alpha_0 e^{\alpha_1 t} A_c^{\alpha_2}(t) Pr_c^{\alpha_3}(t) T^{\alpha_4}(t), \;\; \text{(water stress)},
\end{aligned}
$$

where

- $S_c(t)$ is the total food supply of country c and time t,
- $S_c^D(t)$ is the domestic production of food of country c and time t,
- $W_c(t)$ is the level of water stress of country c at time t, i.e. the freshwater withdrawal in percentage of the available freshwater resources (according to the SDG Indicator 6.4.2[8]),
  - $I_c(t)$ is the GDP of country c at time t defined as $I =$ GDP per capita $\times$ Population,
  - A_c is the area of cropland of country c at time t,
  - $L_c(t)$ is the labour force occupied in the agricultural sector of country c at time t,

[7] http://www.fao.org/

[8] https://www.fao.org/publications/card/en/c/CA8358EN/

– $T_c(t)$ is the (average) temperature in country c at time t,
– $Pr_c(t)$ is the (average) precipitation in country c at time t,

and α_i, β_i and γ_i are constants to be determined from past data. These unknown parameters can be estimated using typical least squares estimation procedures, upon taking logarithms for the model (3). We emphasize that these parameters are country dependent but we omit the subscript c, so as not to clutter the notation.

The following comments are in order concerning model (3): Model (3) is a three stage model, which models total food supply in terms of socioeconomic quantities (e.g. income, labour etc), natural resources (e.g. water stress, cropland etc) and environmental variables (e.g. temperature, precipitation). Population enters model (3) through various routes; it clearly affects the labour force as well as the income. Natural resources as well as environmental resources affect the domestic food production given by model (3b). Finally, the total food supply allows for both domestic or imported food supply, and this depends on the level of domestic supply as well as on economic quantities - GDP - which may play an important role on the ability of the country to acquire imported goods. As already mentioned, the nature of the data available requires a minimal aggregate model, containing explanatory variables which include socioeconomic, natural resources and environmental variables, and the proposed model seems to be well suited in this respect - as will soon be verified in the next section. Moreover, it contains quantities which are scenario dependent hence, it will allow us to assess the effect of various scenarios on the quantities of interest which are related to food security. Finally, the estimation of the model will be done in 3 stages, we will begin with the estimation of water stress using (3a), then we proceed to the estimation of domestic production using (3b), and finally proceed to the estimation of total food supply using (3c).

2.2.3 Scenarios for food security indices

A reasonable choice for a food security index is

$$\mathcal{I}_c(t) := \frac{S_c(t) - C_c^R(t)}{C_c^R(t)}. \tag{4}$$

If $\mathcal{I}_c(t)$ admits positive values then no food shortage is expected and country c is food secure. If on the contrary $\mathcal{I}_c(t)$ admits negative values then country c will face food security issues at time t. Clearly, $\mathcal{I}_c(t)$ is a random variable and scenarios concerning its possible future realizations may reveal important information concerning food security in country c and its dependence on various socioeconomic factors and policy decisions.

Combining the steps and procedures described in the previous sections we may provide probabilistic scenarios compatible with the SSP scenarios for the food security index $\mathcal{I}_c(t)$ as follows:

1. Model (3) is tuned using historical data from the period 1980-2019 for country c to obtain the relevant country dependent parameters $\alpha_i, \beta_i, \gamma_i$.
2. The fitted model (3) is then used for predicting the future total food supply $S_c(t)$ using the probabilistic scenarios (and the relevant trajectories) for the population (see Section 2.1.2) along with projections for the future GDP per capita as obtained from the global macroeconomic model MaGE [4].
3. We use the procedure in Section 2.2.1 to provide future estimates and SSP compatible scenarios for the food demand $C_c^R(t)$.
4. Using the trajectories for $S_c(t)$ and $C_c^R(t)$ obtained in the previous steps we construct trajectories compatible with the various SSP scenarios for the index $\mathcal{I}_c(t)$ and use the trajectories to provide statistical information for the index in the various scenarios.

2.3 Risk quantification for food security: Robust estimates across scenarios

The procedure described in Section 2.2.3 provides trajectories of possible future realizations of the food security index within scenarios, i.e., conditional that a particular SSP scenario has materialized. While this procedure provides important information it has a major drawback: One cannot know in advance which of the possible SSP scenarios will materialize. This drawback has very important policy implications, especially if policy measure related to a long horizon have to be considered. As different measures will be the required ones over different scenarios, a robust policy must be designed which will perform reasonably well over all possible scenarios. This calls for a robust estimation of food security risk, i.e. an estimation of food security risk which will work across scenarios rather than within scenarios (as the one discussed in Section 2.2.3)

The methodological framework of convex risk measures and their robust representations (see e.g. [3]) and in particular the recent developments concerning the construction of convex risk measures or variational utilities that take into account model uncertainty employing the concept of Fréchet mean in Wasserstein space (introduced in [8] see also [9]) turns out to be the ideal setting to treat such questions. Employing the standard framework of risk management, consider a risk L (assumed to be a random variable on a properly selected measurable space) which depends on another set of (possibly vector valued) random variables $\mathbf{Z}$ usually called the risk factors. Fluctuations in the risk factors affect the fluctuations of the risk through the mapping $\mathbf{Z} \to L =: \Phi(\mathbf{Z})$, called the risk mapping. A probabilistic model Q concerning the possible random evolution of the risk factors $\mathbf{Z}$, will induce (through the risk mapping Φ) a probabilistic model Q' for the evolution of the risk L. Otherwise stated, scenarios for the evolution of the risk factors $\mathbf{Z}$ will induce scenarios for the risk L. If a single probabilistic model Q was universally acceptable for $\mathbf{Z}$, this would lead to a single probabilistic model for L, hence, the best estimate for the risk would simply be $\mathbb{E}_{Q'}[L] = \mathbb{E}_Q[\Phi(\mathbf{Z})]$. Within the framework discussed here, if we knew which SSP scenario was to materialize, this will indicate a single probabilistic model for the risk factors $\mathbf{Z}$ and hence the best estimate for the risk L would be $\mathbb{E}_Q[\Phi(\mathbf{Z})]$ (related to the conditional mean for the particular scenario).

However, what often happens in reality is that we do not know of a universally accepted probability model Q for the risk factors $\mathbf{Z}$ so that there is a whole set of probability models $\mathcal{Q} = \{Q_i, \ : \ i = 1, \dots, J\}$, that may provide information concerning the evolution of the risk factors $\mathbf{Z}$. This brings us to the realm of Knightian uncertainty, which requires a better way of providing the best estimate for the risk L. In such cases, robust estimations of the risk L can be proposed. A proposal which is well accepted by the community is the variational form

$$\rho(L) := \sup_{Q \in \mathcal{Q}} [\mathbb{E}_Q[\Phi(\mathbf{Z})] - a(Q)], \tag{5}$$

where $\rho(L)$ called the risk measure of the risk L, considered as an estimation of the risk, and $a : \mathcal{Q} \to \mathbb{R}_+$ is a (convex) penalty function in the space of probability models, which penalizes certain probability models as extreme or improbable. The risk measure defined in expression (5) proposes as an estimation of a risk L, for which one cannot trust a single probability model Q, the worst case expected risk over all probability models, properly weighted by a penalty function which penalizes certain probability models as too extreme. The variational nature of formula (5) gives a robustness flavour to the proposed risk measure, as it no longer depends on the adoption of a single model for L, but rather provides an appropriately weighted estimate for L over the whole universe $\mathcal{Q}$ of plausible models for L.

There are various possibilities for the choice of the penalty function a, leading to an interesting variety of convex risk measures. A recent development in the field ([8] see also [9]) proposed that the penalty function is related to the variability of the plausible models, quantified in terms of the Fréchet function of the set of plausible models $\mathcal{Q}$ considered as an element of the Wasserstein space. This choice, has certain advantages, among which is the possibility of analytic approximation of the risk measure as well as an interesting interpretation of the resulting probability model used for the estimation of $\rho(L)$ as the outcome of an experts agreement procedure and the quantitative connection between risk and uncertainty measures.

This framework can be used to study the problem of food uncertainty by introducing the following analogies. Let us fix a time t. We will set the risk variable as $L = \mathcal{I}(t)$, defined as in (4), which is a random variable, depending on a set of risk factors $\mathbf{Z}(t)$, which in the present context correspond to the population group factors $P_{c,a,m}(t), P_{c,a,f}(t)$, and possibly other socioeconomic factors (e.g. GDP). For the sake of concreteness and without loss of generality we define $\mathbf{Z} = (P_{c,a,m} P_{c,a,f})$ for various (fixed) time instants t. The risk mapping $\mathbf{Z} \to L = \Phi(\mathbf{Z})$ is provided by the composition of the relations described in equations (2), (3) and (4). Each SSP scenario, $\mathcal{S}_i$, corresponds to a probability model Q_i for the risk factors $\mathbf{Z}$, so that we may obtain the set of plausible models (scenarios) $\mathcal{Q} = \{Q_i, \ : \ i = 1, \dots, J\}$, where J is the total number of scenarios. The robust estimation of the food security index $\mathcal{I}$ will correspond to a convex risk measure $\mathcal{I} = \rho(L)$ for the relevant risk variable.

The above procedure is very well suited to the problem at hand for various reasons. On account of the long term nature of the population projections there is a high level of uncertainty (in the Knightian sense) involved in them, hence a robust way of estimating important quantities such as the total future food demand in various socioeconomic scenarios is needed; a methodology that must be based on the probabilistic nature of the demographic projections rather than point estimates of the required quantities. The methodology of convex risk measures, on account of their robust representation over a set of possible probability laws, provides a very good solution to such issues.

Having adopted the fundamental conceptual framework of treating each scenario as a different probabilistic model (probability measure) for the risk factors $\mathbf{Z}$ (mainly population factors in this study) we may now answer the question of robust estimation of the quantity of interest $L = \mathcal{I}_c(t)$, for fixed t, using a convex risk measure $\rho(L)$ of the form (5). The choice of penalty function is important. Since there is need to differentiate between various scenarios (understood as probability measures) the penalty function must be composed of quantities that effectively differentiate between probability measures, as for instance a metric in the space of probability measure. Here we follow up on the suggestion in Papayiannis et al [8] and Petrakou et al [9] and choose

$$a(Q) = \frac{\theta}{2} \sum_{i=1}^{J} w_i d^2(Q, Q_i), \tag{6}$$

where Q_i, $i = 1, \dots, J$ are the probability measures for the risk factors $\mathbf{Z}$ corresponding to each of the scenarios, w_i, $i = 1, \cdots, J$ are (credibility) weights associated to each of the scenarios (these can be subjective and associated to expert opinion or objective i.e. derived from evidence from the data and possible updated through a learning scheme), $d(\cdot, \cdot)$ is a metric in the space of probability measures and $\theta > 0$ is a parameter modeling uncertainty aversion. A suitable choice for d is the Wasserstein metric,

$$d^2(Q, Q_i) = \min_{Z \sim Q, \ Z' \sim Q_i} \mathbb{E}[(Z - Z')^2] \tag{7}$$

arising from optimal transportation, which is directly related to the misspecification error of the random variable L is a different probability model for $\mathbf{Z}$ is chosen in the place of the true model. Morover, the choice (5), with (6) and (7), allows for efficient numerical calculation of the risk measure $\mathcal{I} = \rho(L)$ for a wide class of probability measures for the risk factors $\mathbf{Z}$. In particular, in the case of large values of θ, one may approximate $\mathcal{I}$ by

$$\mathcal{I} = \rho(L) = \mathbb{E}_{Q^*}[\Phi(\mathbf{Z})],$$

where Q^* is the barycentric probability measure over all the scenarios, accordingly weighted with the weights w_i, $i = 1, \dots, J$ defined by

$$Q^* = \arg\min_{Q \in \mathcal{P}} \sum_{i=1}^{J} w_i d^2(Q, Q_i), \tag{8}$$

where the minimization is performed over the space $\mathcal{P}$ of probability measures, metrized by the Wasserstein metric (7). The calculation of the barycenter Q^* if feasible within the class of location-scale distributions with a matrix iteration algorithm which is reasonably easy to implement. The risk measure $\mathcal{I}$ as stated above provides a robust estimate for the food insecurity index, across scenarios, in the limit of deep uncertainty. This is the main approximation we will be using in this work. Further approximations are possible using the Δ- approximation of the risk mapping $\mathbf{Z} \to \Phi(\mathbf{Z})$ (see [8]).

The algorithmic approach in estimating the food insecurity index can be thus summarized as follows:

1. Fix a time t
2. Define the risk factors $\mathbf{Z} = (P_{c,a,f}, P_{c,a,m})$ and using the procedure in Section 2.1.2 obtain probabilistic scenarios for $\mathbf{Z}$ and determine the corresponding probability models Q_i, $i = 1, \dots, J$.
3. By estimating the model (3) and using the expression (2) obtain the risk mapping $\mathbf{Z} \mapsto L =: \Phi(\mathbf{Z})$ for $L = \frac{S_c(t) - C_c^R(t)}{C_c^R(t)}$.
4. Obtain the barycentric scenario Q^*.
5. Using Monte-Carlo simulation estimate $\mathcal{I} = \mathbb{E}_{Q^*}[\Phi(\mathbf{Z})]$.

3 Food security risk assessment for Egypt and Ethiopia under the SSP-RCP scenarios

3.1 Modeling assumptions, data availability and model fitting

We follow the modeling approach presented in Section 2.2 for describing minimum required food consumption (Section 2.2.1) and total food supply (Section 2.2.2) in order to provide estimates for the food security through the food security index is described in Section 2.2.3. For the food consumption component it suffices to use the probabilistic estimates concerning the population evolution for Egypt and Ethiopia, since the population is the only factor. On the other hand, food supply is described in terms of the three-stage nested model (3) with layers water-stress evolution, domestic food production and total food supply.

Effect	Coefficient	Egypt	Ethiopia
Intercept	α_0	4.374	6.845
Year	α_1	-0.007	-0.005
Cropland usage	α_2	1.139	3.991
Precipitation	α_3	-0.041	-0.396
Temperature	α_4	0.288	0.429
R-squared adjusted		0.315	0.959

Table 6: Water stress models estimated coefficients for Egypt and Ethiopia

On the lowest layer, water stress depends on Year, Cropland usage, Precipitation and Temperature. The required data for the model fitting procedure for Egypt and Ethiopia were provided by the publicly open database from FAO. The estimated models parameters are illustrated on Table 6. A negative relation appears between water stress and precipitation which seems to be more intense in Ethiopia. Egypt faces seriously water scarcity issues while annual precipitation levels are about ten times lower than Ethiopia's. As a result is quite meaningful for Ethiopia to be more affected by an increase to the precipitation levels (by reducing more significant the related water stress indicator) than Egypt. Note that latest water stress figures appear to be higher than 110% while Ethiopia's is about to 40%. Cropland usage is also taken into account through the ratio of the land equipped for crops which is actually used to the maximum land area that can be used. The rationale besides using the ratio instead of the land area is to provide limits to our model regarding the natural resources usage. In this direction, for Egypt and Ethiopia we provided a maximum land area that can be used as cropland until 2050 in order to introduce this physical constraint on the model concerning water stress.

Effect	Coefficient	Egypt	Ethiopia
Intercept	β_0	4.815	-9.716
Year	β_1	0.008	0.024
Cropland usage	β_2	0.765	1.902
Labour agriculture	β_3	0.374	1.380
Water Stress	β_4	-0.236	-0.598
Temperature	β_5	-0.345	2.584
R-squared adjusted		0.958	0.983

Table 7: Domestic production models estimated coefficients for Egypt and Ethiopia

On the second layer, domestic food production is assumed to be affected by Year, Cropland usage, Labour force occupied in agriculture, water stress and temperature. Estimated model parameters (Table 7) seem to indicate similar qualitative effects of the factor variables to domestic food production. Possible increase in water stress affects negatively the production quantity while temperature increase affects differently the two domestic production models. Egypt's domestic production is affected negatively by an increase in the temperature while Ethiopia is affected positively, probably due to different climate and terrain morphology conditions between the two countries.

Effect	Coefficient	Egypt	Ethiopia
Intercept	γ_0	9.608	8.682
Domestic Production	γ_1	0.247	0.604
GDP (total)	γ_2	0.262	0.201
R-squared adjusted		0.967	0.994

Table 8: Total food supply models estimated coefficients for Egypt and Ethiopia

Total food supply model estimates are illustrated in Table 8 for both countries. Both for Egypt and Ethiopia total food supply depends positively on domestic food production and total GDP (which contains the population effect internally).

3.2 Food security risk estimation under SSP-RCPs

After the successful modeling task of both food consumption and food supply we can derive estimates for the time period 2020-2050 for Egypt and Ethiopia under each SSP-RCP scenario as described in Table 3. Food consumption depends only on population, while food supply relies on population, labour force occupied in agriculture, GDP, percentage of available land area for crops used, temperature and precipitation. Population projections under each SSP are directly available through the probabilistic model discussed in Section 2.1.2 while labour force amounts are estimated by combining MaGE model participation rates and proper scaling. GDP projections are also provided by MaGE model (non-probabilistic) for each SSP scenario. The environmental quantities temperature and precipitation (in annual basis) are provided for each one of the scenarios in Table 3 from the Climate Change Knowledge Portal online database[9]. The percentage of the available cropland used is estimated through each country's trend in last decade which is restricted by the upper threshold that has been set for each country providing in this manner physical restrictions. In this way, land ratio is provided in the same manner for all scenarios (however it is an easy modification if one desires to provide different scenarios) but the water stress projections, since depend also on temperature and precipitation levels, are different on each SSP-RCP scenario.

Scenario	2020	2025	2030	2035	2040	2045	2050
Egypt (EGY)							
SSP1-1.9	1.06	1.08	1.10	1.12	1.13	1.14	1.17
SSP1-2.6	1.06	1.08	1.11	1.12	1.13	1.14	1.17
SSP2-4.5	1.05	1.05	1.05	1.05	1.04	1.04	1.04
SSP3-7.0	1.03	0.99	0.95	0.91	0.87	0.83	0.80
SSP4-6.0	1.08	1.09	1.10	1.10	1.10	1.10	1.11
SSP5-8.5	1.09	1.12	1.16	1.19	1.22	1.25	1.30
Ethiopia (ETH)							
SSP1-1.9	0.72	0.93	1.16	1.38	1.68	1.94	2.21
SSP1-2.6	0.70	0.91	1.12	1.37	1.64	1.94	2.29
SSP2-4.5	0.67	0.88	1.08	1.32	1.57	1.84	2.16
SSP3-7.0	0.68	0.85	1.00	1.22	1.46	1.69	1.96
SSP4-6.0	0.71	0.89	1.12	1.36	1.62	1.96	2.29
SSP5-8.5	0.70	0.94	1.20	1.47	1.82	2.17	2.54

Table 9: Food security index for Egypt and Ethiopia for the time period 2020-2050 for each SSP-RCP scenario

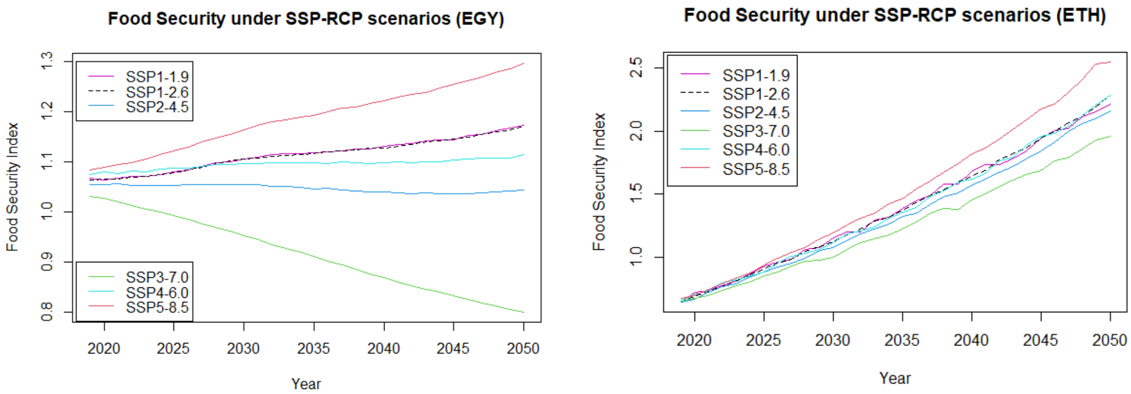

Figure 5: Graphical illustration of the food security index for Egypt (left) and Ethiopia (right) for the time period 2020-2050 under each scenario.

Combining all the above projections we derive our estimations concerning the food security for both countries under each scenario, in terms of the food security index introduced in Section 2.2.3. The results are illustrated in Table 9 and Figure 5. It is evident that neither Egypt or Ethiopia faces serious risk concerning food security for the next thirty years for all SSP-RCP scenarios. However, the food security index evolution presents greater homogeneity for Ethiopia, since all curves are very close indicating an increasing trend. Egypt's index, seems

[9] https://climateknowledgeportal.worldbank.org/

to be close to 1, with the majority of scenarios indicating a slightly increasing trend. SSP5-8.5 scenario indicates the most rapid increase to food security while SSP3-7.0 scenario indicates rapid decline to the food security index, however its value is quite high at the end of the time horizon (2050).

3.3 Food security robust risk estimates across SSP-RCPs

Although it is useful to obtain the food security index estimate for each SSP scenario, this risk evaluation is characterized by the main drawback that it is not robust to the uncertainty concerning to which scenario materializes. Since the policy maker needs to take into account all possible scenarios in order to make a robust decision, we employ the approach discussed in Section 2.3. In terms of a toy example, we provide three different perspectives of decision makers: (a) one who has complete ignorance of the situation, so weights all possible outcomes equally, (b) the optimistic one, who places higher probability to scenarios more favourable for the economy and the environment and (c) the pessimistic one, who allocates higher probability to scenarios that are less favourable for the economy and the environment. These perspectives are illustrated in Table 10.

Expert's Opinion	SSP1-1.9	SSP1-2.6	SSP2-4.5	SSP3-7.0	SSP4-6.0	SSP5-8.5
A. Ignorance	1/6	1/6	1/6	1/6	1/6	1/6
B. Optimistic	1/4	1/5	1/5	1/10	1/5	1/20
C. Pessimistic	1/20	1/10	1/5	1/5	1/5	1/4

Table 10: Realization probabilities of each SSP-RCP scenario according to different perspectives

Practically, several econometric models can be used based on data from the economy, industry, demography, etc. of a region in order to estimate the probability of following a certain SSP direction, or even on expert's opinion. The knowledge of these probabilities allows for a robust risk estimation. For the case of food risk security the probabilities depicted in the above table are used as the weighting strategies that are used to estimate the aggregate probability model as determined in (8).

Perspective	2020	2025	2030	2035	2040	2045	2050
Egypt (EGY)							
Ignorance	1.062	1.067	1.079	1.079	1.080	1.085	1.098
Optimistic	1.062	1.068	1.080	1.081	1.082	1.088	1.102
Pessimistic	1.063	1.067	1.076	1.076	1.075	1.079	1.090
Ethiopia (ETH)							
Ignorance	0.698	0.899	1.112	1.354	1.631	1.923	2.240
Optimistic	0.700	0.899	1.113	1.351	1.622	1.911	2.221
Pessimistic	0.695	0.895	1.108	1.353	1.633	1.931	2.255

Table 11: Food security index for Egypt and Ethiopia for the time period 2020-2050 under different perspectives

In Table 11 and Figure 6 are illustrated the robust estimates per case. For the case of Egypt, although all estimates are similar there is a distinction between them where the optimist perspective drives the curve at the highest level while the pessimist the curve at the lowest level. The distinction in Ethiopia is not that clear since all SSP-RCP are quite homogeneous and the time horizon is probably too short to provide clear differences for this case. However, the robustness of this approach is evident from the fact that even in case where high level of heterogeneity exist on the scenario set, the final risk estimate is quite similar independently on the weighting strategy used.

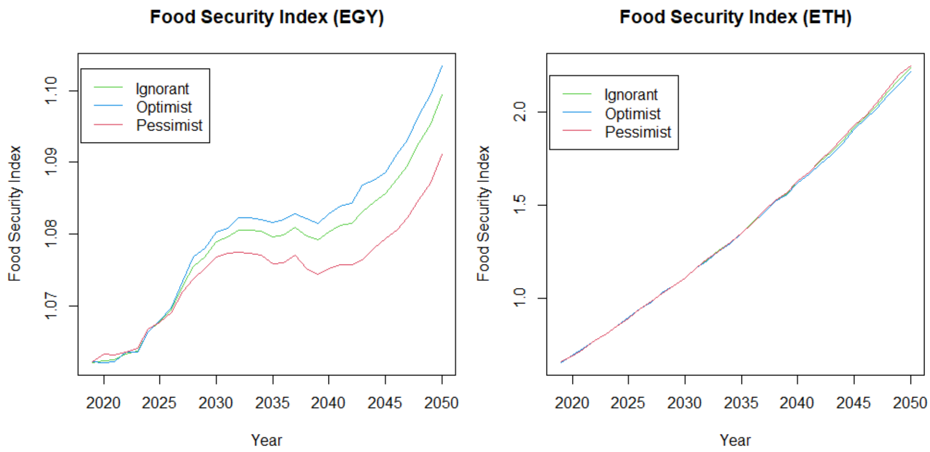

Figure 6: Graphical illustration of the across scenario estimations of the food security index for Egypt (left) and Ethiopia (right) for the time period 2020-2050.

4 Conclusions

In this paper we propose a general methodology for producing probabilistic socio-economic scenarios compatible with the SSP framework. The probabilistic scenarios represent the effects of the inherent uncertainty more efficiently than points estimates and are therefore better suited for projecting important socio-economic quantities into the future. As a possible application of this methodology we consider the issue of food security, we provide a plausible index for its assessment and propose a framework for providing quantitative evaluations per scenario, and future projections. While projections along scenarios is important, in practice there is uncertainty as to which scenario materializes. To address this question, we also propose a framework for providing projections across scenarios using the concept of convex risk measures and their robust representation in the presence of model uncertainty. Our approach is illustrated by an application on two major countries in the upper Nile region, Egypt and Ethiopia.

Acknowledgements

This project has received funding from the European Unions Horizon 2020 partnership for research and innovation in the Mediterranean area programme (PRIMA) under grant agreement No 1942 (AWESOME).

References

[1] Alkema, L., Raftery, A. E., Gerland, P., Clark, S. J., Pelletier, F., Buettner, T., & Heilig, G. K. (2011). Probabilistic projections of the total fertility rate for all countries. *Demography*, 48(3), 815-839.

[2] Azose, J. J., & Raftery, A. E. (2015). Bayesian probabilistic projection of international migration. *Demography*, 52(5), 1627-1650.

[3] Föllmer, H., & Schied, A. (2010). Convex and coherent risk measures. *Encyclopedia of Quantitative Finance*, 355–363

[4] Fouré, J., Bénassy-Quéré, A., & Fontagné, L. (2013). Modelling the world economy at the 2050 horizon. *Economics of Transition*, 21(4), 617-654.

[5] Meinshausen, M., Nicholls, Z. R., Lewis, J. et al. (2020). The shared socio-economic pathway (SSP) greenhouse gas concentrations and their extensions to 2500. *Geoscientific Model Development*, 13(8), 3571-3605.

[6] Lutz, W., Butz, W. P., & Samir, K. E. (Eds.). (2014). World population and human capital in the twenty-first century. *Oxford University Press.*

[7] Lutz, W., Goujon, A., Kc, S., Stonawski, M., & Stilianakis, N. (2018). Demographic and human capital scenarios for the 21st century: 2018 assessment for 201 countries. *Publications Office of the European Union.*

[8] G. Papayiannis and A. N. Yannacopoulos, Convex risk measures for the aggregation of multiple information sources and applications in insurance, *Scandinavian Actuarial Journal*, 2018 (9), 792-822.

[9] E. V. Petracou, A. Xepapadeas & A. N. Yannacopoulos. (2021). Decision Making Under Model Uncertainty: Fréchet-Wasserstein Mean Preferences, *Management Science* (in press) ISSN 0025-1909 (print), ISSN 1526-5501 (online)

[10] Raftery, A. E., Li, N., Sevciková, H., Gerland, P., & Heilig, G. K. (2012). Bayesian probabilistic population projections for all countries. *Proceedings of the National Academy of Sciences*, 109(35), 13915-13921.

[11] Raftery, A. E., Chunn, J. L., Gerland, P., & Sevciková, H. (2013). Bayesian probabilistic projections of life expectancy for all countries. *Demography*, 50(3), 777-801.

[12] Sevciková, H., & Raftery, A. E. (2016). bayesPop: probabilistic population projections. *Journal of Statistical Software*, 75.