63886

Economic Commission for Africa

Population Environment Development Agriculture Model

Technical Manual

Addis Ababa, Ethiopia
May 2007

Table of Contents

1 The PEDA model: an advocacy tool modeling the interrelationships between population, development, the environment and agriculture

1.1 Background

In the last two decades, Africa experienced a severe crisis, manifested in the constant decline of its economic growth rate, among other indicators. Since 1994, the economic situation of the continent has improved steadily, but this resumption is still below the necessary level to have a significant impact on poverty. Although the proportion living in extreme poverty in Africa may have declined with a few percentage points in the last couple of years[1], high population growth rates mean that the absolute number of poor still increases dramatically on a daily basis. Additionally, Africa's economic performance seems to be highly dependent on the international economic environment and on weather conditions - two exogenously determined factors that do not embody any guarantee for future growth (ECA, 1999).

Two basic pressures account for the continued deterioration in the quality of life in Africa. First, the population growth rate exceeds that of economic growth and food production in most African countries (Table 1-1). Additionally, the rapid deterioration of the environment on the continent prevents the desired increase in agricultural productivity. Today, over three-fourths of sub-Saharan African countries produce less food per capita than they did in the 1980s. Daily calorie availability is well below the recommended minimum and as much as 30-40 per cent of the population is undernourished. Malnutrition affects even more people. Food insecurity, rapid population growth and environmental degradation constitute a very important challenge for public policy in Africa today.

Table 1-1: Africa's economic performance vs. population growth, 1965-98

Indicator	Years					
	1965-73	1974-79	1980-85	1986-89	1990-94	1995-1998
Population growth rate	2.6	2.7	2.8	2.7	2.5	2.4
Growth rate of GDP (average)	5.7	3.5	1.8	2.5	1.9	3.3
Growth rate of per capita GDP (average)	3	0.7	-1.1	-0.5	-1.1	0.7
Growth rate of agricultural output (average)	2.7	3	1.5	2.7	2.1	6.2

***Source:** ECA Secretariat*

1 Currently, around 52 per cent of Sub-Saharan Africans are living on less than $ 1 a day (ECA, 2001).

Inspired by the notion of sustainable development as reflected in the series of United Nations conferences during the 1990s, there is increasing understanding of the necessity to go beyond the traditional sectoral approach to national development to one that captures the interaction between sectors and interdependencies between policy objectives. It has been demonstrated that, at least in the medium to long run, a country's economic performance and the food security of its citizens are closely related to its demographic and educational trends as well as to the health of the natural environment. Since these issues are closely interconnected in the real world, they should also be viewed together in national politics and development planning. Not considering this inter-sectoral nexus could have serious repercussions in the future.

The scientific understanding of mutual interdependencies is, however, not yet sufficiently reflected in the political institutions of individual countries. There is tremendous inertia in such systems, partly due to the traditional training of experts that is often characterized by compartmentalization of disciplines, and partly to the fact that the impact of developments in one sector is often invisible in another sector in the short term.

Hence, convincing policy makers and country experts of the negative synergy arising from the interconnections of population growth, environmental deterioration and declining agricultural production is a major objective of the Sustainable Development Division (SDD) of the Economic Commission for Africa (ECA). With that goal in mind, the SDD engaged itself to develop a computer simulation model intended to illustrate the interactions between population changes (P), the environment (E), socio-economic development (D) and agriculture (A).

The first prototype of the model was formally presented to a group of experts including invited scientists from member States and affiliate international organizations in November 1998 at the ECA in Addis Ababa. Provisional data for three countries (Burkina Faso, Madagascar and Zambia) have been prepared to test the assumptions and structure of the model and the evaluation and revision of the model since then has eventually led to the version that is presented herewith. From here onwards, PEDA can be used and customized by researchers, universities and policy makers for policy-making and analysis. The ECA will continue to support the model and any effort for its further application.

1.2 PEDA in brief

PEDA is an interactive computer simulation model (developed for MS-Windows environment), demonstrating the medium- to long-term impacts of alternative national policies on the food security status of the population (Lutz & Scherbov, 2000). Through the manipulation of scenario variables, the model enables the user to project the proportion of the population that will be food secure and food insecure for a chosen point in time. As food security is a factor of developments in the field of population, agriculture, the environment and socio-economic development, the model demonstrates the relationships between these fields as well. The first official version of PEDA software (hereinafter called v1.0) includes the results of the first experiments to introduce an HIV/AIDS component and to illustrate its impact on the other variables in the model. As such the PEDA model is able to give answers to a wide range of policy questions regarding the nexus interactions (Box 1-1).

Box 1-1: Policy questions

"What is the impact of increased education on the environment and land degradation?"

"How does a decrease in fertility rates influence the agricultural production in a country?"

"What is the impact of HIV/AIDS on agricultural outputs?"

"What will be the effect on the food security status of the population if the government took measures to increase fertiliser and machinery use in agriculture by 2% a year?"

"What will be the impact on the food security situation in a country if the educational enrolment rates would immediately be brought up to 75% for both sexes?"

1.3 Two approaches in modeling the nexus interactions

In the field of population-environment models we can distinguish between two kinds of approaches:

(a) Comprehensive models that try to assess the full range of population-environment interactions for a specific region; and
(b) Models that limit their focus to a specific chain of causation and therefore tend to be more focused and theory-driven. Both approaches can contribute to the better understanding of this complex field of studies. Both have their strengths and weaknesses.

The more comprehensive (holistic) approach, which tries to evaluate all relevant factors, can help us to better understand the relative contribution of specific factors to the full picture. The series of PDE (Population-Development-Environment) case studies conducted by the International Institute for Applied Systems Analysis (IIASA, Vienna) in different parts of the world are a good example of such comprehensive studies. Recent applications on Namibia and Botswana, which originally planned to focus primarily on population growth and water scarcity, had to be modified because the HIV/AIDS pandemic turned out to be of paramount importance. Since around one-third of the population of these two countries is estimated to be HIV positive, IIASA has decided to dedicate a big share of the effort to HIV modeling.

The PEDA model follows the other, substantively more focused, strategy. It attempts to quantify one specific causal path, which actually is an assumed loop, or circle that follows a clearly defined theoretical model. It is restricted to portraying factors that are relevant to that specific mechanism, leaving out many others. This approach is more in line with the tradition of economic modeling that tends to make *ceteris paribus* assumptions on all factors that are not directly relevant to the hypothesis studied, even though such factors may be very significant for the future of a country under a more comprehensive approach. For planning purposes and science-policy interactions, both the more comprehensive and the focused approaches have their virtues and shortcomings. In an ideal world a comprehensive super-model may incorporate several focused models, but this is difficult to achieve and may, indeed, suffer from some of the well-known shortages of mammoth models.

Having dwelt on this difference, it is worth noting that both approaches, i.e., the IIASA-PDE approach and PEDA, have important features in common, namely an emphasis on:

(a) Interdisciplinary scientific analysis and projection; and

(b) Science-policy communication.

Interdisciplinary scientific analysis and projection lies at the heart of both approaches. In dealing with crosscutting issues such as the effects of education on fertility and population structure and in turn on agricultural productivity and food security, the model necessarily needs to refer to the state of scientific analysis in a number of different disciplines, ranging from demography to economics, agricultural sciences, land use analysis and even water engineering. By putting information down in quantitative terms and specifying the specific quantitative inter-dependencies such a computer model can also help to contribute to overcome traditional disciplinary boundaries that have been characterized by specific research paradigms and approaches.

Such models can contribute to improved communication between the disciplines by inviting scientists from the different disciplines to add to the model the specific structure and data they consider appropriate without losing the interaction with the other segments of the model. Specific empirical case studies (at the national or sub-national level) seem to be the right strategy to advance this goal and they are also the most useful under a policy perspective when the model is used to produce alternative projections under alternative policy-relevant scenarios.

Science-policy communication. Similar to science being broken down into different disciplines, government policies tend to be compartmentalized according to the competencies of different ministries. This works well for some areas where the issues are limited in scope and require specialized treatment, but it does not work well for crosscutting problems. Issues such as food security have to do with population, the skills of the labour force, agriculture production technologies and environmental issues such as soil quality and water availability. These diverse aspects do not fall into the responsibility of a single ministry in any country of the world. For this reason new ways need to be found to support inter-ministerial connections to reflect the fact that in the real world things are also interconnected.

An inter-sectoral model such as PEDA can help to demonstrate the usefulness and even the necessity for several ministries to work together on these issues. Furthermore, specific quantitative figures showing the outcomes of alternative policy choices over the coming decades are an efficient means of communication between scientists and policy makers. When using such models scientists do not only provide policy makers with vague opinions or unproven recommendations, but they can clearly and quantitatively demonstrate what alternative outcomes are to be expected, given of course that the specific assumptions of the model are accepted. But if the assumptions seem inappropriate, they can be changed and the new results can be compared to the old ones.

1.4 An advocacy tool

As a model with a focus on a specific chain of interactions, the mission attributed to PEDA at its conception was one of advocacy: illustrating the negative development spiral resulting from high population growth, environmental degradation and decreasing per capita agricultural production. It is to demonstrate the magnitude of existing interactions and suggest alternative policy strategies to break this vicious circle in Africa. With this v1.0 of PEDA, the ECA has reached the point where the model can be effectively used as such.

Since its conception and after several rounds of evaluation and review, the different components of the model have grown and have been refined steadily to support ambitions that may exceed its advocacy function. However, the degree to which PEDA can actually be used as a prediction tool to concretely support policies of a given country is not yet clear. For that purpose more sensitivity analysis is needed and more, better and country specific data need to be collected and fed into the model. This is an effort for which the ECA invites research institutes and universities to collaborate in an effort to improve the value of the PEDA model for concrete policy support.

1.5 Two user levels

The model can be used at two different levels. Initializing the model for a new country would ideally be the job of a team of experts with specializations in demography, agriculture and natural resources modeling and with sufficient computer skills. However, once the model has been initialized for a specific country, persons with basic computer skills can easily make projections themselves and test the effect of alternative policy scenarios on the food security status of the population.

The two different user levels are also reflected in the software structure. Most users will restrict themselves to the options available in the standard user interface. For more advanced tasks such as initialising the model or for manipulating some of the parameters, the user needs to go beyond the user interface.

This manual is designed for both users. Those initialising the model for a new country or changing some of its assumptions will find the necessary information and guidelines in the procedures to follow. Others, only interested in carrying out projections for already initialised countries will find in the first three chapters information on the model's assumptions, its possibilities and limitations, the manipulation of the scenario variables and the interpretation of the outputs.

The rest of this manual is divided into a substantive account of the model, a more technical section with a description of the structure and content of the databases and excel spreadsheets underlying the user interface and a chapter that gives some methodological support in initializing the PEDA model for a new country.

2 A substantive description of the PEDA model

2.1 Theoretical foundations and the structure of PEDA

2.1.1 The theoretical inspiration for PEDA

A theoretical construction, often labeled the "vicious circle model," has recently become a very influential paradigm in the discussion around population, poverty, food security and sustainable development. It essentially assumes that high fertility, poverty, low education and status of women are bound up in a web of interactions with environmental degradation and declining food production in such a way that stress from one of the sources can trap certain rural societies, especially those living in marginal areas, into a vicious circle of increasingly destructive responses. One possible illustration of this assumed mechanism is the parable of the firewood (Nerlove 1991). In many countries the collection of firewood takes a lot of time, and more children can help to collect more firewood. But this leads to less firewood near the villages, increasing degradation of the natural resource and the desire for more children to go help collecting firewood from greater distances, also depriving the children of educational opportunities. Dasgupta (1993) presents this argument in a more generalized form.

The condition of poverty and illiteracy of the households concerned prevents substitution of alternative fuel sources or alternative livelihoods. A gender dimension is added through the fact that the low status of women and girls worsen because of the increasing amount of time and effort that they must devote to daily fuel wood gathering (Agarwal 1994; Sen 1994). The education of girls is blocked because girls are kept at home to help their mothers. The result is faster population growth, further degradation of the renewable resource base, increasing food insecurity, stagnating education levels, and yet a further erosion of women's status.

From a theoretical point of view this vicious circle model is a useful contribution towards a more general framework in causally linking fertility, poverty, low female status and environmental degradation. It is also attractive because it explicitly addresses equity concerns. Its multi-dimensional structure helps to view different possible interventions in reproductive health, education, environmental conservation and agricultural efficiency in a unifying context rather than in isolation from each other. Each of the interventions may, under certain conditions contribute to breaking the vicious circle, but a comprehensive strategy viewing all these aspects together and recognizing their interdependencies is likely to be more successful (O'Neill *et al.* 2000).

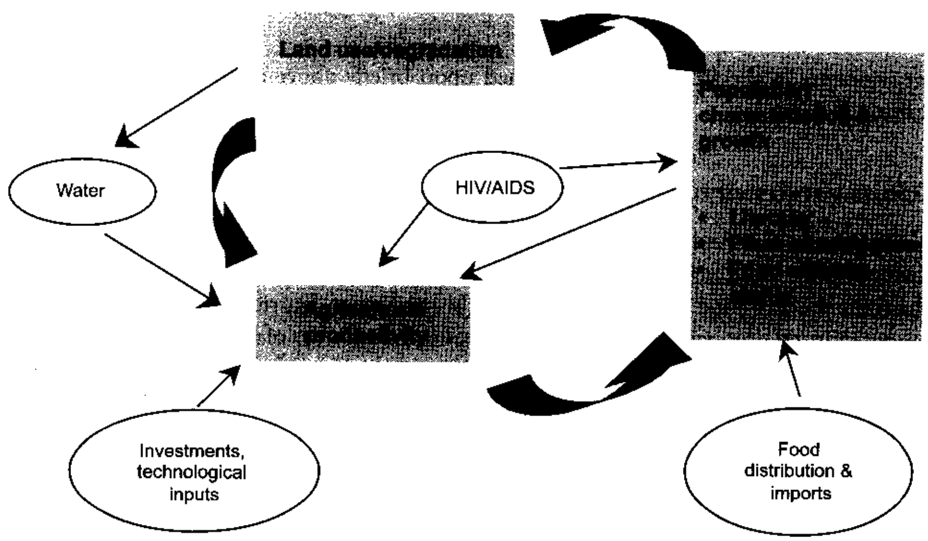
In terms of its empirical relevance, the vicious circle assumption is more controversial. Because the economic reasoning of this model largely operates at the household level, empirical studies on the issue have been mostly confined to that level and reached mixed results. At the macro level of different population segments this model could potentially be very relevant --particularly in the African context-- although some of the assumptions of the stricter version of this model are empirically unconfirmed and controversial. Especially the assumption that environmental degradation may actually lead to increases in fertility is difficult to be defended at a time when fertility rates are rapidly falling all over Africa within a context of degrading environmental resources.

This does not necessarily imply that the assumed effects are entirely absent, but it seems to imply that if they operate, the powerful and dominating processes of the demographic transition overlay them. Hence, it may be reasonable to alternatively assume that food insecurity is associated with a slower decline in fertility, although under certain conditions and in the short run famines may well induce declining fertility. Whatever the position on this issue, the PEDA model as outlined below is general enough to represent alternative assumptions through alternative parameter choices and scenarios.

2.1.2 The structure of the PEDA model

As illustrated in Figure 2-1, the PEDA model consists of three sub modules or segments: a population module, a natural resources module (land and water) and an agricultural production module. Although not immediately visible to the user of PEDA, the model contains an additional food distribution segment that accounts for the inequality in the access to food.

Figure 2-1: The structure of the PEDA model



The vicious circle in PEDA operates through the negative impact of a fast growing (food insecure) population on the natural resources stock, which in its turn decreases agricultural production and that again induces more food insecurity. This negative chain of interactions is, however, not part of the model structure itself. As it will be highlighted in section 2.1.3, the user can easily change the assumptions of the relationships between population growth, the natural resources stock and agricultural production.

In the population segment of PEDA, multi-state population projections are carried out to determine the size of the population by urban-rural place of residence, literacy status and food security status.

The food production and availability in PEDA is influenced by a set of endogenous and exogenous variables. An important resource for agricultural production is land. In the model it is treated in index form combining the quantity and quality of land. Although the user can omit this effect, the model by default assumes a negative impact of population growth on natural resources. The size and qualification level of the labour force, the availability of water, and efforts in irrigation and fertilizer and machinery use further influence agricultural production.

The contribution of water to agricultural outputs is dependent on climatic conditions but also on the status of the land degradation and efforts in irrigation and water management (building of reservoirs).

In addition to the produced food, the model allows to account for post-harvest losses and for food imports and exports to estimate the net food availability. In the last step the available food is distributed over the population following a non-linear food distribution function to determine the proportion of the population that is food insecure.

Recently, an HIV/AIDS component has been added to PEDA to account for its demographic impact through excess mortality and for its negative impact on agricultural production through the reduction of the labour force because of HIV/AIDS related illnesses.

The different segments in the PEDA model touched upon briefly above are treated more in detail in the sections 2.2 through 2.7.

2.1.3 Vicious versus virtuous circle dynamics

Although the PEDA model is flexible with regard to the underlying theoretical assumptions, by default it is set to be compliant with the main principles of the vicious circle at the macro-level described earlier; i.e. that the growth of the (illiterate and food insecure) population in rural areas contributes to the degradation of land, thus lowering agricultural production and further increasing the number of food insecure. If not broken, this vicious circle would lead to ever increasing land degradation and increases in the food-insecure population.

The model does not assume increasing fertility as an automatic response to food insecurity, which is the most problematic part of some vicious circle models. Rather, the food secure and food insecure fractions of the population are assumed to have different fertility levels (subject to exogenously defined trends) and hence, the aggregate fertility level only responds to changes in the food insecurity through the changing weights of the groups in the calculation of the overall fertility.

The vicious circle can be broken through several possible interventions in the field of food production, food distribution, education, environmental protection and population dynamics and thus also allowing for more Boserupian visions on the nexus of population, the environment and agricultural production. The power of the PEDA model is that these assumed positive or negative interactions are all part of the scenarios that can be set by the user and not hard-wired in the model.

There are several ways to carry out optimistic projections. This can, for example, be done by defining scenarios that assume:

- No negative (or even a positive) effect of increasing population densities on natural resources;
- Increasing technological inputs in agriculture (e.g. fertilizer, machinery or irrigation);
- Increased net food imports;

- A greater equality in the food distribution;
- Increased efforts towards human development through better access to education, health care, etc;
- A better natural resource base management in densely populated areas; and
- A lower fertility level for the food-insecure segments of the population.

2.2 The population segment

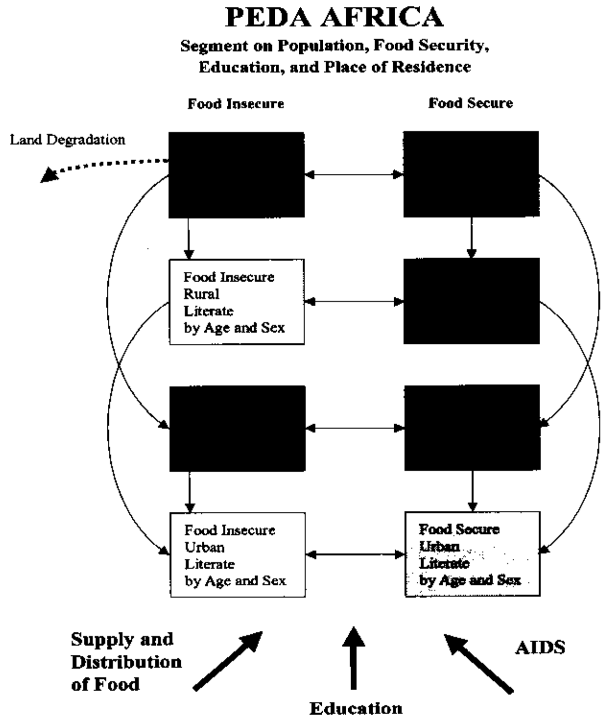
PEDA is different from most macroeconomic models in that it uses a population-based approach. The population-based approach views human beings with their specific characteristics (such as age, sex, education, health, food security status, place of residence, etc.) as agents of social, economic, cultural and environmental change. But the population is also at risk of suffering from repercussions of these changes and of benefiting from positive implications.

In this sense, human beings are seen as a driving force of these changes and the first to be affected by the outcomes and consequences of these changes. Economics, if it comes into the picture, e.g., through the importance of markets in distributing goods, plays only an intermediate role and economic indicators are not seen as an end in itself or the primary objective of the modeling exercise. In this sense, the population-based approach chosen here differs from much of the development-economics literature.

The population-based approach does not assume that population growth or other demographic changes are necessarily the most important factors in shaping our future. Instead, the phenomena that we want to model are studied in terms of different characteristics that can be directly attached to, and (at least theoretically) measured with individual members of the population. Characteristics such as age, sex, literacy, place of residence and even nutritional status can be assessed at the individual level. The sum over these individual characteristics makes up the distribution of the total population. In using these individual characteristics, PEDA distinguishes itself from models that rely on other frequently used economic indicators such as the GNP per capita. Although the GNP per capita is indicative of the average amount of money that an individual has in his/her pocket, it cannot be directly measured at the individual level. It results from aggregated indicators of national accounting with various conceptual and measurement problems.

Although many of the powerful quantitative economic tools (such as general equilibrium models) cannot be applied due to this choice of approach, other very powerful but less well known tools of demographic analysis and projection can be applied. The tools of multi-state population analysis allow for the projection of the population by several characteristics (such as age, sex, education and place of residence) at the same time. Multi-state projections group all individuals of a given population into different sub-populations which are then simultaneously projected into the future, while at each time interval, people can also move from one sub-population to another (e.g., from rural to urban or illiterate to literate for each sex and age group).

Figure 2-2: The population segment



In PEDA, the population of a country under consideration is broken down into eight sub-groups according to urban/rural place of residence, educational and food security status (Figure 2-2). Place of residence and food security status are core elements of the vicious circle reasoning as mentioned earlier. Education, or more precisely literacy status, has been introduced into the model as one of the assumed key sources of population heterogeneity, which is related to both agricultural production and land degradation. Significant educational fertility differentials give the explicit consideration of education in the model a strong rationale, but education has many other important effects. There is abundant literature on the significance of literacy in population-development-environment interactions (e.g. Lutz, ed., 1994). The potential of explicitly including education as a demographic dimension in multi-state population projection models has recently been evaluated, and is strongly recommended in the case of educational fertility and productivity (see Lutz, et al. 1999).

Each of these eight sub-groups is further subdivided by age (in single-year age groups) and sex, i.e., every one of the eight groups has its own age pyramid. During each one-year simulation step, a person will move up the age pyramid by one year within the same sub-group, or move to another sub-group while aging by one year (or die). It is also possible for some people to make multiple transitions within one time step, e.g. from rural/food insecure/illiterate to urban/food secure/illiterate.

Arrows in figure 2-2 show the movements that are possible between groups. For education and rural/urban migration, the model is hierarchical, i.e., people can only move in one direction, from lower to higher education and from rural areas to urban areas. Movement between food security states can happen in both directions, depending on the food conditions in the relevant year and the food distribution function.

This PEDA population module is in itself a useful piece of software for multi-state population projections. As part of the initialization process, the user can set for each of the eight states, age and sex-specific fertility, mortality and educational transition rates. As scenario variables, dynamic future paths can be defined for fertility and mortality. The model automatically adjusts age-specific fertility and mortality patterns according to the levels of the total fertility rate and life expectancy chosen for each year.

The methods of multi-state population projections are well described in the literature (e.g. Rogers, 1975; Keyfitz, 1979 and Scherbov & Grechucha, 1988). Here it is sufficient to mention that the multi-state population projection model is a generalization of the one-dimensional cohort-component model of population projections that takes into account transitions that occur between different states.

Education and rural-urban migration are defined in terms of the proportion of male and female cohorts that will become literate and move from rural to urban areas. While education is concentrated in childhood, rural-urban migration tends to follow a typical pattern with highest migration intensities in young adulthood. For this a standard internal migration schedule has been applied (Rogers and Castro, 1981). In the current version of the PEDA model, both education and internal migration are treated statically. That is, the user can set a specific transition rate and that remains constant throughout the projection period. With respect to education, the model allows for sex and age specific educational transition rates.

Generally, before specific scenarios can be calibrated for a chosen country application, this country first needs to be initialized in terms of setting all empirical data for the starting conditions. This initialization is a significant task to be done outside the PEDA application environment, and requires more demographic skills than the use of the initialized model. More information on the initialization procedures can be found in chapter 5.

Due to the great challenge that the HIV/AIDS pandemic poses to many African countries, the most recent version of PEDA also explicitly deals with this issue. The user can exogenously define assumed future paths in HIV/AIDS morbidity and their impacts on the course of life expectancy for each of the population groups. This procedure considers the fact that the age pattern of mortality is very different for countries with high HIV prevalence. In addition to this direct impact of AIDS on the population segment through excess mortality, AIDS morbidity also affects agricultural production through a reduction of the rural labour force as is outlined in section 2.7.

2.3 Land

Land (arable land and pasture) is a key environmental resource for agricultural production in a given climate zone. It can be positively and negatively influenced by human activity.

Like many of the variables that affect agricultural production, land is treated as an index variable set to 1.0 in the starting year. This value must be seen as describing both the quantity and quality of land. Since the dynamics described below apply to both interpretations, it is not necessary to distinguish between the two aspects at this stage. In many applications one will have to deal with combinations of both aspects (e.g., erosion affecting both the quantity of the total arable land and the quality of the remaining land).

The user should, however, be aware that an imaginary constant value for the land index throughout the projection period may be interpreted in different ways: on the one hand, it could mean that the land under cultivation has been extended while the overall quality of the land has diminished; on the other hand, it could mean that the area under cultivation remained the same while there has not been any reduction in the quality of the land because of proper land management.

In a more general manner, one can label the land variable as 'the natural resources stock'. The change in the stock of natural resources *R(t)* is the result of a combination of indigenous growth or regeneration *g(R(t))* and a reduction through population induced environmental degradation *(D(t)):*

$$R(t) = g(R(t) \ - D(t)$$

Whereby R(t) for the initial year is equal to 1 and indigenous growth or regeneration is defined as:

$$g(R(t) \ = a(\overline{R} - R(t)$$

Herein, parameter a reflects the speed at which the resources recover[1]. However, it is assumed that the pace of resource recovery diminishes as the saturation level ($\overline{R}$) is approached. The saturation level stands for the stationary solution of *R* if the resources are not degraded. The saturation level has to be chosen in accordance with the specific conditions in the country of application and remains constant over the whole projection period.

The degradation (D(t)) of the resource stock depends on the stock of available resources and on the increase in the rural illiterate food insecure population as well as on changes in the overall population density.

As mentioned earlier, the model allows the user to assume that it is especially the rural illiterate food insecure segment of the population that depletes natural resources in their quest for survival. The growth of this subgroup thus leads to the degradation of resources. As this is often used as an assumption in natural resources modeling, this impact diminishes as the stock of resources decreases. In mathematical form this is given by:

$$P_I(t)\frac{R(t)}{R(t)+\eta}$$

---

1 If we consider the area under cultivation, the term *a* ($\overline{R}$ *-R(t))* stands for new cultivated land, where the cultivation is performed with constant intensity independent of any economic or social forces. Bilsborrow notes that the increase in demand due to population growth "can be met by an increase in the land under agricultural use or an increase in the intensity of use of existing land or an increase in both (...) The greater the land resources available and potentially arable, the more likely is extensification instead of intensification." (Bilsborrow, R.E., 1992)

Whereby $P_I(t)$ stands for the relative change in the number rural-illiterate-food insecure population compared to the starting conditions. η is a constant factor that has default value of 1, but can be adjusted by an advanced user of the model if there is good reason to assume a different pattern of diminishing impact.

The scale of environmental degradation is also a function of the change in the overall population density. Mathematically, this is expressed as:

$$\frac{P(t)}{\overline{R}}$$

Whereby P(t) stands for the relative change in the total population compared to the initial year. The denominator reflects the upper limit of the natural resource stock and has a constant value that is country specific and reflects the relative status of the resource stock in the year for which the initial data are prepared.

The complete mathematical expression for environmental degradation is thus:

$$D(t) = \gamma \frac{P(t)}{\overline{R}} P_I(t) \frac{R(t)}{R(t) + \eta}$$

The only parameter not yet described in this function is γ. This parameter is called the land degradation impact factor in PEDA and influences the intensity of the effects described above. The value of γ can be adjusted in the PEDA user interface. Although the value of γ is subject to country specific conditions, by default it is set to equal 0.02. Setting the value of γ to 0.00 implies that there is no assumed negative effect of increasing population densities on the natural resources stock. For experimental purposes, the user could also set a negative value for γ, meaning that increasing population densities have a positive effect on the natural resources stock, but in that case the natural resources upper limit ($\overline{R}$) may be easily exceeded.

The expression of land degradation used here implies that if resources are completely degraded (i.e. R (t)=0), the value of D (t) will be automatically zero as there is nothing to be depleted. Similarly, if the stock of rural illiterate food insecure population is zero, environmental degradation will be zero as well.

In each year, the adjusted index value of current land enters the agricultural production function with the elasticity described below (see section 2.5).

In Table 2-1, we have summarized the different parameters of the land module and their default values. For advocacy purposes, the user could work with these values. For in depth country applications, however, the user should seriously reconsider the values of these parameters. Preferably, they should be determined on the basis of empirical evidence.

Of these parameters, only the land degradation parameter (γ) can be changed from within the user interface. If considered necessary, the other parameters can be manipulated from the Excel spreadsheets underlying the user interface and saved as part of the scenarios (see section 4.3 for instructions on how to change information in the Excel spreadsheets).

Table 2-1: The parameters of the land module and their default values

Indicator	Description	Default value
Land recovery parameter, a	Determines the speed at which the natural resources recover. This parameter also captures the clearing of new land for cultivation.	0.0175
Land degradation impact factor, $D(t)$	Reflects the assumed negative impact of population growth on the natural resources stock. The higher $D(t)$ the higher will be the degradation of the natural resources (D). The value can be adjusted from the PEDA user interface and remains constant over the projection period.	0.02
Natural resources upper limit, $\overline{R}$	A measure of the saturation level of the natural resources illustrating the stationary solution of the natural resources if they are not degraded. This value has to be determined during the initialization process, reflecting the difference between the natural resources stock (R (t)) and the natural resources upper limit in the initial year. By default it has value 1.5, meaning that the actual natural resources stock is delivering a productivity of 2/3 of its potential.	1.50
Natural resources diminishing impact factor, η	This parameter regularizes the behavior of the function. It has a constant value of 4.00 by default but can be cautiously changed by the advanced user. Combined with R (t) in the denominator of the expression, it reduces the degradation of the natural resources (D) if it is higher than 1 and vice versa.	4.00
Natural resources stock, or the current land stock R (t)	This is the output variable of the land module. It is treated as an index variable having value 1, which reflects the conditions in the initial year and describing the quantity and quality of the land under cultivation.	1.00

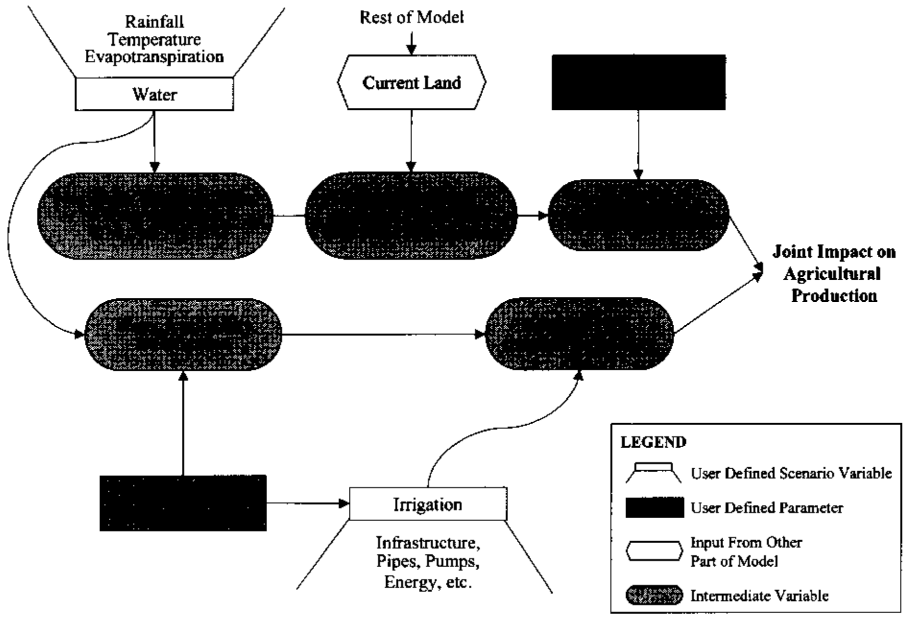
2.4 Water

Water is a key resource for life and human development. Fresh water is a renewable resource that is unevenly spread over the planet. The development of any civilization is closely associated with the management of water, be it in the distribution of scarce water or the management of regular floods. To human life, water is directly necessary for drinking and more indirectly for producing food. Water is also an important factor of personal hygiene and health and a necessary input to many industrial processes in modern society. For its key role in food security and sustainable human development, water deserves particular attention in models dealing with the nexus issues such as PEDA.

In the PEDA model, water is treated in a separate module with an eventual multiplier effect on agricultural outputs. In its current form, the water module contains two externally defined scenario variables (water and irrigation) that can be changed dynamically and two user-defined parameters: the reservoir capacity effect (RCE) and a parameter specifying the impact of land degradation on water availability (subsequently labeled as the water impact factor, WIF).

The value of the latter two parameters should ideally be determined during the process of initialization although an advanced user can still define different values as part of the scenario settings. The water segment in PEDA also relies on the input from other variables in the model such as the quantity and quality of land and calculates a number of intermediate variables. This input is completely endogenously determined and cannot be manipulated by the user.

Figure 2-3: The water module



The water module essentially has two parts, one referring to rain fed agriculture and the other to irrigated agriculture. The scenario variable water, W (t), is relevant for both segments and covers the general climatic conditions in year t particularly with respect to rainfall and evapo-transpiration. It therefore can be used to simulate both short-term or cyclical droughts and longer-term climate change. Unlike most of the other scenario variables, water will not be set to 1.0 in the starting year, but its initial value will be defined in terms of its position on a nonlinear curve that describes the relationship between water availability and agricultural production (Figure 2-4). This definition of the initial value depends on the specific climate conditions of the country in the initial year and is to be part of the initialization procedure.

2.4.1 Rain- fed agriculture

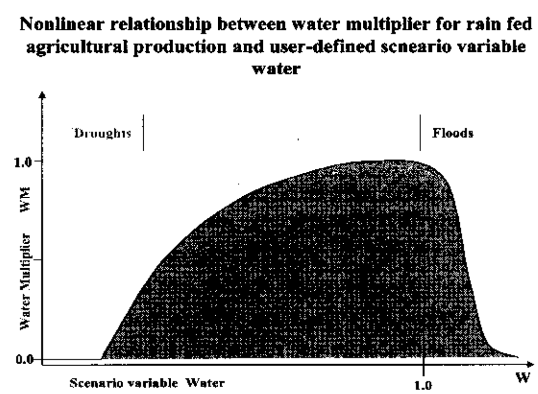
For calculating the impact of the scenario variable 'water' on rain-fed agriculture, a nonlinear transformation into a water multiplier (WM (t)) is introduced because an increase of one unit of water does not always have an equal impact on agricultural production. The assumed relationship that is derived from the hydrological and agricultural literature is presented in stylized form in Figure 2-4. The specific shape of this non-linear relationship greatly depends on local conditions and the kind of crops and/or livestock under consideration. For in-depth applications of PEDA, the definition of this curve requires serious attention. In the current version of PEDA a hypothetical curve has been assumed. Its general features are that in case of serious drought nothing can grow, but after this point small increases in water availability can produce great returns. With further increases in water availability the curve flattens to eventually reach a saturation level, starting from which more water will adversely affect agricultural outputs. Beyond this point, flooding starts to be harmful to production and may ultimately destroy all production.

The assumed non-linear relationship of water on agricultural outputs, however, only holds if other relevant determinants of soil moisture remain constant over time. Unfortunately, land degradation and land erosion tend to increase the runoff of rainwater and therefore decrease the moisture that will be stored in the soil thus negatively affecting agricultural productivity. Since land degradation is explicitly modeled in other parts of PEDA, the impact of land degradation can be directly taken into account. To do this in quantitative terms, another user defined input parameter has been introduced, i.e. the water impact factor of land degradation or short WIF. WIF is defined in the form of an elasticity applied to current land (defined as R (t) in the previous section). This results in an effective water multiplier (EWM) for rain fed agriculture:

$$EWM(t) = WM(t)*R(t)^{WIF}$$

Values bigger than 1.0 for WIF, increase the effect of the status of the land or resources in the calculation of the effective water multiplier; values smaller than 1.0 decrease its elasticity.

Figure 2-4: The water saturation curve for rain fed agriculture



2.4.2 Irrigated agriculture

For irrigated agriculture the dynamics in which water availability and irrigation efforts have a joint effect on agricultural outputs are even more complex. The functional relationships defined here are a great simplification but should still be able to capture the most important dynamics. The formula given below creates an intermediate variable, effective irrigation (EIR), which enters the agricultural production function as a multiplier. To determine the value of EIR, PEDA makes a distinction between the situation where water availability is already at or above the saturation level of 1.0 and when it is still below (see Figure 2-4). If the water supply is above the saturation level, efforts in irrigation do not make any difference in agricultural production. If, on the other hand, the water supply is below the saturation level, the positive effect of irrigation efforts further depends on the level of the water supply, and the reservoir capacity effect. The reservoir capacity parameter stands for the potential to stock water for later use in irrigated agriculture.

In the mathematical expression, W (t) stands for the value of the scenario variable water at time t and IR (t) for the value of the irrigation variable. Frt is the elasticity of fertilizer in the production function that is applied to irrigation as well because of the lack of better data on the effect of irrigation on agricultural production.

$$EIR(t) = \begin{cases} 1; if\ W(t) \geq 1 \\ \left[RCE * \dfrac{R\ (t)}{R\ (0)} * \dfrac{W(t)}{W(0)} \right]^{Frt} & 1; if\ W(t) < 1 \end{cases}$$

The second line of the above formula may require some explanation since it is an approximation of several more complex mechanisms. The main reasoning is that even under high irrigation efforts, there needs to be water available in order to have any effect and that even high reservoir capacity does not help if it does not rain for a long time. Ideally, this would require the calculation of cumulative water storage effects over time. A simple approximation is obtained through the multiplication with a ratio of current water over initial water in case that current water availability is still below saturation level. This also implies that if there are no extra irrigation efforts but simply more water, it also generates positive returns in terms of agricultural outputs. This can be interpreted as an additional direct effect of water on production that is weighted by improvements in irrigation and reservoir capacity.

Both effective irrigation (EIR (t)) and the effective water multiplier (EWM (t)) are then added to the agricultural production function as additional multipliers. It is worth noting that in this setup, the two water-related variables and current land enter the agricultural production function twice but in two different forms and after transformations as outlined above. The water variable enters both through rain fed and irrigated agriculture as discussed. Land degradation as captured by current land enters as a regular production factor with a given elasticity (as described in the next section) and through the impact of land degradation on increasing runoff and therefore decreasing soil moisture. Depending on the other settings of the model and the specific parameter choice, this indirect effect of land degradation through declining soil moisture may be even more relevant than the direct effect through the production function.

In Table 2-2, we have summarized the different variables and parameters of the water module, and their default values. For advocacy purposes, the user could work with these default values. For in depth country applications, however, the user should seriously reconsider the values of these variables and parameters. The list summarized in Table 2-2 contains dynamic scenario variables, intermediate variables and parameters. The value of the scenario variables can be changed for each year in the projection period from within the user interface. The intermediate variables exist for calculation purposes only and cannot be manipulated by the user. The value of the other (static) parameters can be manipulated from the Excel spreadsheets underlying the user interface and be saved as part of the scenarios (see section 4.3 for instructions on how to change information in the Excel spreadsheets).

Table 2-2:The parameters and variables of the water module and their default values

Indicator	Description	Default Value
Water, W (t)	Scenario variable. Reflects the change in weather/climate conditions. This variable has an impact on both rain fed and irrigated agriculture.	0.7
Irrigation, IR(t)	Scenario variable. Stands for relative changes in irrigation efforts. Set by default as 1.0 for the starting year. This value enters a function to determine the impact of irrigation on agricultural outputs.	1.0
Water multiplier, WM	Endogenously determined and intermediate variable. It is the output of the non-linear transformation of the water scenario variable (W (t)) through the water saturation curve (see Figure 2-4). The underlying logic is that an increase or decrease of one unit water does not always have an equal impact on agricultural production.	
Water impact factor of land degradation on rain fed agriculture, WIF	A measure of the impact of land degradation on water availability. The value is predefined as 1.0, but can be changed by the advanced user. Values bigger than 1.0 increase the elasticity of land degradation, values smaller than 1.0 decrease the effect of land degradation (on rain fed agriculture).	1.0
Effective water multiplier for rain fed agriculture, EWM (t)	An intermediate variable that represents the effect of water availability on rain fed agriculture. It enters the agricultural production function as a multiplier.	
Reservoir capacity effect, RCE	Stands for the capacity to stock water for later use in irrigated agriculture. This is a static model parameter can be changed from the Excel spreadsheets.	1.0
Effective irrigation, EIR(t)	An intermediate variable that represents the effect of irrigation on agricultural outputs (see section 2.4.2). It enters the agricultural production function as a multiplier.	

2.5 Agricultural production segment

The total agricultural production in one year, measured in total calories produced (in index form), is calculated in PEDA through a Cobb-Douglas type agricultural production function. Many agricultural production functions exist, but most of them do not consider the labor force and the skills of the labor force as a production factor. In stead, they largely focus on physical and financial inputs. The chosen production function is a notable exception. Based on pooled data sets of time series for most countries of the world, Hayami and Ruttan (1971) estimated a large number of Cobb-Douglas type production functions with different combinations of input factors for different groups of countries.

The equation that seemed most appropriate for PEDA Africa is the Principal Components Regression for developing countries including educational variables. These estimated elasticities are given as default values in PEDA. Should the user have more recent estimates or ones that are more appropriate for the country under consideration, the elasticities should be changed accordingly.

The total production is a result of the inputs in terms of the human labor force by educational level, land and technological investments in fertilizer use mechanization, etc. In PEDA, the other segments in the model endogenously determine some of these inputs and others are treated in terms of externally defined scenario variables because these factors are not assumed to depend directly on other variables of the PEDA model (Table 2-3).

The population by age and sex in the eight defined categories affects total agricultural production in two different ways. First, the population projections produce an estimate of the size of the rural labour force. Their productivity is affected by the proportion of literates within that category (endogenously determined) in addition to their technical training (externally defined scenario variable). The values for all these variables directly enter the agricultural production function as discussed below.

The other chain of causation is a reflection of the vicious cycle reasoning: the factor land is degraded as a function of the relative change in the rural illiterate food insecure segment of the population and of the population density in general as has been discussed above. Other main factors influencing the agricultural output of a country such as mechanization and fertilizer use need to be specified in externally defined scenario variables. If, however, a user wants to make for example the rate of technological investments in agriculture dependent on population growth in either a positive or negative way, it is not difficult to define corresponding scenarios and study the alternative results. As discussed in section 2.4, water is treated in a separate module with an eventual multiplier effect on the total agricultural outputs.

Table 2-3: The elasticities and the specific variable definitions in agricultural production function as taken from Hayami and Ruttan (1971, p. 145, Q 19):

Variable	Description	Default elasticities
Labour force, LF(t)	Total rural adult population aged 15-60, calculated from combining the appropriate age groups in all four rural sub-groups	0.534
Total agricultural land, R (t)	Endogenously determined by the land segment	0.088
Fertilizer use, FERT (t)	Exogenous scenario variable	0.162
Mechanization, MECH (t)	Exogenous scenario variable. In the original production function of Hayami and Ruttan, this variable stood for the use of tractors; in the PEDA model it is considered to cover mechanization in general.	0.072
Literacy of the labour force, LITLF (t)	Specified as the proportion literate of the total rural population aged 10-45, (both male and female). The value is determined endogenously by combining both the food-secure and the food-insecure rural literate sub-populations.	0.276
Technical education , TE(t)	In addition to literacy, this variable represents the technical education of the labour force. In the current version this variable is still treated as an exogenous scenario variable. In a later stage it may be related to the educational efforts parameter.	0.158

All these inputs to agricultural production are considered on a relative scale, i.e., their values are set to 1.0 in the starting year and change over time as a result of effects emanating from the other segments in the model or as defined in the scenario settings for the exogenous variables. If, for example, we assumed an increase in fertilizer use of 20 percent by 2005, this would mean that the value of that variable is set to increase to 1.2 by that year.

$$ProdIndex = LF(t)^{0.534} * R(t)^{0.088} * FERT(t)^{0.162} * MECH(t)^{0.072} * LITLF(t)^{0.276} * TE(t)^{0.158} * EIR(t) * EWM(t)$$

In sum, total agricultural production in terms of total calories produced is calculated in the following multiplicative manner, in which the last two factors are water multipliers that have been discussed earlier:

Although the variables of the agricultural production function and their elasticities are by default distributed with the model and applicable to all countries, a thorough initialization of the model for a particular country should include estimating new elasticities or even consider the inclusion of other variables. Much will depend on the mix of crops under production and the share of livestock in the total outputs.

2.6 Food availability and distribution

2.6.1 Food availability

Not all food that is produced in a country will be consumed by its citizens. A fraction will be lost during the harvest, transportation, storage and processing and a part of it may be intended for use as seeds in the next cropping season or for export. On the other hand, imports may complement the food that is produced locally. Mathematically, the net food available (FA) can be expressed in the following manner:

$$FA_t = FP_t * (1 - LT_t) + imports_{t_0} * FIE_t$$

In this expression, the net food production is equal to the gross production (FP) times the proportion that remains after the deduction of the post-harvest losses and seedlings (LT). The post-harvest losses are expressed as a proportion of the total production and can be manipulated in a dynamic scenario variable.

The net food imports are added to the net food production to obtain the net food availability. For the initial year of the projection period, absolute values of the net food production and the net food imports have to be specified by the user in terms of daily Kcal per capita. This is usually done as part of the initialization process. FIE is a dynamic scenario variable that has the default value of 1 for the initial year and that allows the user to account for fluctuations in the volume of imports.

2.6.2 Food distribution

After the correction for post-harvest losses and net imports, the estimated amount of available food is distributed over the population in two steps. First, an 'urban bias factor' (external scenario variable) determines which fraction of the available food is consumed by the rural and urban population respectively. As with the other scenario variables this is done on a relative scale, with 1.0 reflecting a condition of equality between rural and urban areas in the access to food. A value of 1.1 means that urban areas get 10% more than they would have been attributed if food had been distributed proportionally to the size of the population living in urban areas.

In addition to the urban bias factor, the PEDA model also accounts for the inequality in the access to food within urban and rural areas. The distribution of food is often unequal because some persons simply have more purchasing power than others or have privileged access to food by other means. This will result in the fact that some people remain food insecure even when the average total amount of food reaching the population is theoretically sufficient to provide the necessary minimum diet for everybody. As the access to food is usually more unequal in urban than in rural areas, PEDA works with two food distribution functions, one for rural and one for urban areas.

There is abundant theoretical and empirical evidence indicating that the inequality in the food distribution is at least as important as the total production of food in explaining food insecurity. Especially the work of Amartya Sen (1994) demonstrated that some of the worst famines occurred under conditions in which there would have been enough food for everybody, if the distribution had been appropriate. For this reason it is evident that a model focusing on food security without paying attention to the distributional aspects would be incomplete, if not misleading. The main problem with considering such distributions, however, lies in the fact that hardly any empirical data exist on distributive mechanisms in African countries of today, and that theoretical distributions are hardly appropriate because conditions tend to vary significantly from one country to another. PEDA opts for another solution to approximate the inequality in the access to food through household income distribution functions that exist for a number of African countries on the basis of household expenditure surveys[1]. Hence, PEDA relies on the assumption that the inequality in the access to food follows a similar distribution to the inequality in the access to income in respectively urban and rural areas.

In each one-year step of the projections, food is allocated to rural and urban areas following the urban bias factor and within each of these areas the food distribution function determines the new sizes of the food secure and insecure sub-populations. Figure 2-5 gives an example of such a food distribution function. It is a Lorenz Curve with the cumulated proportion of the population on one axis and the cumulated calories available for distribution on the other. The available food is then distributed from right to left along the curve. In an ideal situation, food should be distributed equally among each individual and the Lorenz curve would coincide with a straight line, i.e. 45°-diagonal.

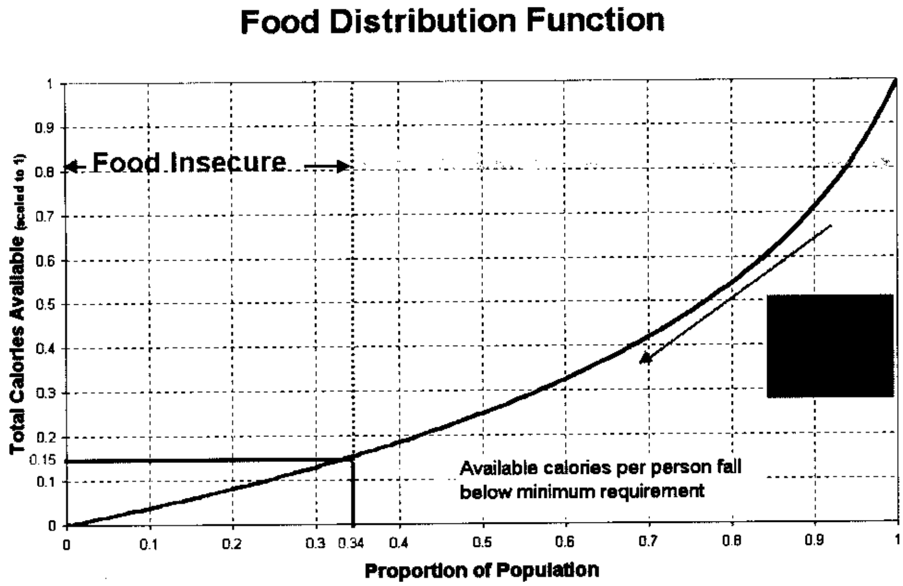
It means that the slope of the curve is equal to 1.0 and everyone in the economy will have access to the mean available calories; i.e. it describes a situation of perfect equality. However, because in the real world food is not equally distributed, it would be realistic to assume that a curve will be plotted above or below the straight line. The convexity of the curve therefore measures the degree of inequality.

The given curve in Figure 2-5 indicates that in this case, the first (most privileged) 10 per cent of the population consume 30 per cent of the available food. Going further down the curve, about 23 per cent of the population consumes half of the food, and half of the population uses 75 percent of the food. The borderline between the food-secure and the food-insecure population is then established by applying an externally defined minimum calorie requirement per person. The iterations are carried out in steps of 1 percent of the population. At the point where the food allocation of the percentile falls below the minimum food requirement specified, the borderline for the population considered to be food insecure is established. In this example, the least privileged 34 per cent of the population has access to only 15 per cent of the food.

As this fraction of the remaining food is considered to be insufficient to fulfill the necessary minimum daily food intake (on the basis of a threshold level set by the user in a scenario parameter, not shown in the illustration), 34 per cent of the population is considered food insecure. Over time the proportions food insecure may change as a consequence of changes in the population size and food availability, or, possible changes in the assumed food distribution function. In the current version of the PEDA model, however, the food distribution function is assumed to remain constant over the whole projection period. Defining the food distribution functions for rural and urban areas is part of process of initializing the model for a new country. The food distribution functions cannot be stored as part of the scenarios; each new definition of the food distribution functions would result in a new initialization.

1 World Bank African Development Indicators 1998/99. Section 15

Figure 2-5: The food distribution function

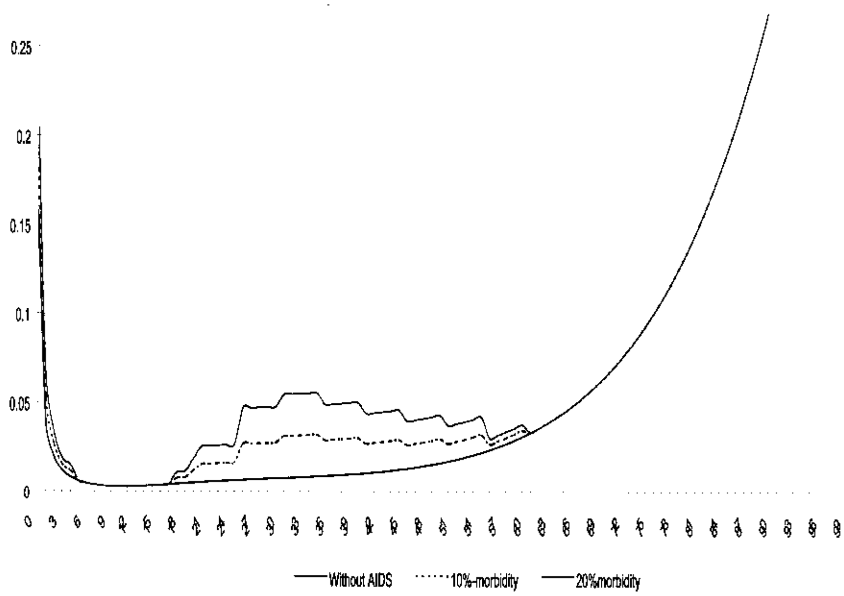


2.7 HIV/AIDS in PEDA

As a model demonstrating the interactions between population, education, environment and food security, PEDA has not been designed to explicitly cover the possible sustainable development and food security consequences of HIV/AIDS. However, since the pandemic has become such a challenge for many African countries and because it is not meaningful to talk about the future development path without explicitly considering HIV/AIDS, this section will discuss the way in which HIV/AIDS is incorporated in PEDA. As there is still relatively little empirical knowledge on the trends of the pandemic and especially not on its effect on the different sectors of the economy, the model is necessarily experimental in its conception and treatment of HIV/AIDS.

The most obvious impact of HIV/AIDS on the nexus is through excess mortality and thus indirectly through a reduction of the labour force. As the age pattern of AIDS mortality differs from the one observed in a population without AIDS, it was opted to include that effect via a user specified scenario variable. In that scenario variable the user needs to specify AIDS related morbidity rates, which is similar to prevalence information but with lower values than the widely available HIV prevalence rates (cfr. infra). The morbidity rates are translated into age specific mortality rates, which are added to the age specific mortality pattern of each of the eight different subgroups in the population. The additional age specific mortality pattern due to AIDS has been estimated on the basis of a simulation model for Botswana. This age-pattern of AIDS mortality has a typical shape with very high rates in the early adulthood, and due to vertical transmission also in early childhood. The specified age specific excess mortality due to AIDS is then scaled up or down depending on the morbidity level set by the user in the HIV/AIDS scenario variable and superimposed to the mortality from other causes. If a user wants to specify age specific excess mortality on the basis of country specific data, this has to be implemented during the process of initialization. Figure 2-6 gives an illustration of the age specific mortality rates (m_x) for one of the subgroups in an imaginary population without AIDS and with a 10 and 20% morbidity level. The figure clearly illustrates the above-mentioned concentration of AIDS related mortality in the early childhood and young adulthood.

Figure 2-6: Age specific mortality rates (mx) in a population without and with HIV/AIDS



The user should also be aware that one cannot account for the demographic impact through manipulation of this scenario variable alone since it only changes the age pattern of mortality but not the level. A consistent scenario should include setting lower life expectancies than one would have assumed in the absence of AIDS and the development of life expectancy over time will have to be set in such a way that it reflects the recent and anticipated future trend of HIV prevalence and the resulting AIDS mortality (see Table 2-4 for some guidelines in setting such scenarios) under the assumed absence of efficient medication.

Table 2-4: Reductions in life expectancy under different HIV/AIDS morbidity levels (as specified in the top row)

Life ex-pectan-cies	HIV/AIDS adult morbidity level										
	0.00	0.02	0.04	0.06	0.08	0.10	0.12	0.14	0.16	0.18	0.20
40.0	40.0	38.2	36.4	34.6	32.8	31.0	29.9	28.7	27.6	26.4	25.3
41.0	41.0	39.1	37.3	35.4	33.6	31.7	30.5	29.3	28.2	27.0	25.8
42.0	42.0	40.1	38.2	36.2	34.3	32.4	31.2	30.0	28.7	27.5	26.3
43.0	43.0	41.0	39.0	37.1	35.1	33.1	31.9	30.6	29.4	28.1	26.9
44.0	44.0	42.0	39.9	37.9	35.8	33.8	32.5	31.2	30.0	28.7	27.4
45.0	45.0	42.9	40.8	38.7	36.6	34.5	33.2	31.9	30.5	29.2	27.9
46.0	46.0	43.8	41.7	39.5	37.4	35.2	33.8	32.5	31.1	29.8	28.4
47.0	47.0	44.8	42.6	40.3	38.1	35.9	34.5	33.1	31.7	30.3	28.9
48.0	48.0	45.7	43.4	41.2	38.9	36.6	35.2	33.7	32.3	30.8	29.4
49.0	49.0	46.7	44.3	42.0	39.6	37.3	35.8	34.3	32.9	31.4	29.9
50.0	50.0	47.6	45.2	42.8	40.4	38.0	36.5	35.0	33.4	31.9	30.4
51.0	51.0	48.5	46.1	43.6	41.2	38.7	37.1	35.6	34.0	32.5	30.9
52.0	52.0	49.5	46.9	44.4	41.8	39.3	37.7	36.1	34.6	33.0	31.4
53.0	53.0	50.4	47.8	45.2	42.6	40.0	38.4	36.8	35.1	33.5	31.9
54.0	54.0	51.3	48.7	46.0	43.4	40.7	39.0	37.4	35.7	34.1	32.4
55.0	55.0	52.3	49.5	46.8	44.0	41.3	39.6	37.9	36.2	34.5	32.8
56.0	56.0	53.2	50.4	47.6	44.8	42.0	40.3	38.5	36.8	35.0	33.3
57.0	57.0	54.1	51.2	48.4	45.5	42.6	40.8	39.1	37.3	35.6	33.8
58.0	58.0	55.1	52.1	49.2	46.2	43.3	41.5	39.7	37.8	36.0	34.2
59.0	59.0	56.0	53.0	49.9	46.9	43.9	42.1	40.2	38.4	36.5	34.7
60.0	60.0	56.9	53.8	50.8	47.7	44.6	42.7	40.8	38.9	37.0	35.1
61.0	61.0	57.8	54.7	51.5	48.4	45.2	43.3	41.3	39.4	37.4	35.5
62.0	62.0	58.8	55.6	52.3	49.1	45.9	43.9	41.9	40.0	38.0	36.0
63.0	63.0	59.7	56.4	53.1	49.8	46.5	44.5	42.5	40.4	38.4	36.4
64.0	64.0	60.6	57.2	53.9	50.5	47.1	45.0	43.0	40.9	38.9	36.8
65.0	65.0	61.6	58.1	54.7	51.2	47.8	45.7	43.6	41.4	39.3	37.2
66.0	66.0	62.5	59.0	55.4	51.9	48.4	46.2	44.1	41.9	39.8	37.6
67.0	67.0	63.4	59.8	56.2	52.6	49.0	46.8	44.6	42.5	40.3	38.1
68.0	68.0	64.3	60.6	57.0	53.3	49.6	47.4	45.2	42.9	40.7	38.5
69.0	69.0	65.2	61.5	57.7	54.0	50.2	47.9	45.6	43.4	41.1	38.8
70.0	70.0	66.2	62.3	58.5	54.6	50.8	48.5	46.2	43.8	41.5	39.2

When interpreting this table one has to be aware of the fact that the AIDS morbidity rate of the adult population aged 15-49 for a given population tends to be much lower than the frequently cited HIV prevalence rate based on Sentinel Surveys. Due to the fact that the incubation period (HIV positive without symptoms) tends to be 5-10 years in Africa whereas the symptomatic period before death is much shorter (1-3 years) the morbidity level tends to be less than a third of the HIV prevalence level. Due to this delay there may be cases where AIDS morbidity is still low while HIV prevalence is growing rapidly. This dynamism of the disease can be captured in PEDA through dynamic assumptions of life expectancy.

In the current version of PEDA, the scenario variable wherein the user specifies the AIDS morbidity rates is applicable to the population as a whole. As of now, different morbidity rates (as implied by different morbidity rates) cannot be applied to the eight sub groups separately. However, the user can make different assumptions regarding the expected reduction in the level of life expectancy and thus account for a potential different impact of HIV/AIDS in rural and urban areas, among literates and illiterates etc.

HIV/AIDS not only induces increased mortality with its impact on the population structure; it also seriously affects agricultural production through different mechanisms. Most of these effects are not endogenized in PEDA and need to be dealt with in the form of consistent user-defined scenarios. Similar to many other possible future developments, such as droughts, climate change, wars, heavy new investments in technical education or in irrigation, etc., the user of PEDA has to choose the various model parameters and scenario variables in such a way that they describe consistent "stories" or scenarios of possible future trends.

If AIDS is assumed to be a major issue in a country, then AIDS mortality and morbidity should be assumed to reduce the educational enrollment and technical education due to AIDS orphanhood, fewer qualified teachers (they may die) and AIDS-induced financial and economic constraints. These likely negative consequences of AIDS on the incomes of affected households may also reduce investments in fertilizer, irrigation and mechanical inputs in agriculture. The only endogenized feedback from the AIDS morbidity rates, as set by the user, on agricultural production is the reduction in the productivity of the labour force.

Again, very limited empirical evidence on the issue exists, therefore an experimental solution had to be chosen. By default the model assumes a linear decline in the productivity of the labour force at a rate of the morbidity level. Put more simply, people who are sick, i.e. symptomatic with HIV/AIDS are subtracted from the labor force. The rationale is that the capacity of sick people to work is greatly reduced and that some healthy people will have to look after their sick relatives instead of working and thus reduce their productivity in agriculture.

Taken together these different effects may have a drastic impact on agricultural outputs and the development of a country, but at the moment, hardly any systematic empirical evidence exists on these issues. Hence, the assumptions need to be highly speculative at this point. If more empirical information on these effects becomes available in the future, then PEDA could be expanded into a model that quite comprehensively captures the impacts of AIDS on human development and food security.

While PEDA explicitly deals with HIV/AIDS, other diseases like malaria, tuberculosis, measles are not accounted for through distinct variables or relationships. The reason is that the latter diseases exist for a longer time on the continent, and they are already captured in the standard age-sex specific mortality patterns of African populations. Since the HIV/AIDS pandemic is relatively recent and because it has a very specific mortality pattern with children and young adults being the most vulnerable, it had to be treated in a special manner.

If necessary, the user could still capture an explosion of one or the other epidemic, as long as its effects do not deviate too much from the standard mortality patterns, through the reduction of the assumed future life expectancies for each of the sub-populations separately.

3 An overview of the user-defined scenario variables and output variables

As mentioned earlier in this manual, working with PEDA can be done at two different levels. The initialization of PEDA for a new country requires significant effort by -ideally- an interdisciplinary group of experts. Some methodological support for initializing PEDA for a new country is given in chapter 5 of this manual. In addition to the initialization, the need may arise to change some more advanced model parameters that cannot be manipulated from within the PEDA user interface. Guidelines on how to locate and change information in the Excel spreadsheets underlying the interface can be found in section 4.3.

Most users of PEDA will, however, work with already initialized countries and possibly pre-defined scenarios. For them there is no need to go beyond the PEDA user interface. They can focus on a number of model parameters and scenario variables that can be set and adjusted in an easily accessible manner. In this chapter, an overview is given of the parameters and variables that can be accessed through the PEDA user interface. In a later stage the standard output variables are summarized.

3.1 Model parameters and scenario variables

3.1.1 General model settings

These include some of the basic model specifications that are set at the beginning of each simulation. They are considered constants that cannot be changed over time or across sub-populations. In other words, these parameters apply to the whole country and throughout the projection period (except the net food production and the net food imports that apply only to the initial year). Figure 3-1 presents a screenshot of the window wherein these parameters can be changed. In Table 3-1, a brief description of each of these parameters is given.

Figure 3-1: The general model settings window

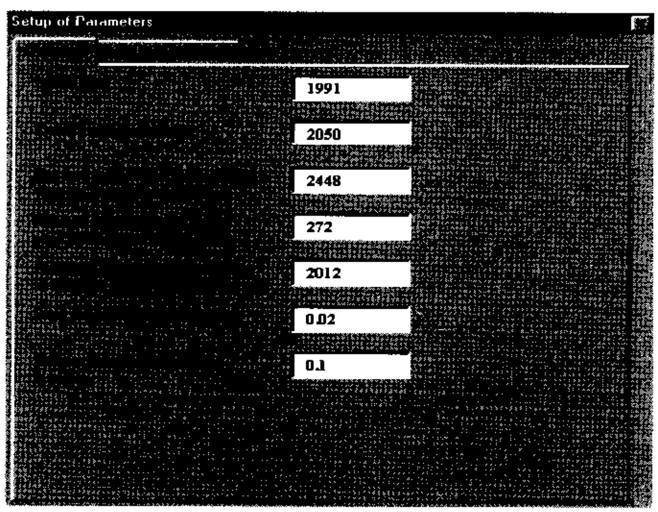


Table 3-1 : Parameter definitions of the general model settings

Parameter name	Description
Initial year	The starting year of the projections. This is the year to which the baseline data apply. Although this value could be changed, it is better not to do so as it corresponds to the initial data.
End of projection period	The end of the projection period. All simulations will be run in single year steps up to the year specified here and all results will be stored up to the end of the specified period. Although there is no direct limit set to the value of this parameter, projection periods of longer than 50 years will slow down calculations and are subject to increasing uncertainty.
Net food production in the initial year (in daily per capita Kcal)	Refers to the net daily per capita amount of food produced in the starting year. Food production in PEDA is expressed in terms of its energy equivalent or calories. Note that one has to specify this 'net' food production, that is the total production after deduction of post-harvest losses and a fraction of the production, reserved to be used as seeds in the next cropping season. From this information and the level of post harvest losses in the initial year, PEDA automatically estimates the gross food production. Once the gross production is determined for the initial year, PEDA projects it into subsequent years of the projection period in single year steps using the agricultural production function (see section 2.5).
Net food imports in the initial year (in daily per capita Kcal)	Refers to the net daily per capita Kcal imports in the starting year. Combined with the net food production, it results in the net food available for consumption (see section 2.6.1). Fluctuations in the net food imports during the projection period can be accounted for by manipulating the scenario variable 'Food imports and exports' (see Table 3-3).
Assumed minimum consumption in kcal per capita to be food secure	In the PEDA model, calorie (energy) requirements are used as a proxy for food requirements. "The minimum energy requirement is the amount of energy that is required on average in a population to satisfy the basic physiological needs and the needs for light activities of adults and the normal energy needs of children and adolescents (including the extra needs for the growth). Two main factors determine the estimation of the energy requirement of a population: the distribution by age and sex; and the body weight."* The value of this variable may thus vary under different national conditions, or the user can set the value to define different thresholds to evaluate its effect on the model interactions.
Land degradation impact factor	This variable reflects the assumed negative impact of population growth on the natural resource stock. See section 2.3, p.14 for more detailed information on the definition of this parameter.
Proportion of cohort moving to cities	This variable enables the user to set net rural-urban migration rates. A value of 0.2 means that of every cohort born in rural areas, 20% will move permanently to urban areas over their lifetime. These movements are distributed over the different ages according to standard age-specific migration schedules (see section 2.2).

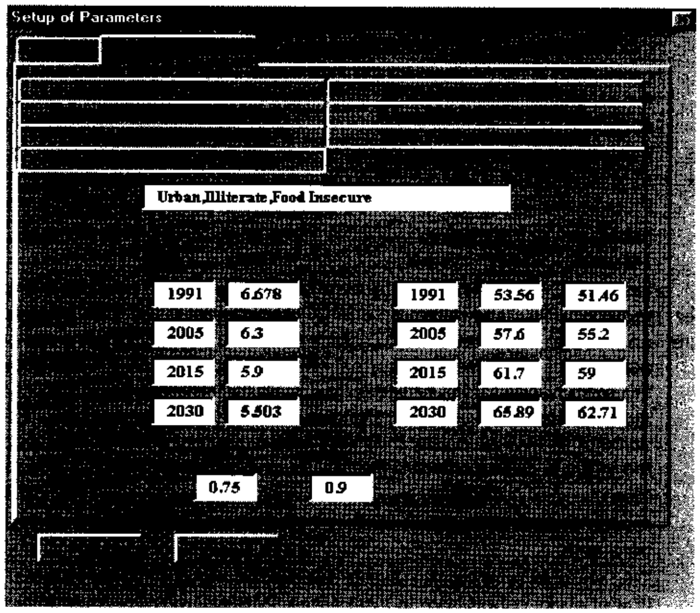
* This definition is taken from the African Nutrition Database Initiative (ANDI) web site, at http://www.africanutrition.net/

3.1.2 Sub-population parameters

These variables allow the user to make assumptions with respect to the future path of fertility (through the Total Fertility Rate, TFR), mortality (through life expectancy, e_0) and literacy (through educational transition rates). All these variables are directly related to the population projections segment in the model and can be specified for the eight different subgroups independently. All these variables are summary measures that, after adjustment, will influence age specific transition rates as defined during the process of initialization. Both fertility and mortality are treated dynamically. Educational transition rates are assumed to remain constant over the whole projection period.

The Total Fertility Rate (TFR) is the mean number of children a woman would get throughout her reproductive life (if she survived to age 50 and was exposed to the age-specific fertility rates observed or assumed for a given year). TFR is treated dynamically in the model, i.e. it can be assumed to change over time. The pattern of age-specific fertility rates is automatically adjusted accordingly. In addition to the TFR for the starting year, the user can set two intermediate levels and one final assumed level of fertility to be reached at the end of the projection period. If the assumed values are different, the model will calculate a gradual change from the first to the second value that is linear over time. The TFR set for the starting year should ideally reflect the situation in the year for which the data were initialized.

Figure 3-2: The sub-population parameter settings window



Life expectancy at birth (e_0) can be set for men and women separately for each of the eight sub-groups. Just as fertility, life expectancy is treated dynamically. Note that, although the user may keep the TFR and e_0 constant for experimental reasons, their aggregated values in the total population may change due to changing weights of the different subgroups over time (this of course is only valid if different assumptions regarding fertility and mortality are made for the subgroups in the population).

In addition to fertility and mortality, the user can make assumptions regarding the educational transition rates (by default to be considered the transition rates from illiterate to literate status). The educational transition rates can be set for men and women separately and is treated as a constant over time. The number entered in the model is the proportion of girls and boys in each birth cohort that will move over their lifetime from the illiterate to the literate state. In the present form of the model, these transitions are concentrated around age 15 although more advanced users may specify a particular age pattern for the transition from the illiterate to literate status. Such an age specific pattern for the transition from illiterate to literate status (as opposed to an assumed total proportion to become literate) will influence the other variables in the model only to a very limited degree.

As mentioned before, educational transition rates assumed by the user, remain constant over the whole projection period, but a user may assume higher transition rates than those reflecting the conditions in the initial year and these will then be applied from the first year in the projections and thus adding a dynamic aspect to the treatment of education. In any case the future educational composition of the population of the next few decades is mostly a function of past changes in educational enrollment reflected in the educational distribution by age in the starting year. Since education is concentrated in young age, it will take quite some time to have changes in enrollment reflected in the working age population. For the sub-groups that are already educated, no educational transition rates can be set because PEDA doesn't consider movements back from literate to illiterate.

3.1.3 HIV/AIDS

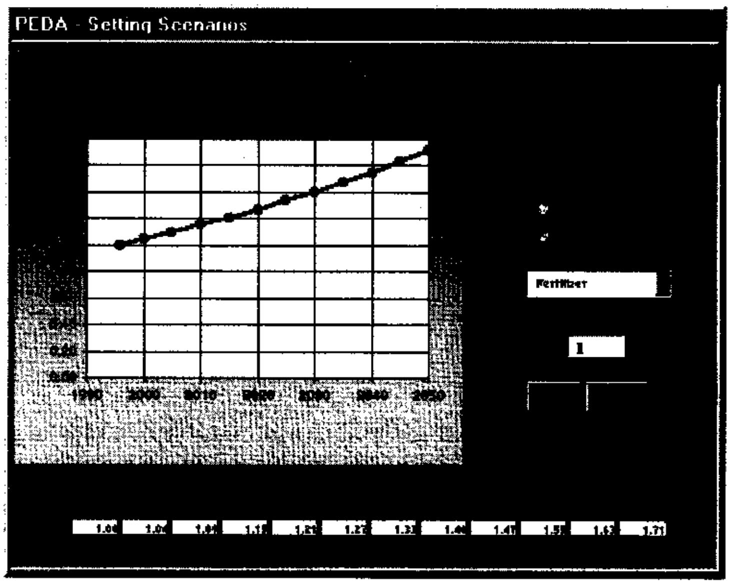
HIV/AIDS morbidity rates can be set in a user specified scenario variable that adds an age specific mortality pattern to the mortality schedules for each of the eight subgroups. Just as mortality, HIV/AIDS is treated dynamically but unlike life expectancy, this special age-specific mortality pattern can be specified only for the population as a whole. Variations in the impact of HIV/AIDS on the different subpopulations can only be accounted for through setting different life expectancy trends in these subpopulations.

In the HIV/AIDS scenario variable, the user is expected to set an HIV/AIDS related morbidity pattern over time that is defined in terms of proportions of the young adult population, aged 15-49. A value of 0.02 for this variable means that 2 per cent of the young adult population is assumed to be sick to an AIDS related complaint. In any population this proportion tends to be lower than the HIV prevalence rate because the period of morbidity tends to be shorter than the incubation period. See also section 2.7 for more details on setting comprehensive AIDS scenarios.

3.1.4 Parameters for food supply

Most of the (external) factors associated with food production and supply are treated in index form, which means that their value in the starting year is set to be 1.0 and their levels in all the subsequent years are defined on a relative scale. This treatment of scenario variables allows using the model even if one does not have empirical information about the exact quantitative level of each factor since one only has to specify the relative change as compared to the starting year. The variables considered in PEDA with respect to the food supply are summarized in the tables below. The first set of variables will enter the agricultural production function (see Table 3-2); the second set includes variables to account for post-harvest losses, food imports and exports, and potential differences between urban and rural areas in the access to food (see Table 3-3).

Figure 3-3: The parameters for food supply settings window



The model endogenously determines some of the variables that influence the agricultural production in a country. Since the user through external scenario variables cannot directly manipulate their values, they are not elaborated upon here. See section 2.5 for a more complete definition of the agricultural production function and its components. The scenario variables related to the availability of water are -strictly speaking- not part of the agricultural production function, but are part of the PEDA-water segment with an eventual multiplier effect on agricultural outputs (see section 2.4).

Table 3-2: Variables of the agricultural production function

Variable name	Description
Size of the rural labour force	Endogenously determined variable
Literacy of the labour force	Endogenously determined variable
Land	Endogenously determined variable
Fertilizer	The amount and productivity of fertilizer used in agriculture. The user needs to set a value that expresses a relative improvement/worsening with respect to the conditions in the starting year.
Machinery	The amount and productivity of machinery used in agriculture. The user needs to set a value that expresses a relative improvement/worsening with respect to the conditions in the starting year.
Technical education	In addition to literacy, the user can specify the technical capacity of the rural labour force for agricultural purposes. As with Fertilizer and machinery use, one needs to give a value that expresses a relative improvement/worsening of the conditions in a particular year as compared to the starting year.
Water	This covers the change in weather/climate conditions. Its initial value depends on the climate conditions of the country at the time of initialization. The range of acceptable values for the water scenario variable depends very much on the definition of the water saturation curve. In the current version of PEDA, values for the water scenario variable can range from 0 to 1.5, with 1 describing a situation of optimal rainfall conditions for agriculture. Values higher than 1 indicate a surplus of water and have a negative impact on agricultural outputs.
Irrigation	This covers the efforts made in irrigation, including pipelines, pumps, energy etc. The value of this scenario variable expresses a relative improvement/worsening of the conditions in a particular year as compared to the conditions in the year of initialization.

With the exception of the three endogenously determined variables and the water scenario variable, all variables in Table 3-2 are treated as indexes having a default value of 1 for the initial year. In setting scenario assumptions, all values greater than 0 are allowed with values greater than 1 reflecting a relative increase in the inputs as compared to the initial year.

Table 3-3: Other variables influencing the availability of food

Variable name	Description
Loss in transport and storage	Individuals will not consume all the food that is produced. Some of the food will be lost during the harvest; the transport, storage or processing, and some of it will be used as seeds for the next cropping season. For the initial year, one has to specify the proportion of the gross production that is not available for consumption for one of the reasons specified above. For the subsequent years in the projection, the user can assume changes in the post-harvest losses by manipulating this variable. The range of possible values for this scenario variable vary between 0 to 1, with 0 describing a situation where there are no post harvest losses and 1 a situation where none of the locally produced food is available for consumption. Although one has to specify the net production in the initial year as part of the initialization process, PEDA automatically recalculates the gross production. The latter is the basis for the estimation of the food production in the following years of the projection period and from which the post-harvest losses are deducted (see Table 3-1). Note that post-harvest losses are only applied to locally produced food and not to net imports.
Urban bias factor	This variable enables the user to allocate food disproportionally between urban and rural areas. A value of 1.0 means that the food is distributed proportionally to the size of the population living in urban and rural areas respectively (a state of 'No bias'). A value of 1.1 means that urban areas receive 10% more food than would be allocated under conditions of equal access. The remaining share of the food is then attributed to rural areas. By giving a value lower than 1, the user can assume a 'rural-bias' in the access to food.
Food imports and exports (net trade)	This variable allows the user to take changes in food imports and exports into account. During the initialization process, one has to specify an absolute value for the net food imports (see Table 3-1). The dynamic scenario variable food imports/exports has a default value of 1 reflecting the conditions in the initial year. By manipulating the value of this scenario variable, the user can make different assumptions regarding the net imports/exports. A value of 2 for example means that the amount of food imported has doubled compared to the initial year.

3.2 Output variables

3.2.1 Multiple interfaces to access the output of the projections

PEDA has large numbers of output parameters and variables that can be accessed in different formats and for different subgroups in the population. For each one-year step in the simulation every age and sex specific population number for each of the eight different subgroups is stored in a database and can be accessed by any user in a relatively easy manner. The same is also true for most of a wide range of other scenario and output variables. A description of these databases is given in the following chapter. In addition, the PEDA model contains a number of integrated and standardized output modules including (dynamic) population pyramids, and graphs showing trends for a number of variables over time. The user can specify to compare sub-populations given for a particular scenario and indicator or he/she can compare scenarios for a given indicator. Please refer to the *PEDA User's Manual* for more information on these standard output modules.

3.2.2 List of indicators

Hereunder, we have listed the output variables available to the user with their explanation wherever they are not self-explanatory. If no specification is given, the indicators can be generated for the eight different subpopulations separately. Otherwise it is specified that the indicator is only available at the country level. The standard output modules in the PEDA model allow the user to request the values for most of the scenario variables as well. This option was included for the user to be able to verify his/her assumptions for a given scenario while being in the output module. In the table below, these variables are labeled as input variable.

Table 3-4: Standard output variables in PEDA

Variable	Variable description
Available food	This is the sum of the total amount of food produced in the country in a particular year minus the post-harvest losses, +/- food imports and exports. It stands for the net available food to the population for each year in the projection period and it is expressed in terms of calories. This indicator is only available for the country as a whole.
Births	Total number of births
Current land	The combination of the quantity and quality of land for each year of the projection period. Index value summarizing the effects of land degradation and regeneration. This indicator is only available for the country as a whole.
Deaths	Total number of deaths
Life expectancy (e0)	When requested for each of the eight sub-groups separately and sex specific, this is the graduated value of the assumptions of the user. At the country level, however, it is also influenced by changes in the relative weights of the subgroups in the population (provided that different life expectancies have been set for the different subgroups).
Literate Life Expectancy (LLE)	Number of years a person is expected to live in a literate status from the age of 15 onwards. See section 3.2.3 for a more detailed description.
Fertilizer	Is an input variable.
Food import/exports	Is an input variable.
Food production	This is the gross production for each year of the projection period, and it is expressed in terms of calories. This indicator is similar to food availability, but it does not account for post-harvest losses and imports/exports. This indicator is only available for the country as a whole.
HIV/AIDS morbidity rates	Is an input variable
Irrigation	Is an input variable
Loss in transport/storage	Is an input variable
Machinery	Is an input variable
Proportion food insecure	In addition to the population size that can be generated by urban/rural place of residence, literacy status and food security status; the model has an extra output variable that portrays the proportion food insecure in the country for any year of the projection period. This indicator is only available for the country as a whole.
Total population	Population size. It can be generated for each of the eight subpopulations separately and for both sexes separately. There is also a possibility to extract age specific information directly from the databases (see section 4.2).
Technical education	Is an input variable

Variable	Variable description
TFR	When requested for each of the eight sub-groups separately, this is the graduated value of the assumptions of the user. At the country level, however, it is also influenced by changes in the relative weights of the subgroups in the population (provided that different fertility rates have been set for the different subgroups)
Urban bias factor	Is an input variable
Water	Is an input variable

3.2.3 Literate Life Expectancy (LLE)

As it is a less well-known indicator, Literate Life Expectancy deserves particular attention in this manual. LLE stands for *the average number of years a man or a woman lives in the literate state; i.e. he/she is able to read and write under the prevailing mortality and literacy conditions.* Literate Life Expectancy (LLE) has recently been proposed as an indicator of social development and quality of life. LLE is perfectly suited to fit the structure of PEDA which gives projections of the population by age, literacy status and mortality levels (depending among others on the food security status) and it may even have a number of advantages over another widely spread development related indicator, namely the Human Development Index (HDI).

A single indicator that comprehensively describes a population's quality of life and can be used for comparative purposes can be very useful if it has a clear interpretation. In the past, GNP per capita has almost exclusively been used for this purpose, although it remains a highly problematic indicator for comprehensively measuring quality of life and development. United Nations Development Programme offered a more comprehensive alternative by introducing the Human Development Index (HDI). This index combines indicators of life expectancy, educational attainment and income in one figure on a scale between 0 and 1. HDI served an important purpose in giving more attention to the social aspects of development but also has some shortcomings.

The fact that measures of income (through whatever concept of national accounting) are not reflected in LLE can be seen as a benefit in terms of purity rather than a deficiency. Finding the right level of mortality and of literacy is a largely empirical issue once an operational definition of literacy is applied. In practice there are of course all kinds of measurement problems, but figures for material wealth do not only have the measurement problem.

It is to a much higher degree dependent on the accounting framework applied, whether only the formal economy is considered, whether the depletion of natural capital, whether real purchasing power, or distributional aspects are considered. It may be wiser not to mix these two very different kinds of indicators (one based on individual characteristics, the other on an abstract entity called the economy). Let GNP per capita describe economic development in the formal economic sectors and take literate life expectancy to describe social development.

Literate Life Expectancy as a summary indicator of social development is based on two underlying indicators: age-specific mortality rates and age-specific proportions literate. There are convincing arguments for taking individual survival probabilities and empowerment through basic education as two of the most important and least ambiguous aspects of human quality of life. Below is an attempt to highlight some of the underlying reasoning which in this brevity will certainly be incomplete.

Personal survival to a mature age and the survival of immediate family and close friends are universal human aspirations. It is one of the few values likely to be shared by a Buddhist monk, a Wall Street broker, and a street child in Sao Paulo. Individual survival is the necessary prerequisite for enjoying any kind of quality of life. In more economic terms, increases in life expectancy at the level of society also increase the expectation of returns from investments ranging from education to housing, consumer durables, production sites, etc. This expectation is a basic prerequisite for any kind of development.

But increasing life expectancy not only facilitates development, its level also reflects social advancement and quality of life in a very comprehensive and sensitive manner. Life expectancy responds positively to most of the things that we consider important ingredients of quality of life (good diet, efficient health care, good housing, benign technologies, good social and economic infrastructure, healthy working conditions, education, intellectual stimulation, etc.) and negatively to things we want to avoid (armed conflict, malnutrition, poverty, hazardous work, stress, depression, etc.).

One might even go as far as to say that happiness tends to reduce adult mortality, while unhappiness increases the risks; a lot of psychosomatic evidence points in that direction. Given all the problems involved in directly measuring health (not to speak of happiness), longevity is still the best proxy of health and possibly even telling of emotional conditions. On a societal level the mortality crisis in Eastern Europe, for instance, clearly reflects a social, economic, environmental and psychological crisis.

Basic education as an indicator of empowerment also has an individual and a societal component. On the individual level, reading and writing skills are basic prerequisites for almost every kind of professional advancement and improvement of living conditions. It is also important for securing one's basic entitlements. Not being able to read and write in most societies means being excluded from progress, and carries the danger of being further disempowered. Especially for the empowerment of women in society and in her family, basic education is the key variable.

On the level of society it makes a big difference whether educational budgets are spent towards achieving universal primary education or for the higher education of small elites, while large proportions of the population remain illiterate. Historical analysis shows that countries that worked towards universal basic education do significantly better in several aspects ranging from population stabilization to economic development and social equity. More generally, human capital --of which education is the most essential ingredient--seems to be by far the most important factor of economic development in the long run.

A look at the historical developments of European countries over the last century shows that some initially very backward countries without any significant natural resources or financial capital (such as Finland) made it to the top simply by pushing for education and achieving near to universal literacy as early as the beginning of the century. Finland thereby passed other central and southern European countries that originally had much more sophisticated cultures and significantly more physical and financial capital. In short, literacy not only shows the current level of social development, it also characterizes a country's potential for future development.

The calculation of LLE requires empirical data on age-specific mortality rates and age-specific proportions literate, which are both readily available from PEDA. The calculation of LLE follows the regular life table method that is used to calculate the mortality-based life expectancy, adding only that the number of person-years at each age is weighted with the age-specific proportion literate. Literate Life Expectancy is always somewhat lower than regular life expectancy because early childhood is always an illiterate state. Since data on the literacy status of children varies for different PEDA applications, literacy is only considered from the age 15 onwards. The data given on LLE do still refer to life expectancy at birth, considering child mortality, but do not count the years some children may be literate before age 15.

In PEDA, time series of LLE for the country and scenario under consideration are a standard output feature in the presentation of results segment. Together with the total proportion food insecure LLE may be used as a major criterion to see whether one development path should be considered preferable to another. In fact this additional indicator is a very necessary complement to the proportion food insecure, which taken alone may be misleading. There can be scenarios in which the food insecure population declines due to very high mortality (e.g. through epidemics). This is clearly not a desirable scenario, although it would not be apparent if one only looks at the food insecure population or even the proportion food insecure. LLE in this case would clearly reveal the dramatic deterioration in quality of life.

Box 3-1: A methodological note on the calculation of literate life expectancy

Necessary input data: Age-specific mortality rates (mx)
ge-specific proportions literate (PLx)

Method: In a regular life table, the Lx column (total number of person-years lived in age group x) is multiplied by PLx to generate the LLx column (literate person-years lived). Like in a regular life table, literate life expectancy (Le0x) is then obtained by dividing the cumulative literate person-years (LTx) by the lx column.

Example: LLE of rural men in Egypt, 1986

	Regular Life Table				Literate Life Table			
Age	mx	lx	Lx	e0x	PLx	LLx	LTx	Le0x
<1	1.041	100.000	93340	56.60	0.00	0	2382889	23.8
4	.081	90.106	353413	64.00	0.00	0	2382889	26.4
5-9	.017	87.232	434190	62.06	0.42	183203	2382889	27.3
10-14	.010	86.494	431434	57.57	0.84	364130	2199686	25.4
15-19	.012	86.062	429077	52.84	0.68	290486	1835556	21.3
20-24	.017	85.546	426000	48.15	0.78	333556	1545071	18.1
25-29	.021	84.824	421991	43.54	0.48	202978	1211513	14.3
30-34	.027	83.938	416986	38.97	0.48	200570	1008535	12.0
35-39	.032	82.812	410805	34.46	0.38	156966	807964	9.8
40-44	.035	81.496	404094	29.98	0.38	154364	650999	8.0
45-49	.060	80.084	393900	25.48	0.30	119170	496635	6.2
50-54	.121	77.398	375934	21.26	0.30	112780	376466	4.9
55-59	.240	72.824	344335	17.43	0.25	85851	263686	3.6
60-64	.252	64.580	304529	14.32	0.25	76219	180633	2.8
65-69	.572	56.925	250441	10.89	0.20	50088	109415	1.9
70-74	.662	42.681	183505	8.08	0.20	36713	55327	1.3
75+	1.625	30.297	186136	6.35	0.10	18614	18614	0.6

4 Structure of the software, underlying data bases and spreadsheets

4.1 Introduction

The PEDA model is developed for a Windows environment and written in Visual Basic. Both the input and output are stored in Microsoft Access databases and the calculations are carried out in Microsoft Excel. When running the PEDA software, the user is not immediately aware of the manipulations of databases and data in spreadsheets. The interface or shell acts as a buffer between the commands (e.g. read a previously defined scenario X for country Z) and the operations that are carried out in the background (e.g. open the database with scenario X for country Z, load the information into Excel and display the values of the scenario variables in the user interface). This, however, implies that the professional version of Microsoft Office has to be installed on the computer from which one wants to run the PEDA model (refer to the PEDA User's Manual for more details on the installation procedure and system requirements).

The first prototypes of the software consisted of a package wherein the mathematical model, country data and scenario assumptions were integrated. Although this had been a viable option for testing the main model assumptions, the design turned out to be inflexible when applying the model to multiple countries or use different sets of baseline data for one country as each of them would result in a different application. The original approach also complicated the task of implementing adjustments in the mathematical model to the different applications as they had to be carried out as many times as there were applications.

To cope with the potential inconveniencies as the applications of PEDA to different countries expand, the mathematical model has been separated from a number of databases containing the country data and scenario assumptions. In v1.0 of the PEDA model, there is only one version of the software containing the main user interface and the mathematical definition of the model. From within that interface, the user can choose between already initialized countries to carry out projections or to see the results of already carried out simulations.

The following sections of this technical manual contain a presentation and description of the databases and spreadsheets whereupon the PEDA model relies. It is not so much dealing with the user interface, but it is explaining what happens behind that interface. Hence, what follows will be much more of interest to those who want to go further than simply carrying out projections with already existing data or make use of non standardized output features. This may be the case during the process of initializing the model for a new country or when the elasticities of some of the assumed effects want to be changed, or, what may be most commonly the case, when a user wants to generate output that is not by default provided by the software. In these explanations we assume that the model has been installed on the computer. For installation instructions and guidelines to use PEDA to carry out simulations, we refer the reader to the *PEDA User's Manual.*

4.2 The structure of the application and databases

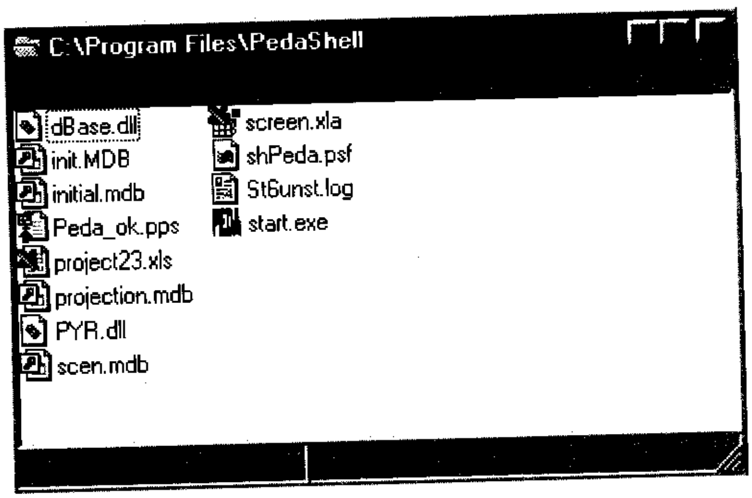
There are basically three distinct operations a user may want to carry out with PEDA. The first and most sizable task is to initialize the model for a new country. Further, the user may probably want to carry out simulations and eventually see the results of these projections. The PEDA application is built around these three distinct logical steps and this is also reflected in the structure of the databases underlying the user interface. The management of these databases is usually carried out in the background: for each particular task or command, the PEDA application opens the relevant database, selects the appropriate information from within that database, loads it into Microsoft Excel which takes over to carry out calculations if required, or, displays the requested information in the user interface.

If the command includes performing calculations, the results are later again stored in the appropriate database. The structure of the Excel spreadsheets wherein the calculations are carried out will be discussed in section 4.3. First we'll describe the database files and how they are used in PEDA. Users wanting to manipulate any of the information stored in these databases without passing through the PEDA user interface, should be sufficiently familiar with Microsoft Access and the PEDA model as a whole to do so.

4.2.1 Four databases

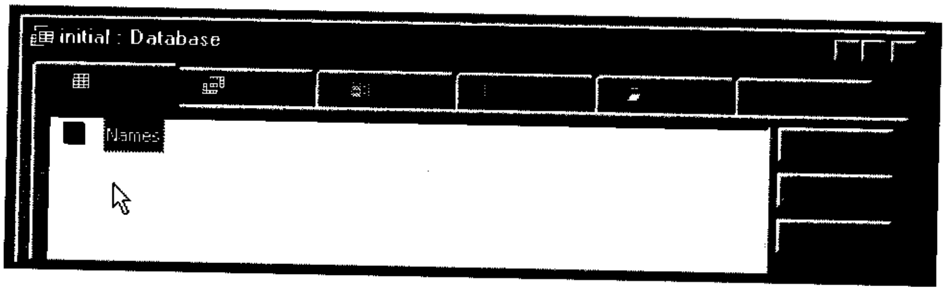
Each installed version of the PEDA model relies on four database files (Microsoft Access files in this case). These files can be found in the PEDA programme folder (Figure 4-1) on the hard drive. By default the PEDA program folder is located in 'c:\Program Files\'.

Figure 4-1: Screen shot of the PEDA program folder



The files 'initial.mdb' and 'init.mdb' contain information on the status of the initialization of the PEDA model for a particular country and the baseline data for already initialized countries. Although the initialization of the model will be dealt with in detail in chapter 5, some comments on the content of these databases may be useful at this stage. The database 'initial.mbd' contains only one table 'Names' and keeps track of the countries which are already initialized and which are not, together with fields presenting standard UN and World Bank codes for the countries in question. This table should not be manipulated manually.

Figure 4-2: The 'Initial.mbd' database

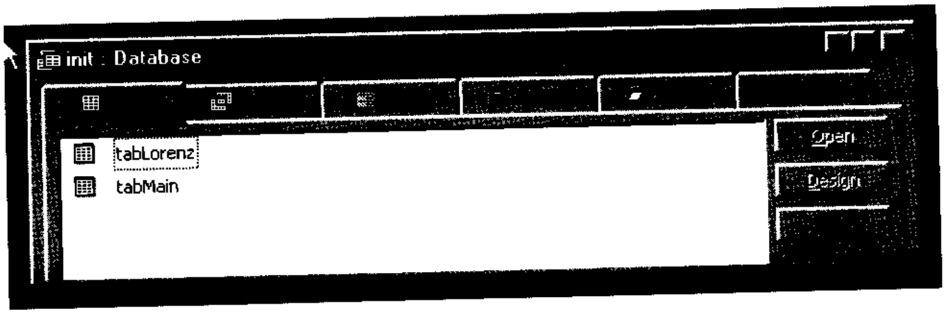


The 'init.mdb' file (Figure 4-3) contains the baseline data for all the initialized countries. These baseline data are made up of the year to which the data apply (i.e. the starting year in the projections); the population size by age and sex; age and sex specific mortality schedules; age specific fertility rates and age and sex specific transition rates from the illiterate to the literate status. All this information is defined for each of the eight different subgroups in the population separately. These demographic baseline characteristics are stored in the table 'tabMain'. The other table in the 'init.mdb' database 'tabLorenz' contains the food distribution function for the rural areas and the urban areas separately.

Note that this is the only information that is not changeable for a particular initialized country. All the other information (e.g. the elasticities of the agricultural production function, the curve defining the relationship between water availability and agricultural production, etc.) is stored as part of the scenario definitions. This has the advantage that the user can for example make different assumptions regarding the elasticities of the agricultural production function, store these as different scenarios for the same country and compare the results directly.

On the other hand, all the information in the 'init.mdb' database cannot be changed once the country has been initialized. Therefore different assumptions regarding the food distribution function, for example, will lead to two different initializations for one country.

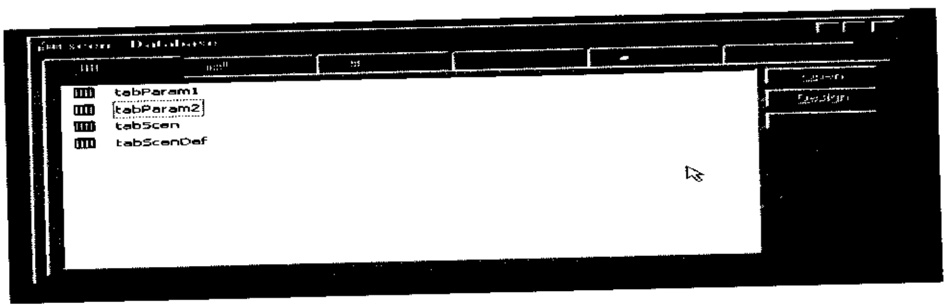
Figure 4-3: The 'init.mbd' database



Except during the initialization of PEDA for a new country, the content of the 'init.mdb' and 'initial.mdb' databases does not change. The information in these databases is used by the application whenever a user selects a country to define new scenarios and carry out projections.

The scenario database ('scen.mdb') contains all the user defined scenarios for all initialized countries. This database consists of four tables (Figure 4-5) and the user in the PEDA application interface has specified most of the information stored here. This database gradually expands as the user specifies more scenarios. The PEDA software will be distributed with at least one (baseline) scenario for each of the initialized countries.

Figure 4-4: The scenario database ('scen.mdb')



In the scenario database, the table 'tabScenDef' keeps track of all the scenarios that have been defined for the initialized countries with their description. For each scenario and each country, the table 'tabParam1' stores the scenario assumptions regarding fertility, mortality and education. A screenshot of the content of that table is shown in Figure 4-7. The second column identifies the country and the third column the scenario name. In the third column the name of the indicator is given and the fourth column further specifies that indicator; usually whether the values are applicable to females or to males. From the fifth column onwards the values of the different variables are given for each of the eight subgroups in the population (st1, st2, ... st8). These values are only applicable to this particular scenario and were originally read from the values that the user entered in the PEDA application interface. Some rows in this table are reserved for future use (e.g. row 9-12); some rows are used to store parameters for calculation purposes (e.g. row 3-6)

Figure 4-5: The 'tabParam1' table in the scenario database

ETH	Baseline	Poor	f				
ETH	Baseline	Poor	m				
ETH	Baseline	C-T	ttr	5	5	6	5
ETH	Baseline	C-T	k	0.534164172007	0.534164172007	0.534164172007	0.534164172007
ETH	Baseline	C-T	a0	12.68725739228	12.68725739228	12.68725739228	12.68725739228
ETH	Baseline	C-T	m	0.244025585470	0.244025585470	0.244025585470	0.244025585470
ETH	Baseline	e0	f	50	42	49	43
ETH	Baseline	e0	m	48	40	47	41
ETH	Baseline						
ETH	Baseline						
ETH	Baseline						
ETH	Baseline						
ETH	Baseline	Education	f			0.95	0.95
ETH	Baseline	Education	m			0.95	0.95
ETH	Baseline						
ETH	Baseline	Year e0 0	f+m	1995	1995	1995	1995
ETH	Baseline	Year e0 1	f+m	2005	2005	2005	2005
ETH	Baseline	Year e0 2	f+m	2015	2015	2015	2015
ETH	Baseline	Year e0 3	f+m	2030	2030	2030	2030
ETH	Baseline	e0 0	f	64.73	59.83507092352	57.54296965804	53.15294492254
ETH	Baseline	e0 1	f	70.90274053183	65.53213191714	63.02179341202	58.21378239696
ETH	Baseline	e0 2	f	77.05429591514	71.21773075086	68.48959408689	63.26443775670
ETH	Baseline	e0 3	f	84.84557567101	78.41885117209	75.41486128399	69.66136770468
ETH	Baseline	e0 0	m	60.42678954275	55.7741414958	53.53828029178	49.38490522503
ETH	Baseline	e0 1	m	66.46020313724	61.34300368537	58.88389918999	54.31582346183
ETH	Baseline	e0 2	m	72.40861498798	66.83340896338	64.15420633962	59.17727246005
ETH	Baseline	e0 3	m	79.88040060537	73.72989364185	70.77422629585	65.28372668952
ETH	Baseline	Year TFR 0	f	1995	1995	1995	1995
ETH	Baseline	Year TFR 1	f	2005	2005	2005	2005
ETH	Baseline	Year TFR 2	f	2015	2015	2015	2015
ETH	Baseline	Year TFR 3	f	2030	2030	2030	2030
ETH	Baseline	TFR 0	f	4.282043238269	4.282043238269	4.831023140611	4.831023140611
ETH	Baseline	TFR 1	f	3.568369365224	3.568369365224	4.025852617176	4.025852617176
ETH	Baseline	TFR 2	f	2.854695492179	2.854695492179	3.220682093741	3.220682093741
ETH	Baseline	TFR 3	f	2.283756393743	2.283756393743	2.576545674993	2.576545674993

The table 'tabParam2' contains another set of important parameters. Unlike those in the 'tabParam1' they are not applicable to each of the eight subgroups separately, but to the country as a whole. These indicators are summarized in Figure 4-9 and are mostly self-explanatory. Most of these indicators are static scenario parameters, i.e. they do not change over time. Some of these parameters do not appear in the PEDA user interface, and if they need to be changed, it has to be done from within the Excel spreadsheets (cfr. infra). It is better not to change the value of any of the indicators in the databases.

Figure 4-6: The 'tabParam2' table in the scenario database

ETH	Baseline	Initial year	Initial year	1995
ETH	Baseline	Proportion of females born	prF	0.48780488
ETH	Baseline	Projection year	Projection year	2050
ETH	Baseline	Minimal consumption of Ckal	minKcal	1700
ETH	Baseline	Initial production of Kcal	initialKcal	1756.8
ETH	Baseline	initial imports of Kcal	initialKcal	73.2
ETH	Baseline	Auxilary	AX2	0
ETH	Baseline	Auxilary	AX3	0
ETH	Baseline	Labor	Labor	0.2
ETH	Baseline	Land	Land	0.088
ETH	Baseline	Fertilizer	Fertilizer	0.162
ETH	Baseline	Machinery	Machinery	0.072
ETH	Baseline	General education	GE+PAR	0.276
ETH	Baseline	Technical education	TE PAR	0.158
ETH	Baseline	Water saturation curve par 1	EWM 0.6	0.1
ETH	Baseline	Water saturation curve par2	EWM 0.75	0.85
ETH	Baseline	Auxilary	AX6	0
ETH	Baseline	Auxilary	AX7	0
ETH	Baseline	Gross migration rate	GMR	0.06
ETH	Baseline	Land recoverey	a	0.0175
ETH	Baseline	Land degradation param.	g	0.03
ETH	Baseline		n	4
ETH	Baseline	Natural resources upper limit	R	1.5
ETH	Baseline	Water Impact Factor	WI	1
ETH	Baseline	Reservoir Capacity Effect	RCE	1

Among the indicators or variables that do not appear in the PEDA user interface are the proportion of the females born (the fraction of each birth cohort that is female), the elasticities of the agricultural production function, and a number of parameters applying to the land and water modules. See the respective sections on land and water for a detailed description of these parameters (see sections 2.3 and 2.4).

For each scenario there are a number of auxiliary indicators. This space is reserved in case future versions of the model would require the inclusion of new variables.

The table 'tabScen' (Figure 4-11) contains the values applying to the dynamic scenario variables of the agricultural production function, the availability of food and the HIV/AIDS morbidity rates. Just as the variables or parameters in the 'tabParam2', these variables are defined at the level of the country (thus not for each of the eight subgroups in the population separately), but unlike the parameters in the 'tabParam2' table, these are dynamic parameters. Different values can be assumed for different years throughout the projection period.

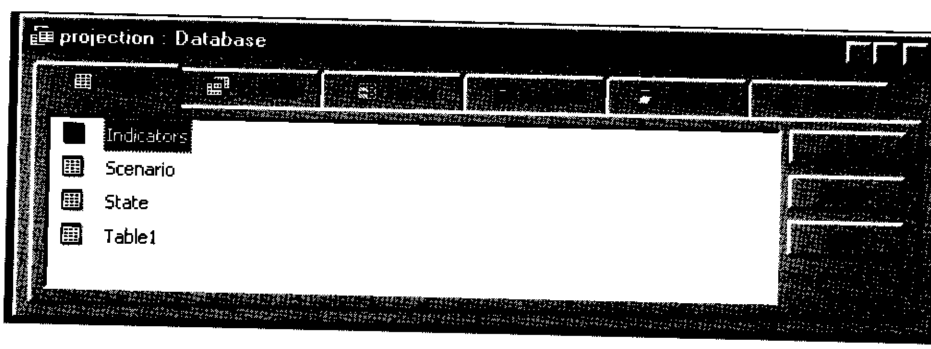
Figure 4-7: The 'tabScen' table in the scenario database

ETH	Baseline	year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
ETH	Baseline	LT	0.15	0.1485	0.14702	0.14554	0.14409	0.14265	0.14122	0.13981	0.13841	0.13703
ETH	Baseline	FIE	1	1.02	1.0404	1.06121	1.08243	1.10408	1.12616	1.14869	1.17166	1.19509
ETH	Baseline	UB	1	1	1	1	1	1	1	1	1	1
ETH	Baseline	HIV	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011
ETH	Baseline	F	1	1.05526	1.11051	1.16577	1.22103	1.27628	1.34680	1.41733	1.48785	1.55837
ETH	Baseline	M	1	1.005	1.01003	1.01508	1.02015	1.02525	1.03038	1.03553	1.04071	1.04591
ETH	Baseline	TE	1	1.015	1.03023	1.04568	1.06136	1.07728	1.09344	1.10984	1.12649	1.14339
ETH	Baseline	W	0.88	1	0.84	0.89	0.78	0.77	0.91	0.85	0.88	1
ETH	Baseline	IR	1	1.01	1.0201	1.03030	1.04060	1.05101	1.06152	1.07214	1.08286	1.09369

The values for the variables stored in this table are the loss of food in the harvest, transport and storage (LT); food imports and exports (FIE); the urban bias factor (UB); the HIV/AIDS morbidity rate (HIV); fertilizer use in agriculture (F), machinery use in agriculture (M); the technical education of the labour force (TE), climate (W); and irrigation (IR). With the exception of HIV/AIDS morbidity rates, the post-harvest losses, the urban bias factor and water, these variables are treated as indexes having value 1 for the starting year. Other factors that affect the availability of food but are endogenously determined by the model (e.g. the size and literacy of the labour force) cannot be found in this table.

The projection database ('projection.mdb', Figure 4-8) stores all projection results. It consists of four tables. In the projection database, the 'indicators' table only contains a list of the indicators that are part of the standardized output modules in PEDA with their abbreviations. In that table, those indicators marked with a '*' are only available at the level of the country as a whole. The 'scenario' table keeps track of all scenarios for which simulations have been carried out and their description. It is different from the 'tabScenDef' in the scenario database in the sense that scenario will only appear here once the simulation for that particular scenario and country has been carried out and stored in the projection database. The table 'State' only contains a description of the eight different subgroups in the population and their abbreviation.

Figure 4-8: The projection database ('projection.mdb')



'Table1' (Figure 4-9) is by far the most bulky table in this database and contains all the projection results for all scenarios and all countries. For each year in the projection period of each scenario for each country, there are currently about 116 records. They keep track of the values of all possible output variables and if appropriate, for each of the eight subgroups in the population and 5 year age intervals. Although projections in PEDA are carried out by single years of age, the results are only stored by 5-year age intervals.

This will be the database table that users will most commonly refer to. It is particularly useful in case the user wants to present projection results in another format than the ones predefined in the PEDA application (see section4.2.2.2). This may for example be the case when one wants to generate age-specific time series. It is fairly easy to extract that information from this table using standard database query techniques.

Figure 4-9 presents a fragment of the 'Table1' in the projection database. The rows labeled 54 to 63 present the population (both sexes) in each of the eight different states for the age groups 55 to 100 for the year 2010 for a scenario called 'mali1'. The next line gives the expected number of births for the year 2010 under the scenario assumptions specified in 'mali1', etc.

Figure 4-9: Fragment of Table1 in the projection database

54	MLI	popr	mali1	2010	t	55	27817.0982437	13781.5942472	17436.562693	9342.92195629
55	MLI	popr	mali1	2010	t	60	27760.5322739	14096.8885179	17628.5909993	9562.67629205
56	MLI	popr	mali1	2010	t	65	28035.1302679	14531.8890149	17497.9042811	9544.9278721
57	MLI	popr	mali1	2010	t	70	24299.5809714	12747.9729850	17684.0617680	9657.06979064
58	MLI	popr	mali1	2010	t	75	15889.2780047	8304.36880038	10481.0253395	5724.84733974
59	MLI	popr	mali1	2010	t	80	7478.00181620	3884.55768437	4484.34037474	2464.88398555
60	MLI	popr	mali1	2010	t	85	2147.33766319	1122.02207493	1562.80762561	870.595483414
61	MLI	popr	mali1	2010	t	90	372.750926558	199.437249908	345.851846294	196.928120449
62	MLI	popr	mali1	2010	t	95	0	0	0	0
63	MLI	popr	mali1	2010	t	100	0	0	0	0
64	MLI	birth	mali1	2010	f	tot	58730.2088709	38524.7128537	106341.726988	51174.6123154
65	MLI	ptotr	mali1	2010	f	tot	673945.504926	353222.224531	1279704.10363	719841.044551
66	MLI	ptotr	mali1	2010	m	tot	696116.126649	358806.250634	1378477.06229	730503.742764
67	MLI	ptotr	mali1	2010	t	tot	1370061.63157	712028.475165	2658181.16591	1450344.78732
68	MLI	tfr	mali1	2010	f	tot	4.2	5	6	5

4.2.2 The integration of the databases in the PEDA application

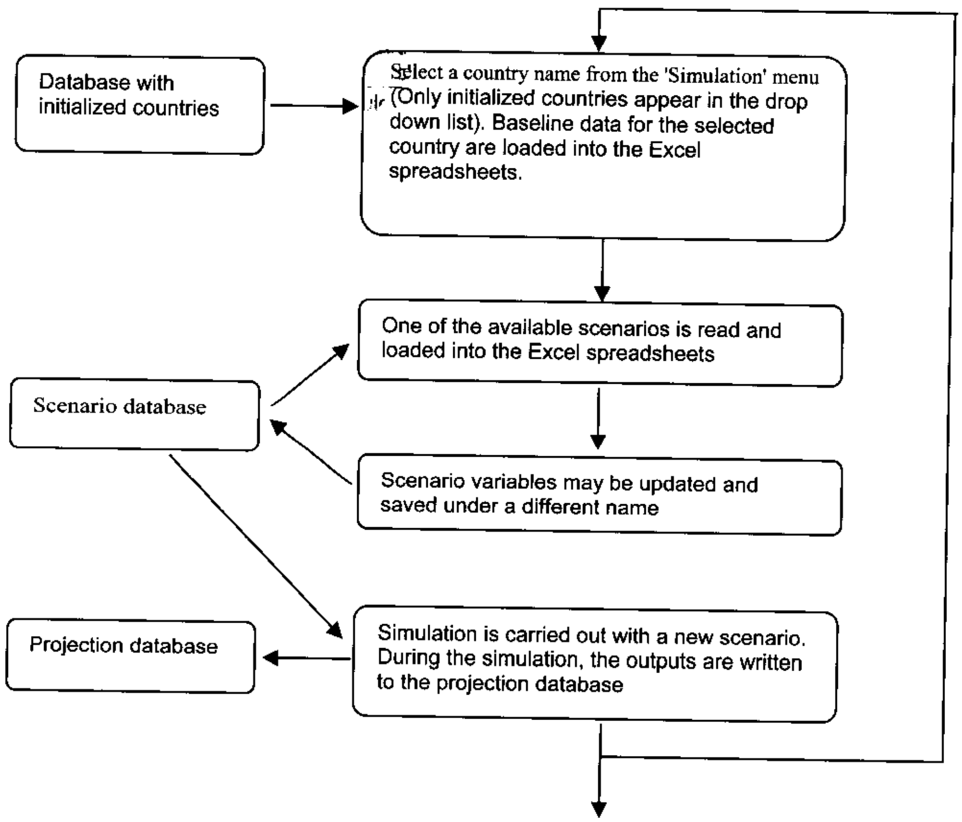
4.2.2.1 In making projections

To carry out projections, the user first has to select a country from the simulation menu of the PEDA application. Only if the country has already been initialized it appears in the list. Once a country is selected, the initial data are loaded from the database into the model or spreadsheet. After the baseline data have been loaded, the user selects a scenario for this particular country. At least one scenario exists for each initialized country. By default this is called the baseline scenario. Reading a scenario loads the values for the model parameters and scenario variables into the spreadsheets.

Once a scenario is selected and loaded into the model the user may begin working with the model. From here onwards, the user can change some of the parameters of the scenario that has been loaded, or he/she may want to carry out projections with the loaded scenario. A scenario that has been changed needs to be saved under a different name before simulations can be carried out. Once simulations are performed, the results are stored in the projection database. It is not required that each scenario that is stored in a scenario database has the output parameters stored in the projection database. Scenarios can be made in advance and stored in a scenario database to carry out simulations later. Scenario definitions alone take less space than the full set of projection results.

The operations involved in carrying out projections are given in Figure 4-10. It may seem as if many operations are involved, but most of them are carried out behind the user interface and can be done in a few mouse clicks.

Figure 4-10: Schematic overview of the steps involved in carrying out projections

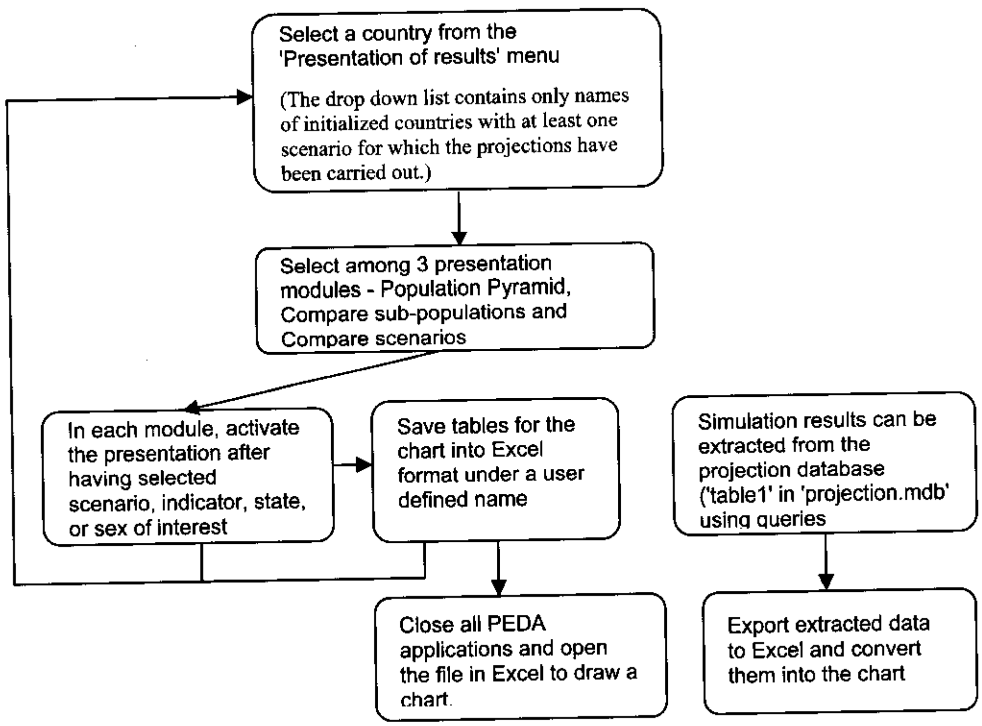


After the simulation has been performed the user may want to repeat the whole process for another country or another scenario.

4.2.2.2 *In presenting the results*

There are three ways in which the user can obtain the data from the projections (Figure 4-11). Of course, this can only be done for countries that have already been initialized and for scenarios for which the projections have been carried out. The first option for the user is to make use of the standard output features of the PEDA application. The latter contain population pyramids and time series. The tables for these charts can be saved from within the PEDA application into Excel format to be reworked later. Refer to the PEDA User's Manual for more details on these procedures. The third option is to extract simulation results from the projection database ('table1' in the 'projection.mdb' file) using queries. As mentioned earlier, this will be most useful to chart characteristics for specific age groups.

Figure 4-11: Schematic overview of the steps involved in presenting the results of projections



4.2.2.3 In managing the information

In order to be able to maintain a well-structured and up to date database, it will be necessary to delete data for particular scenarios and/or countries from the projection database from time to time. Similarly data can also be deleted from the scenario database. Since all databases are MS Access databases these procedures could be easily performed directly from MS Access (see Annex 1).

After the deletion of scenarios or projection results the databases could be compacted in order to reclaim the space that corresponds to the deleted data. This can be done by standard Microsoft Access commands.

4.3 The Excel spreadsheets

The projections and calculations are always carried out in Excel spreadsheets. Just as is the case for the operations with the different databases, this is usually not directly visible to the user of the model but happens behind the user interface. In most cases, a user will not need to go into these spreadsheets to manipulate information or define scenarios. However, a good knowledge of these spreadsheets may enable the user to change some of the more advanced parameter settings if required. In the following paragraphs, the content of the different Excel spreadsheets will be discussed with indications where the user can intervene and adjust parameter settings. Again, the user should be sufficiently familiar with the PEDA model and Excel to do so.

Before changing information in the Excel spreadsheets, you must first start the PEDA application, choose a country from the drop down list and click on the simulation button. Now the baseline data for that country are loaded into the Excel spreadsheets underlying the interface. None of the other information pertaining to that country has been loaded yet. To do so, you have to read a scenario first (baseline, or any other scenario that has been defined for that country). Only in very exceptional cases you may need to go to the Excel spreadsheets without reading a scenario first. This is for example the case when you have just initialized a country for which no scenarios have been defined. Once a scenario is read you can go to the Excel spreadsheets by pressing the <ctrl> <shift> and <R> keys simultaneously. This action resets the main workbook and allows the user to go directly to the PEDA projections workbook for viewing and editing. If settings want to be changed for another scenario than the baseline scenario, this scenario needs to be loaded first (through clicking on the 'Read predefined scenario').

The workbook consists of 17 worksheets. The simulation window in the PEDA application is one of them and is active after pressing the <ctrl> <shift> and <R> keys. This worksheet is called 'Main'. By clicking on the tabs at the bottom of the screen, the user can switch from one worksheet to the other (see Figure 4-19).

Figure 4-12: The different worksheets in the PEDA projections workbook

Hereunder a description is given of each of the 17 worksheets with indications where information may be changed. Whenever information has been changed in any of these worksheets, the user MUST return to the 'Main' worksheet (the one selected in Figure 4-19) and click on the 'save scenario' button to save the changes made under a different scenario name.

In the tables describing the worksheets, those ranges that MAY NOT be changed by the user in any case are labeled with a ' ! '; those ranges that contain information that can ONLY be changed in the worksheets and not through the PEDA user interface are labeled with ' * '; ranges that contain information that can both be changed through the user interface and in the spreadsheets are not labeled. To avoid problems it is preferable to change the information through the user interface wherever possible. Wherever ranges are not mentioned in the description that follows, they should not be touched in any case.

Some of the worksheets where changes are allowed are protected. To unprotect a worksheet, choose the 'Protection -> Unprotect sheet ...' command from the 'Tools' menu.

4.3.1 The 'Params' worksheet

This worksheet contains some major scenario parameters, including fertility and mortality assumptions, the food distribution function and the water saturation function.

Table 4-1: The content of the 'Params' worksheet

Range	Description	
B1:D1	The message displayed in the Simulation/Parameter settings window in the PEDA application interface	*
D5:M16	This range is used for testing purposes and for future use.	
D17:M39	Scenario settings for educational transition rates and the dynamic settings for fertility and mortality for each of the eight subgroups	
A48:CX52	The definition of the food distribution curve for rural and urban areas	!
AO23:AR30	The definition of the water saturation function for the calculation EWM. In the current version of PEDA, two of the water saturation curve parameters can be changed in the 'Params1' worksheet (see section 4.3.2)	!
Y7:AA16	Definition of the names of the subpopulations	!
AH7:AI16	Description of the dynamic scenario variables and their abbreviations	!
AO4:AW17	Information used in the definition of the graphs	!

4.3.2 The 'Params1' worksheet

This worksheet contains some important model and scenario parameters, including the elasticities of the agricultural production function, the parameters for the land and water modules and HIV/AIDS related mortality schedules.

Table 4-2: The content of the 'Params I' worksheet

Range	Description	
E4:G4	The initial year of the projections. This is the year to which the baseline data apply and is defined as part of initializing the model for a new country.	
E5:G5	The proportion of females born. This parameter refers to the sex ratio at birth. As it can change a little between countries a value can be specified here. This is the only place where this parameter can be adjusted.	*
E6:G6	The end year of the projection period.	
E7:G9	Parameter values for some of the general model settings, including the assumptions regarding the amount of daily per capita calories produced and imported in the initial year, the minimum daily per capita food requirements. All these values can be changed through the user interface.	
E10:G11 & E20:G21	Ranges not used in the current version of the PEDA model. This space is reserved in case future versions of PEDA would require the inclusion of new variables.	*
E12:G17	Elasticities of the agricultural production function.	*
E18:G19	In the new version of PEDA, two of the water saturation curve parameters can be changed here by the user and stored as part of the scenarios. A reference to these cells is made from appropriate cells in the 'Params' worksheet (see section 4.3.1).	*
E22:G22	Parameter for the proportion of the cohort moving from rural to urban areas during their lifetime. This value can also be changed through the user interface.	
E23:G26	Parameters of the land module. Of these, only the land degradation impact factor can be changed through the user interface. (g stands for gamma, n for eta and R for Rbar, See section 2.3 for their full description).	*
E27:G28	Two parameters of the water module. See section 2.4 for their full description.	*
L3:W105	Definition of the age specific additional mortality rates (mx) for different levels of HIV/AIDS morbidity rates.	*

4.3.3 The 'SetScen' worksheet

This worksheet contains the definition of the user interface for setting the dynamic scenario variables. It should not be changed.

4.3.4 The 'NDScen' worksheet

In this worksheet the values for the dynamic scenario variables are stored. It is recommended to change these values through the user interface. However, as the user interface only allows assumptions to be made by five-year periods or a gradual annual increase or decrease, the user can define in these worksheets scenario assumptions with a precision of one year.

Table 4-3: The content of the 'NDScen' worksheet

Range	Description	
D6:DE15	Definition of the values of the dynamic scenario variables	*
All other ranges	All other ranges on this worksheet are used for calculating purposes and should not be touched	!

4.3.5 The 'DistribF' worksheet

In this spreadsheet, most calculations are carried out concerning the food production and the calculation of the food secure and food insecure fraction of the population.

Table 4-4: The content of the 'DistribF' worksheet

Range	Description	
B2:C9	Elasticities of the food production function and two parameters of the water module. They cannot be changed here. (See section 4.3.2)	!
E1:F7	Range wherein the gross daily food production and the net daily food imports pertaining to the total population are calculated from the assumptions made on the net daily per capita food produced and imported. They relate to the initial year only and the results are used as input for calculation of the food production and availability during the projection period. No change is allowed.	!
A11:DC26	Range wherein the food production and availability are calculated for the whole projection period. Different assumptions can be made in this range regarding the variables that affect food production (e.g. the endogenous feed back of HIV/AIDS morbidity on the labour force input into the agricultural production function). Only users that are very well acquainted with PEDA should engage in making changes in this range. Changes made here cannot be stored as part of the scenarios. They should be saved through the 'save' option in the 'File' menu of the workbook. Note that, once saved, the new mathematical definition of the production function will be applied to all future simulations and that previously carried out simulations will not correspond to the new definition.	*
A36: DD152	Range used for calculation of the population per sub group for each year in the projection period. No changes should be made.	!

4.3.6 The 'St1' through 'St8' and 'Total' spreadsheets

These spreadsheets are used for the calculation of the population by age and sex for each of the eight different subgroups in the population. In the 'Total' spreadsheet, the data are aggregated for males, females and the total population. This spreadsheet is also used for the calculation of literate life expectancy (LLE). No changes are allowed in these spreadsheets.

4.3.7 The 'IPop' spreadsheet

This worksheet contains the initial or baseline data for each country. These data are automatically read from the 'init.mdb' database that is updated with data for each country during the process of initialization. It is not advisable to change the information in these spreadsheets.

Table 4-5: The content of the 'Ipop' worksheet

Range	Description	
E8:L98	The total population by single years of age and subgroup, females	!
E130:L222	The total population by single years of age and subgroup, males	!
E254: L357	Age specific mortality rates by subgroup, females	!
E378: L481	Age specific mortality rates by subgroup, males	!
E502:L525	Age specific transition rates from illiterate to literate status, females	!
E527:L550	Age specific transition rates from illiterate to literate status, males	!
E556:L596	Age specific fertility rates	!
A5:C596	Auxiliary range used for performing calculations. No changes are allowed.	!

4.3.8 The 'Main' worksheet

This is the worksheet that has to be activated if one wants to return to the PEDA application user interface. If any changes have been made in any of the worksheets; these can be saved ONLY by clicking on the save scenario button in this worksheet. Do not try to save the changes through the File/Save command as you would do in any other software package.

4.3.9 The 'Results' worksheet

This worksheet regroups the outputs of the projections before they are stored into the projections database. This worksheet is of little use to the user as it only contains temporary results of the projections for one year at the time (the values of the output variables for the last year of the projection period in the scenario under consideration). No changes are allowed in this worksheet.

5 Methodological support for the initialization of PEDA for new countries

5.1 Introduction

Since PEDA is a multi-sectoral model, the initialization of PEDA for a new country should ideally be done by a team of experts from various disciplines. An in depth initialization of the model for a particular country includes among others the preparation of the demographic baseline data, the estimation of the agricultural production function elasticities (if necessary, it may even entail the definition of a new agricultural production function), the estimation of the water saturation curve, the estimation of the land degradation and recovery parameters, the estimation of the food distribution curves, the estimation of age specific rural-urban migration patterns, and possibly even the inclusion of new variables that prove to be important in the population, environment, development and agriculture interactions for that particular country.

However, there is a more minimal approach to the initialization that is satisfactory if the PEDA model is to be used for advocacy purposes only. This approach entails the preparation of the demographic baseline data and the estimation of the food distribution curve. For all the other relationships, the model will then rely on the theoretical distributions as described in the previous sections of this manual. Here, you will find some guidelines in the preparation of these baseline data. These are only to be seen as possible steps to be followed as the data available for different countries may come in different formats and may thus require different estimation techniques etc.

These baseline data contain the necessary information to be fed into the 'init.mdb' database and are the technical minimum requirement to make PEDA operational for a new country. Although it is the minimum, it is not necessarily the end of the initialization process. Further improvements to the data and relationships can be gradually implemented as they become available. As mentioned before most of the other relationships and distributions are stored in databases as part of the scenarios and not as part of the initial data per se.

If a country has already been initialized and a better empirical description of the water-saturation function becomes available at a later stage, it can be easily added to PEDA to replace the current theoretical function and stored as part of a new scenario. The PEDA model thus allows for a gradual improvement of the data and the country applications.

The necessary baseline data include the distribution of the population by sex and single years of age for each of the eight subgroups in the population (states); the food distribution functions for urban and rural areas separately; fertility and mortality schedules for each of the eight states; and the age specific educational transition rates for the non literate states. Hereunder the different steps in the initialization process will be illustrated with data for Ethiopia.

The preparation of the baseline data is most likely to be done in Excel, and after all initial data are collected and prepared, they are fed into a small utility that comes with the PEDA model that automatically updates the databases (see section 5.6).

5.2 The food-distribution function

As we assume that all the citizens of a country do not have equal access to the available food in a country, a food distribution function has to be defined to reflect that inequality. This food distribution function is of crucial importance in the process of estimating the fraction of the population that is considered food secure and food insecure. Adequate data on such a food distribution functions do, however, not exist. The solution applied in PEDA is to use household expenditure curves that exist for a number of African countries. The model thus relies on the assumption that the inequality in the expenditure curve is equal to the inequality in the access to food.

This relationship does not seem to be immediately obvious, but as we know that savings are relatively low in Africa, the step to assume that income equals expenditure is only a small one. What we are left with is the assumption that the inequality in the access to income is similar to the inequality in the access to food. The knowledge that income inequality is usually different in rural and urban areas is extended into PEDA through the use of two food distribution curves, one for rural areas and one for urban areas.

In our example of Ethiopia, no data are available on expenditures or income. Therefore, we used household expenditure data for Uganda as a proxy. These data are available from the World Bank[1]. The data give household expenditures in national currency from the poorest quintile to the richest. These figures include the value of own produce consumed. As we are thinking in terms of income or food distribution in PEDA, we will consider these figures to represent income from now on.

Table 5-1: Preparation of the food distribution function

			Urban			
Pop quintile	**Mean income**	**% of the total income**	**cum income**	**Mean income**	**% of the total income**	**cum income**
			0.0			0
1	4	7.0	7.0	7	5.7	5.7
2	7	12.3	19.3	12	9.8	15.4
3	9	15.8	35.1	17	13.8	29.3
4	13	22.8	57.9	25	20.3	49.6
5	24	42.1	100.0	62	50.4	100.0
	57			123		

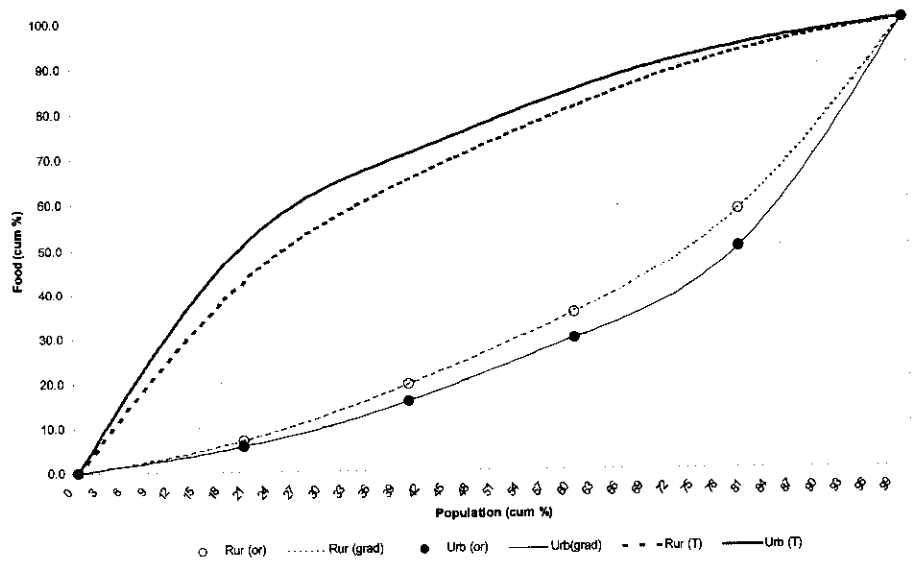
As we do not need national currency as a measurement unit, but simply the relative share of each quintile in the total income, we first need to transform the actual income into fractions of the total income that is obtained by each quintile of the population (see Table 5-1). These fractions are then cumulated to obtain the total income available to the total population (in rural and urban areas separately). Read from the bottom to the top, it means that for the rural population 100 per cent of the income/food is available to 100 per cent

1 African Development Indicators 1997, World Bank, Washington, D.C., p. 387-431

of the population, and 58 per cent of the income/food is available to 80 per cent of the population. In this case, the richest 20 per cent of the rural population has already consumed 42 per cent of the food, leaving the rest for the less privileged. Similarly, 65 per cent of the food is consumed by the most privileged 40 per cent of the population, leaving the remaining 35 per cent of the food for the 60 least privileged per cent of the population.

PEDA determines the fraction and number food insecure people in a subgroup on the basis of these food distribution curves and the amount of available food in a particular country and year. If the remaining amount of food per capita is not sufficient to fulfill minimum food requirements, the remaining fraction of the population is considered to be food insecure. To determine this we need two additional pieces of information; i.e. the amount of per capita food supply in a country and the assumed minimum requirement. The FAO usually collects and publishes these data[1].

Figure 5-1: Original data points, the graduated and transposed food distribution curves for rural and urban areas, Ethiopia



The data, as given in Table 5-1 are, however, not sufficient as the PEDA model works with a precision of 1 percentage point. Both curves thus need to be graduated before they are of any use to PEDA. The graduation of these and similar curves can be done, using a variety of techniques. In this example we used Spline[2]. The result of the graduation is shown in Figure 5-1. The cumulative proportion of the population is given on the X-axis ordered from least to most privileged, the cumulative share of the food on the Y-axis. The circles stand for the original data points, the dotted and full lines for the graduated curves. Before these Lorenz curves can be entered in PEDA, they need to be transposed so that the most privileged section of the population is

1 A good source of these data is the 'Africa Nutrition Database Initiative Website' (http://www.africanutrition.net).

2 A small Excel application for Spline has been developed by S. Scherbov that simplifies this task (see Annex 2).

situated towards the origin of the X-axis. In Figure 5-1, the transposed curves as they have to be entered in the PEDA model are indicated with (Rur (T) & and Urb (T)).

To come back to our example of Ethiopia, the daily average per capita food supply in 1995 is to be situated around 1830 kcal. With an assumed threshold value of 1500 kcal[1] as the minimum daily food required for an individual to be food secure, a proportional distribution of the available food along the food distribution curves as defined above for urban and rural areas would result in a proportion of 53 per cent of the rural population and 67 per cent of the urban population being food insecure.

5.3 The age and sex distribution of the population for each of the eight states

PEDA relies on multi state demographic techniques whereby eight different subgroups in the population are projected at the same time while transitions from one group to the other at each step in the projection period are possible. The eight different states are defined on the basis of three dichotomous characteristics: rural/urban place of residence, literacy status and food security status. The eight states are:

st1: Urban, literate, food secure
st2: Urban, literate, food insecure
st3: Urban, illiterate, food secure
st4: Urban, illiterate, food insecure
st5: Rural, literate, food secure
st6: Rural, literate, food insecure
st7: Rural, illiterate, food secure
st8: Rural, illiterate, food insecure

Note that literacy status[2] has been used so far as a cut-off point for educational status in the initialized countries. However, if there are enough reasons to believe that other distinctions (e.g. that between complete primary education and not complete primary education) may be of more relevance in explaining differences in fertility rates, productivity or in the relationship of human beings with the environment, these alternative classifications can be used as well.

For each of the eight states, a population distribution by single years of age (0-100) and sex needs to be defined. This information is usually not readily available from any document. A census report or UN publications are good sources to find population distributions by age, sex and often also urban/rural place of residence. Sometimes these distributions are even given by single years of age, sometimes only in five year age groups.

---

1 Note that we have chosen in this example for an extremely low value. Usually the desirable daily per capita consumption is considered to be around 2000 Kcal or even higher.

2 Following the UNESCO definition, a person is considered literate if he/she can both read and write with understanding a short simple statement on his/her everyday life (see the statistics component of the UNESCO website for full description, www.unesco.org).

Be aware that the population distribution by single years of age reported in census monographs often concern the observed population and are not necessarily the best data to work with. Population projections require the mid-year population and observed population distributions are often subject to many errors such as age misreporting.

Therefore, it is preferable to use the population distributions at the mid year following the year of the census and tables giving the population distribution in five-year age groups and graduate them later to obtain the population distribution by single years of age. These corrected population distributions are usually reported in census monographs. The starting table would thus look like the one in Figure 5-2.

Figure 5-2: The population distribution for Ethiopia (1995) by age, sex and rural-urban place of residence

Pop by age sex and place of residence, July 1995 (CSA, 1999:302)					
	Urban		Rural		
age-group	Male	Female	Male	Female	
0-4	435,140	415,519	4,087,055	3,980,497	
5-9	466,269	485,985	3,615,734	3,459,602	
10-14	486,684	540,154	3,138,687	2,951,401	
15-19	453,512	512,854	2,633,003	2,506,687	
20-24	392,116	443,216	2,049,295	2,058,120	
25-29	329,663	370,252	1,653,812	1,722,678	
30-34	252,162	278,391	1,291,759	1,422,906	
35-39	204,111	217,067	1,072,736	1,183,876	
40-44	167,350	161,023	929,828	970,159	
45-49	133,362	125,469	789,829	789,509	
50-54	98,783	98,739	657,738	614,389	
55-59	75,796	78,924	544,319	484,922	
60-64	57,417	64,182	439,758	374,701	
65-69	42,298	48,992	340,619	278,014	
70-74	28,972	34,270	244,639	188,808	
75-79	17,449	20,013	151,804	107,080	
80+	21,541	29,025	195,380	133,110	
total	3,662,625	3,924,075	23,835,995	23,226,459	**54,649,154**

Age, sex and rural-urban specific literacy rates can usually be obtained from census reports as well. One thus has to apply the proportions literate to each of the subgroups in Figure 5-2 to obtain the population by sex, age, literacy status and rural-urban place of residence. This gives us the distribution of four subgroups by (five year) age (groups) and sex.

These four subgroups or states still have to be split up by food security status to obtain the eight states required in PEDA. On this issue no data are available at all. The only information we have (estimated ourselves under section 5.2) is the proportion of food insecure people in urban areas and rural areas separately. It is then up to the user whether to apply these proportions to both states of literacy, sex and age groups equally or not.

As we do not have any empirical evidence of a potential unequal distribution of food insecurity over the population by sex, literacy status and age, earlier initializations of PEDA have only accounted for rural-urban differences in the food security status of the population. The proportions food secure and food insecure obtained under section 5.2 will thus be applied to both literate and illiterate states to obtain the eight states by sex and (five year) age (groups).

Because PEDA carries out projections by single years of age and uses 100 age groups, these data still need to be graduated. Again, Spline is an ideal technique to do so (see Annex 2).

5.4 Age- and sex-specific death rates (mx) for each of the eight states

In order to carry out the population projections, age and sex specific death rates need to be prepared for each of the eight states. Since these are usually not available for each of the subgroups separately, they have to be estimated. The first step in such a procedure is to estimate life expectancy at birth (e_0) for each of the eight states by sex. Usually census reports give the value of e_0 for both sexes separately and by another characteristic such as rural/urban place of residence or literacy status.

In addition, census reports often present a table with differences in e_0 by a number of background characteristics that may also be of use in the estimation process. No standard techniques exist for such an estimation procedure. They are necessarily experimental in their nature.

In the case of Ethiopia, e_0 is given for males and females by place of residence (CSA, 1999: 238). These are the four 'observed' and aggregated e_0's from which we will estimate the life expectancies for each of the eight states. Life expectancies for literates and illiterates in Ethiopia are only partly available from the census report (CSA, 1999: 242), and for food secure and food insecure they are not available at all. In estimating life expectancy for the food secure and food insecure groups, we for example assumed a deviation of two years from the national level e_0. These estimates are then to be turned into a multiplier effect of being in the category literate, illiterate, food secure and food insecure (see Figure 5-3). These multipliers are applied to the four observed values in the population to obtain the e_0 for the eight states by sex.

Figure 5-3: Illustration of the estimation of e0 by state and sex

Observed					
				M	F
overall LE	50.7		Rural	48.8	51
			Urban	52.7	56
First guess			contribution of literacy		
literate	56		1.104536		
illiterate	49.7		0.980276		
			contribution of food security		
food secur	52.7		1.039448		
food insec	48.7		0.960552		
First guess of LE by sex and state					
				Fsecure	F insecure
male	rural	literate		56.0	51.8
		illiterate		49.7	46.0
	urban	literate		60.5	55.9
		illiterate		53.7	49.6
female	rural	literate		58.6	54.1
		illiterate		52.0	48.0
	urban	literate		64.3	59.4
		illiterate		57.1	52.7

To obtain the e_0 for the rural, literate and food secure males; we thus multiplied the cells F8, E12, and E15 in Figure 5-3. These are our first estimates of life expectancy at birth by sex for each of the eight states.

In the next step, the MATCH module of MORTPAK (UN-DESA, 1988) can be used to generate age specific probabilities of dying (q_x), based on the e_0 for each of the different subgroups. In the case of Ethiopia, we used the Coale-Demeney West model life table.

Since PEDA works with single years of age, the output of Match needs to be graduated using UNABR, another module of MORTPAK. UNABR also gives Heligman-Pollard parameters that can be used to extend these life tables to the age of 100 (PEDA works with 100 single year age groups). Another method of extending the life tables can be done through a linear transformation and regression.

From here onwards, we could proceed to calculate the age specific death rates (m_x) that are needed as part of the initial data. However, it is recommended to test whether our estimates for the life expectancies for each of the eight states by sex and the transformations carried out afterwards are still concordant with the aggregated values for e_0 as given in the census reports.

In the case of Ethiopia, we had four 'observed' values for e_0 (e_0 by sex and rural urban place of residence). Let's take for example the case of urban females. Their 'observed' e_0 was 56 and starting from this value, we estimated the values for the urban females by literacy and food security status. If we build life tables around the four nested subgroups of urban females (starting from q_x we obtained as an output of UNABR and the population by age, sex and state that was obtained following the procedures in section 5.3), we can estimate the number of age specific deaths in each subgroup.

The population and age-specific deaths can be summed up to obtain the aggregated population distribution and age specific deaths for the urban females. These two columns can be used as the basis of a new life table to calculate the aggregated life expectancy for the urban females (Figure 5-4). If that value is close to the 'observed' value, the estimates are plausible. If the aggregated life expectancy is far from the 'observed' value, the estimation procedure as illustrated in Figure 5-3 needs to be repeated with different assumptions.

Figure 5-4: Testing the plausibility of the estimated life expectancies through aggregation and comparison with the original value

alpha:	0							
female urban								
Px	Dx	mx	qx	l_x	d_x	LL_x	T_x	e_x
74897	7914	0.105663	0.1003607	100000	10036	94982	5543211	55.4
76948	2040	0.026518	0.02617071	89964	2354	88787	5448229	60.6
81052	1135	0.014005	0.01390722	87610	1218	87000	5359442	61.2
87207	780	0.008942	0.00890259	86391	769	86007	5272442	61.0
95415	600	0.00629	0.00626978	85622	537	85354	5186435	60.6
104186	492	0.004718	0.0047067	85085	400	84885	5101082	60.0

In Figure 5-4, the columns Px and Dx are the sum of the respective columns of the nested life tables for urban females. From these two columns a life table can be built to obtain an aggregated e_0 for urban females. In our case the value is 55.4, which is very close to the value of 56.0 as reported in the census monographs. This means that our estimates are plausible and that we can use the m_x values of the nested life tables as an input to PEDA.

If desirable, it is fairly easy to adjust the nested m_x values proportionally, to obtain an aggregated value that precisely equals the observed e_0 of 56. In Excel it can be done through creating new m_x columns for each of the nested life tables that are equal to the old m_x column times 1+ alpha. This is illustrated in Figure 5-7.

Figure 5-5: The " * (1+alpha)" transformation of the mx column for urban illiterate food secure females

Illiterate, Food Secure				
pop	Mx	mx*(1+alpha)	Dx	e_x
24716	0.08967	0.087297	2158	57.5
25393	0.02095	0.020396	518	61.7
26747	0.01107	0.010777	288	62.0
28778	0.00712	0.006932	199	61.7
31487	0.00505	0.004916	155	61.1
34381	0.00382	0.003719	128	60.4

The new m_x column has to be used as the one to build the life table of each of the nested groups and to calculate the aggregated number of age specific deaths etc. With the 'Goal seek' function in Excel it is easy to scale the mortality schedules of the nested subgroups proportionally up or down with the objective of obtaining, in this example, the aggregated e_0 of 56 for urban females. The new m_x values will then be the adjusted mortality schedules that are to be used as the part of the baseline data for Ethiopia. The new values for e_0 also need to be introduced as those reflecting the situation in the initial year.

The example given here is to be repeated four times as we had four observed values for e_0 with each four nested values that must be estimated and subsequently tested through aggregation. If only two observed values exist for e_0 the procedure needs to be repeated twice but for eight nested subpopulations.

5.5 Age-specific fertility rates (ASFR) for each of the eight states

The procedure to follow in estimating age specific fertility rates (ASFR) for each of the eight states is analogous to the one used in estimating age specific mortality rates. Starting from the observed and aggregated TFRs, we estimate those for the nested subgroups. To obtain the age specific fertility rates for each of these subgroups, we scale the reported and aggregated ASFR up or down to make them consistent with the nested TFRs. From the estimated ASFR, we work backward to calculate the aggregated TFRs and control whether they are consistent with the observed/reported values.

In the case of Ethiopia, the census report gives observed and adjusted values for the TFR by rural/urban place of residence (CSA, 1999: 221-233). As fertility is often underreported (e.g. through underreporting children that passed away), adjusted TFRs are to be used preferably. To obtain fertility estimates for each of the eight states, we again have to rely on some estimation techniques. One possible way to do that is illustrated in Figure 5-6.

In addition to the observed and adjusted TFR on the national level and by rural urban place of residence (range B7:D8), the census report also gives an indication of the fertility differentials by educational level (range B10:C11). Fertility differentials by food security status are not available.

The adjusted TFRs by place of residence will be the starting point to make estimates for the nested TFRs. From the fertility differentials by educational status, we can compute a tentative multiplier (range G5:G6). To calculate a similar multiplier for food security status we do not have any empirical evidence. However, we can rely on the literature (see section 2.1.1) to assume that the fertility of the rural food insecure is somewhat higher than that of the rural food secure as may it be a reaction of the rural poor that more children are needed to participate in food production activities.

In urban areas, no such discrepancy in fertility rates of the food-secure and food- insecure segments of the population are assumed (range F9:H10). Note that the multipliers calculated here are relatively arbitrary. For several reasons, one could decide to make other assumptions regarding the fertility differentials by food security status.

From these multipliers and the observed TFRs by rural-urban place of residence, we can make our initial estimates of the TFR for the nested subgroups. The value of 4.08 for the Urban Literate Food Secure for example, results from multiplying cells D8, G6 and G9.

Figure 5-6: Illustration of the estimation of TFR by state

M5 =D8*G6*G9

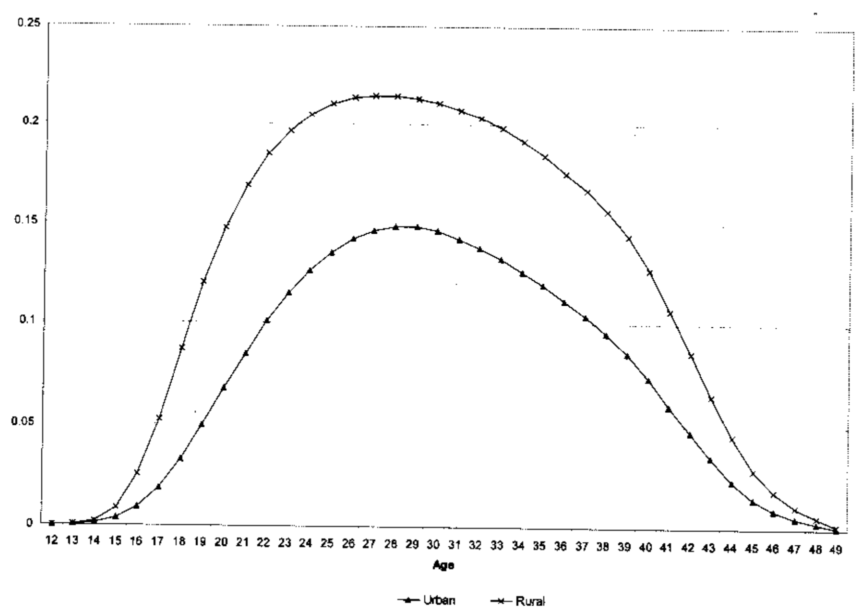
TFR from census report (CSA, 1999: 229)	obs	adj
country	4.3	6.74
rural	4.6	7.19
urban	2.8	4.5
illiterate	4.4	
literate	3.9 ?	

Estimated multipliers for literacy and food security		
multipl for illiteracy	1.023256	
multipl for literacy	0.906977	
	urban	rural
multipl for food sec	1	0.930233
multipl for food insec	1	1.069767

First estimates of TFR by state	
Urban literate Food Secure	4.08
Urban Literate Food Insecure	4.08
Urban Illiterate Food Secure	4.60
Urban illiterate Food Insecure	4.60
Rural Literate Food Secure	6.07
Rural Literate Food Insecure	6.98
Rural Illiterate Food Secure	6.84
Rural Illiterate Food Insecure	7.87

As an input, PEDA needs both, the TFR for each of the nested subgroups and the ASFR. To obtain ASFR for each of the eight states, we can use ASFR reported in census monographs. For Ethiopia, the census report gives ASFR by five-year age groups for urban and rural areas separately (CSA, 1999: 221). Because PEDA works with single years of age, these age specific fertility patterns need to be graduated. In this case this has been done using the Coale-Trussell function (Coale & Trussell,1974): and the result is shown in Figure 5-7. The estimation of parameters was performed using a Gauss-Newton like algorithm (Scherbov & Golubkov, 1986).

Figure 5-7: Graduated ASFR for urban and rural areas, Ethiopia 1994



These two patterns of ASFR still need to be scaled up or down to make them consistent with the TFR that we estimated for the nested subgroups.

Just as we did after estimating the age specific death rates for each state, we can now work backwards to check whether our estimates for the ASFR of the nested subgroups are still concordant with the aggregated values for the TFR reported in the census monographs. Here, we will illustrate this procedure for rural women.

We started from the (adjusted) TFR for rural women of 7.19, and from that value we made an initial guess for the TFR of the nested subgroups. In the last step, we computed the age specific fertility pattern for each of the four rural subgroups.

Given the age distribution of women in each of the four rural states and the ASFR, we can calculate the age specific number of births in each of the subgroups. These births will be summed up, and from these numbers and the aggregated number of women we can again calculate the ASFR for the rural women. If the sum of these values is close to the adjusted TFR of 7.19, our estimates are consistent with the reported value in the census monograph (though not necessarily correct). If not, the estimation procedure illustrated in Figure 5-6 needs to be repeated with different assumptions.

Similar to the procedure illustrated for the age specific death rates and e_0, it is fairly easy to adjust the nested ASFR to obtain an aggregate value for the TFR that is identical to the reported value of 7.19. This is illustrated in Figure 5-8.

Figure 5-8: The "*(1+alpha)" transformation of the ASFR for rural, illiterate, food insecure women and the calculation of the aggregated ASFR for rural women

(a) Before the transformation

V9 =U9*(1+$Z$2)

						alpha:	0	
	St 8: rural, illiterate, food insecure					**Aggregate rural**		
	TFR	7.87	**7.87**				TFR	**7.31**
Age	Women	ASFR	ASFR*(1+alpha)	Births		Women	Births	ASFR
12	282609	0.000019	0.000019	5.40		585,527	10	0.000018
13	268743	0.000400	0.000400	107.49		563,603	209	0.000371
14	256648	0.003352	0.003352	860.25		546,433	1,697	0.003106
15	246070	0.015008	0.015008	3,692.91		532,944	7,398	0.013881
16	235987	0.042976	0.042976	10,141.82		518,835	20,593	0.039692

(b) After the transformation

V9	=U9*(1+$Z$2)							
						alpha:	-0.0166	
	St 8: rural, illiterate, food insecure				Aggregate rural			
	TFR	7.87	7.74				TFR	7.19
Age	Women	ASFR	ASFR*(1+alpha)	Births	Women	Births	ASFR	
12	282609	0.000019	0.000019	5.31	585,527	10	0.000017	
13	268743	0.000400	0.000393	105.71	563,603	206	0.000365	
14	256648	0.003352	0.003296	845.98	546,433	1,669	0.003054	
15	246070	0.015008	0.014759	3,631.63	532,944	7,275	0.013651	
16	235987	0.042976	0.042263	9,973.51	518,835	20,252	0.039033	

For each of the nested subgroups, we first need to create a new column with ASFR that equals the old ASFR times (1+alpha). This new column is then used to calculate the age specific births and these will be the basis to calculate the aggregated number of age specific births and ASFR for rural women. With the 'Goal seek' function in Excel it is then easy to scale the ASFR for each of the nested subgroups up or down with the objective of obtaining a TFR for rural women of exactly 7.19. The new ASFR for the nested subgroups are then to be used as part of the baseline data. The example given here, is to be repeated two times as we had two reported TFR values in the census monograph with four nested subgroups each.

5.6 Initializing the model for a new country

Once the baseline data are prepared following the guidelines under section 5.1 through 5.5, the model can be initialized for that particular country. Note that these are only the minimum requirements to initialize PEDA for a new country. A thorough initialization would require the definition and estimation of many other functions and model parameters regarding the agricultural production function, the land and water modules etc. As mentioned before, many of them can be adjusted at a later stage and saved as part of the scenarios.

To facilitate the initialization, a small utility has been developed that automatically updates the necessary databases. It is an Excel file that comes with the PEDA CD-ROM, and by default it is called 'dataIni.xls' (the name may be changed if considered necessary). Note that this file contains macros, and when opening the file a warning message is displayed, indicating that files with macros may contain viruses. To use the utility, you have to select the option in the message window that enables macros. Additionally, the utility will only function properly if it is located in and opened from the PEDA program folder on the hard disk. As 'dataIni.xls' is not part of the standard PEDA software at the time of the installation (see Figure 4-1), you are required to copy manually the utility from its original source into the PEDA program folder by using the 'Copy' and 'Paste' functions on the 'Edit' menu in 'My Computer' or 'Windows Explorer'.

As illustrated in Figure 5-9, the 'dataIni.xls' file contains three worksheets: 'Control', 'tabMain' and 'tabLorenz'.

Figure 5-9: The three worksheets of the 'dataIni.xls' utility

The 'Control' worksheet contains the country code and three buttons (see Figure 5-15). In this worksheet the World Bank country code has to be specified in cell C2 (for Ethiopia that is 'eth'). A complete list of World Bank country codes can be found in the 'Names' table of the 'initial.mbd' database (see section 4.2.1).

Figure 5-10: The control worksheet of the 'dataIni.xls' utility

From the three buttons on the Control worksheet, the 'Append to Tab' button will be of most use. After having arranged the necessary data in the two other worksheets properly (cfr. infra), you can click on that button to initialize the model for the specified country and to add the data for that country to the set of already initialized countries.

The 'Create New Tabs' button deletes all formerly initialized countries from the PEDA application installed on your computer and replaces them by the country for which the data have been prepared in the 'dataIni.xls' utility.

The third button, deletes the country for which the code is specified in cell C2 from the databases of the installed version of PEDA.

The 'tabMain' worksheet is designed to contain all the necessary demographic baseline data needed to initialize the model for a new country. One will have to arrange and copy all the baseline data from the Excel spreadsheets wherein the preparation of the data has been done into this worksheet following the format described in Table 5-2. For the 'dataIni.xls' utility to work properly, none of the content other than the ranges specified in the table may be changed.

Table 5-2: The content of the 'tabMain' worksheet of the 'dataIni.xls' utility

Range*	Description
G6:N107	Female population by age and state**
G109:N210	Male population by age and state**
G212:N313	Age-specific death rates (mx) by state, females
G315:N416	Age-specific death rates (mx) by state, males
G418:N438	Age-specific transition rates from the illiterate to the literate status by state, females***
G440: N460	Age-specific transition rates from the illiterate to the literate status by state, males***
G462:N499	Age-specific fertility rates (ASFR) by state
Column D	Year to which the initial data apply

* In each range name below, the row numbers may vary according to the number of age groups of the data.

** PEDA assumes that everybody is born in the illiterate status. Some of the states (1,2, 5 and 6) will thus not contain any person in the first age groups.

*** Although the user can set an age specific pattern for the transition from illiterate to the literate status (i.e. the proportion of the cohort that is expected to become literate at that particular age), the assumption can also be made that everybody that is expected to become literate will do so at the same age. As long as this is happening at a young age, this will not influence the projection results much. For the literate states (1,2,5 and 6) no transition rates have to be set. These columns remain empty.

The 'tabLorenz' worksheet contains the food distribution functions for rural and urban areas separately with a precision of one-percentage point intervals. In this worksheet only the values of the cell range D6:E106 may be changed. Note that the format in which the food distribution function needs to be entered in PEDA is not the one in which the Lorenz Curve is most often presented (see section 5.2). Column C stands for the cumulative proportion of the population, ordered from most to least privileged in terms of their access to food. In columns D and E, the cumulative share of the consumed food has to be specified for rural and urban areas respectively. For example, for value 0.1 in column C, one would thus read in columns D and E the share of the food that is consumed by the most privileged 10 percent of the population in rural and urban areas.

After all the initial data have been entered in both the 'tabLorenz' and 'tabMain' worksheets, one can return to the 'Control' worksheet and click on the 'Append to Tab' button to initialize PEDA for that country and add the data to the existing databases. The PEDA application must be closed while doing so.

The next time the PEDA application is opened, the newly initialized country will appear in the list of countries under the simulation button. However, the initialization is not yet complete at this stage. A baseline scenario still needs to be defined and a number of other parameters of the model need to be adjusted. To start implementing these last changes, the country has to be selected from the drop down list under the simulation button, followed by clicking on the simulation button itself. In Table 5-3, the additional adjustments are summarized that are the minimum requirement for completing the initialization and defining a baseline scenario. In this example the baseline scenario is a constant rates scenario, i.e. a scenario wherein all parameter and variable values are assumed to remain constant over the projection period. Note that these are the absolute minimum requirements. Preferably, all variables and parameters should be evaluated and their values revised if necessary.

Table 5-3: Minimum requirements to complete the initialization and to define a (constant rates) baseline scenario

Location	Variable or parameter name *	Action
General settings and population parameters, General tab	Initial year	By default it has value 1995. It has to be changed to the year to which the initial data apply
	Net food production in the initial year (in daily per capita Kcal)	This value has to be looked up in the literature and entered here. Note that it is the net production, thus after deduction of the post-harvest losses.
	Net food imports in the initial year (in daily per capita Kcal)	This value has to be looked up in the literature and entered here. Note that this is the net trade, i.e. imports minus exports expressed in daily per capita calories.
	Assumed minimum consumption of Kcal per capita to be food secure	As the minimum food requirements change under different national conditions, this value has to be looked up in the literature.
General settings and population parameters, Sub-populations tab (for all the necessary states)	TFR	The values of the TFR for the initial year, estimated in section 5.5 have to be entered here as well as the year to which the initial data apply. For a constant rates scenario, one should keep the initial level of the TFR constant for the subsequent data points.
	Life expectancy (e0)	The e0 values by sex for the initial year, estimated in section 5.4 have to be entered as well as the year to which the initial data apply. For a constant rates scenario, one should keep the value of e0 constant for the subsequent data points.
	Education	The educational transition rates need to be specified for the illiterate states only. For a constant rates scenario, this means the same values as those specified in the 'tabMain' of the 'dataIni.xls' utility. Other, higher values may be specified if there are reasons to assume improvements in the access to education in the medium term.
Dynamic parameter settings	AIDS morbidity	By default, the value for this scenario variable is 0, though the user can incorporate the impact of HIV/AIDS into the baseline scenario through defining the proportion of adults being sick due to an HIV/AIDS related illness. See sections 2.7 and 3.1.3 for more details on setting consistent scenarios that incorporate HIV/AIDS.
	Water	The value of the water scenario variable should be defined in such a way that it reflects the situation of the initial year with respect to the position on the water saturation curve.
	Loss in transport/storage	For the initial year, one has to give an estimate of the proportion of the production that is lost during the harvest, storage, transport and processing of food as well as the share of the production that is used as seeds for the next cropping season. This figure is used to determine the gross production in the initial year. That gross production is eventually used as a basis for the calculation of the agricultural production in the following years of the projection period.

Location	Variable or parameter name *	Action
Excel spreadsheets, 'NDScen' worksheet (See section 4.3 for instructions to go beyond the user interface)	Year range of the projection period for the dynamic parameter settings	In cell E6, the year to which the initial data apply has to be specified. The range F6:DE6, is reserved for the following years of the projection period and these need to be adjusted accordingly
Excel spreadsheets 'Params1' worksheet	Water saturation curve parameter 1,2	In cells G18 and G19, two water saturation curve parameters should be defined (see section 2.4 and section 4.3.2).

* See section 3.1 for more details on the meaning of these parameters and variables.

Once the values of these parameters have been changed, the -constant rates- baseline scenario can be saved from within the user interface. Additional changes to scenario variables or model parameters can be easily implemented at a later stage and saved under a different scenario name. The instructions listed above only apply to the absolute minimum requirements for the initialization of PEDA for a new country.

6 References

Agarwal, B. (1994); The gender and environment debate: Lessons from India, In L. Arizpe, M.P. Stone, and D.C. Major, eds., Population and the Environment: Rethinking the Debate. Boulder, CO: Westview Press, p. 97-124.

Bilsborrow, R.E. and H.W.O. Okoth Ogendo (1992); Population-driven changes in land use in developing countries, Ambio, Vol.21, No.1, p. 37-45**

Cleaver Kevin M. (1994); Direction in Development: Reversing the Spiral, The Population, Agriculture, and Environment Nexus in Sub-Saharan Africa, Washington D.C., World Bank.

CSA (1999); The 1994 Population and Housing Census of Ethiopia. Results at the Country Level: Volume II, Analytical Report. Addis Ababa, Central Statistical Authority, Office of Population and Housing Census Commission, 375 p.

Coale, A.J. and Trussell, T.J. (1974); Model fertility schedules: variations in the age structure at childbearing in human populations. *Population Index*, 40, 185-258. Erratum: ibid., 41, (1975), 577.

Dasgupta, P.S. (1993); An Inquiry into Well-Being and Destitution. Oxford, U.K.: Oxford University Press.

Economic Commission for Africa (ECA), 1999. Economic Report on Africa 1999: The Challenge of Poverty Reduction and Sustainability, Addis Ababa, UNECA.

Economic Commission for Africa (ECA), 2001. Economic Report on Africa 2000: Transforming Africa's Economies: Overview, Addis Ababa, UNECA.

Hayami, Y. and V. Ruttan (1971); Agricultural Development: An International Perspective. Baltimore, MD: Johns Hopkins University Press.

Keyfitz, Nathan (1979); Multidimensionality in Population Analysis, in: "Sociological Methodology 1980", Karl F. Schuessler (ed.), San Francisco.

Lutz, W., Ed. (1994). Population-Development-Environment: Understanding their Interactions in Mauritius. Berlin: Springer Verlag.

Lutz, W., Goujon, A. and Doblhammer-Reiter, G. (1999); Demographic dimensions in forecasting: Adding education to age and sex. Pages 42-58 in W. Lutz, J.W. Vaupel, and D.A. Ahlburg, eds., Frontiers of Population Forecasting. A Supplement to Vol. 24, 1998, Population and Development Review. New York, NY: Population Council.

Lutz, W. and Scherbov, S. (2000); *Quantifying vicious circle dynamics: The PEDA model for population environment development and agriculture in African countries,* in E.Dockner, R.Hartl, M.Luptacik and G.Sorger eds. Optimization, Dynamics and Economic Analysis, Springer/Physica,

Nerlove, M. (1991); Population and the environment: A parable of firewood and other tales. American Journal of Agricultural Economics 73:1334-1347.

O'Neill, B.C., F.L. MacKellar, and W. Lutz. (2000); Population and Climate Change. Cambridge, U.K.: Cambridge University Press.

Rogers, A. (1975); Introduction to Multiregional Mathematical Demography, New York: John Wiley.

Rogers, A. and Castro, L (1981); Model Migration Schedules, Vienna, International Institute for Applied Systems Analysis, RR-81-30

Sanderson, Warren (forthcoming); Modeling AIDS in Botswana, IIASA Report 2001

Scherbov, S. and Golubkov, V. (1986); Derivative Free Gauss-Newton-like Algorithm for Parameter Estimation. Laxenburg, Austria, International institute for Applied Systems Analysis, IIASA Working Papers WP-86-63

Scherbov, S. and Grechucha, V. (1988); *"DIAL" - A System for Modeling Multidimensional Demographic Processes*, WP-88-36, Laxenburg, Austria: International Institute for Applied Systems Analysis.

Sen, A. (1994); Women, poverty, and population: Issues for the concerned environmentalist. In L. Arizpe, M.P. Stone, and D.C. Major, eds., Population and the Environment: Rethinking the Debate. Boulder, CO: Westview Press, p. 67-86

UN-DESA (1988); Mortpak-Lite : the United Nations Software Package for Mortality Measurement, Population Studies n° 104 (ST/ESA/SER.A/104), New York, United Nations, Department of Economic and Social Affairs, 1988, 121p.

Annex I: Deleting scenarios and projection results

In the current version, the PEDA model does not foresee an automated deletion of scenarios and projection results from the databases. If the user wants to delete any of these two, it must be done directly from the Microsoft Access files using standard database techniques.

Deleting scenarios, and especially the results of simulations may be desirable to regain disc space. Storing the outputs for one simulation may easily expand the size of the database by 1MB.

The user has the option to delete only the simulation results and keep the scenario definitions, or, to delete both. To delete the simulation results, one has to open the 'projection.mbd' database (see section 4.2.1 and Figure 4-8). In this database there are two tables that need to be manipulated: *'Scenario'* and *'Table1'*. In both tables, the scenario(s) to be erased must be selected and deleted (using a query or any other database technique). This procedure erases the output of the simulations for the specified scenarios but keeps the scenario definitions intact. This means that, if necessary, new projections could be carried out again for the same scenario(s).

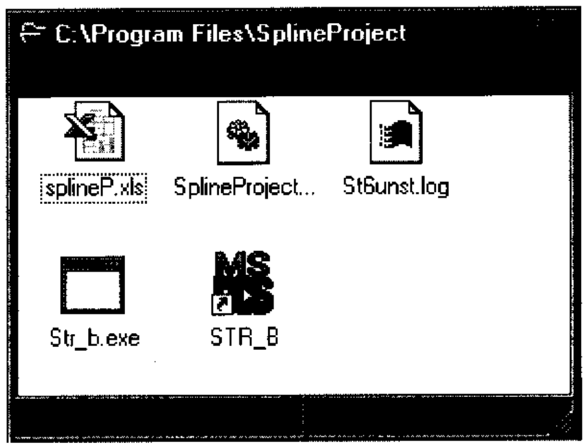
To delete the scenario definitions as well, they have to be deleted in an analogous way from the scenario database ('scen.mdb', see Figure 4-5, p.52). This database consists of four tables, and the records for the redundant scenario(s) need to be filtered out and deleted for each of these tables.

Annex 2: Using Spline for interpolating the population age structure

The PEDA CD-ROM comes with a Spline utility developed by S. Scherbov that can be used to interpolate the population age structure as well as to smooth numerous other curves. A lot of literature on Splines is available both in mathematical textbooks and on the Internet. Here, we will illustrate the use of this particular utility to smooth the population age structure.

The Spline utility comes as a separate installation package on the CD-ROM that has to be installed on the computer following standard Windows procedures.

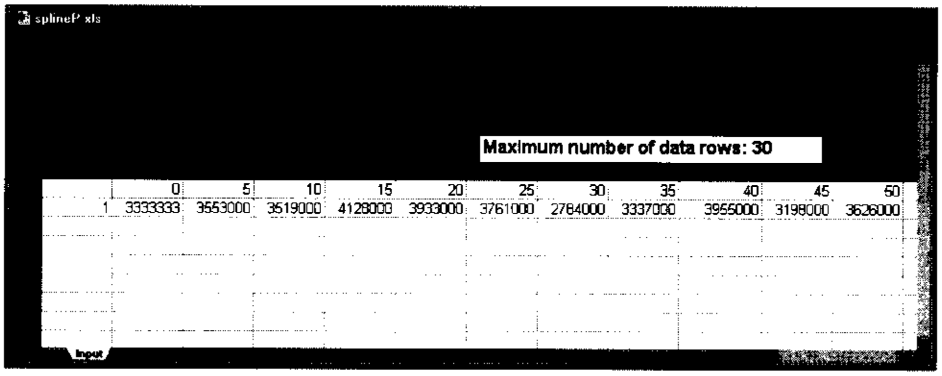
Figure 6-1: The SplineProject program folder



The installation package does not automatically add the Spline utility to the Start menu. It has to be opened from the programs folder. The folder is called SplineProject and its content is illustrated in Figure 6-1. Double click on the 'splineP.xls' file to start the utility.

The Spline utility is an Excel file containing macros. To use it, you need to enable the use of macros when opening the file. The file contains two worksheets: 'input' and 'output' (see Figure 6-2).

Figure 6-2: the SplineProject utility input worksheet

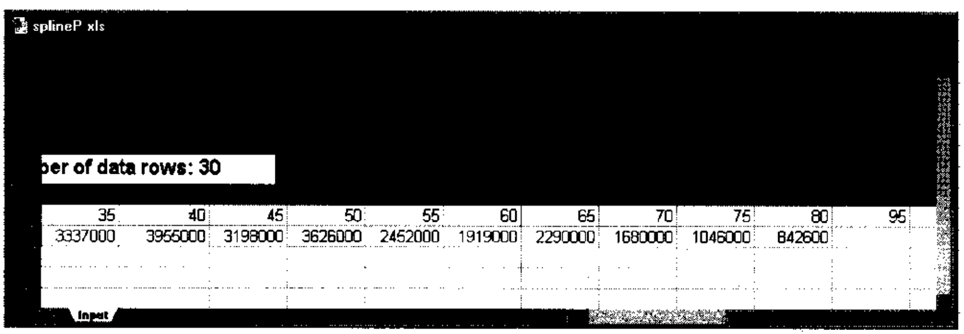


In row 7, the age groups are specified, and in row 8 the population effective for the five-year age group intervals. By clicking on the 'Graduate' button, the population effectives are graduated over the interval for which the upper limit is specified in row 7 of the following column. For the first effective (cell B8 in Figure 6-2), this means that the total population (3,333,333) is distributed over the ages 1 through 5. Similarly the effective of cell C8 is distributed over the ages 6 through 10. This logic also implies that we have to specify an upper limit for the interpolation of the last effective.

Note that this does not necessarily have to be the same (age) interval as the ones used in the previous groups. In the example in Figure 6-4, 95 is specified as the upper age limit, meaning that the effective from cell R8 will be graduated over the ages 81-95. Note that this is an interesting feature to extend the population distribution to the desired age because population distributions are often closed with an open category (e.g. 80+) and that is not always the desired age group for projection purposes. However, this extension of the age distribution is not unlimited. It is only possible to the extent that Spline returns positive values as output.

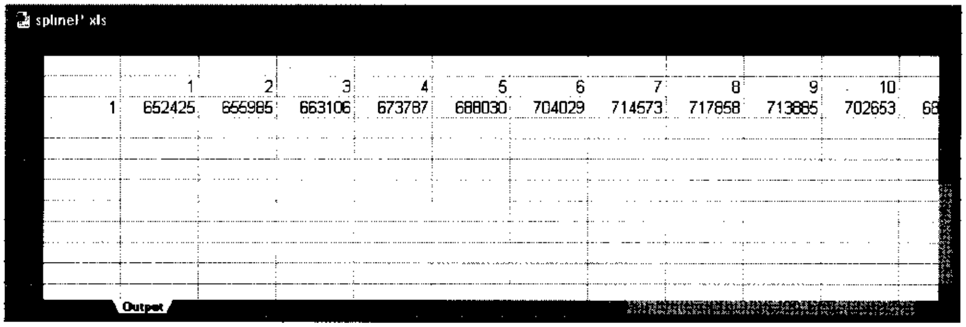
Note that this Spline utility can manage the graduation of up to 30 distributions simultaneously. In Figure 6-2 only one distribution is given as an input. More distributions could have been listed from row 9 through 37. Column A is reserved for indexing the different distributions.

Figure 6-3: Specifying the input in the Spline utility



The output is presented by single years of age in the 'Output' worksheet. This is illustrated in Figure 6-4.

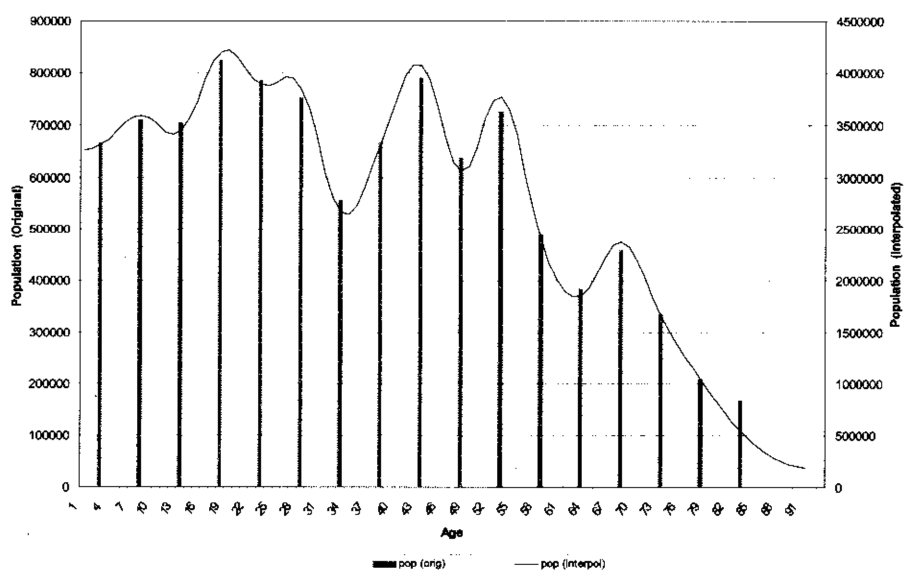
Figure 6-4: The output of the Spline utility



	1	2	3	4	5	6	7	8	9	10	
1	652425	655985	663106	673787	688030	704029	714573	717858	713885	702653	68

A comparison of the original and interpolated data points is given in Figure 6-5.

Figure 6-5: Spline interpolation of a (fictive) population age distribution



Note that this Spline utility can also interpolate other than population distributions, although one limitation is that it works with large numbers (since it was developed for interpolating population distributions). One way around this, and to use it for smaller numbers is to multiply the original data points first by a big number and divide the output again by the same number.