A DEMOGRAPHIC SIMULATION MODEL FOR HEALTH CARE, EDUCATIONAL, AND URBAN SYSTEMS PLANNING

Philipp F. Schweizer
Westinghouse Research Laboratories
Pittsburgh, Pennsylvania 15235

Abstract

This paper describes the development, application, and digital computer simulation of a demographic model suitable for long term planning. The simulation model is based on the "cohort survival methodology" and projects population characteristics (population numbers for each region, age group, sex, year and racial or income group) for a planning period less than or equal to twenty years.

Demonstration of how the simulation model is applied to problems in Health Care, Educational and Urban Systems Planning are presented.

1. Introduction

This paper discusses the development and application of a demographic model for use in long term planning.

The need for demographic information became apparent from previous work concerned with the development of planning tools for the Westinghouse Health Systems Department and the Westinghouse Learning Corporation. Population forecasts were necessary to determine the future demand on health care facilities and to predict enrollment in planning educational facilities.

An initial investigation was conducted to determine the demographic information available from local planning groups or the U.S. Census Bureau and whether or not this information was sufficient for the intended planning purposes.

An examination of the information available from the U.S. Census Bureau uncovered the following difficulties: 1) Population forecasts for the U.S. and most states were available, however, forecasts for counties and local areas were rare. When and where these forecasts for local area existed, quite often, they were merely ratios of aggregate projections for the state or county. The forecasts did not account for local influences on the population. 2) The population projections are commonly given for 10 or 15 year periods. However, many planners must make decisions on a yearly or even monthly basis. In an area where rapid change prevails interpolation of the projections may be difficult. 3) In most cases a single population forecast was given which did not yield a sensitivity of the population to various factors such as fertility rates, mortality rates, migration, employment, housing development, etc. This sensitivity information is a valuable asset to the planner and could be provided with a family of forecasts, however, these again when provided were usually with respect to a single factor. Recognizing that the factors of interest to various planners would be different, a scheme was needed for producing forecasts which were a function of local influencing factors.

In studying information from some local planning groups, it was apparent that the quantity and quality of available information was highly variable and depended on the size and sophistication of the specific planning group. Since the program goal was to provide a planning tool that would be applicable to any local area or state it was assumed to be too risky to rely on a local planning group for sufficient information.

The conclusion drawn from the above was that although the Census Bureau could provide regional population forecasts and local planning groups could provide many of the "necessary bits and pieces" of the demographic picture, this information alone would not be sufficient for local and even state planning needs. A demographic model would be necessary to manipulate this data into a more usable form.

The next stage of the investigation involved the resolution of whether a demographic model should be developed or whether an existing model could be used.

Models operating on past history were available (Ref.5). Significant work in the area of statistics has allowed planners using regression techniques to build models based on past history alone. The projections from this type of model have provided valuable information for the short term. Caution certainly must be exercised in applying these models for long term planning. In many cases even though the short term projections are accurate, the long term projections are misleading. A better approach appeared to be a technique which made use of past history but also weighed the planner's subjective judgments about the future (i.e., housing development, land use, future employment).

Investigating various available models (Refs. 4,5,6) showed that some were aimed at very general studies (i.e., population forecasts for U.S., the world, India, China, etc.) while others at very specific applications. None of those uncovered seemed directly appropriate to the problem of forecasting population for a rapidly developing suburban area or new town which was one of the primary intentions of this work. It was concluded that a model should be developed since modification of existing models would require as much or more effort.

The approach finally taken was one based on the "cohort survival method" which has been adopted for use by the U.S. Census Bureau. In general, the "cohort survival method" begins with the detailed distribution of a population obtained in a base year (most likely a census year), and moves that population through time applying to it various population changing factors, according to a set of assumptions about those factors. A model was constructed using this methodology but with modifications for including local population influencing factors.

Refs. 1 through 4 provide background information for the decisions made and the models developed in this study. Although none of the information from these references were explicitly used, considerable insight for the modeling problem was obtained from them.

The remaining sections of this paper present the mathematical model (Mathematical Model Description), discuss the computer code used in implementing the model (Computer Code), and demonstrate applications (Applications) of the simulation model.

2. Mathematical Model Description

The model classifies the total population according to five factors: 1) geographical location (region), 2) age groups, 3) year of existence, 4) race or income group, and 5) sex. A variable (or state) is assigned to represent the number of people possessing any possible combination of the above five factors.

The total population is divided into age groups or cohorts. Diagrams depicting the four basic considerations, aging, mortality, fertility and migration, that are modeled for each cohort are shown in Figures 1 and 2.

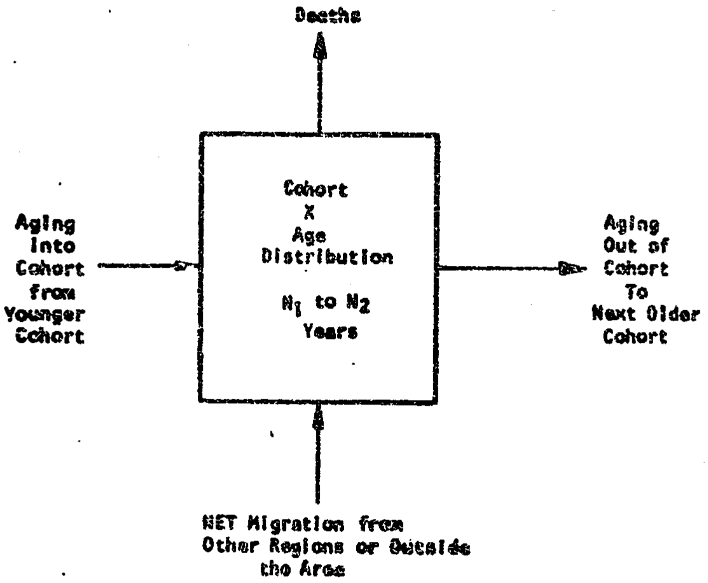


Fig. 1—Single Cohort (age group) model

The dynamics are incorporated in the model by changing the number of people that belong to the cohort each planning period (usually each year). A certain number of people are removed from the cohort to represent those that have aged to the next older cohort, those that have died during the planning period and those that have migrated from the area. Numbers of people are added to the cohort to represent those that are aging from a younger cohort and those that are migrating into the area.

The above modeling procedure may be mathematically expressed for each cohort by the following expression.

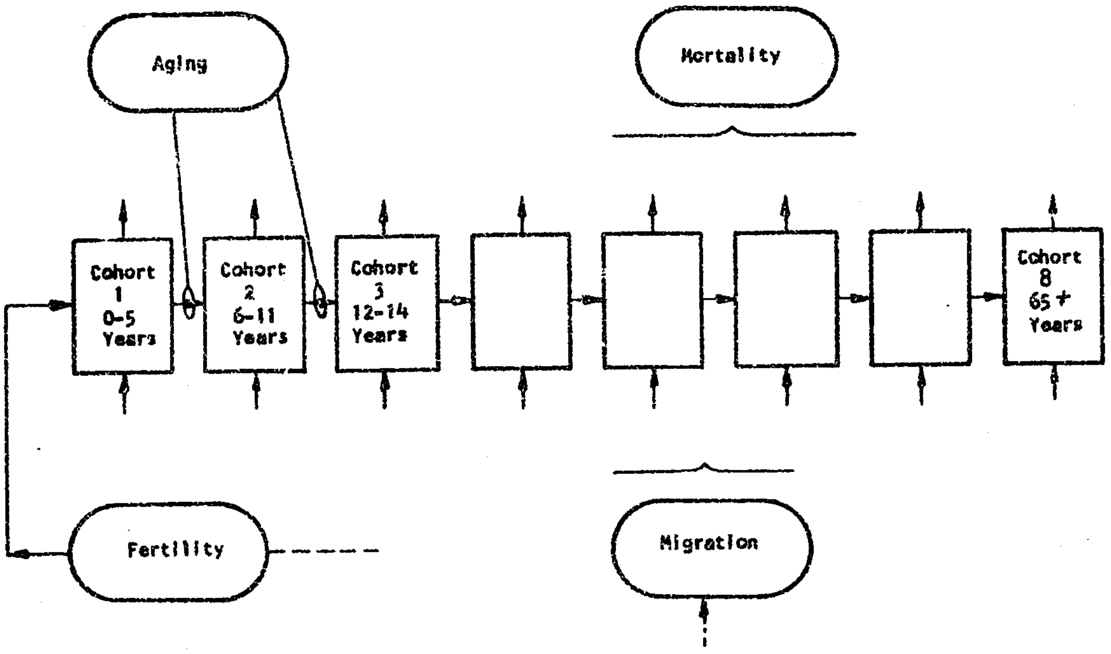


Fig. 2—Functional diagram of Demographic Model depicting the four primary considerations

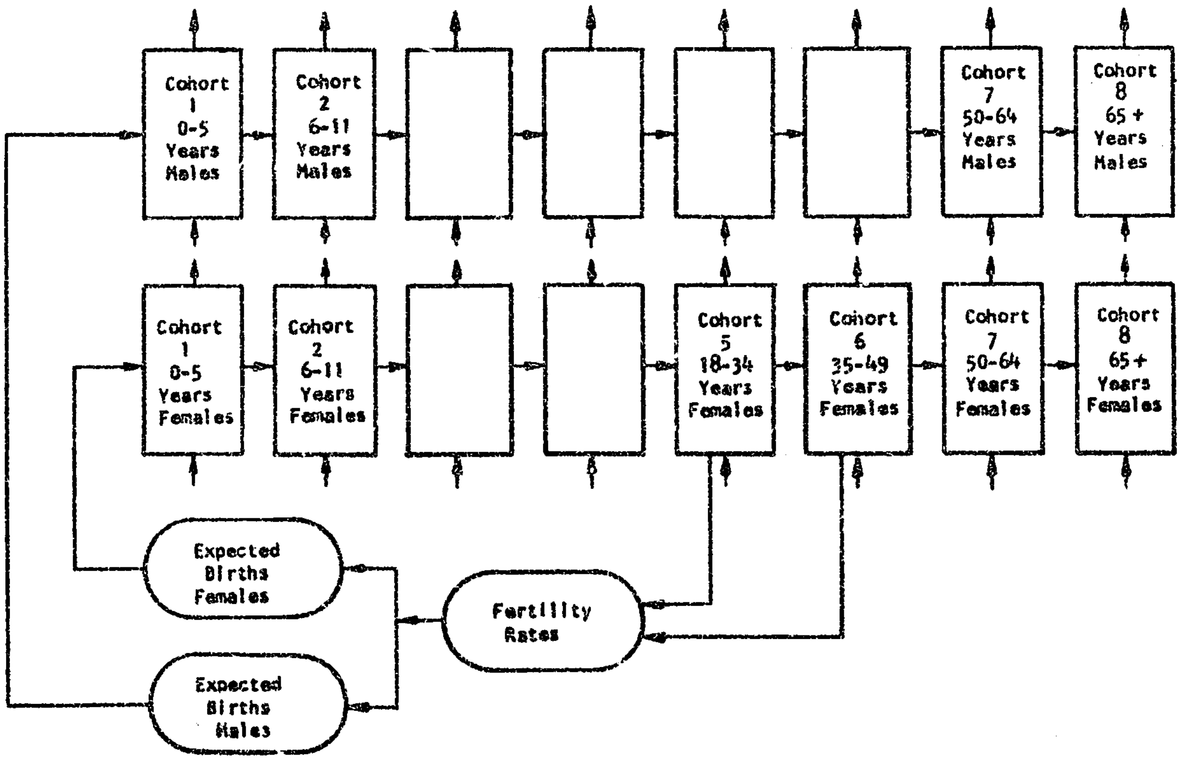


Fig. 3—Functional diagram depicting the feedback effect of female population through expected births

$$x^{m,n}_{i,j,k+1} = (1-A^1_{j,k})\, x^{m,n}_{i,j,k} + A^2_{j,k}\, x_{i,j-1,k} + y^{m,n}_{i,j,k} + F^m \sum_{j=JCB1}^{JCB2} \alpha_j\, x^{m,1}_{i,j,k} + M^{m,n} D^n_j \sum_{\substack{i,j,m\\n}} x^{m,n}_{i,j,k} + H_{j,p}\, v_{i,p,k} \qquad (1)$$

where

$x^{m,n}_{i,j,k}$ represents the number of people in a particular population cohort

with subscripts

i designating the demographic region

j designating the age grouping of the cohort

k designating the year.

and with superscripts

m designating the race or income group

n designating the sex.

Each term in Equation (1) will subsequently be discussed with regard to its contribution to the total expression.

The first term, $(1-A^1_{j,k})x^{m,n}_{i,j,k}$ represents the difference between the cohort population in the k^th year and the number of people who will leave this cohort during the k^{th} year because of aging.

The aging parameter $A^1_{j,k}$ is computed from

$$A^1_{j,k} = CD_{j,k}/CS_{j,k} \qquad (2)$$

where

$CS_{j,k}$ is the span of cohort j in the k^{th} year

$CD_{j,k}$ is determined from the age distribution of cohort j in the k^{th} year.

The second term, $A^2_{j,k}x^{m,n}_{i,j-1,k}$ represents the number of people who will enter the cohort during the k^{th} year because of aging. The aging parameter $A^2_{j,k}$ is computed from

$$A^2_{j,k} = CD_{j-1,k}/CS_{j-1,k} \qquad (3)$$

where terms in Equation (3) are as defined in Equation (2).

The third term $y^{m,n}_{i,j,k}$ represents the migration into or out of the i^{th} region for the j^{th} cohort and k^{th} year.

The fourth term, $F^m \sum_{j=JCB1}^{JCB2} \alpha_j\, x^{m,1}_{i,j,k}$ models the expected births during the k^{th} year. The parameters used in this term are defined by

F^m	the fertility rates for each race, m
JCB1	the first cohort with women of child-bearing age (15-44 years)
JCB2	the last cohort with women of child-bearing age
α_j	fraction of the cohort of child-bearing age.
$x^{m,1}_{i,j,k}$	female population in cohort j, region i, year k, and race m.

Figure 3 depicts the effect of this fertility term on the total population. Females in the child-bearing cohorts are multiplied by fertility rates to determine expected births for males and females.

The fifth term, $M^{m,n}D^n_j \sum_{\substack{i,j\\m,n}} x^{m,n}_{i,j,k}$ models the expected deaths during the k^{th} year (Fig. 4). The parameters of this term are defined by

$M^{m,n}$	the mortality rates for each race and sex
D^n_j	the mortality distribution (the portion of total deaths that are in the j^{th} cohort with sex n)

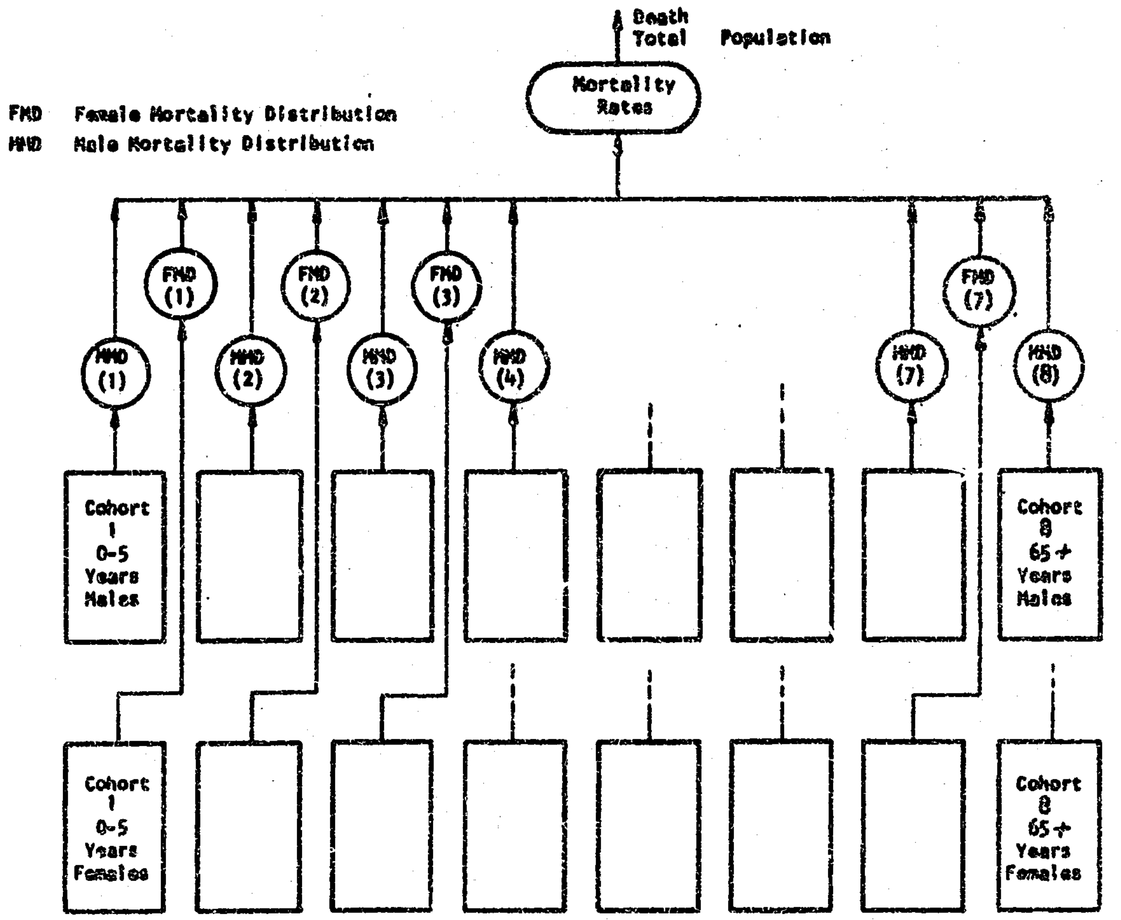


Fig. 4—Functional diagram showing the interaction of mortality rates and distribution with each cohort

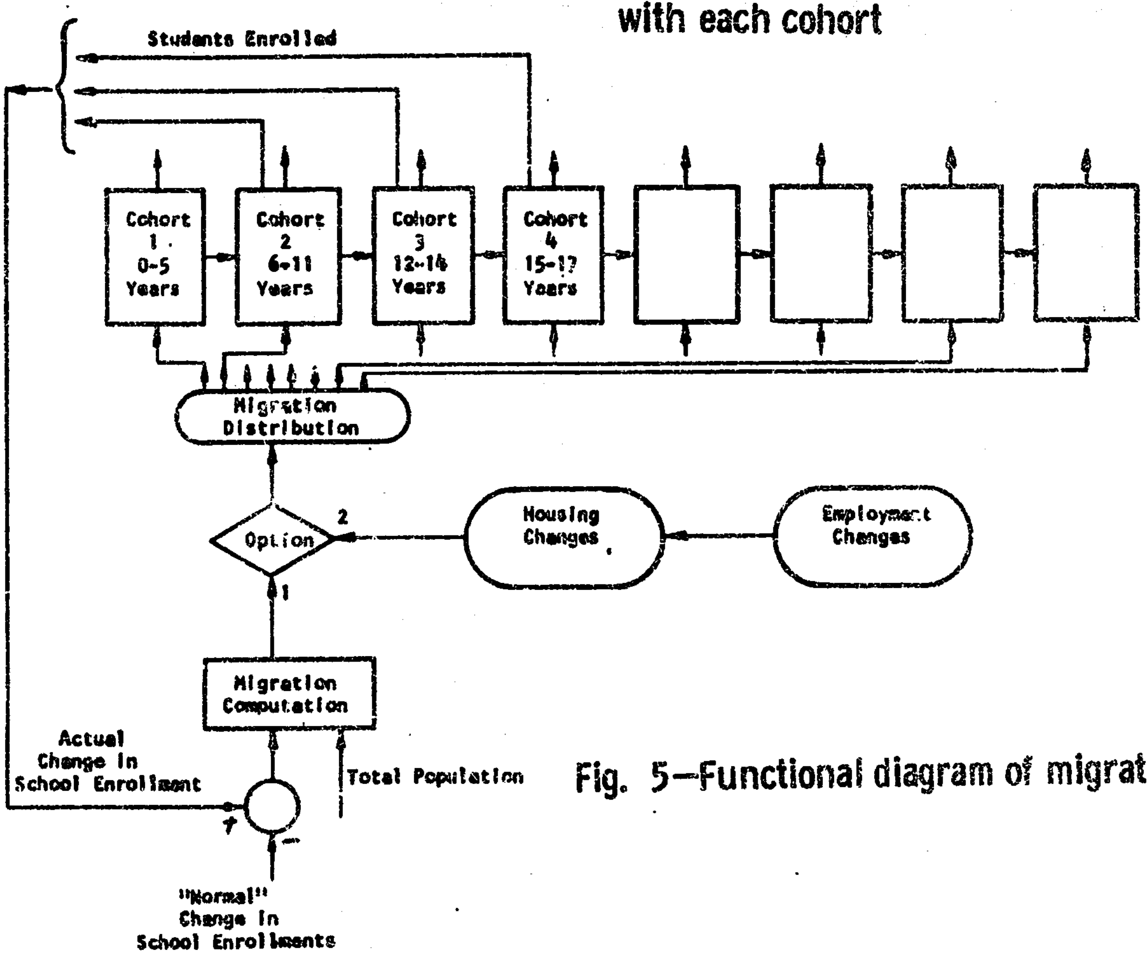


Fig. 5—Functional diagram of migration model

$\sum_{\substack{i,j\\m,n}} x^{m,n}_{i,j,k}$ represents the total population (the summation over all regions, cohorts, races, and sex).

The last term in Equation (1) models the effect of the population because of changes in types and numbers of dwelling units. The coefficients $H_{j,p}$ represent the average number of people belonging to cohort j and living in a dwelling unit of type p. The variable $v_{i,p,k}$ represents the number of p type dwelling units in region i in year k. These dwelling units could be considered as premium, choice, and economy houses; luxury, convenience and low income apartments; townhouses and mobile homes. The distinction between premium, choice and economy housing is made by considering the lot size and living space area. The distinction between luxury, convenience, and low income apartments is based on living space area. These classifications of dwelling units are actually arbitrary and may be changed by the planner if desired.

The scholastics defined as those students enrolled in kindergarten through grade twelve, are determined from the general population by the following expression

$$S_{i,r,k} = ED_{i,r,k}\, x_{i,j,k} \qquad (4)$$

where

$S_{i,r,k}$ represents the scholastics in region i, grade r, and year k.

$ED_{i,r,k}$ is the fraction of the j^{th} cohort in grade r for region i and year k.

The migration term in Equation (1) may be known from other considerations or may be computed from the following expression

$$y^{m,n}_{i,j,k} = SECA^{m}_{i}\,(AFS/ANSAC)(x^{m,n}_{i,j,k}/\sum_{j=1}^{NC} x^{m,n}_{i,j,k} \qquad (5)$$

with

$$SECA^{m}_{i} = (\sum_{k=1}^{NYEH} SEC^{m}_{i,k})/NYEH$$

where

$SECA^{m}_{i}$ = average school enrollment change over the past for region i, and race m

$SEC^{m}_{i,k}$ = actual school enrollment change for region i, year k, and race m

NYEH = number of years of school enrollment history

AFS = average family size

ANSAC = average number of school age children per family

NC = number of cohorts

The reasoning behind Equation (5) is as follows. First, the number of immigrating new school enrollments is estimated based on past history, SECA. This number is then divided by ANSAC to determine immigrating families and subsequently multiplied by AFS to determine the immigrating population. To allocate this population to appropriate cohorts, the ratio in Equation (5) is used.

The model for migration (Fig. 5) allows the planner to use one of two options (term 3 or 6 in Equation (1)). Option 1 uses Equation (5) and computes migration from a knowledge of past and present school enrollment changes by race, the average family size and average number of school age children. Option 2 uses the sixth term in Equation (1) to compute migration based on the type and number of dwelling units being constructed or removed from the area.

In cases where an exact number of future dwelling units are unknown the units for the planning period may be dynamically represented by

$$v_{i,j,k+1} = v_{i,j,k}\, C_i \qquad (6)$$

where

C_i is an estimated annual rate of change of dwelling units for region i.

When estimates of future employment are available, the rate of change, C_i is determined from

$$C_i = KE_i \qquad (7)$$

where

K is a proportionality constant

E_i represents the annual rate of change in basic employment in region i.

In making the population projections, the maximum population or saturation condition for each region must constrain the population numbers for each region. This constraints is enforced by the following expression.

$$0 \le x^{m,n}_{i,j,k} \le x^{m,n}_{SATi,j} \qquad (8)$$

where

$x^{m,n}_{SATi,j}$ is the maximum expected popuatlion in the j^{th} cohort region i for race m and sex n.

This saturation population for each region is computed from an assumed set of characteristics for the neighborhood. These characteristics include type and number of dwelling units, and the average number of people per type dwelling unit. The expression used in the computation is given by

$$x^{m,n}_{SATi,j} = (ANP_p)\,(PR_i)\,(TA_i)/ALS_p \qquad (9)$$

where

ANP_p represents the average number of people living in a p type dwelling unit for cohort j

PR_i represents the percentage of region i that will be devoted to residential development

TA_i is the total area in region i

ALS_p is the average area occupied by a p type dwelling unit.

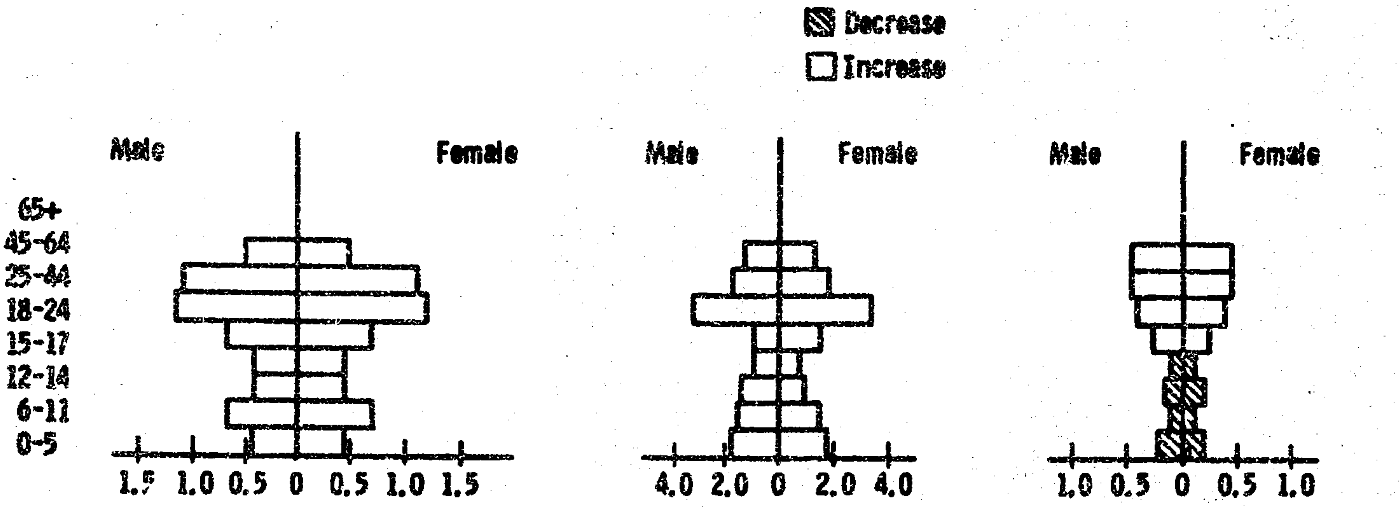
Equations (1) through (9) are used to make the population charactieristics projections for the planning period.

3. Computer Code

The model equations presented in the preceding section have been implemented in a computer code which is written in Fortran V and is operational on the Univac 1106 computer at the Westinghouse Research Laboratories. A version of the code in Fortran IV is also available at other Westinghouse locations and on a time sharing basis. A description of this code will not be presented here because of space limitations and because the primary intent of this paper is to demonstrate its application to long range planning.

4. Applications

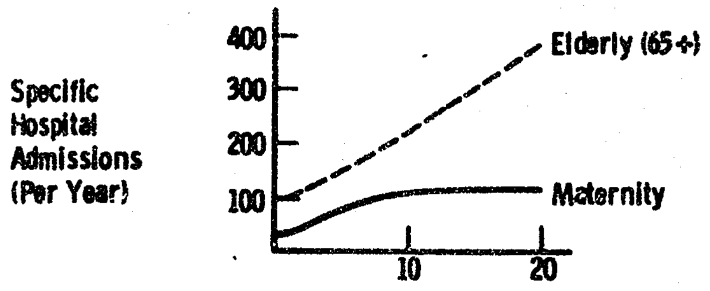
The purpose of this section of the paper is to briefly demonstrate how this simulation model has been applied so that the reader might perceive of ways of applying the methodology



(a) Base Year Population By Age, and Sex (1000's)

(b) Population Difference From Base Year by Age and Sex (1000's)

(c) Population Difference From 10 Year Period by Age and Sex (1000's)



(d) Planning Period (Years)

Fig. 6—Health care planning example

presented here to his particular problems. The examples demonstrated (Figs. 6 through 10) have been simplified to avoid the details that were obtained in thorough studies of these problems.

Applications are discussed in three areas: 1) Health Systems Planning; 2) Educational Planning; and 3) Urban Planning.

4.1 Health Systems Planning

The population characteristics projected for the planning period when used with the known incidence of disease for race, sex or income group provide a future demand forecast for health care facilities. This information is then used in determining the size and location of new facilities or the modification of existing facilities.

Consider the problem of forecasting hospital admissions for elderly and maternity care for a 20-year planning period as shown in Fig. 6. It is assumed that this hospital is servicing a rapidly growing suburban area in which developers are building 400 or 500 units per year over the first five to eight years of the planning period. From the eighth to the twentieth year moderate or little developer activity is assumed and the area grows at its natural rate (based on assumed fertility and mortality rates).

Results from the demographic model simulation are shown in Fig. 6. In (1) a graph depicting the population composition in the base year is shown. In (b) incremental changes from the base year are shown and (c) incremental changes from the 10-year projection are shown. Hospital admissions for both elderly (65+, 100 per 1000 pop. per year) and maternity (assuming 75 births/1000 females, ages 15-44) are plotted versus the planning period.

At least two areas of significance appear from an examination of the results. First, the planner should be cautious in overstaffing or building for maternity admissions early in the planning period and secondly, one must not delay too long in planning for elderly care to avoid a crisis situation late in the planning period.

This demonstrates just one of the many population related problems in health care planning that might be examined through simulation.

4.2 Educational System Planning

The projection of scholastics for the planning period is of direct value to educational planners in determining future enrollments which dictate the location, size, staff and material requirements for educational facilities. The projections may also be used indirectly to determine the financial resources or size of bond issue necessary for future facility construction and operation. In addition they can provide information which shows such situations as a peak in grade or middle school enrollments followed by a sharp decline. Situations like this may favor portable modules for schools rather than permanent construction.

Consider the problem of forecasting educational costs and resulting school tax burdens in a rapidly growing suburban area as defined in the previous section. Results obtained directly from a demographic simulation for this problem are shown in Figure 7. In (a), (b), and (c), the

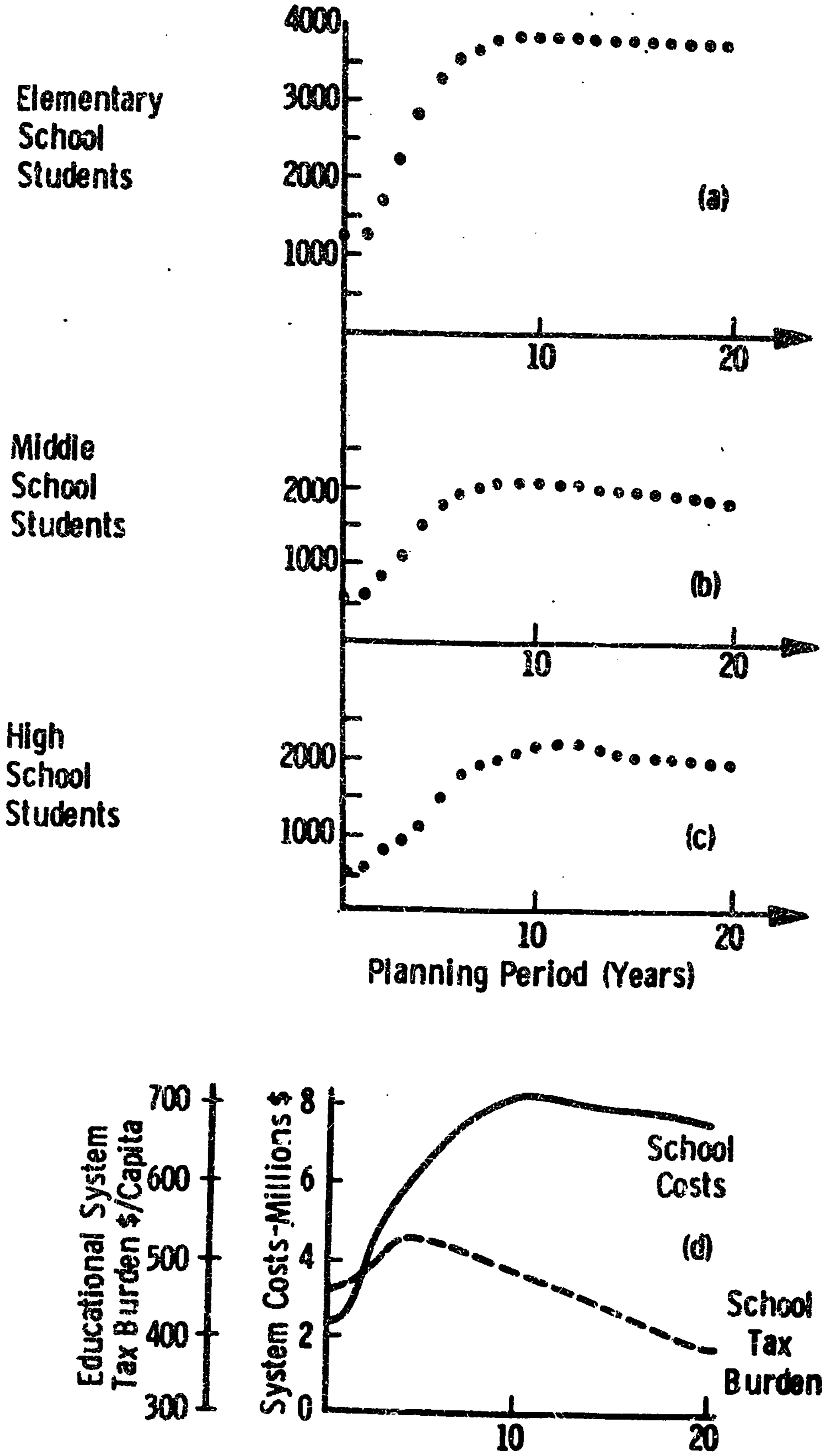


Fig. 7—Educational planning example

elementary, middle and high school students, as obtained from the demographic code, are plotted versus the planning period. In (d) educational system costs (assuming $1000 per year per student) and resulting tax burden are plotted for the planning period. The tax burden is approximated by dividing the educational costs by the sum of the cohort populations age 25 and above. In an educational system based entirely on real estate taxes this may be slightly inaccurate with regard to exact numbers but the trends as shown in Fig. 9 should still be valid. The inflation factor has also not been explicitly accounted for in the dollar values. It has been assumed that inflation has equal impact on the tax burden and school costs.

Examination of Fig. 7d shows that the tax burden rises during the development period (1-8 year planning period) but once the community has matured the burden actually falls below initial levels. This dynamic change in tax burden demonstrates the importance of a well planned development schedule.

4.3 Urban Planning

This digital computer demographic simulation should be helpful in investigating many urban problems. Some particular applications might include: 1) showing the effects of eliminating one type of dwelling unit (i.e. single family housing) and replacing them with another type (i.e. low income apartments), 2) effects of zoning law changes, 3) population shifting impact of new towns on existing communities, 4) transportation planning, 5) recreational facilities planning, and 6) low cost housing programs.

Some of the results from applying this simulation model to a study of the population and school enrollment for the Alief Independent School District, Harris County, Texas, are shown in Figs. 8, 9 and 10.

For this study the Alief area was divided into five regions as depicted by the simple map shown in Fig. 10. These regions are explicitly defined by the following.

Region 1 - north boundary, Alief Independent School District; west boundary, Barker Reservoir; south boundary, Fort Bend-Harris County line; east boundary, feeder to Katy Highway.

Region 2 - north boundary, Alief Independent School District; west boundary, feeder to Katy Highway; south boundary, Fort Bend-Harris County line; east boundary, Synnot Road.

Region 3 - north boundary, Alief Independent School District; west boundary, Synnot Road; south boundary, Alief Jeanetta Road; east boundary, Alief Independent School District.

Region 4 - north boundary, Alief Jeanetta Road; west boundary, Synnot Road; south boundary, Bissonnet Road; east boundary, Alief Independent School District.

Region 5 - north boundary, Bissonnet Road; west boundary, Synnot Road; south boundary, Fort Bend-Harris County line; east boundary, Alief Independent School District.

Figure 8 shows a high, average and low population projection corresponding to three different sets of input data to the demographic code for a planning period of 12 years beginning in 1968 and continuing through 1980. Some of the necessary input information for computing the saturation population of the area is shown in Figure 9.

Figure 10 shows the distribution of the population which was constructed from the simulation output for two years, 1975 and 1980. This output has been used by Alief planners in estimating future community needs.

6. Conclusions

The development, application and simulation of a demographic model suitable for long range planning has been described.

Applications of the simulation model to problems in health care, educational and urban systems planning have been presented.

7. References

1. U.S. Bureau of the Census, "Census Use Study: General Description," Report No. 1, Washington, D.C., 1970.
2. U.S. Bureau of the Census, "Americans at Mid Decade," Series P-23, No. 16, Jan. 1966.
3. Department of Administration, Bureau of State Planning, "Wisconsin Population Projections," distributed by Document Sales, State Office Building, Madison, Wisconsin, April 1969.
4. Keyfitz, Nathan, Introduction to the Mathematics of Population, Addison-Wiley, Reading, Massachusetts, 1968.
5. Benrud, C. H., "Systematic Procedures for Population Estimates and Projections for North Carolina," Vols. 1 and 2, Research Triangle Institute, N.C., PB 190 224, February 1969.
6. Benrud, C. H., "A Review of Population Estimation and Projection Procedures with Special Reference to Small Areas," Working Paper No. 1, Project SY-388, Research Triangle Institute, December 1968.
7. Lowry, I. S., "A Model of Metropolis," Rand Corporation Memorandum RM-4035-RC, August 1964.

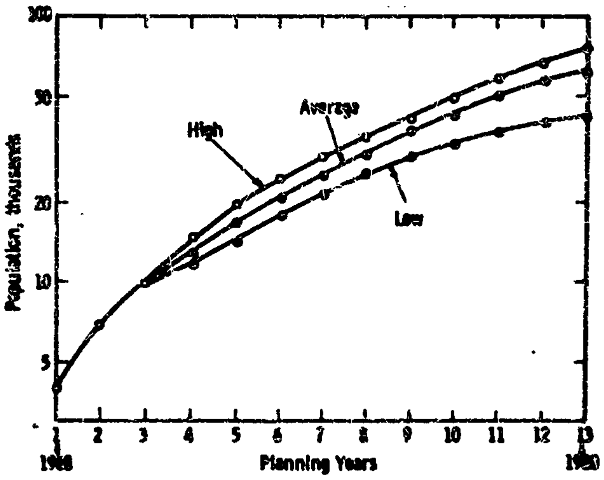
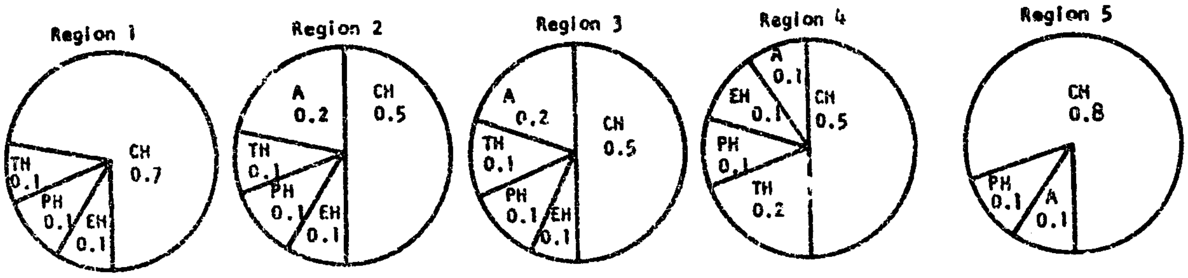


Fig. 8--Alief Area, Harris County, Texas. Population Projection 1968-1980

Demographic Region	Percent of Area Available for Residential Development
1	60
2	50
3	20
4	60
5	60



A- Apartments
TH- Townhouses
PH- Premium Houses
CH- Choice Houses
EH- Economy Houses

Fig. 9--Development input data for demographic modeling code

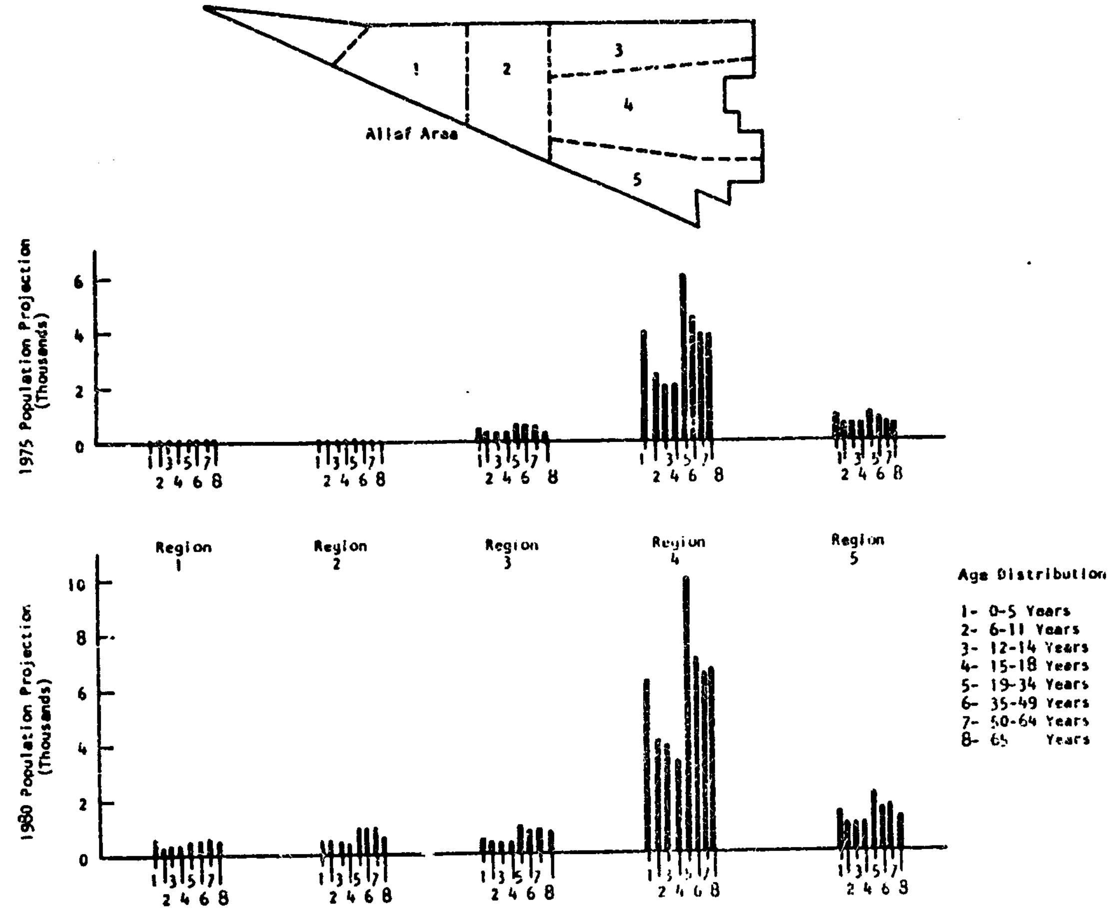


Fig. 10 --Demographic regions & population distribution for allef area, Harris County, Texas