**Mathématiques et sciences humaines**
Mathematics and social sciences

**186 | 2009**
**186, Varia**

# The PICRATE model for fitting the age pattern of first marriage

*Ajustement du schéma par âge d'entrée en première union : le modèle PICRATE*

**Alan P. Matthews, Pauline M. Leclerc et Michel L. Garenne**

**Édition électronique**
URL : http://msh.revues.org/11056
DOI : 10.4000/msh.11056
ISSN : 1950-6821

**Éditeur**
Centre d'analyse et de mathématique sociales de l'EHESS

**Édition imprimée**
Date de publication : 12 octobre 2009
Pagination : 17-28
ISSN : 0987-6936

**Référence électronique**
Alan P. Matthews, Pauline M. Leclerc et Michel L. Garenne, « The PICRATE model for fitting the age pattern of first marriage », *Mathématiques et sciences humaines* [En ligne], 186 | Été 2009, mis en ligne le 15 octobre 2009, consulté le 14 octobre 2016. URL : http://msh.revues.org/11056 ; DOI : 10.4000/msh.11056

Ce document est un fac-similé de l'édition imprimée.

© École des hautes études en sciences sociales

*Math. Sci. hum / Mathematics and Social Sciences* (47e année, n° 186, *2009*(2), p. 17-28)

# THE PICRATE MODEL FOR FITTING THE AGE PATTERN OF FIRST MARRIAGE

Alan P. MATTHEWS[1], Pauline M. LECLERC[2], Michel L. GARENNE[2]

RÉSUMÉ – Ajustement du schéma par âge d'entrée en première union : le modèle PICRATE
*Nous présentons un nouveau modèle qui décrit le schéma par âge d'entrée dans une nouvelle situation. L'introduction du taux de recrutement rend l'utilisation de ce modèle très simple, en particulier pour simuler l'évolution d'une population au cours du temps. Le modèle est défini par trois paramètres : $a_0$, l'âge minimal à l'événement, $p_{max}$, la proportion maximale de la population qui connaîtra l'événement, et $r_{max}$ le taux maximal de recrutement. Ce modèle est testé sur l'entrée en premier mariage, et est comparé au modèle de Coale-McNeil. Nous avons appliqué ce modèle au cas de la Zambie, ainsi qu'à divers pays africains ayant des schémas différents. Le modèle peut s'appliquer à divers processus, tels que les premiers rapports sexuels, la première naissance, le premier emploi, la première migration adulte, etc.*

MOTS-CLÉS – Afrique, Âge au premier mariage, Analyse biographique, Enquête EDS, Modèle de Coale-Mcneil, Modélisation mathématique

SUMMARY – *We present a new model for fitting the age pattern of entry into a new situation. The formulation of the marriage recruitment rate is simpler than in other available models, and particularly useful for computer simulation of the time-evolution of a population. The model has three parameters: $a_0$, the starting age at event, $p_{max}$, the maximum proportion of the population at risk p(a), and $r_{max}$, the maximum value of the recruitment rate r(a). This model was tested on the entry into first union, and was compared with the widely-used Coale-McNeil model. We applied this model to the case of Zambia, and to a variety of African countries with different features. This model could also be used for fitting a variety of processes, such as first sexual intercourse, first birth, first job, first adult migration, etc.*

KEYWORDS – Africa, Age at first marriage, Coale-McNeil model, DHS survey, Event history, Mathematical model

## 1. INTRODUCTION

Processes describing an entry into a new situation are very general, and can be found in various domains such as demography, economics, sociology and epidemiology. This is the case, for instance, of first sexual intercourse, first marriage, firth birth, first job, first adult migration. These processes have common features: they start at a given age, are concentrated within a relatively narrow age group, are cumulative, and concern only a fraction of a population. The proportion of persons who have experienced the event at age (x) follows a sigmoid function, that can be modelised for a variety of purposes:

[1] School of Physics, U niversity of KwaZulu-Natal, Durban 4041, South Africa, matthewsa@ukzn.ac.za
[2] Epidemiology of Emerging Diseases Unit, Institut Pasteur, 25-28 rue du Docteur Roux, 75015 Paris, leclercpauline@yahoo.fr, mgarenne@pasteur.fr

smoothing data, comparing populations, calculating mean and variance of age at event, and computer projections and simulations.

The model that we are presenting here was first applied to the case of first marriage, but has a much broader scope. This research came out of computer modelling of the dynamics of the HIV-AIDS epidemic in Southern Africa. This exercise required functions that model the age-specific "recruitment rate" into first marriage in a cohort, a common indicator in demography . The recruitment rate $r(a)$ determines the first marriage rate $f(a)$ and hence the whole distribution of ages at marriage.

Several models were already available, although all required complex calculations. The most widely used model in demography is that of Coale and McNeil (1972), who presented a formula in closed form for a female cohort. This model has an empirical dimension, since it is based on the age pattern of first marriage in 19th century Sweden. Another formula in closed form had been proposed at about the same time by Hernes [1972], though with more problems for practical applications. The Coale-McNeil model was further analysed and improved by a variety of authors [Feeney, 1972; Ewbank, 1974; Rodriguez and Trussell, 1980]. The Coale-McNeil model is based on a double exponential, which has been recently re-analysed, simplified and made more systematic by several authors [Kaneko, 1991; Liang, 2000; Kaneko, 2003]. In the latest formulation by Kaneko (2003), the model became a generalised Log-Gamma function, still requiring complex calculations for practical purposes.

The aim of this paper is to provide a simpler model, fitting empirical data at least as well as the Coale-McNeil reference model, independent on any empirical pattern, and easier to program for computer simulations. In section 2, we provide the reader with background information on other models. In section 3, we present the Picrate model. In section 4 we compare our approach with the Coale-McNeil model. In section 5 we present an empirical application on a case study.

## 2. BACKGROUND

The model developed by Coale and McNeil was based on a previous finding by Coale [1971] that a standard function, when scaled and shifted, was a good fit to the age patterns of first marriage across a range of populations. The Coale-McNeil model is defined by a probability density function composed of a double exponential (see below section 4). This model implies that the distribution of waiting times is composed of a series of exponentially distributed waiting times. In the original model, this was interpreted as a series of waiting times as in real life, such as: time to entry into marriageable state, time to meet a possible mate, time for engagement, and eventually waiting time for marriage.

More generally, when a process is controlled by a constant rate $r$, the mean time for the event to happen is simply $T = 1/r$. Applied to first marriage, such a simple model where the recruitment rate $r_c$ that commences at age $a_0$, is constant, results in the proportion ever-married given by $p(a) = p_{\max}\left\{1 - \exp\left[-r_C\left(a - a_0\right)\right]\right\}$. Although this provides a good fit to data for later ages, it fails to model the initial exponential rise that is observed. We discovered that a much better fit could be obtained if the age-specific recruitment rate $r(a)$ was phased-in over a time scale $T$, from 0 at age $a_0$ to a maximum value $r_{\max}$. Its rate of change is defined to be $dr/da = (a - a_0)(r_{\max} - r)/T^2$. We took

*dr*/*da* as a Weibull distribution function, and it turned out that a good fit was obtained when $T = 1/r_{\max}$, which has the great advantage of reducing the number of parameters by one. This was the rationale for our new model. The acronym PICRATE describes this choice: Phased-In-Constant Rate.

## 3. DESCRIPTION OF THE PICRATE MODEL

We define the following quantities for an age-cohort of women:

- $a$ age (continuous, measured in years)
- $a_0$ age at which marriages begin
- $p(a)$ proportion ever married at age $a$.
- $p_{\max}$ proportion that will eventually marry (maximum value of $p$)
- $r(a)$ age-specific first marriage recruitment rate (nuptiality rate among susceptible women)
- $r_{\max}$ maximum value of $r(a)$
- $f(a)$ first marriage rate, or marriage schedule, the proportion of the cohort that marries per time unit. In a population, $f(a)$ is equal to $dp/dt$, and in a stable population to $dp/da$. The $f(a)$ function is the probability density function of first marriages.

The ever-married proportion *p(a)* increases from 0 at $a_0$ to a maximum value $p_{\max}$, and $r(a)$ rises from 0 at $a_0$ to a maximum value $r_{\max}$. At a given age, the proportion of the cohort that has yet to marry is $(p_{\max} - p)$. Before $a_0$, the functions $r$, $f$, and $p$ are strictly zero.

The model's equations are described here in scale-independent "standard" form, by defining the dimensionless standardised age as

$$\mathrm{x} = \mathrm{r}_{\max}\left(\mathrm{a} - \mathrm{a}_0\right) \tag{1}$$

where $T \equiv 1/r_{\max}$ is the time-scale for the process. We also defined normalised functions as:

$$\hat{p}(x) = \frac{p(x)}{p_{\max}}, \qquad \hat{f}(x) = \frac{f(x)}{p_{\max}} \qquad \text{and} \qquad \hat{r}(x) = \frac{r(x)}{r_{\max}}. \tag{2}$$

The first two functions are respectively the cumulative probability distribution function and the probability density function.

The rate of new marriages for $a > a_0$ (or $x > 0$) is given by:

$$\frac{d\hat{p}}{dx} = \hat{r}\left(1 - \hat{p}\right), \tag{3}$$

and we set

$$\hat{r}(x) = 1 - \exp\left(-x^2/2\right). \tag{4}$$

Consequently

$$\frac{d\hat{r}}{dx} = x \exp\left(-x^2/2\right), \text{ a Weibull distribution function.} \quad (5)$$

The choice of a Weibull function is justified by the sigmoid shape of the recruitment rate in empirical data. Other functions which are more symmetrical, such as the Logistic function, would not have permitted to fit the empirical data. Furthermore, the Weibull function is widely available, and easy to calculate numerically [Weibull, 1951].

From equations (3) and (4) we derive an explicit formula for $\hat{p}(x)$ by separation of variables and integration to obtain:

$$\hat{p}(x) = 1 - \exp\left\{ -\int_0^x \hat{r}(u)\, du \right\}. \quad (6)$$

We evaluate the integral:

$$\int_0^x \hat{r}(u)\, du = x - \int_0^x \exp\left(-\frac{1}{2}u^2\right) du \quad (7)$$

to obtain a final expression for $\hat{p}(x)$:

$$\hat{p}(x) = 1 - \exp\{ -x + I(x) \}. \quad (8)$$

where

$$I(x) = \int_0^x \exp\left(-\frac{1}{2}u^2\right) du = \sqrt{2\pi}\left[N_C(x) - \frac{1}{2}\right], \quad (9)$$

and $N_C(x)$ is the cumulative normal distribution function with mean 0 and standard deviation 1.

Equation (9) is not a formula in closed form, but it is nevertheless easily evaluated with the use of special functions that are provided by most mathematical computer packages. For instance, Abramowitz and Stegun [1964] provide the following nearly single-precision (error < $1\times10^{-5}$) approximation for $N_C(x)$, where we first set $t$=1/(1+0.33267 $x$), and then

$$N_C(t(x)) \approx 1 - (0.4361836\, t - 0.1201676\, t^2 + 0.9372980\, t^3) \exp(-t^2/2)/\sqrt{2\pi}\ . \quad (10)$$

Other related special functions may be used in place of $N_C(x)$.

Furthermore, the function presented in (9) can be integrated numerically, to obtain the median, mean and variance in units of standardised age: *MED(X)* = 1.8693, $E(X)$ = 2.090, $V(X)$ = 1.292; the standard deviation is then $\sigma(X)$ = 1.137. These estimates can be used to calculate directly, in units of years and years$^2$, the mean and variance of the distribution of age at first marriage, simply by inversing the change in variables:

$$MED(A) = a_0 + \frac{1.8693}{r_{\max}}$$

$$E(A) = a_0 + \frac{2.090}{r_{\max}} \tag{11}$$

$$V(A) = \frac{1.292}{r_{\max}^2}. \tag{12}$$

Finally, directly from (3), (4) and (8), we obtain:

$$\hat{f} = \frac{d\hat{p}}{dx} = \left(1 - e^{-\frac{1}{2}x^2}\right) \exp\left\{-x + I(x)\right\}. \tag{13}$$

Straightforward derivations provide the needed functions after the change in variable is inversed from (1). Expanding from the normalised functions and standardised age, we have the marriage recruitment rate from (4):

$$r(a) = r_{\max}\left\{1 - \exp\left[-\frac{1}{2} r_{\max}^2 (a - a_0)^2\right]\right\} \tag{14}$$

the schedule of first marriages from (13):

$$f(a) = \frac{dp}{da} = p_{\max}\, r_{\max}\left(1 - e^{-\frac{1}{2} r_{\max}^2 (a-a_0)^2}\right) \exp\left\{-r_{\max}(a - a_0) + I\left(r_{\max}(a - a_0)\right)\right\} \tag{15}$$

and the ever-married proportion from (8):

$$p(a) = p_{\max}\left\{1 - \exp\left[-r_{\max}(a - a_0) + I\left(r_{\max}(a - a_0)\right)\right]\right\} \tag{16}$$

The curve for $p(a)$ describes the history of first-marriages in the cohort, but for a stable population it also describes the age-specific ever-married proportion for a population. Note that the curve is determined by the three parameters $a_0$, $r_{\max}$ and $p_{\max}$.

## 4. THEORETICAL COMPARISON WITH THE COALE-McNEIL MODEL

In this section we present a simple formulation of the Coale-McNeil model – including the recruitment rate – based on the original age-distribution function given by Coale and McNeil [1972], a modified version by Rodriguez and Trussell [1980], and an expression for the cumulative age-distribution given by Kaneko [2003]. We then formally compare the PICRATE and Coale-McNeil models.

The Coale-McNeil [1972] distribution of first marriages (or marriage rate) in a female cohort is given by the following formula:

$$\hat{f}_{CM}(a) = \frac{\beta}{\Gamma(\alpha/\beta)} \exp\left[-\alpha(a - \mu) - \exp\left\{-\beta(a - \mu)\right\}\right]. \tag{17}$$

where $\Gamma(\theta)$ is the gamma function. Note that this is the distribution of ages at first marriage, excluding those who will never marry. The age-specific schedule (proportion of the entire cohort that marries at a given age) is given by $f(a) = p_{\max}\, \hat{f}(a)$.

We define the following notation:

$$\theta \equiv \alpha/\beta \quad \text{and} \quad z(a) \equiv \exp\left[-\beta(a-\mu)\right]. \tag{18}$$

Then (17) may be written as

$$\hat{f}_{CM}(a) = \frac{\beta}{\Gamma(\theta)} z^{\theta} e^{-z}, \tag{19}$$

and the cumulative distribution as

$$\hat{p}_{CM}(a) = 1 - \frac{\gamma(\theta,z)}{\Gamma(\theta)}, \tag{20}$$

where the lower incomplete gamma function is defined here as

$$\gamma(\theta,z) = \int_0^z e^{-u} u^{\theta-1} du. \tag{21}$$

It is easily verified that (20) is the cumulative distribution function by differentiating:

$$\hat{f}_{CM}(a) = \frac{d\hat{p}_{CM}}{da} = -\frac{1}{\Gamma(\theta)} \frac{dz}{da} \frac{d\gamma}{dz} = \frac{\beta}{\Gamma(\theta)} z^{\theta} e^{-z}.$$

Finally, the Coale-McNeil first marriage recruitment rate (or hazard) is given by

$$r_{CM}(a) = \frac{\hat{f}_{CM}(a)}{1 - \hat{p}_{CM}(a)} = \frac{\beta z^{\theta} e^{-z}}{\gamma(\theta,z)}. \tag{22}$$

It is straightforward to show that as $a \to \infty$, $r_{CM} \to \alpha$. So, as in the PICRATE model, the Coale-McNeil first marriage recruitment rate tends asymptotically to a constant value at late ages, after an initial exponential rise.

We further compared formally the PICRATE model with the Coale-McNeil model. The Coale-McNeil age-distribution $\hat{f}_{CM}(a)$ is given by (17). A standard two-parameter version of this model was provided by Coale and McNeil [1972], but for this comparison between the models we used the adjusted standard version by Rodriguez and Trussell [1980]. For mean $a_c$ and standard deviation $\sigma$, for which we also define $k_c = 1/\sigma$, we define standardised age as

$$y = \frac{(a - a_c)}{\sigma} = k_c(a - a_c). \tag{23}$$

We then set $\alpha = 1.145$, $\beta = 1.896$ and $\mu = -0.805$, to obtain

$$\hat{f}_s(y) = (1.2813/\sigma)\exp\left[-1.145(y + 0.805) - \exp\{-1.896(y + 0.805)\}\right]. \tag{24}$$

The comparison between the Coale-McNeil and the PICRATE models was done by computing the Coale-McNeil functions (19), (20) and (22) for the mean and standard deviation of the PICRATE distribution: $\mu = 2.090$ and $\sigma = 1.137$, in units of PICRATE

standardised age $x$ (1). The PICRATE functions plotted were (4), (8) and (13). Special functions were computed by algorithms in Press *et al* [1992].

The curves are plotted in Figure 1 below. The maximum absolute difference between the age-distributions was 0.015, between the cumulative distributions 0.004, and between recruitment rates 0.02. (Note that in this comparison, PICRATE $r_{max} = 1$, and C-M $r_{max} = \alpha/\sigma = 1.007$.) The two distributions match closely, and since each model may be scaled and located with three parameters, the PICRATE model serves as a close approximation and hence as an alternative to the Coale-McNeil model.

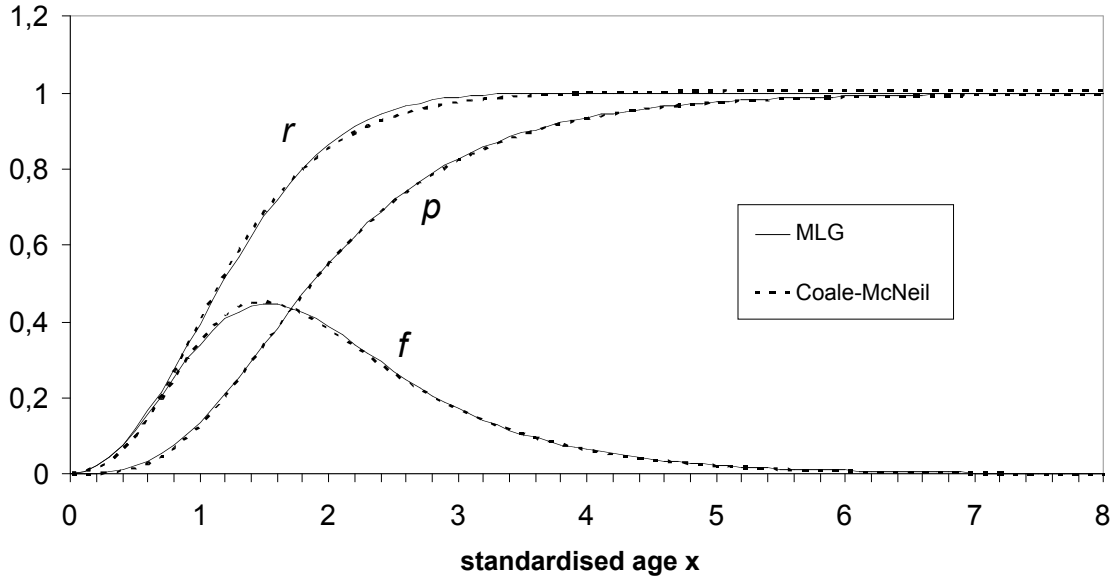



FIGURE 1. Theoretical comparison of the PICRATE model with the Coale-McNeil model. Values of $r$, $\hat{p}$, and $\hat{f}$ are plotted against standardised age $x$, with the PICRATE curves as solid lines, and the Coale-McNeil curves as dotted lines.

## 5. EMPIRICAL VALIDATION OF THE PICRATE MODEL

We have applied the PICRATE model to the case of Zambia, and compared it to the Coale-McNeil model. Empirical data were provided by the Demographic and Health Surveys (DHS), conducted in 1992, 1996 and 2001, which were combined for increasing sample size and reducing random fluctuations. These data provided at the same time the proportion ever-married and the age at first marriage, which allowed us to compute the empirical values of $p(a)$ and $f$(a). Fitting the PICRATE model was done by minimising the squared-distance between the empirical $p(a)$ and the model $p'(a)$, using the Solver module in Excel, and a similar procedure was used for the Coale-McNeil model.

Results give the following values for the PICRATE model: $a_0 = 11.697$, $r_{max} = 0.298$, and $p_{max} = 0.988$, which results in a mean age at marriage = 18.72 years, and a standard deviation = 3.82 years. Comparison with the empirical values, for $p(a)$ and $f(a)$, are displayed in Figure 2, and shows the goodness of fit of the model. The proportion ever married fitted closely the empirical data, except below age 14 and above

age 40, both differences which could be explained by sample size fluctuations (less than 5 % of marriages occurred prior to age 14, and only 21 marriages occurred above age 35 out of a sample of 22739 women), and possibly by cohort changes or by misreporting of age at first marriage. For the marriage rate, the fitting was also good, including at young ages (below age 20). The age pattern was well described by the PICRATE model, although empirical values tended to be somewhat below the fitted curve between age 20 and age 30 and somewhat higher from age 30 to age 40 (cf. Figure 2).

The Coale-McNeil model fitted equally well the Zambian data, and was always very close to the PICRATE model (cf. Figure 2). Discrepancies with $p(a)$ were the same at very young and very old ages, as well as discrepancies with $f(a)$.

We also tried to use other models on the Zambian case. The Kaneko model (the Log-Gamma function) did not fit the $p(a)$ function any better. In addition, it tended to over-estimate the mode of the $r(a)$ function around age 20, and tended to decline faster than the other models afterwards, leading to an increasing bias above age 30. The Hernes model was much worse, since it tended to produce a very symmetrical $f(a)$ function and missed the sharp increase in the $p(a)$ function at young ages.

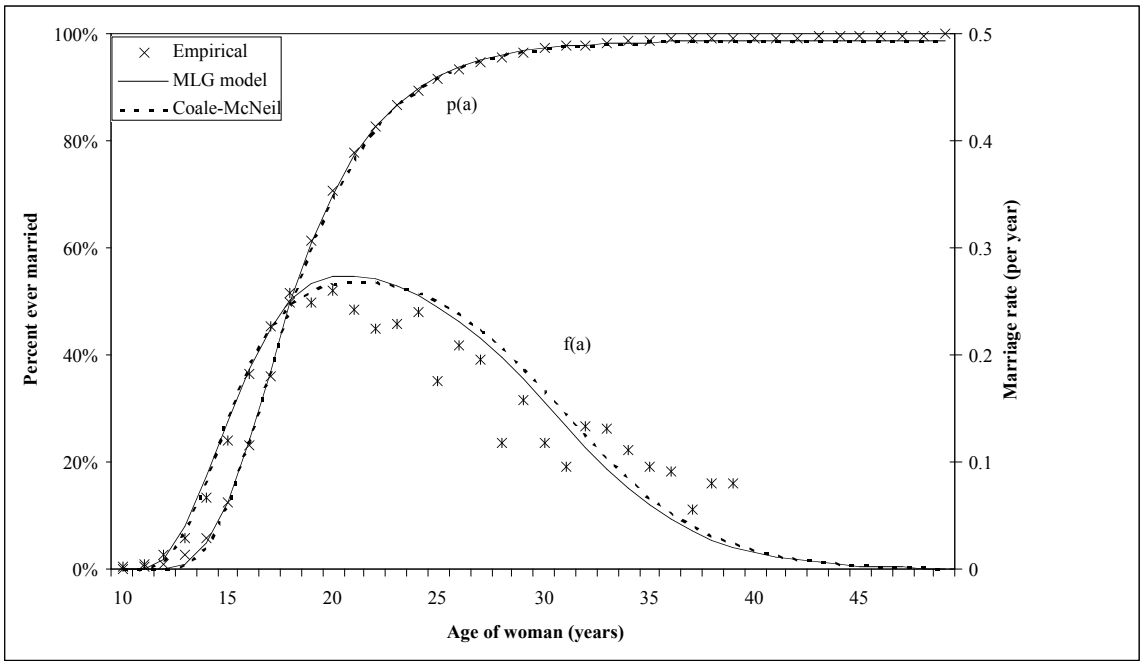



FIGURE 2. Fitting empirical data with the PICRATE model and comparison with Coale-McNeil model, Zambia DHS surveys, female population

## 6. APPLICATION TO AFRICAN DATA

To verify how the model behaved in a variety of situations, we applied the PICRATE model to three case studies using African DHS datasets: one with very low age at marriage (Niger, 1992), one with average age at marriage (Tanzania, 1999), and one with very high age at marriage (Botswana, 1988), which represent the maximum variations in age pattern of first marriage from African data.

Table 1 and Figure 3 show the goodness of fit with the two models. Distances to empirical data were basically the same in the three cases, and estimates of mean age,

median age, and variance in age at marriage were also similar. There is therefore no doubt that the PICRATE model performed well, and at least as well as the Coale-McNeil model, for fitting African data on age at first marriage.

TABLE 1. Fitting the PICRATE model, and comparison with the Coale-McNeil model in countries where first marriage is early (Niger), average (Tanzania), and late (Botswana). The Coale-McNeil parameters below are related to the parameters in (17) by $\alpha = 0.174/k^{CM}$, $\beta = 0.2881/k^{CM}$, and $\mu = a_0{}^{CM} + 6.06\, k^{CM}$.

| Country | | Niger | Tanzania | Botswana |
|---|---|---|---|---|
| *Year of survey* | | 1992 | 1999 | 1988 |
| Coale-McNeil parameters | $a_0{}^{CM}$ | 11.165 | 12.784 | 11.677 |
| | $k^{CM}$ | 0.420 | 0.601 | 1.137 |
| | $p_{\max}$ | 0.999 | 0.986 | 0.817 |
| Coale-McNeil results | Mean age | 15.44 | 19.10 | 23.81 |
| | Median age | 14.38 | 17.87 | 23.92 |
| | Variance | 7.65 | 15.57 | 46.69 |
| | Distance from empirical data | 0.0109 | 0.0028 | 0.0103 |
| PICRATE parameters | $a_0$ | 10.433 | 11.870 | 10.400 |
| | $r_{\max}$ | 0.419 | 0.290 | 0.153 |
| | $p_{\max}$ | 0.999 | 0.986 | 0.823 |
| PICRATE results | Mean age | 15.42 | 19.08 | 23.99 |
| | Median age | 14.40 | 17.90 | 23.93 |
| | Variance | 7.42 | 15.40 | 53.19 |
| | Distance from empirical data | 0.0098 | 0.0026 | 0.0114 |
| Distance between both models | | 0.00012 | 0.00009 | 0.00011 |

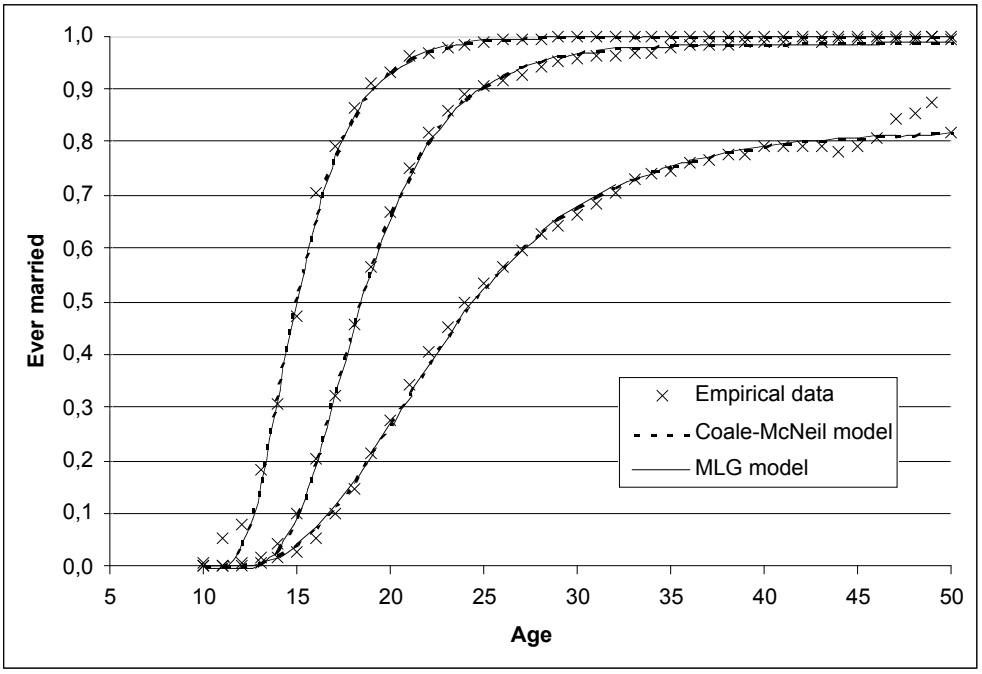



FIGURE 3. Fitting empirical data of proportions of ever married women with the PICRATE model, and comparison with the Coale-McNeil model, in three extreme situations: Niger 1992, Tanzania 1999 and Botswana 1988.

## 7. DISCUSSION

The PICRATE model was developed to fit the age pattern of first marriage in a variety of situations, with the main focus on African countries. The PICRATE model appears as a close approximation to the Coale-McNeil model, offers a simpler formulation of the recruitment rate, and seems to fit empirical data at least as well.

A practical interpretation of the PICRATE model is that the marriage process is determined by an age-specific recruitment rate $r(a)$ from the susceptible population that will eventually marry, that is phased-in from a starting age $a_0$, before which the proportion ever-married and the recruitment rate are zero, to a maximum rate $r_{max}$ at later ages. Like the Coale-McNeil model, the PICRATE model has only three parameters ($a_0$, $p_{max}$ and $r_{max}$), all of which have a simple interpretation: earliest age at marriage, proportion that will eventually marry, and maximum rate of marriage in the susceptible population.

The PICRATE formulation is particularly useful for computer simulations of populations that follow the marriage process through time, since the first marriage recruitment rate $r(a)$ has a simple formula in closed form. Also, the simulated process starts at a definite age $a_0$, with all relevant variable quantities set to zero prior to this point. Furthermore, $r(a)$ tends to $r_{max}$ as age increases, which suggests that $r_{max}$ could be a suitable value for the remarriage rate, although this is a matter for further investigation.

Both $p(a)$ and $f(a)$ may be expressed as formulae, almost in closed form, although including the cumulative normal distribution function, which is available in most

mathematical software packages, and for which approximations exist in closed form, such as equation (10) in this article.

The following features of the PICRATE model appear as new and useful. Firstly, the PICRATE recruitment rate has a simple, easily evaluated formula (14), compared with the equivalent Coale-McNeil formula (22). If $r(a)$ has to be calculated often in a computer program, then the PICRATE formulation has a definite advantage. In an individual-based population computer simulation, it is the probability for each susceptible individual that is necessary to determine transition to first marriage. The marriage rate and the cumulative distribution are needed to verify that the simulation produces acceptable results, but it is the recruitment rate that is needed for the time-evolution.

Secondly, the PICRATE model has zero first marriages before the starting age $a_0$, whereas the Coale-McNeil has a finite – though admittedly very small – proportion of first marriages at ages before $a_0$, extending to zero age and even to negative ages. More significantly, the Coale-McNeil process requires an initial "seed" starting value of the married proportion, whereas the PICRATE model starts with the proportion at zero. A first marriage model that theoretically allows marriages at all ages, even negative ones, cannot be strictly based on realistic behavioural assumptions. The fact that the Coale-McNeil model does allow such marriages to occur, whereas the PICRATE model does not, suggests that the PICRATE model may have, embedded in its mathematical structure, a more realistic basis than the Coale-McNeil model. This is not surprising, since the Coale-McNeil model was derived as a fit to empirical data, rather than as the result of a behavioural process. The relative simplicity of the PICRATE model formulation suggests that it could turn out to be the result of an underlying theoretical process that is a good model for first marriages.

Our model seems to apply to many similar situations, since its three parameters are very general. Since its development, we have applied it successfully to other processes, such as entry into sexual life (first intercourse), entry into fertile life (first birth), for men as for women. It could also be tried on other similar processes, such as entry into working life (first job), exit from family home (first adult migration), and to many other processes in demography, economy, sociology or epidemiology.

The shape of the Weibull function is simple and asymetrical, and corresponds to $k = 2$, that is the Rayleigh distribution. One could also select other shapes for the recruitment rate, with different parameters of the Weibull function, which imply different asymmetry, to fit other processes such as those which happen at the end of life (end of working life, last migration, etc.).

*Acknowledgements*. Alan Matthews acknowledges financial support from the South African Centre for Epidemiological Modelling and Analysis (SACEMA), a scholarship from the International AIDS Society through the HIV Research Trust, an invitation from the Pasteur Institute, and facilitation of this project by the University of KwaZulu-Natal. The three investigators also acknowledge the financial support for the Franco-South African consortium for scientific exchange.

# REFERENCES

ABRAMOWITZ M, STEGUN I. A. (eds), "Handbook of mathematical functions", *Applied Mathematics Series* 55, Washington (DC), National Bureau of Standards, US Government Printing Office, 1964.

Central Statistical Office [Zambia]. Census 2001. Central Board of Health, and ORC Macro. Zambia Demographic and Health Survey 2001-2002, Calverton, Maryland, USA, Central Statistical Office, Central Board of Health, and ORC, 2003.

COALE A. J., "Age Patterns of Marriage", *Population Studies* 25(2), 1971, p. 193-214.

COALE A. J., McNEIL D. R., "The Distribution by Age of the Frequency of First Marriage in Female Cohort", *Journal of the American Statistical Association* 67, 1972, p. 743-749.

EWBANK D., *An Examination of Several Applications of the Standard Pattern of Age at First Marriage,* Princeton University, Dept. of Economics, 1974, [unpublished Ph.D. dissertation].

FEENEY G., *A Model for the Age Distribution of First Marriage*, Working Paper No. 23, East-West Population Institute, 1972.

HERNES G., "The Process of Entry into First Marriage", *American Sociological Review* 37, 1972, p. 173-182.

KANEKO R., "Demographic analysis of First Marriage Process", *The Journal of Population Problems* 47(3), 1991, p. 3-27.

KANEKO R., "Elaboration of the Coale-McNeil Nuptuality Model as the Generalized Log Gamma Distribution: A New Identity and Empirical Enhancements", *Demographic Research* 9(10), 2003, p. 221-262.

LIANG Z., *The Coale-McNeil model. Theory, generalisation and application,* Amsterdam, Thela thesis, 2000.

PRESS W.H., TEUKOLSKY S.A., VETTERLING W.T., FLANNERY B.P., *Numerical Recipes*, second edition. Cambridge University Press, 1992.

RODRIGUEZ G., TRUSSELL J., "Maximum Likelihood Estimation of the Parameters of Coale's Model Nuptuality Schedule from Survey Data", *World Fertility Survey*, Technical Bulletins, No. 7, 1980.

WEIBULL W., "A statistical distribution function of wide applicability", *J. Appl. Mech.-Trans. ASME* 18(3), 1951, p. 293-297.

# AJUSTEMENT DU SCHEMA PAR AGE D'ENTRÉE EN PREMIERE UNION : LE MODELE PICRATE

Alan P. MATTHEWS[1],
Pauline M. LECLERC[2],
Michel L. GARENNE[3]

RESUME – *Nous présentons un nouveau modèle qui décrit le schéma par âge d'entrée dans une nouvelle situation. L'introduction du taux de recrutement rend l'utilisation de ce modèle très simple, en particulier pour simuler l'évolution d'une population au cours du temps. Le modèle est défini par trois paramètres : $a_0$, l'âge minimal à l'événement, $p_{max}$, la proportion maximale de la population qui connaîtra l'événement, et $r_{max}$ le taux maximal de recrutement. Ce modèle est testé sur l'entrée en premier mariage, et est comparé au modèle de Coale-McNeil. Nous avons appliqué ce modèle au cas de la Zambie, ainsi qu'à divers pays africains ayant des schémas différents. Le modèle peut s'appliquer à divers processus, tels que les premiers rapports sexuels, la première naissance, le premier emploi, la première migration adulte, etc.*

MOTS CLES – Analyse biographique, Modélisation mathématique, Age au premier mariage, Modèle de Coale-McNeil, Enquête EDS.

ABSTRACT – Fitting the age pattern of entry into first union: the Picrate model.
*We present a new model for fitting the age pattern of entry into a new situation. The formulation of the marriage recruitment rate is simpler than in other available models, and particularly useful for computer simulation of the time-evolution of a population. The model has three parameters: $a_0$, the starting age at event, $p_{max}$, the maximum proportion of the population at risk p(a), and $r_{max}$, the maximum value of the recruitment rate r(a). This model was tested on the entry into first union, and was compared with the widely-used Coale-McNeil model. We applied this model to the case of Zambia, and to a variety of African countries with different features. This model could also be used for fitting a variety of processes, such as first sexual intercourse, first birth, first job, first adult migration, etc.*

KEY WORDS – Event history, Mathematical model, Age at first marriage, Coale-McNeil model, DHS survey.

[1] School of Physics, Université du KwaZulu-Natal, Durban 4041, Afrique du Sud
Courriel: matthewsa@ukzn.ac.za
[2] Doctorante, Université Paris VI, Paris & Institut Pasteur, Epidémiologie des Maladies Emergentes. Courriel : leclerc@pasteur.fr
[3] PhD, Institut Pasteur, Epidémiologie des Maladies Emergentes & Institut de Recherche pour le Développement, Paris. Courriel : mgarenne@pasteur.fr

## INTRODUCTION

Les phénomènes d'entrée dans une nouvelle situation sont assez généraux, et se retrouvent dans de nombreux domaines, tels que la démographie, l'économie, la sociologie ou l'épidémiologie. Pour donner quelques exemples on peut citer divers événements tels que : les premiers rapports sexuels, la première union, la première naissance, le premier emploi, la première migration adulte. Ces phénomènes ont des points communs : ils sont concentrés dans un groupe d'âge, sont cumulatifs, et ils touchent *in fine* une fraction plus ou moins élevée de la population. La proportion de personnes ayant connu l'événement à l'âge (x) suit donc une courbe de type sigmoïde, qu'il est intéressant de pouvoir modéliser pour plusieurs raisons : lisser des données, comparer des populations, calculer les moyennes et variances des âges à l'événement considéré ou simuler les histoires de vie au sein d'une population.

Le modèle que nous proposons ici s'applique d'abord au premier mariage, mais il a une portée plus générale. Cette étude a son origine dans un projet plus vaste qui consiste à modéliser les dynamiques d'épidémies du VIH/SIDA dans les pays d'Afrique sub-Saharienne. Pour ce faire, il était nécessaire d'approcher au mieux les comportements de la population étudiée et donc de pouvoir décrire l'entrée en premier mariage. Ceci impliquait de définir une fonction que les démographes appellent le « taux de recrutement au premier mariage », noté *r(a),* qui est défini ci-dessous et qui détermine le taux d'entrée en premier mariage *f(a)* et donc la distribution des âges au mariage.

Plusieurs modèles de ce type sont déjà disponibles dans la littérature, mais ils exigent souvent des hypothèses fortes, des paramètres arbitraires et des calculs complexes. Le plus communément utilisé en démographie est le modèle de Coale et McNeil [Coale et McNeil, 1972], qui propose une formulation exacte pour une cohorte féminine. Ce modèle est très empirique, puisque ses coefficients dépendent de la distribution des premiers mariages en Suède au XIX° siècle. La même année, une autre formulation de la distribution des âges au mariage a été proposée par Hernes [Hernes, 1972], mais elle pose plus de problèmes pour les applications pratiques. Par la suite, le modèle de Coale-McNeil a été repris par plusieurs auteurs [Feeney, 1972 ; Ewbank, 1974 ; Rodriguez et Trussell, 1980]. Le modèle de Coale-McNeil repose sur une double exponentielle qui a été récemment ré-analysée et simplifiée par plusieurs auteurs [Kaneko, 1991 ; Liang, 2000 ; Kaneko, 2003]. Dans cette dernière version proposée par Kaneko, le modèle devient une fonction log-Gamma généralisée, restant tout de même difficile à utiliser en pratique.

Le but de cet article est de proposer une formulation plus simple et plus générale pour les applications pratiques, approchant les données au moins aussi bien que le modèle de Coale-McNeil, indépendante de tout paramètre spécifique d'une situation particulière, et plus facile à utiliser pour la programmation informatique. Nous détaillerons tout d'abord certains modèles approchant le schéma par âge du premier mariage cités précédemment, puis nous présenterons notre modèle. Nous le comparerons ensuite au modèle de Coale-McNeil et, enfin, nous présenterons une application à des données empiriques issues d'enquêtes réalisées en population générale en Afrique.

## LES BASES DU MODELE

Le modèle développé par Coale et McNeil est basé sur le fait qu'une fonction simple, après mise à l'échelle, permet un bon ajustement du schéma par âge d'entrée au premier mariage [Coale, 1971]. Le modèle de Coale-McNeil est défini par une fonction de densité de probabilité composée d'une double exponentielle, ce qui sous-entend que la distribution des temps d'attente est composée d'une suite de temps d'attente distribués de façon exponentielle. Dans le modèle original, trois temps d'attente sont utilisés : attente avant d'être en âge de se marier, puis attente avant de rencontrer un partenaire, et enfin attente avant le mariage.

Plus généralement, lorsqu'un processus est contrôlé par un taux constant $r$, le temps moyen d'attente avant qu'un événement arrive est $T=1/r$. Appliqué aux premiers mariages, un modèle simple, dans lequel le taux de recrutement $r_c$, non nul pour des âges à partir de $a_0$, est constant, donne, dans les notations définies ci-dessous, la proportion suivante de personnes déjà mariées : $p(a) = p_{\max}\left\{1 - \exp\left[-r_c(a - a_0)\right]\right\}$. Bien qu'il donne un bon ajustement des données pour les âges élevés, il modélise mal l'évolution initiale de l'exponentielle. Nous avons trouvé qu'un meilleur ajustement pouvait être obtenu si le taux de recrutement $r(a)$ augmente progressivement avec l'âge sur une période de temps $T$, allant de 0 pour l'âge $a_0$ initial à une valeur maximale $r_{max}$. Il varie selon la différentielle $dr/da = (a-a_0)(r_{max}-r)/T^2$. Nous ajustons $dr/da$ par une fonction de Weibull, et une bonne approximation est obtenue pour $T=1/r_{max}$, qui présente le grand avantage de réduire le nombre de paramètres. Cette observation est la principale justification de notre nouveau modèle, nommé PICRATE, selon son acronyme en anglais (Phased-In Constant RATE).

## DESCRIPTION DU MODELE

Définissons les notations suivantes pour une cohorte d'hommes ou de femmes :

- $a$ âge (en continu, mesuré en années)
- $a_0$ âge minimal au premier mariage
- $p(a)$ proportion de la population d'âge $a$ ayant déjà été mariée
- $p_{max}$ proportion ultime de la population qui se mariera, la valeur maximale de $p(a)$
- $r(a)$ taux de recrutement au premier mariage parmi les femmes susceptibles de se marier à l'âge a
- $r_{max}$ valeur maximale de la fonction $r(a)$
- $f(a)$ taux de primo-nuptialité, ou d'entrée au premier mariage, soit la proportion de la cohorte d'âge $a$ qui se marrie par unité de temps. Dans une population stable $f(a)$ est égal à $dp/da$. La fonction $f(a)$ est la densité de probabilité des premiers mariages.

La proportion de la population ayant déjà connu le mariage $p(a)$ varie de 0 à l'âge $a_0$ à une valeur maximale $p_{max}$, et $r(a)$ varie de 0 à l'âge $a_0$ à une valeur maximale $r_{max}$. À un âge donné, la proportion de la cohorte qui n'est pas encore mariée est $(p_{max}-p)$, qui est celle des femmes susceptibles de se marier. Avant $a_0$, les fonctions $r$, $f$ et $p$ sont nulles.

Les équations du modèle sont décrites en définissant l'âge standardisé (sans dimension) de la façon suivante :

$$x = r_{\max}(a - a_0) \tag{1}$$

$T \equiv 1/r_{\max}$ est l'échelle de temps du processus. On définit également les fonctions normalisées suivantes :

$$\hat{p}(x) = \frac{p(x)}{p_{\max}}, \hat{f}(x) = \frac{f(x)}{p_{\max}} \text{ et } \hat{r}(x) = \frac{r(x)}{r_{\max}} \tag{2}$$

Les deux premières fonctions correspondent respectivement aux fonctions de distribution cumulée et la densité de probabilité d'entrée en premier mariage.

Le taux de nouveaux mariages pour des âges supérieurs à $a_0$ (ou pour des $x$ supérieurs à 0) est donné par :

$$\frac{d\hat{p}}{dx} = \hat{r}(1 - \hat{p}) \tag{3}$$

Puis

$$\hat{r}(x) = 1 - \exp(-x^2/2) \tag{4}$$

ce qui donne

$$\frac{d\hat{r}}{dx} = x \exp(-x^2/2), \text{ qui est une fonction de Weibull.} \tag{5}$$

Le choix de la fonction (4) est dicté par sa forme sigmoïde, et par le fait qu'elle produit des taux de recrutement qui peuvent être asymétriques, comme cela est observé empiriquement. Le choix d'une fonction symétrique, comme la fonction Logistique, n'aurait pas permis d'ajuster correctement les taux empiriques. En outre, la fonction de Weibull est une fonction classique, et simple à calculer numériquement [Weibull, 1951].

On peut alors déduire des équations (3) et (4) une formule explicite pour $\hat{p}(x)$, par séparation de variables et intégration :

$$\hat{p}(x) = 1 - \exp\left(-\int_0^x \hat{r}(u)du\right). \tag{6}$$

Après calcul de l'intégrale:

$$\int_0^x \hat{r}(u)du = x - \int_0^x \exp\left(-\frac{1}{2}u^2\right)du \tag{7}$$

nous obtenons pour $\hat{p}(x)$ l'expression suivante :

$$\hat{p}(x) = 1 - \exp\{-x + I(x)\}. \quad (8)$$

où

$$I(x) = \int_0^x \exp\left(-\frac{1}{2}u^2\right) du = \sqrt{2\pi}\left(N_C(x) - \frac{1}{2}\right), \quad (9)$$

et $N_C(x)$ est la fonction de distribution cumulée de la loi normale centrée réduite.

Le calcul numérique des valeurs prises par cette dernière fonction ne peut faire qu'approximativement. Mais les logiciels de calculs fournissent des approximations acceptables. Ainsi, Abramowitz et Stegun [Abramowitz et Stegun, 1964] fournissent une approximation de $N_C(x)$, avec une erreur inférieure à $10^{-5}$ :

Avec t=1/(1+0,33267$x$), on obtient :

$$N_C(x) \approx 1 - (0{,}4361836\, t - 0{,}1201676\, t^2 + 0{,}9372980\, t^3) \exp(-t^2/2)/\sqrt{2\pi} \quad (10)$$

D'autres approximations de $N_C(x)$ peuvent être utilisées.

De plus, la fonction (9) permet d'obtenir, par intégration numérique, la médiane la moyenne et la variance en unité d'âge normalisée. Les calculs donnent : Med[X]= 1.8693, $E[X]$ = 2,090, $Var[X]$ = 1,292, soit un écart-type de $S[X]$ = 1,137. Ces estimations peuvent être utilisées pour calculer directement la médiane, la moyenne (en années) et la variance (en années$^2$) des âges aux premiers mariages correspondant à des données empiriques ajustées par le modèle :

$$Med(A) = a_0 + \frac{1.8693}{r_{\max}}$$

$$E(A) = a_0 + \frac{2.090}{r_{\max}} \quad (11)$$

$$V(A) = \frac{1.292}{r_{\max}^2} \quad (12)$$

ce qui fournit en outre une relation linéaire simple entre la médiane et la moyenne :

$$E(A) = Med(A) + \frac{0.2207}{r_{\max}}$$

Finalement, à partir de (3), (4) et (8) nous obtenons une expression de $\hat{f}$ :

$$\hat{f} = \frac{d\hat{p}}{dx} = \left(1 - e^{-\frac{1}{2}x^2}\right)\exp\{-x + I(x)\} \tag{13}$$

En substituant à x sa valeur donnée par (1), on obtient alors l'expression suivante pour le taux de recrutement au premier mariage :

$$r(a) = r_{\max}\left[1 - \exp\left[-\frac{1}{2}r_{\max}^2(a - a_0)^2\right]\right] \tag{14}$$

Puis la fonction de distribution des âges au premier mariage :

$$f(a) = \frac{dp}{da} = p_{\max} r_{\max}\left[1 - e^{-\frac{1}{2}r_{\max}^2(a-a_0)^2}\right]\exp\{-r_{\max}(a - a_0) + I(r_{\max}(a - a_0))\} \tag{15}$$

Et ainsi la proportion de la cohorte d'âge *a* ayant déjà connu le mariage est donnée par :

$$p(a) = p_{\max}\{1 - \exp[-r_{\max}(a - a_0) + I(r_{\max}(a - a_0))]\} \tag{16}$$

La courbe représentative de *p* décrit alors l'histoire des premiers mariages de la cohorte. Dans une population stable, elle décrit également, pour chaque âge, la proportion de la population ayant déjà été mariée. Notons que cette fonction de distribution ne dépend que de trois paramètres : $a_0$, $r_{max}$ et $p_{max}$.

## COMPARAISON AVEC LE MODELE DE COALE-McNEIL

Dans cette section, nous présentons une formulation simple du modèle de Coale-McNeil, incluant le taux de recrutement, basée sur la fonction de distribution par âge originale donnée par Coale et McNeil [Coale et McNeil, 1972], dans sa version modifiée par Rodriguez et Trussell [Rodriguez et Trussell, 1980], et avec une expression pour la distribution cumulée par âge donnée par Kaneko [Kaneko, 2003]. Puis nous comparons ce modèle à notre modèle.

La distribution des premiers mariages de Coale et McNeil (ou taux de nuptialité) dans une cohorte féminine est donnée par la formule suivante :

$$\hat{f}_{CM}(a) = \frac{\square}{\square(\square/\square)}\exp[\square\square(a - \square) - \exp\{-\square\square(a - \square)\}] \tag{17}$$

où _(□) est la fonction Gamma. Notons que cette distribution est celle des âges au premier mariage, en excluant les femmes qui ne se marieront jamais. La proportion de la cohorte qui se marrie à un âge donné est $f(a) = p_{\max}\hat{f}(a)$ (proportion calculée cette fois sur toutes les femmes de la cohorte).

Nous définissons alors :

$$\lambda \equiv \beta/\alpha \quad \text{et} \quad z(a) \equiv \exp[\lambda(a - \mu)] \tag{18}$$

Puis (17) peut être écrite de la façon suivante

$$\hat{f}_{CM}(a) = \frac{\lambda}{\Gamma(\alpha)} z^{\alpha} e^{-z} \tag{19}$$

et la distribution cumulée comme suit :

$$\hat{p}_{CM}(a) = 1 - \frac{\Gamma(\alpha, z)}{\Gamma(\alpha)} \tag{20}$$

où la fonction gamma incomplète est définie de la façon suivante :

$$\Gamma(\alpha, z) = \int_0^z e^{-u} u^{\alpha-1} du \tag{21}$$

En différentiant, la fonction de densité standardisée est donnée par :

$$\hat{f}_{CM}(a) = \frac{d\hat{p}_{CM}}{da} = -\frac{1}{\Gamma(\alpha)} \frac{d\Gamma}{dz} \frac{dz}{da} = \frac{\lambda}{\Gamma(\alpha)} z^{\alpha} e^{-z}$$

Finalement, le taux de recrutement au premier mariage de Coale-McNeil a l'expression suivante :

$$r_{CM}(a) = \frac{\hat{f}_{CM}(a)}{1 - \hat{p}_{CM}(a)} = \frac{\lambda z^{\alpha} e^{-z}}{\Gamma(\alpha, z)} \tag{22}$$

Lorsque *a* tend vers l'infini, $r_{CM}$ tend vers _, le taux asymptotique de recrutement.
Ainsi, comme dans notre modèle, le taux de recrutement au premier mariage de Coale et McNeil tend asymptotiquement vers une constante, pour les âges élevés.

La distribution par âge $\hat{f}_{CM}(a)$ de Coale-McNeil est donnée par l'équation (17). Une version de ce modèle avec deux paramètres a été donnée par la suite par Coale et McNeil [Coale et McNeil, 1972], mais, pour comparer ce modèle au nôtre, nous utiliserons la version standardisée développée par Rodriguez et Trussell [Rodriguez et Trussell, 1980]. Pour une moyenne $a_c$ et un écart-type _, nous définissons $k_c = 1/$ _ ainsi que l'âge standardisé suivant :

$$y = \frac{(a - a_c)}{\sigma} = k_c(a - a_c) \tag{23}$$

Pour _ =1,145, _ = 1,896 et _ = -0,805, on obtient

$$\hat{f}_s(y) = (1{,}2813/\square)\exp[\square 1{,}145(y + 0{,}805) \square \exp\{\square 1{,}896(y + 0{,}805)\}] \qquad (24)$$

La comparaison entre les deux modèles a été réalisée en calculant les fonctions de Coale-McNeil (19), (20) et (22) pour les moyenne et écart-type de la distribution de notre modèle : _=2,090 et _ =1,137 (en unité d'âge standard *x* pour notre modèle (1)). Nous traçons alors les fonctions (4), (8) et (13) de notre modèle en utilisant, pour les fonctions spécifiques, les algorithmes de Press [Press et al., 2002].

Les courbes sont données sur la figure 1. La différence maximale absolue entre les deux distributions par âge est 0,015, entre les deux distributions cumulées de 0,004 et entre les taux de recrutement 0,02 (nous remarquons que, dans notre modèle, $r_{max}$=1 et dans le modèle de Coale-McNeil, $r_{max}$ = _/□ =1,007). Les deux distributions sont très proches l'une de l'autre et dépendent toutes les deux de trois paramètres. Ainsi, notre modèle donne un bon ajustement des distributions par âge d'entrée en premier mariage et peut servir d'alternative au modèle de Coale et McNeil.

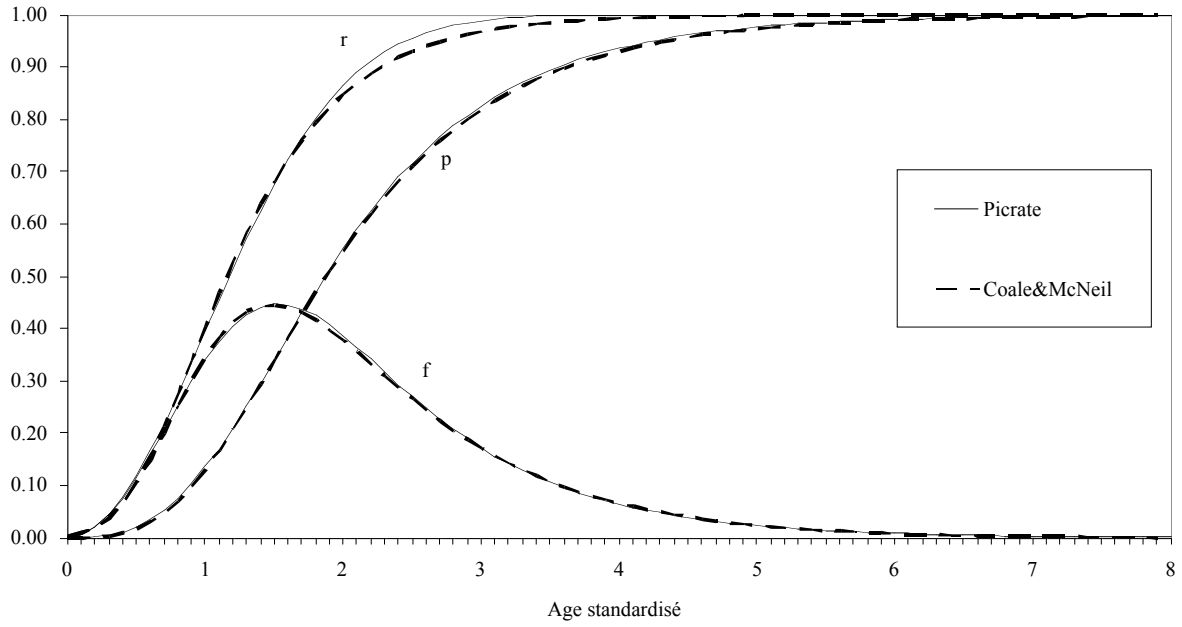



**Figure 1 : Comparaison théorique du modèle PICRATE avec le modèle de Coale-McNeil. Les valeurs de $\hat{r}$, $\hat{p}$, et $\hat{f}$ sont tracées en fonction de l'âge standardisé *x*. Les courbes correspondant à notre modèle sont tracées en ligne pleine et celles correspondant au modèle de Coale-McNeil en pointillés**

## VALIDATION EMPIRIQUE DU MODELE

Nous avons appliqué notre modèle au cas de la Zambie, et l'avons comparé avec le modèle de Coale-McNeil. Les données empiriques sont issues des Enquêtes Démographiques et Sanitaires (EDS) [Central Statistical Office, 2003] réalisées en Zambie en 1992, 1996 et 2001, qui ont été fusionnées pour augmenter la taille de l'échantillon et réduire ainsi les fluctuations d'échantillonnage. Ces données fournissent, pour tous âges, la proportion de la population déjà mariée, et pour chaque personne l'âge au premier mariage, ce qui permet de calculer les valeurs empiriques de *p(a)* et *f(a)*. Pour ajuster ces données, nous avons minimisé la distance des moindres carrés entre la proportion empirique *p(a)* et la proportion *p'(a)* obtenue avec notre modèle. Pour cela, nous avons utilisé le module Solveur du logiciel Excel. Nous avons fait de même pour le modèle de Coale-McNeil.

Les résultats obtenus pour notre modèle sont les suivants : $a_0$=11,697, $r_{max}$ = 0,298 et $p_{max}$ = 0,988, ce qui donne un âge moyen au premier mariage de 18,72 ans avec un écart-type de 3,82 ans. La figure 2 donne la comparaison de *p(a)* et *f(a)* avec les valeurs empiriques. On peut voir que l'approximation obtenue est bonne. La proportion de la population ayant déjà été mariée s'ajuste bien aux données empiriques, à l'exception peut-être des âges inférieurs à 14 ans d'une part, et des âges supérieurs à 40 ans d'autre part, ces différences pouvant s'expliquer par des fluctuations d'échantillonnage (moins de 5% des mariages ont lieu avant 14 ans et seulement 21 mariages ont eu lieu pour des personnes âgées de plus de 35 ans sur un échantillon de 22739 femmes). Pour le taux de mariage, l'ajustement reste satisfaisant, même aux âges inférieurs à 20 ans. Le schéma par âge est bien décrit par notre modèle, bien que les valeurs empiriques se trouvent légèrement en dessous de la courbe d'ajustement pour des âges compris entre 20 et 30 ans et au dessus pour des âges compris entre 30 et 40 ans (Figure 2).

Le modèle de Coale-McNeil s'ajuste également bien aux données de la Zambie et donne des résultats très proches de ceux obtenus avec notre modèle (Figure 2). De faibles différences s'observent également pour *p(a)* et *f(a)* pour les faibles âges ainsi que pour les âges élevés.

Nous avons également testé d'autres modèles sur le cas de la Zambie. Le modèle de Kaneko (la fonction Log-Gamma) ne s'ajuste pas mieux à la fonction empirique *p(a)*. De plus, il surestime légèrement l'âge modal de la fonction *r(a)* aux alentours de 20 ans et décroît plus rapidement que les autres modèles après cet âge, ce qui engendre un biais autour de 30 ans. Le modèle de Hernes ne donne pas de meilleurs résultats, car il tend à produire une fonction *f(a)* symétrique et ne s'ajuste pas à l'accroissement de *p(a)* aux âges les plus faibles.

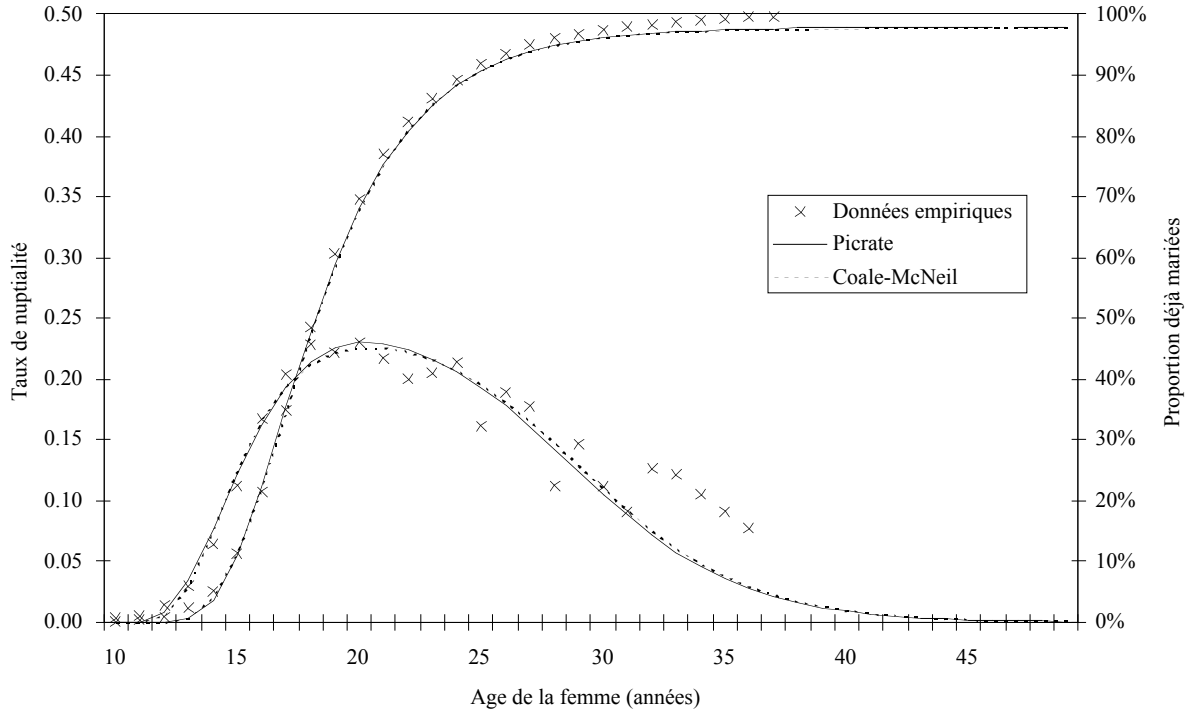



**Figure 2 : Ajustement des données empiriques et comparaison avec le modèle de Coale-McNeil, enquête EDS Zambie, population féminine.**

## APPLICATION AUX DONNEES AFRICAINES

Pour tester la robustesse de notre modèle, nous avons voulu vérifier comment il se comportait dans différentes situations. Nous l'avons appliqué à un jeu de données provenant des enquêtes EDS de trois pays africains : un pays où le mariage est précoce (Niger 1992), un pays où l'âge moyen au premier mariage se situe dans la moyenne africaine (Tanzanie 1999), et un pays où le premier mariage est tardif (Botswana 1988), représentant ainsi les variations extrêmes des données africaines disponibles.

Le tableau 1 et la figure 3 montrent que les deux modèles (le modèle de Coale-McNeil et le nôtre) s'ajustent bien aux données, dans les deux cas extrêmes tout comme dans le cas moyen. Les distances aux données empiriques, définies comme la somme des carrés des différences, sont les mêmes dans les trois cas et les estimations des moyennes et des variances des âges au premier mariage sont également très proches de celles des données empiriques. Ainsi, dans ces trois situations contrastées, notre modèle semble au moins aussi performant que le modèle de Coale-McNeil pour ajuster les distributions des âges au premier mariage.

**Tableau 1 : lkfklfdqlkjfqdijqAjustement avec notre modèle, et comparaison avec le modèle de Coale-McNeil pour des pays où le premier mariage est précoce (Niger), moyen (Tanzanie) et tardif (Botswana). Les paramètres de Coale-McNeil sont liés aux paramètres de l'équation (17) par $\alpha$ = 0,174/$k^{CM}$, $\beta$ = 0,2881/$k^{CM}$, et $\gamma = a_0^{CM} + 6{,}06\ k^{CM}$.**

| | Pays : | Niger | Tanzanie | Botswana |
|---|---|---|---|---|
| | **Année de l'enquête :** | 1992 | 1999 | 1988 |
| ***Modèle de Coale-McNeil*** | | | | |
| **Paramètres** | $a_0^{CM}$ | 11,165 | 12,784 | 11,677 |
| | $k^{CM}$ | 0,420 | 0,601 | 1,137 |
| | $p_{max}$ | 0,999 | 0,986 | 0,817 |
| **Résultats** | **Age moyen** | 15,44 | 19,10 | 23,81 |
| | **Age médian** | 14,38 | 17,87 | 23,92 |
| | **Variance** | 7,65 | 15,57 | 46,69 |
| | **Distance aux données empiriques** | 0,0109 | 0,0028 | 0,0103 |
| ***Modèle Picrate*** | | | | |
| **Paramètres** | $a_0$ | 10,433 | 11,870 | 10,400 |
| | $r_{max}$ | 0,419 | 0,290 | 0,153 |
| | $p_{max}$ | 0,999 | 0,986 | 0,823 |
| **Résultats** | **Age moyen** | 15,42 | 19,08 | 23,99 |
| | **Age médian** | 14,40 | 17,90 | 23,93 |
| | **Variance** | 7,42 | 15,40 | 53,19 |
| | **Distance aux données empiriques** | 0,0098 | 0,0026 | 0,0114 |
| **Distance entre les deux modèles** | | 0,00012 | 0,00009 | 0,00011 |

NB : La distance est définie comme la somme des carrés des différences entre les deux courbes.

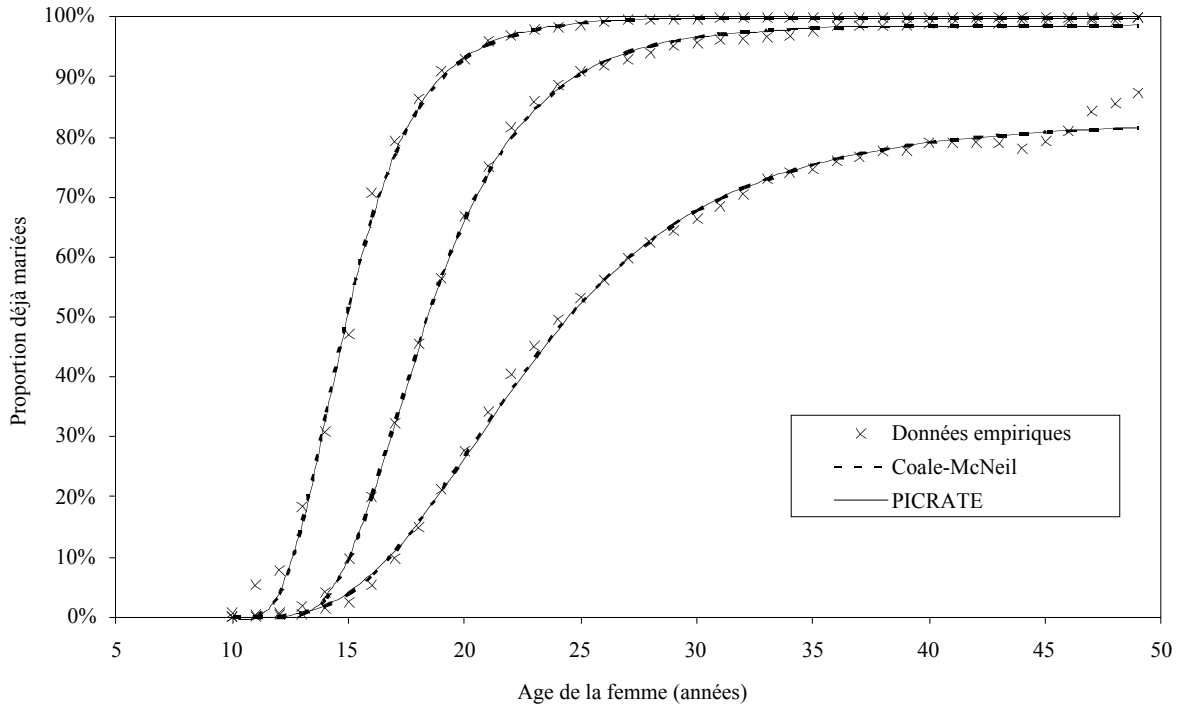



**Figure 3 : Ajustement des proportions de femmes déjà mariées dans trois situations différenciées : Niger 1992 (mariage précoce), Tanzanie 1999 (mariage moyen) et Botswana 1988 (mariage tardif).**

## DISCUSSION

Notre modèle a été développé pour ajuster les schémas par âge d'entrée en premier mariage pour une grande variété de situations, principalement pour des pays d'Afrique sub-Saharienne. Il permet une bonne approximation du modèle de Coale-McNeil, et semble s'ajuster aux données empiriques au moins aussi bien. En outre, il offre une formulation plus simple du taux de recrutement.

Une interprétation de notre modèle est que le processus d'entrée en premier mariage est déterminé par un taux de recrutement $r(a)$ dépendant de l'âge, qui s'applique à une population de femmes susceptibles de se marier. Ce taux croît progressivement de l'âge minimal $a_0$, avant lequel aucun mariage n'a lieu et donc avant lequel la proportion de la population déjà mariée est nulle, jusqu'à un taux maximal $r_{max}$ pour les âges élevés. Tout comme le modèle de Coale-McNeil, ce modèle est déterminé par trois paramètres ($a_0$, $p_{max}$ et $r_{max}$), qui ont tous une interprétation simple : le plus jeune âge au premier mariage avant lequel aucun mariage n'a lieu, la proportion de la population qui finira par se marier et le taux maximal de mariage dans la population de susceptibles.

La formulation que nous avons choisie pour notre modèle est particulièrement utile pour la programmation informatique de la simulation de l'entrée en premier mariage. En effet, le taux de recrutement est exprimé de façon simple et facilement évaluable. De plus, la simulation du processus de mariage peut débuter à un âge défini $a_0$, toutes quantités étant nulles avant cet âge. Enfin, $r(a)$ tend vers $r_{max}$ lorsque les âges augmentent, ce qui suggère que

$r_{max}$ pourrait être une bonne valeur pour le taux de remariage. Cela pourra être le sujet de recherches ultérieures.

Aussi bien $p(a)$ que $f(a)$ peuvent être exprimées par des formules utilisant des fonctions disponibles dans la plupart des logiciels de calculs, incluant notamment une fonction de distribution normale cumulée pour laquelle une approximation simple existe (équation 10).

Les particularités suivantes du nouveau modèle semblent avoir des applications pratiques. Le taux de recrutement s'écrit de manière simple, et sa formule (14) est plus facilement évaluable que celle de Coale-McNeil (22). Si $r(a)$ doit être calculé à plusieurs reprises au cours d'une simulation, notre modèle apporte alors un réel avantage en terme de temps de calcul. Dans une micro-simulation, centrée sur l'individu, il est en effet nécessaire de connaître la probabilité de se marier pour chaque susceptible pour définir les transitions vers le premier mariage.

Enfin, notre modèle n'accepte aucun mariage avant l'âge $a_0$, alors que le modèle de Coale-McNeil en accepte une proportion finie, bien que faible, et même à l'âge 0 ou à des âges négatifs. De plus, le modèle de Coale-McNeil exige une proportion initiale non nulle de population mariée, alors que notre modèle commence avec une proportion nulle. Le modèle original était aussi basé sur le schéma suédois d'entrée en mariage, ce qui garde un caractère arbitraire. Un modèle de mariage qui accepte les mariages à tout âge, même aux âges négatifs, n'est pas totalement satisfaisant pour l'esprit, même s'il marche bien empiriquement. Le fait que le modèle de Coale-McNeil permette de tels mariages, alors que notre modèle ne le permet pas, montre que ce dernier est plus proche des comportements réels.

Notre modèle semble pourvoir s'appliquer à de nombreuses situations similaires, puisque les trois paramètres sont très généraux. Depuis son développement, nous l'avons appliqué sans difficulté à l'entrée en vie sexuelle (premiers rapports), et à l'entrée en vie féconde (première naissance). On pourra aussi l'appliquer à d'autres processus similaires, tels que l'entrée en vie active (premier emploi), la sortie du nid familial (première migration adulte), etc. Il pourrait donc s'appliquer à de nombreux processus en démographie, en sociologie, en économie, en épidémiologie, voire dans d'autres domaines.

La forme de la fonction de Weibull choisie est simple, et correspond au paramètre $k = 2$, c'est-à-dire à la distribution de Rayleigh. On peut aussi imaginer des formes différentes pour le taux de recrutement, avec des asymétries différentes en prenant des paramètres différents pour la fonction de Weibull, pour ajuster d'autres phénomènes, tels que ceux qui se produisent en fin de vie (fin d'activité professionnelle, dernière migration etc.).

REMERCIEMENTS

Alan Matthews souhaite remercier pour son soutien financier les organismes suivants : « South African Centre for Epidemiological Modelling and Analysis » (SACEMA) ; « International AIDS Society » ; « HIV Research Trust ». Il souhaite également remercier l'Institut Pasteur de Paris pour son séjour en accueil, et l'université du Kwazulu-Natal en Afrique du Sud pour avoir facilité cet échange. Les trois auteurs remercient également le programme d'échange scientifique « Franco-Sud Africain » (projet PROTEA), pour son appui financier.

BIBLIOGRAPHIE

Abramowitz M., Stegun I. A., *Handbook of mathematical functions*. Washington, DC: National Bureau of Standards, Applied Mathematics Series – 55, US Government Printing Office, 1964.

Central Statistical Office [Zambia]. (2003). Census 2001. Central Board of Health, and ORC Macro. *Zambia Demographic and Health Survey 2001-2002.* Calverton, Maryland, USA: Central Statistical Office, Central Board of Health, and ORC, 2003.

Coale A. J., "Age Patterns of Marriage". *Population Studies,* 1971, 25 (2): 193 – 214.

Coale A. J. and McNeil D. R., "The Distribution by Age of the Frequency of First Marriage in Female Cohort". *Journal of the American Statistical Association,* 1972, 67: 743 – 749.

Ewbank D., *An Examination of Several Applications of the Standard Pattern of Age at First Marriage.* Unpublished Ph.D. dissertation, Princeton University, Dept. of Economics, 1974

Feeney, G., "A Model for the Age Distribution of First Marriage". Working Paper No. 23, East-West Population Institute.

Hernes G., "The Process of Entry into First Marriage". *American Sociological Review,* 1972, 37: 173 – 182.

Kaneko R., "Demographic analysis of First Marriage Process". *The Journal of Population Problems*, 1991, 47 (3): 3 – 27.

Kaneko R., "Elaboration of the Coale-McNeil Nuptiality Model as the Generalized Log Gamma Distribution: A New Identity and Empirical Enhancements". *Demographic Research,* 2003, 9 (10): 221 – 262.

Liang Z., *The Coale-McNeil model. Theory, generalisation and application.* Amsterdam, The Netherlands: Thela thesis, 2000.

Press W.H., Teukolsky S. A., Vetterling W. T., and Flannery. B. P., *Numerical Recipes, second edition.* Cambridge University Press, 2002.

Rodriguez G. and Trussell J., *Maximum Likelihood Estimation of the Parameters of Coale's Model Nuptuality Schedule from Survey Data*. In: World Fertility Survey, Technical Bulletins, No. 7, 1980.

Weibull W., "A statistical distribution function of wide applicability". *J. Appl. Mech.-Trans. ASME,* 1951,18(3): 293-297.