

# Fermi statistics method applied to model macroscopic demographic data

Giuseppe Alberti  
Independent Researcher

**ABSTRACT:** *The paper presents a recursive function able to mimic demographic mortality curves. Unlike many and more common demographic mortality models, the model presented here is not based on empirical assumptions or time series fitting methods. The model function depends only from one parameter that has a precise meaning in a cellular automaton model. This automaton is also described. For the function definition, the Fermi statistics method of calculation has been used, resulting in similarities with known statistical distribution curves. A continuous representation of the recursive equations is also provided. Implications with demographic mortality rate curves and more general life cycle concepts are outlined. A correlation with a recent study using a multi-omics approach is pointed out.*

**KEY WORDS:** *Demographic mortality, recursive algorithm, cellular automata, Diophantine equations, differential equations, mechanical statistics, demographic Life Tables, mortality rate, lifespan*

-----  
Date of Submission: 14-10-2024 Date of acceptance: 30-10-2024  
-----

## I. INTRODUCTION

A great deal of study has been done on the topic of demographic mortality particularly at advanced ages. Accurate estimates of life span prediction have obvious social and economic motivations. Many references were made in these studies to concepts coming from various scientific disciplines, such as references to reliability theory, computer theories, biological models, entropy theory, and many others. Ref. [1-6]. The present study aims to bring a point of view that -in the author’s opinion- has some new aspects. The original features of the work are the use of statistical mechanics methods and the presence of only one parameter in the model equations. The basic finding comes from the study of a cellular automaton model, when a recurring equation is found as follows:

$$m(TC, r) = (2^r - \sum_{t=1}^{r-1} 2^{r-t} m(TC, t)) / (1 + 2^{-r}(TC - 2)); m(TC, 1) = 0; r \geq 2; \dots\dots\dots(1)$$

where the TC (“Total Cases”) symbol is a parameter, and *r* represents the recursion step. If we associate the above *m(TC, r)* function with some mortality statistical data vs time intervals (addressed by the *r* step) and if we assume the total cases sample as TC, we found the Fig.1 with TC=100 000 . Note that the TC value corresponds to the area under both curves. The “m” curve is generated by the (1) function, while the other curve corresponds to the mortality table data available at the ISTAT web site (Ref. [7]). for e.g. Italy 2019 when the *r* steps correspond to a five years interval count and the vertical axes represents the deaths in the *r* interval for a total sample of 100 000 people. The two curves show a quite evident similarity, having a peak close feature and in particular a similar descending slope. The major difference is in the leading part where the curves show different slopes. It must be noted that the “m” curve is not a fitting attempt on the ISTAT curve but it is just an independent deterministic computation of the (1) function when TC=100 000 and the *r* step runs the positive integer range.

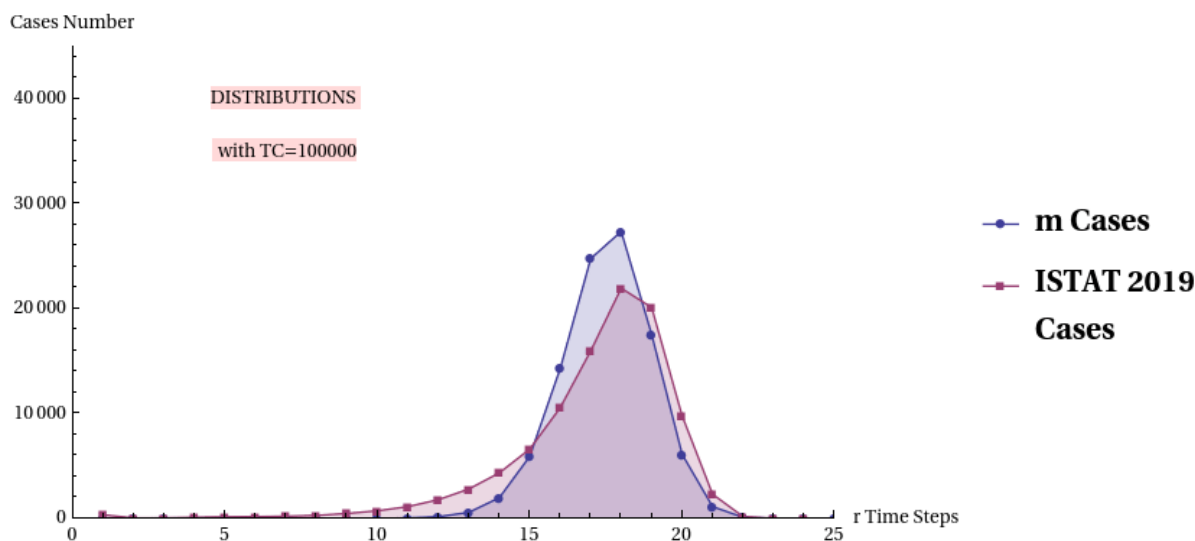


Fig. 1 - m function & ISTAT 2019 mortality distribution

The next section will introduce the cellular automaton at the origin of the above recursive equation. The third section will discuss the mathematics of the object, and the subsequent sections will present a Fermi’s statistics method application leading to the proposed recursive algorithm. A comparison with demographic curve data will then be provided. To further explore the features and capabilities of the model, an additional section will introduce continuous equations instead of recursive finite difference equations. Finally, the features of the model will be discussed leading to possible explanations of the shape of the real demographic curves.

## II. THE ARBITRARY OSCILLATOR

To build the equation (1) we need to introduce a particular cellular automaton. Consider the Fig. 2.

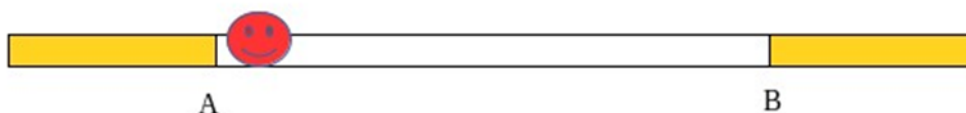


Fig.2 – Arbitrary oscillator

In the figure a “bug” can move on a horizontal x axis exclusively between two stable positions points: A and B. The decision to move (or to stay) is free for the “bug”, therefore not predictable. In the boundary positions the “insect” encounters the outside world that can determine a “safe” condition (food, reproduction, ...) or a “deadly” condition (predators, danger, disease, ...). These occurrences are also not predictable. At each step of time the “bug” decides what to do: stay/move (say “0” for stay and ”1” for move). A “path-space” for this Arbitrary Oscillator (ArbO) can be envisaged. It includes a space and time coordinate and a “d” path index, collecting all the possible decisions sequence. Fig. 3 shows the scheme. The figure depicts a sequence of four time steps with binary decisions tagged “0” or ”1” and amounting in 16 total possible paths (or “orbits”).

The sequence starts at a common initial spatial position. Note, at the sides of the figure, the “extreme” decision possibilities: a) {1,1,1,1} -> ever change or b) {0,0,0,0}-> stay still.

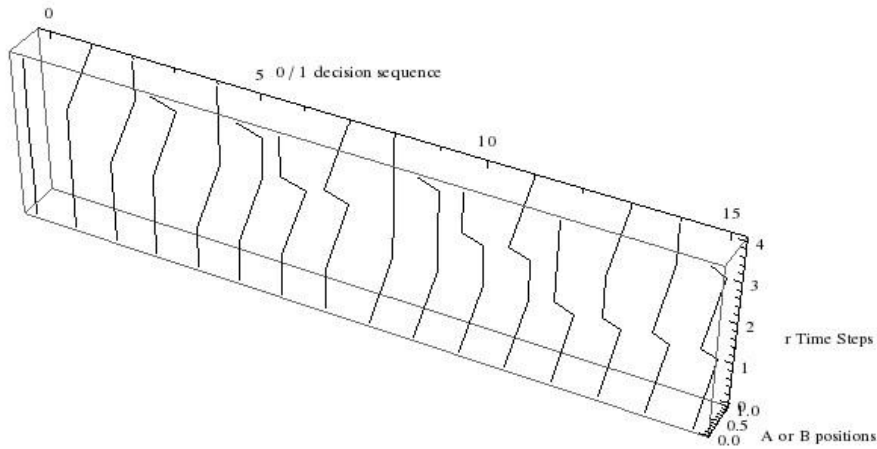


Fig. 3 - path-space of the Arbitrary Oscillator for 4 steps of time decisions

The same situation is shown in Fig. 4 where the spatial coordinate is omitted and a ‘cellular’ structure is given. At each decision step, the rectangular “cell” splits into two subsequent underlying cells. By convention, the left new cell can represent a “no move” or “0” decision, while the right new cell results from a “move” or “1” decision. The total final possibilities are 16 different sequences classified by the “d” index shown at the bottom of the last group of cells. The d index label can follow any coding preferred rule. The cycle starts at an initial zero step, common to all subsequent paths, where the next decision is drawn, resulting in the two different cells of the first step and so on.

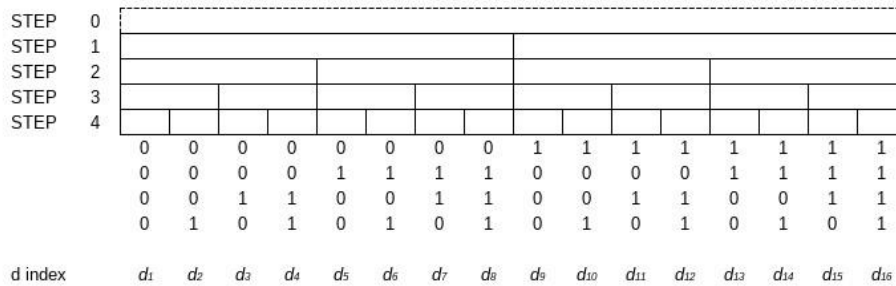


Fig. 4 - Possible choices sequences

In the following Fig. 5, we introduce -for our ArbO- the risk of a End of Life (EoL) event. In this case, the crossed cells mean the occurring of a EoL event. Of course in this case no subsequent cells can be expected and the final “d” sequence is truncated. The EoL events are randomly distributed and can arise at any available decision cell, one per cell, on one or many cells in the level.

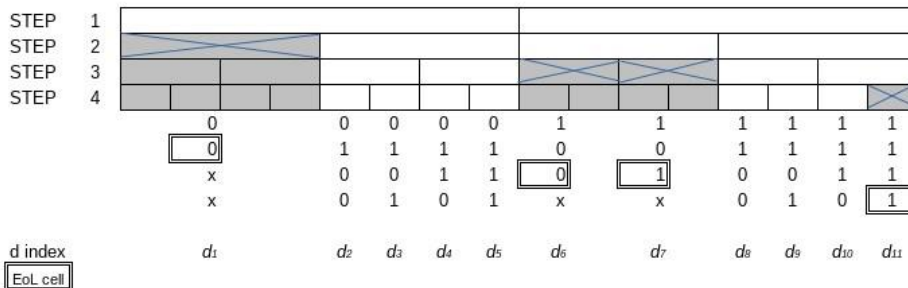


Fig. 5 - EoL events

With the EoL event possibility, an end of the full “life cycle” of the ArbO can be considered as per Fig. 6. Here the four step case is again presented. In the figure all the “path-space” is covered by sequences terminating with and EoL event. Note that, for the four step case in this example, we find a five ( $d_1, d_2, \dots, d_5$ ) different path history.

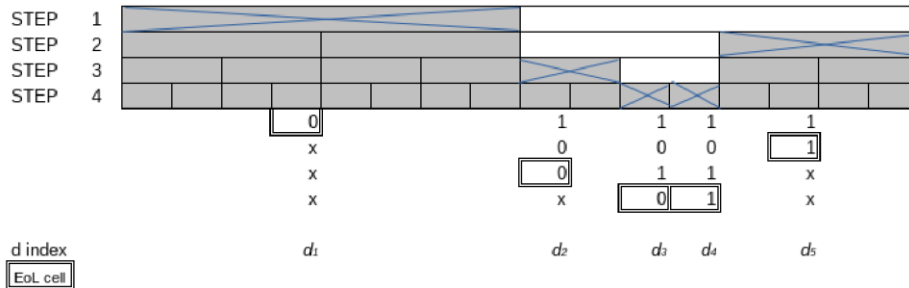


Fig. 6- A fully ended life cycle of the arbitrary oscillator

### III. THE ARBITRARY OSCILLATOR EQUATIONS

The above processes can be mathematically described assuming that:

- (i) The EoL events cover only one cell per event. Their occurrence is not predictable, in number and position, for any time step level and over the available level cells.
- (ii) The number of possible ArbO decision sequence paths is limited to a max Total Cases (TC).

The (i) statement formalizes the ArbO behaviour described in the previous Section 2. The (ii) statement means simply that the step sequence cannot be “perpetual” i.e. unlimited. Consider indeed that after 116 steps (without EoL events) the number of possible sequences in the path space becomes about  $10^{35}$ . In this case -supposing L a fixed length in the d space dimension-the mean distance between the paths will be  $L \cdot 10^{-35}$ . With (i) and (ii) assumptions, we can then define the following formulas and conditions:

$$m_r + v_r = 2 v_{r-1}; 0 \leq m_r, v_r \leq 2^r \dots\dots\dots (2)$$

$$m_r, v_r \text{ Integers}; v_0 = 1, r \geq 1; \dots\dots\dots (3)$$

We define  $m_r$  as the number of EoL events at the step r and  $v_r$  the remaining “safe” decision cells at the same step level. Referring e.g. to Fig. 5 at step 2 level we have  $m_2 = 1$  and  $v_2 = 3$ . The factor of 2, in the Eq. (2), comes from the “binary” decision rule. Solving the recursion equation (2) vs the m variables, we find:

$$v_r = 2^r - \sum_{t=1}^r 2^{r-t} m_t \dots\dots\dots (4)$$

From the statement (ii) it follows that any sequence of decisions must end with an EoL event. The total number of these terminated paths must be TC. Then:

$$\sum_{r=1}^{Rmax} m_r = TC \dots\dots\dots (5)$$

$$Rmax = TC - 1 \dots\dots\dots (6)$$

The presence of the  $v_r$  variables in the above formulas can be eliminated if we consider the “final” level of the life cycle of our ArbO: there, it must be  $v_{Rmax} = 0$ . Then factoring the term  $2^r$  and renaming t index as r index the eq. (4) leads to the (7) equations, with equivalent expressions:

$$\sum_{r=1}^{Rmax} 2^{-r} m_r = 1 \dots\dots\dots (7)$$

$$\sum_{r=1}^{Rmax} 2^{Rmax-r} m_r = 2^{Rmax} \dots\dots\dots (7.1)$$

The eq. (5), (6) and (7) build up a system of diophantine equations that we call “ $S^{TC}$  systems” (see Ref. [8]). The number of solutions  $\{m_r\}$  for this system varies with TC and quickly grows exponentially. No explicit solution formula of an  $S^{TC}$  system is known to the author. By computer analysis one can retrieve the solutions

for TC values of some tenths, in a reasonable machine time. For example with TC = 26, the number of solutions is computed as 565168. With TC = 5 and Rmax = 4 the S<sup>5</sup>system equations and solutions are (last solution as per Fig. 6):

$$8m_1 + 4m_2 + 2m_3 + m_4 = 16$$

$$m_1 + m_2 + m_3 + m_4 = 5$$

$$\{ \{m_1 \rightarrow 0, m_2 \rightarrow 3, m_3 \rightarrow 2, m_4 \rightarrow 0, v_0 \rightarrow 1, v_1 \rightarrow 2, v_2 \rightarrow 1, v_3 \rightarrow 0, v_4 \rightarrow 0\},$$

$$\{m_1 \rightarrow 1, m_2 \rightarrow 0, m_3 \rightarrow 4, m_4 \rightarrow 0, v_0 \rightarrow 1, v_1 \rightarrow 1, v_2 \rightarrow 2, v_3 \rightarrow 0, v_4 \rightarrow 0\},$$

$$\{m_1 \rightarrow 1, m_2 \rightarrow 1, m_3 \rightarrow 1, m_4 \rightarrow 2, v_0 \rightarrow 1, v_1 \rightarrow 1, v_2 \rightarrow 1, v_3 \rightarrow 1, v_4 \rightarrow 0\} \}$$

Note that, given a set of solutions  $\{m_r\}$ , the corresponding  $\{v_r\}$  set is obtained by eq. (4). Finally it can be shown that, defining:

$$Q_r = m_r + v_r \dots \dots \dots (8)$$

the following formulas hold:

$$\sum_{r=1}^{Rmax} v_r = TC - 2 \dots \dots \dots (9)$$

$$\sum_{r=1}^{Rmax} Q_r = 2(TC - 1) \dots \dots \dots (10)$$

Note that the  $Q_r$  set can be viewed as the set of the ArbO “critical decisions” whose result can be a safe or “deadly” end.

#### IV. THE APPLICATION OF THE FERMI STATISTICS METHOD

There are some formal analogies between the arbitrary oscillator model above described and the quantum models studied in the last century by E. Fermi to derive its statistics for ideal gas of molecules and electrons. In both cases we deal with each other indistinguishable objects, that can be arbitrarily allocated into available empty positions over certain predefined levels and imposing only one item per position. In both cases, any possible allocation must comply with some “boundary” conditions. In the next rows we follow the approach of the great scientist even using the same notations and rationale. This Fermi method is well described in his book “Molecules, Crystals and Quantum Statistics”, Ref. [9]. The rationale is as follows:

- we seek for the most probable solution of a S<sup>TC</sup> system
- to achieve this, we consider the number of ways to distribute some  $m$  events over the available  $Q$  cells at each step level
- the most probable solution will have the maximum number of the above said ways, say  $\Pi$
- we look for this maximum, looking at the maximum of  $\text{Log}(\Pi)$  -that is the same thing- using the method of Lagrange multipliers
- finally, we consider the boundary conditions to fix the unknown parameters.

Considering our ArbO model, the above said  $\Pi$  value will be:

$$\Pi = \binom{Q_1}{m_1} \binom{Q_2}{m_2} \dots \binom{Q_r}{m_r} \dots$$

where the  $m_r$  and, accordingly, the  $Q_r$  are solutions of a S<sup>TC</sup> system and the symbol  $\binom{Q}{m}$  represents the binomial coefficient. If we pass to the natural logarithms  $\text{Log}()$  function and using the Stirling approximation formula, we will have:

$$\text{Log}(\Pi) = \sum_{r=1}^{Rmax} (Q_r - \text{Log}(Q_r) - m_r \text{Log}(m_r) - (Q_r - m_r) \text{Log}(Q_r - m_r))$$

we now consider the eq. (4), where we extract the  $m_r$  term from the  $\sum$  symbol, obtaining:

$$v_r = 2^r - m_r - \sum_{t=1}^{r-1} 2^{r-t} m_t \dots \dots \dots (11)$$

now, remembering the (8), we have:

$$Q_r = 2^r - \sum_{t=1}^{r-1} 2^{r-t} m_t \dots \dots \dots (12)$$

We see that, in the above eq., the  $Q_r$  variable is independent from  $m_r$ , thing that is reasonable since  $Q_r = 2v_{r-1}$ . Now we can move on to the search of the maximum of the  $\text{Log}(\prod)$  searching the maximum of the expression:  $\text{Log}(\prod) - a C1 - b C2$ , where a, b are undetermined constant coefficients (according to Lagrange multipliers method), and C1, C2 are the first member of eq. (5) and (7), also constant in total value. We search then for the null condition of the derivative vs  $m_r$  of the expression :

$$\text{Log}(\prod) - a \sum_{r=1}^{Rmax} m_r - b \sum_{r=1}^{Rmax} 2^{-r} m_r \dots \dots \dots (13)$$

After some mathematical steps, we find -as the null condition- the expression:

$$m_r = \frac{2^r - \sum_{t=1}^{r-1} 2^{r-t} m_t}{1 + e^{a+2^{-r}b}} \dots \dots \dots (14)$$

Recalling the (12), we have:

$$m_r = \frac{Q_r}{1 + e^{a+2^{-r}b}}$$

To find the values of a and b coefficients, we sum over r and, imposing the boundary conditions (5), (9) and (10), we obtain, after some algebra, the following expression leading to the same eq. (1) presented in the first section:

$$m_r = \frac{2^r - \sum_{t=1}^{r-1} 2^{r-t} m_t}{2^{-r}(\text{TC}-2)+1}$$

If we introduce the parameter  $rF = \log(\text{TC}-2)/\log(2)$ , we obtain the following set of equations that represent - in a way- the most “probable” solution set  $\{\mathbf{m}_r, \mathbf{v}_r, \mathbf{Q}_r\}$  of a diophantine  $S^{\text{TC}}$  system. We use here the bold character to distinguish this particular solution set to the other possible solutions.

$$\begin{aligned} rF &= \frac{\text{Log}(\text{TC}-2)}{\text{Log}(2)} \\ \mathbf{m}_r &= \frac{2^r - \sum_{t=1}^{r-1} 2^{r-t} m_t}{2^{rF-r} + 1} \\ \mathbf{v}_r &= 2^r - \mathbf{m}_r - \sum_{t=1}^{r-1} 2^{r-t} m_t \\ \mathbf{Q}_r &= 2^r - \sum_{t=1}^{r-1} 2^{r-t} m_t \end{aligned} \quad (15)$$

### V. PROPERTIES OF THE MOST PROBABLE SOLUTION

From the eq.s (15) we see that:

$$m_r/Q_r = \frac{1}{1 + 2^{rF-r}} \dots \dots \dots (16)$$

The form of the (16) leads to a Cumulative Logistic Distribution shape. This is shown in the following Fig. 7 for  $\text{TC}=100\,000$ , where the vertical line intercepts the curve at 0.5 level and the  $r$  axis at  $rF$  point, with  $rF=16.6096$ .

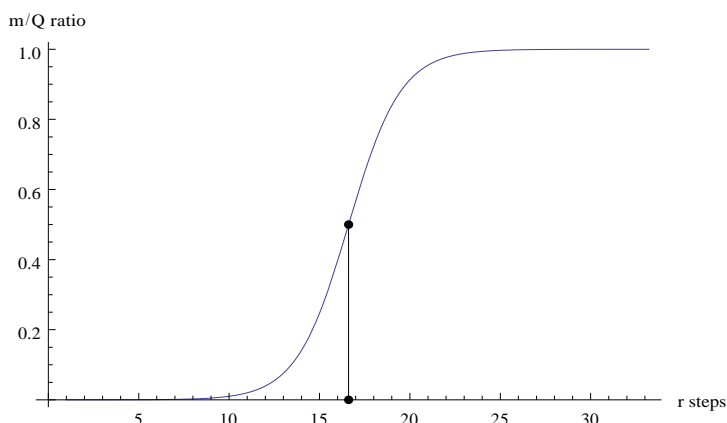


Fig. 7 – The  $\frac{m_r}{Q_r}$  Logistic curve shape for TC=100 000 with rF point shown

The above curve says that, with 100 000 decisions cases, the most probable  $\{m_r\}$  solution of the  $S^{TC}$  system will have the initial  $m$  values at zero, then, quite soon around  $rF$ , the  $m$  values become similar to the  $v$  values and -finally- all the  $m$  (EoL) values will saturate the available  $Q$  levels, meaning a life cycle end for the cellular automaton. It is seen that, even if  $R_{max}=100\ 000-1$ , the most probable solution set will involve a number of non-zero variables limited to few tenths. This is indeed explained by the logarithmic dependence of  $rF$  vs  $TC$  (see eq. (15)). It is noteworthy to consider that solutions will exist with about 100 000  $m$  non zero variables, e.g. like the banal solution  $\{m_1 = 1, m_2 = 1, \dots, m_{99998} = 1, m_{99999} = 2\}$  but these will have far less probability to arise. Let's now consider the following Table 1, where, for the  $TC=100\ 000$  case, the values of the  $m, v, Q$  (rounded to integer value) are presented together with the mortality figures for years 1974 and 2019 in Italy (plus 2019 q<sub>x</sub> data, Ref. [7]). The calculation for the  $m, v, Q$  is done by computer, via iteration across the  $r$  steps, according to eq.(15).

Years interval	r	m	v	Q	ISTAT - 2019 1000 x q <sub>x</sub>	ISTAT-2019 dx	ISTAT-1974 dx
Up to 4 years	1	0	2	2	3.3453	335	2735
5-9	2	0	4	4	0.36563	36	180
10-14	3	0	8	8	0.43792	44	170
15-19	4	0	16	16	0.99425	99	334
20-24	5	0	32	32	1.429	142	385
25-29	6	0	64	64	1.63059	162	368
30-34	7	0	128	128	1.9993	198	505
35-39	8	1	255	255	2.86143	283	686
40-44	9	3	507	509	4.69355	463	1128
45-49	10	10	1003	1014	7.33727	721	1863
50-54	11	40	1966	2007	11.46247	1118	2799
55-59	12	155	3778	3933	18.46051	1780	4470
60-64	13	572	6984	7556	29.60728	2801	6265
65-69	14	1966	12002	13968	47.28093	4341	9089
70-74	15	5924	18079	24003	75.57279	6611	12542
75-79	16	14315	21843	36158	130.66618	10566	16581
80-84	17	24780	18906	43686	227.36915	15984	17535
85-89	18	27370	10441	37811	404.2388	21956	13839
90-94	19	17537	3345	20882	621.66989	20117	6806
95-99	20	6107	582	6690	798.55747	9776	1591
100-104	21	1112	53	1165	931.49213	2297	127
105-109	22	104	2	106	986.30673	167	2
110-114	23	5	0	5	998.26306	2	0
115-119	24	0	0	0	999.85919	0	0
Total		100001	100000	200002		99999	100000

Table 1.  $m, v, Q$  values with ISTAT 1974 & 2019 data for 100000 Total Cases

The Tab.1 data for  $m$ ,  $v$ ,  $Q$  are graphically shown in Fig. 8 here under, utilizing a suitable computer interpolation algorithm to generate continuous lines. The dashed line start at  $rF$  and intercepts -as expected- the crossing of  $m$  and  $v$  curves. It can be demonstrated that the peaks of the three curves conserve their relative allocation around  $rF$  value and between themselves, for any  $rF$  value.

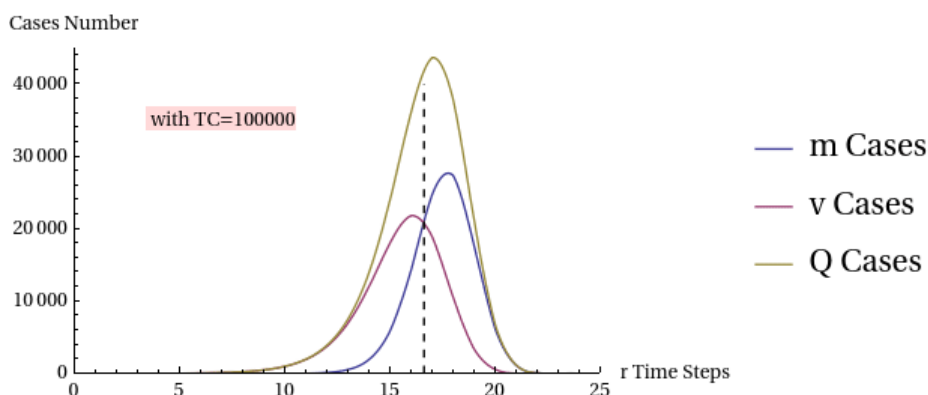


Fig. 8 Interpolated  $m$ ,  $v$ ,  $Q$  curves, with  $rF$  dashed line

### VI.COMPARISON WITH DEMOGRAPHIC DATA

From Table 1 and Fig. 9, we see that the  $m$  data have the same peak interval and trailing slope of the ISTAT dx 2019 data. It is also evident that the infant and mid-age mortality components are not present in the  $m$  model. The “improvement” in the life span between years from 1974 and 2019 is also remarkable.

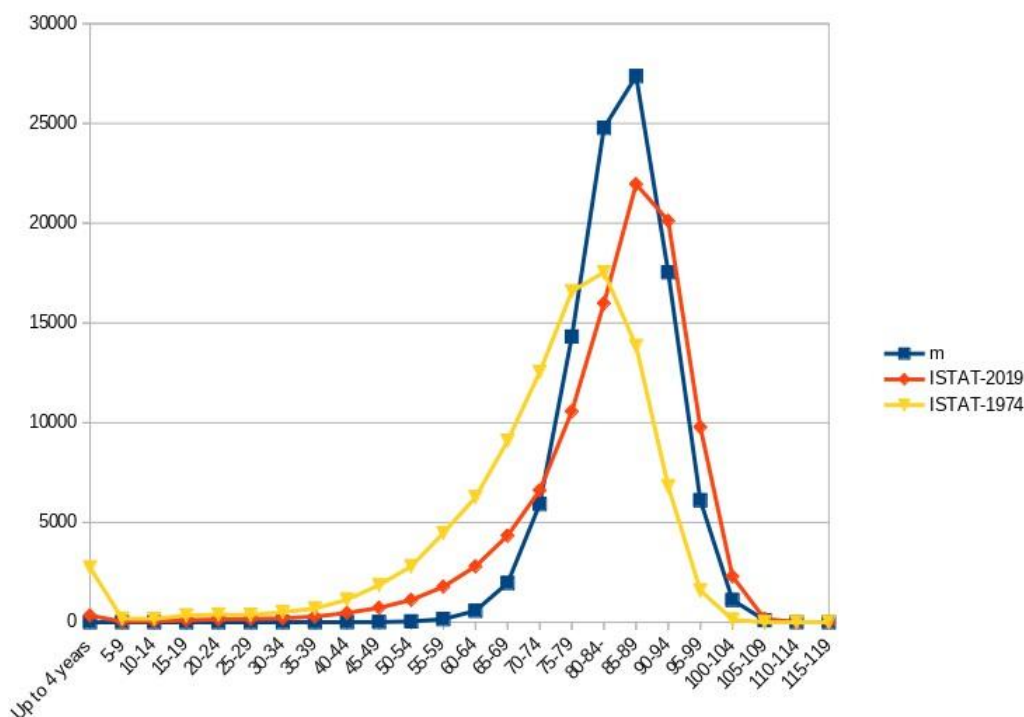
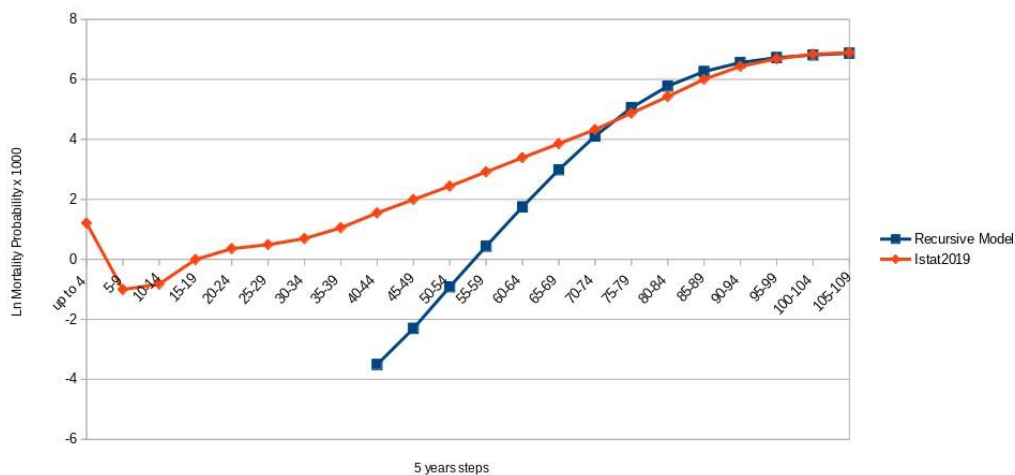


Fig. 9 Graphic comparison of  $m$ , dx ISTAT74, dx ISTAT19 curves vs five years intervals

To appreciate the meaning of the above comparisons, we must recall that the  $m$  curve in Tab.1 represents the most probable single solution of the  $S^{100000}$  system. In other words - for a single ArbO subject- a plenty of EoL events can happen over the overall available  $Q$  events, accordingly with the various possible "space paths". If we now consider a big number of samples of 100 000 ArbO objects, the various (most probable) alternatives in the life cycle paths will statistically materialize over the plotted  $m$  curve. Therefore this  $m$  curve becomes comparable with the standard mortality curves detected by the demographic survey relevant to the same normalized sample number of elements. This comparison is applicable if a particular size of age interval is chosen, namely the five years interval, corresponding to the unitary  $r$  step of the recursive model.





**Fig. 10 Log mortality probability given by recursive model and ISTAT2019 data**

From the Table 1, another Fig. 10 can be derived, comparing the mortality probability (x 1000 on the log. scale ) derived from the ArbO recursive algorithm with the values measured with ISTAT2019. The mortality probability ( $q_x$ ) is defined here as the ratio between the number of deaths in a time interval and the surviving elements entering at the beginning of the interval and it is usually defined as  $q_x$  in demographic Life Tables, Ref. [3][7]. The Fig. 10 shows a linear shape of the log. curves in most part of the curve range (Ref. [10-11]). At initial and middle ages -however- a strong difference is evident (thing that is also evident on above Fig. 9 for mortality). Moreover, we see a common flattening (for both model and ISTAT data) of the curves slope at the oldest ages .

## VII. THE CONTINUOUS MODEL EQUATIONS

### 7.1 Introducing the continuous functions

The above (15) equations are finite difference equations, that require -for a numerical evaluation- a recursive computation. We want to transform these in explicit expression of the  $m$ ,  $v$ ,  $Q$  variables as functions of time and TC parameter. This can be useful for a better model mathematical handling and simulation. We look therefore for the differential equations possibly associated to the (15) equations and their solution. In doing this we introduce the  $m(t)$ ,  $v(t)$ ,  $Q(t)$  functions of a continuous time  $t$  variable such that:

$$m_r = m(r); v_r = v(r); Q_r = Q(r);$$

The above conditions imply that the continuous curves  $m$ ,  $v$ ,  $Q$  pass through the discrete points  $(r, m_r)$ ,  $(r, v_r)$ ,  $(r, Q_r)$  respectively, individuated by the recursive equations (15) (\*). We keep the bold face of the function symbol meaning that we are again handling the particular (most probable) solutions set of (15) equations. Since the ArbO model is essentially a discrete binary model, the above defined continuous functions do not have a “physical” meaning (like e.g. number of choice cells, etc) but they can have a statistic meaning if we imagine a plenty of ArbO objects starting at the same time and evolving independently. In this case the  $m$ ,  $v$ ,  $Q$  curves can quantify the mean number of the corresponding random choices and events along a sliding time interval. Considering the equation (2), we know that, in a continuous  $t$  domain and with  $m(t) = 0$ , the equation becomes equal to  $v' = kv$ , giving the solution  $v(t) = 2^t$  with  $k = \text{Log}(2)$  where  $\text{Log}()$  represents the natural logarithm function. Extending the analysis including  $m(t) \neq 0$ , we consider the differential equation:

$$v' = k(v - m) - g m'; g = \text{const} > 0$$

This equation still comes to (2) if  $m(t) = 0$ . For the general case  $m(t) \geq 0$ , we try some possible expressions of the  $g$  constant looking for a possible best fitting solution vs the points calculated with our recursive model. We find that a possible good fitting is reached with a value such that (*Assumption (iii)*):

$$g = (1 - k)$$

giving the differential equation:

$$v' = k(v - m) - (1 - k)m' \dots \dots \dots (17)$$

(\*) Note: the constraint assumed for the continuous curve implies a possible shift (about 2.5 years) to the right of the simulated curve with respect to the real curve. This is not critical for our analysis, as we are looking for structural effects (not influenced by an offset component) and not for an absolute age effect.

**7.2 Solving the equations**

Equation (17) involves two unknown time functions  $v(t)$  and  $m(t)$ . Recalling formula (8) and formula (16), which has no recursive form, we can extend this formula to the continuous domain. We thus obtain a dependence between  $v$  and  $m$  such that:

$$v(t) = 2^{rF-t}m(t) \dots \dots \dots (18)$$

Using such a formula, we can substitute the  $v$  terms in the (17) diff. equation thus obtaining a diff. equation in only one unknown function  $m(t)$ . To avoid excessive reader burden, we jump directly to the computed solutions of the combined (17) and (18) equations. For the solution computation we define a boundary condition, in consistency with the eq. (7), such as:

$$\int_0^\infty m(t)2^{-t} dt = 1$$

These computations are performed with the aid of a well known commercial mathematical computer application. These solutions have the following form:

$$\begin{aligned} m(t) &= 2^{2t - \frac{(1+2g)\text{Log}(2^{rF} + 2^t g)}{gk}} c1 \\ v(t) &= 2^{rF-t}m(t) \\ Q(t) &= m(t) + v(t) \end{aligned} \quad (19)$$

The  $g, k$  terms are numbers while  $c1$  and  $rF$  are constants depending only from the TC parameter:

$$\begin{aligned} g &= (1 - k), k = \text{Log}(2) \\ c1 &= (1 + g)(2^{rF} + g)^{(1+1/g)} k \\ rF &= \frac{\text{Log}(TC-2)}{k} \end{aligned} \quad (20)$$

With these assumption the (19) and (20) equations can be used to characterize the particular ArbO model presented in the Sect. 4 and 5, giving  $m, v, Q$  as functions only of time and of the TC parameter.

**7.3 The continuous  $qx(t)$  mortality probability definition**

In the case of the discrete ArbO model we compute the  $qx_r$  probability values with the same method of the standard demographic tables, Ref. [3]-[7]:

$$qx_r = m_r / (TC - \sum_{j=1}^r m_j + m_r)$$

that means to compare the current  $r$  interval mortality figure  $m_r$  with the survived subjects entering at the beginning of the  $r$  interval, i.e. the denominator value in the above formula. In our case the  $m_r$  figures are given by the ArbO discrete algorithm, while for the standard demographic Life Tables one uses the  $dx$  values sampled in the demographic survey. Note again that our  $qx_r$  calculation is an independent process from demographic data, to which it can be compared assuming a same demographic total sample value and a suitable  $r$  scale such

that  $x$  [years] =  $5 r$ . With the above considerations in mind and the 7.2 subsection equations, we can also define by similarity the  $qx(t)$  continuous function as:

$$qx(t) = \frac{m(t)}{N(t)}$$

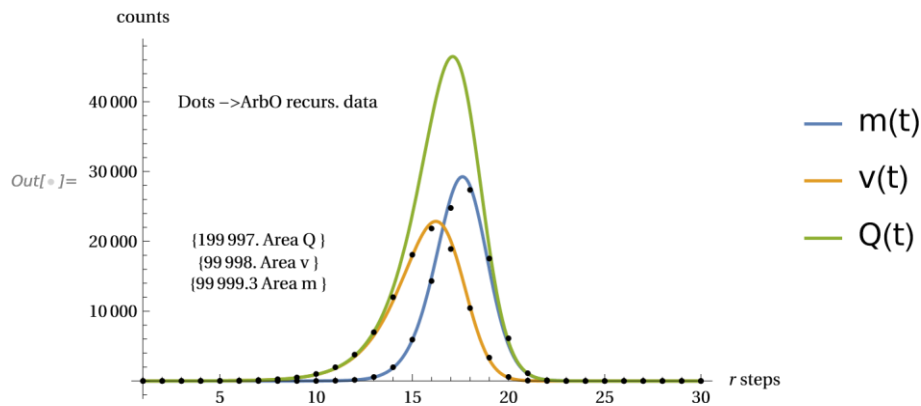
where  $N(t)$  is the survival subjects number at instant  $t$  (*Assumption (iv)*):

$$N(t) = TC - \int_0^t m(\tau) d\tau + ggm(t) \dots \dots \dots (21)$$

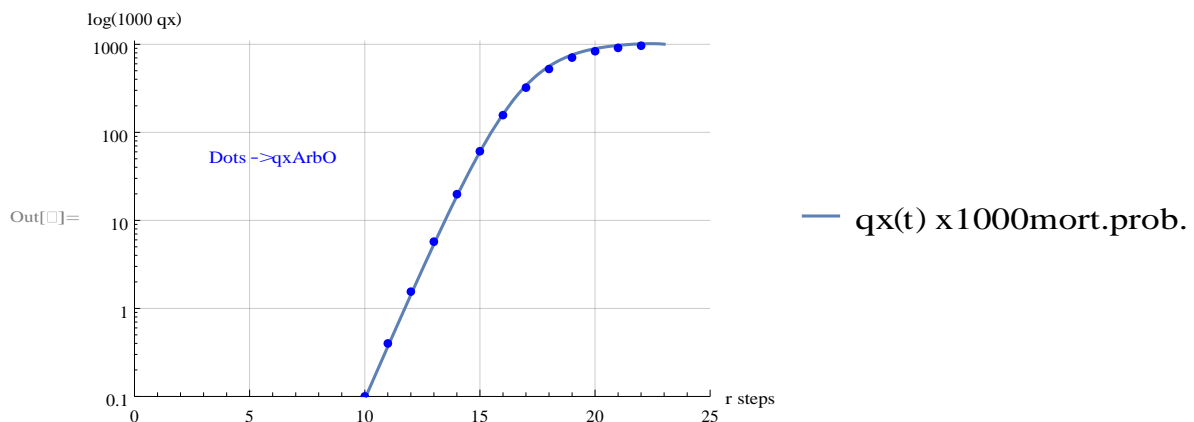
This equation comes from both the above discrete definition of  $qx$  and the introduction of a "trimming" term with a  $gg$  ( $\neq g$ ) constant. The trimming term is due to the need to compensate the numeric precision when passing from the 'Sum' function to the 'Integral' function. The  $gg$  value will result, in practical calculations, as  $gg = 0.5$ . In Ref. [8] an explicit form of  $N(t)$  is given. This form has been obtained solving the integral in (21) using the (19) formulas.

**7.4 The simulated continuous curves**

With the above foundations, we show some graphics (see Fig.11 and subs.) with the defined  $m(t)$ ,  $v(t)$ ,  $Q(t)$ ,  $qx(t)$  functions over-imposed to points computed with the recursive algorithm based on (15) formulas, when a reference TC value of 100 000 is assumed. The  $t$  values run continuously along the  $r$  integers steps range. The dots coincide, except for minor deviations, with the  $m(t), v(t), Q(t), qx(t)$  curves. We see that the interpolation shapes of the numerical data given in Tab. 1 with Fig. 8 are quite confirmed, and also that the area under the curves results to be close to the expected sizes (eq. (5) & (9)). The "dotting" of Q curve is omitted to avoid graphic burden to the Figures.



**Fig. 11 Comparison between the continuous curves  $m,v,Q$  and the recursive ArbO discrete data, TC=100000**



**Fig. 12 Comparison between the continuous  $qx$  curve and the ArbO discrete data, TC=100 000**

We see also from Fig. 12 that the  $qx(t)$  simulated curve fits well with the computed ArbO model  $qx$  dots (same blue dots on the Fig. 10). The shape of the theoretical  $qx(t)$  confirms also a flattening at high ages.

### 7.5 The $m, v, Q, qx$ curves features vs different TC parameter

Until now we considered a reference fixed value of  $TC = 100\ 000$ . It could be interesting to look at the effects on the curves of a variable  $TC$ . We found that the curves confirm the same shape and aspect, as per Fig. 13 and Fig. 14, with e.g.  $TC = 10\ 000$  and  $TC = 1000\ 000$  respectively. We also see that the curves slide along the x-axis due to their different  $rF$  value (located under the crossing of  $m$  and  $v$  curve) and that these have accordingly different areas as per the  $TC$  specific value. A similar analysis can be done for the theoretical  $qx(t)$  magnitudes resulting in the Fig. 15. These show a bending at advanced ages of the  $qx$  probabilities curves. The wide span between the curves is due to the three order of magnitudes considered for the  $TC$ s. All these figures show theoretical -continuous & discrete- data coming from the ArbO model.

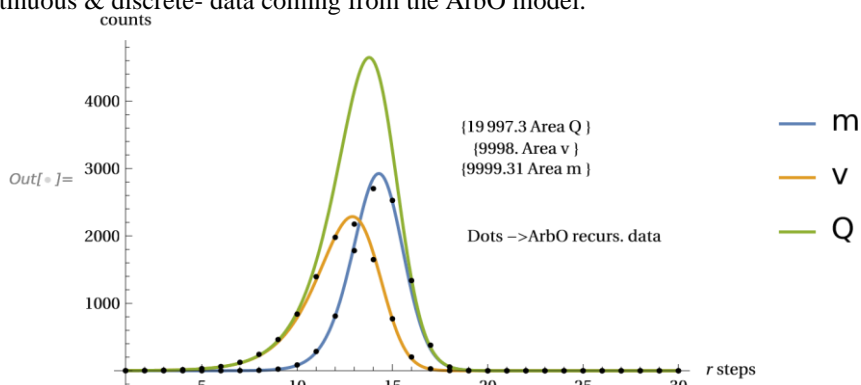


Fig. 13 Comparison between the continuous curves  $m, v, Q$  and the recursive ArbO discrete data,  $TC=10000$

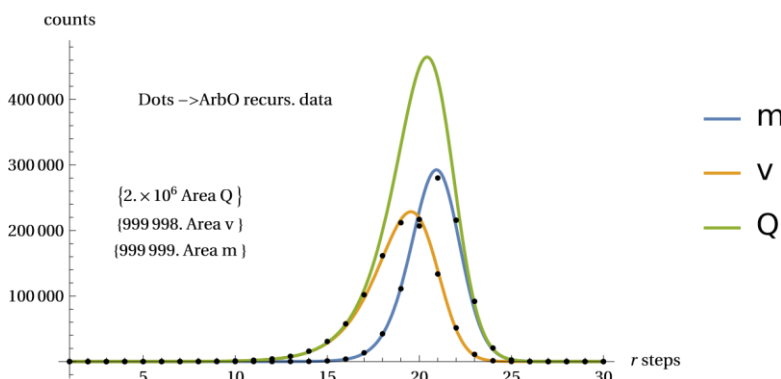


Fig. 14 Comparison between the continuous curves  $m, v, Q$  and the recursive ArbO discrete data,  $TC=1000000$

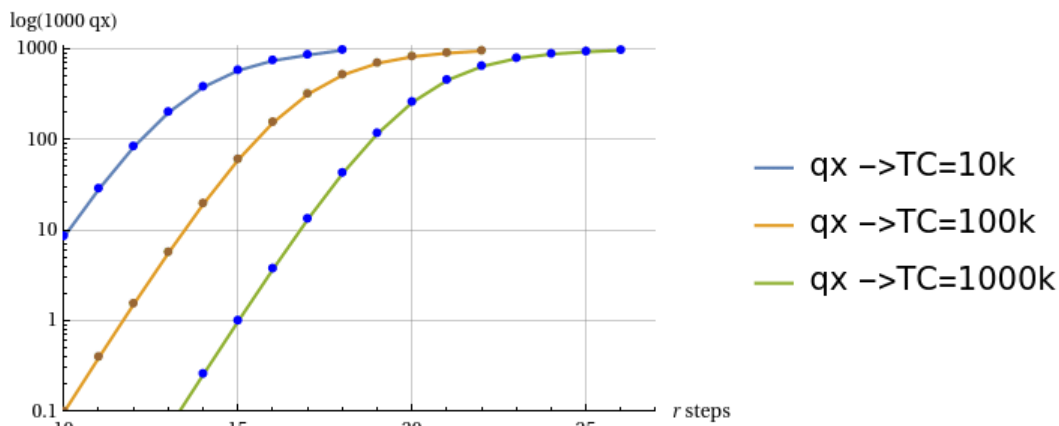


Fig. 15 Theoretical  $qx(t)$  values computed with the ArbO model for different  $TC$  parameter values

It is also interesting to consider the question of how the "peaks" of these curves can change along different TC values. This is done calculating the curves null derivative position on the  $r$  step axis (giving a fractional real  $r$  value). The result is shown in Fig.16, where the positions of the curves peaks (presented in years units) are plotted vs TC range. It is seen that the relative peaks and  $rF$  level positions distances result constant. The figure shows also the logarithmic shape of the bundle since  $rF = \text{Log}(TC-2)/\text{Log}(2)$ . The Fig.16 includes also a line called "lifespan" defining conventionally the position of the  $r$  value when the  $m$  curve reach the 10% (at the "older ages" right side) of the peak value. It happens that this "delta" value to the  $m$  peak is also constant resulting in about 2.77  $r$  intervals (i.e. about 13.8 years from the max. mortality peak position).

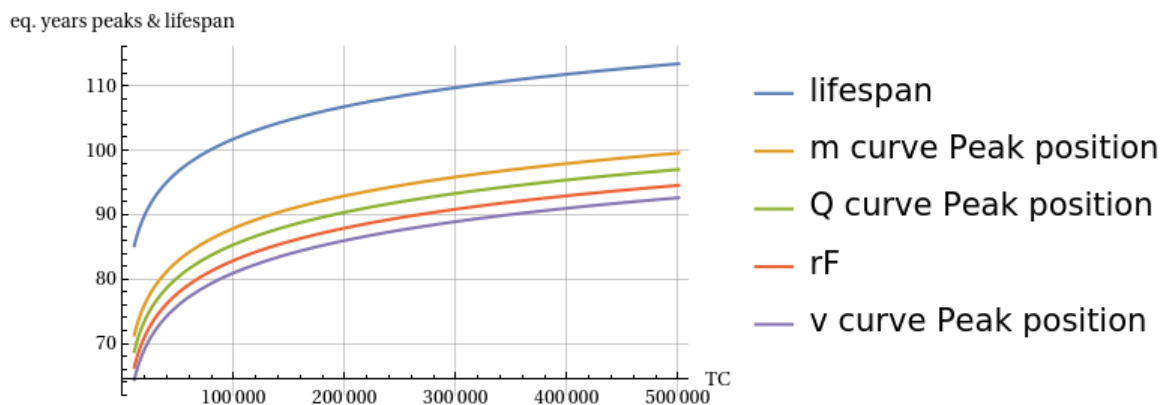


Fig. 16 Age peak positions for the continuous curves  $m, v, Q$  vs TC values

### 7.6 A possible interpretation of the real demographic mortality curves

With the above theoretical foundations, if we look to Tab. 1 and Fig. 9, we see that the demographic mortality tables show two different peaks for ISTAT1974 and ISTAT2019 data respectively. Now, considering the ISTAT2019 case, we assume that the actual data curve is composed of the mixture of two groups of subjects with different TC values coexisting in the demographic survey. Figures 17 and 18 show the total combined effect and the decomposition into individual curves to be summed to obtain the total effect. For this exercise we consider two TC values groups with different percentage weights on the total normalized area of 100 000 events. These  $w_j$  weights can be defined as under follows, where the  $p_j$  numbers represent the share portions of the groups "labeled" as  $TC_j$  that constitute the overall total T of subjects; these  $w_j$  values therefore will "normalize" the total sample under analysis to 100 000 cases:

$$\sum_{j=1}^n p_j TC_j = T; \quad w_j = (100\,000 p_j) / T; \quad \sum_{j=1}^n w_j TC_j = 100\,000$$

thus the "mixed" mortality profile will be given, in the case of two components, by:

$$mmix(t) = w_1 m(t, TC1) + w_2 m(t, TC2)$$

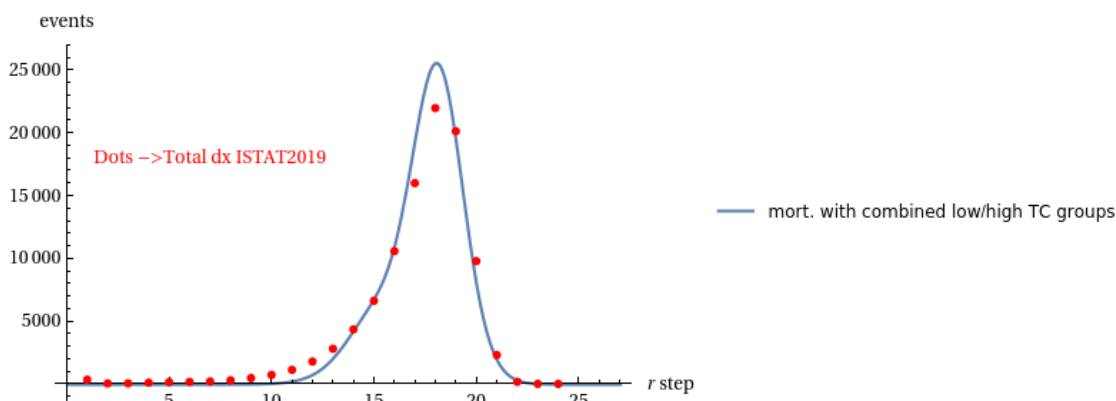
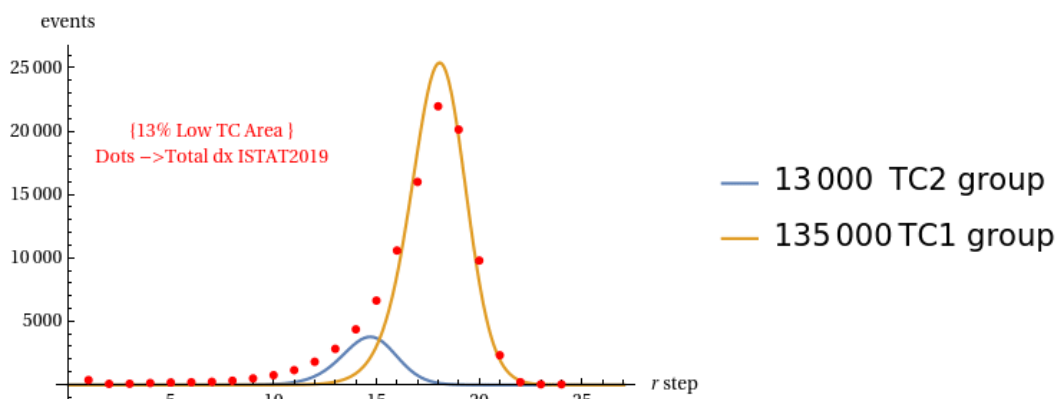
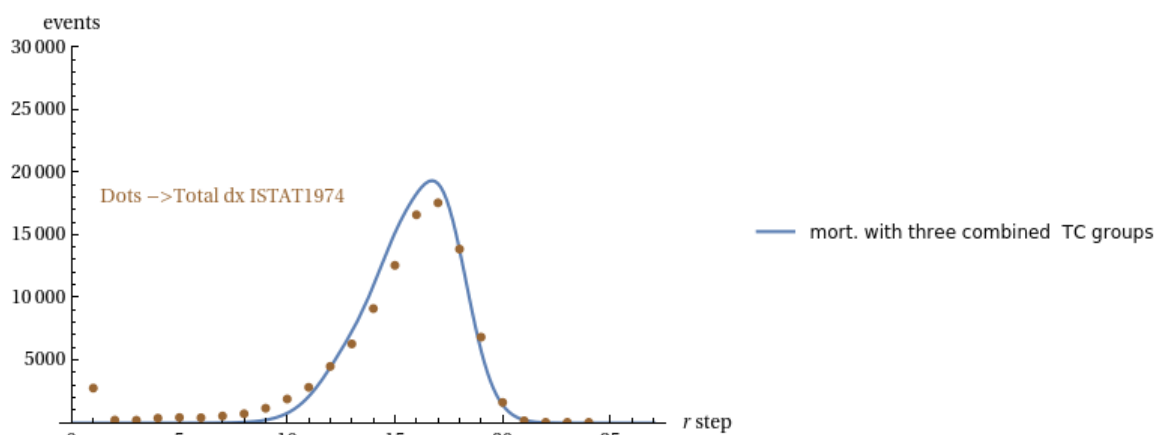


Fig. 17 ISTAT2019 mortality data over a curve combined by  $m(t)$  sub-groups with different TC

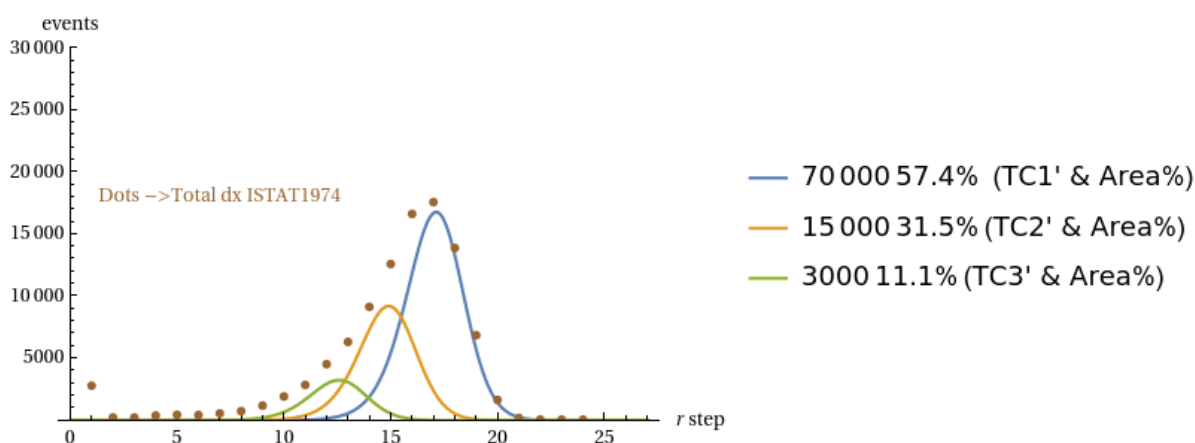


**Fig. 18** ISTAT2019 mortality data and the two  $m(t)$  sub-groups curves to be summed to give the combined total curve

The same approach is applied to the more elongated curve of ISTAT1974. With the aid of three combined curves with different TC values and groups weights we give the Fig. 19 and 20.



**Fig. 19** ISTAT1974 mortality data over a curve combined by  $m(t)$  sub-groups with different TC

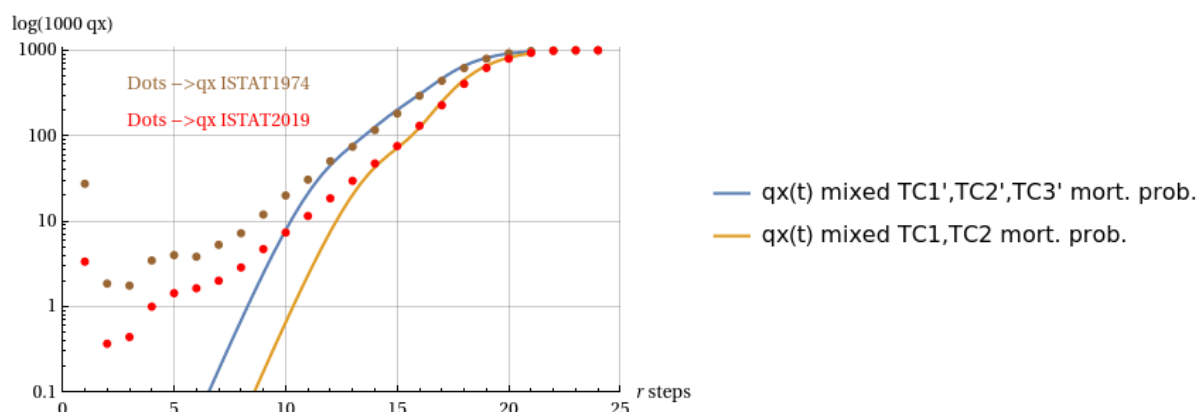


**Fig. 20** ISTAT1974 mortality data and the three component  $m(t)$  sub-groups curves to be summed to give the combined total curve

From the previous simulation tests we see that one possible explanation for the shape of the real mortality demographic curves is the overlap of mortality from different component groups. The same as above considerations can be extended to the  $qx$  probability as result of a mix of sub-groups components. For this task we define the mixed  $qx$  curve when in presence of e.g. three major components, as:

$$qx_{mix}(t) = \frac{w_1 m(t, TC1) + w_2 m(t, TC2) + w_3 m(t, TC3)}{w_1 N(t, TC1) + w_2 N(t, TC2) + w_3 N(t, TC3)}$$

The formula is proposed for the three components case of Fig. 19 (ISTAT1974 data) using the same “weights”  $w_j$ . The  $m(t,TC)$  and  $N(t,TC)$  functions are those obtained before with (19), (20), (21) applying the relevant TC parameter. The result is shown in Fig. 21 hereunder, where we found a good fitting of the curves from ages higher then about 55 years ( $r = 11$ ). In the Fig. 21 we also added the result of the same exercise for ISTAT2019 data with a  $qx_{mix}(t)$  function calculated as above but with only two components (as in Fig .17 situation).



**Fig. 21 ISTAT1974 and 2019 mortality probability compared with  $qx(t)$  mixed curves coming from combining three groups or two groups respectively (same quotas as in Fig.20 and 18)**

The above hypothesis of a multi-component structure of the curves derived from the  $dx$  and  $qx$  data in the demographic Life Tables was studied in more detail by the author in the work presented in Ref. [12].

## VIII. FINAL COMMENTS AND FUTURE ANALYSIS DEVELOPMENT

### 8.1 General aspects of the model

The present work considers and mathematically describes an abstract object (the Arbitrary Oscillator - ArbO) that generates events that evolve over time and may also include instances of "death." To find the most probable distribution of such events, the Fermi mechanical statistics is used, obtaining a recursive computational algorithm that provides the solution as dependent on a single parameter TC. By means of scaling, the mortality predictions generated by the algorithm can be compared with actual demographic mortality data. This comparison shows similarities of the theoretical curves generated by the algorithm with the shape of the real data.

### 8.2 Specific aspects of the model

- The mortality pattern that emerged from the study does not appear to have an absolute fixed limitation to life span; this limit is set if the TC parameter is fixed but, as the TC parameter changes, the limit of maximum life expectancy shifts with logarithmic law (Fig. 16)
- Increasing the TC parameter lengthens the life span limit accordingly, but the critical phase narrows in percentage terms. This is due to the constant time distance (~14 years) expected between peak mortality and the end point of lifespan: for example, if the peak is at 80 years, one can hope to still live to 94 years (17% more); if the peak is at 100 years, one can hope to still live to 114 years (14% more).
- Flattening of the curve in the probability of death in old age is a matter of debate as to whether it is real or an effect of measurement methods and the sparsity and incompleteness of data (Ref. [11]). In the case of the proposed model, flattening is confirmed theoretically by the evolution of both recursive data and continuous simulation curves (Fig. 15).
- Actual mortality curves may be due to a mixture of components (sub-groups of subjects) with different TCs, these mixing of groups can explain the real curves shapes at high ages for both mortality and mortality probability (Fig.17, 19, 21).
- If we look also to the evolution of the ISTAT curves, from past to today (Fig. 9), we see that the spread in the mortality peak is sharpening becoming closest to a  $m(t)$  curve. The model would seem indeed to predict

that in more advanced social situations the possible different TCs groups tend to approach a single final TC value, while more backward social situations may be explained by a broader distribution of TCs groups.

### **8.3 Open points and possible future developments**

When we associate an abstract object such as ArbO with real living subjects we can assume direct analogies with e.g. mortality events but less clear, in the analogy, is the meaning of the TC parameter and also of the  $\nu$  and  $Q$  variables. For the latter two variables we can give interpretations such as : number of “critical” decisions/events in a time interval ( $Q$  variable) and number of “safe-ending” decisions/events in the same interval ( $\nu$  variable). For the TC parameter (“Total Cases”), there is no clear reference to the reality addressed in the analogy. In the ArbO model, the TC parameter limits the maximum number of paths in the “path space.” Limit reached which, all possible events must lead to the end of the life cycle. In sociological terms, our TC could represent a generic social target parameter such as ‘maximum of longevity’, limiting the lifespan, and whose improvement corresponds to a similar improvement in life expectancy. In scientific terms, however, in the analogy with living subjects, one would have to find a correlation with a quantity similar to TC, if one exists. A further possible extension of our model could be considered in the case of growth by splitting of biological or physical objects. In this case the model can be adapted to that of a system growing by splitting of components in the presence of a random risk of death and an absolute limit on the number of elements to be met (TC). On the latter aspect, i.e., population growth and the patterns that can govern it, references such as Ref.[12-15] can be considered and studied in correlation with future developments of our model.

## **IX. DEVELOPMENTS SINCE THE FIRST PUBLICATION OF THIS WORK**

This article was prepared in May 2022 and has had several minor revisions at later times. Starting from the ArbO model presented here, the author subsequently developed a conjecture about the trend in demographic mortality at older ages. This conjecture was originally detailed in Ref. [16] and also led to a further study of the components of demographic mortality (developing an insight discussed in subsection 7.6 of this paper). Finally, in a further paper, the author hypothesised the presence of absolute limits on mortality curves in demographic life tables. These last two papers are cited in Ref. [17-18] respectively.

### **9.1 Correlation with a recent study using a multi-omics approach**

In a paper published in August 2024 in the journal Nature Aging (Ref. [19]), researchers highlighted a non-linear disease risk that accumulates with the biological ageing process. “The analysis revealed consistent nonlinear patterns in molecular markers of aging, with substantial dysregulation occurring at two major periods occurring at approximately 44 years and 60 years of chronological age” (cit. from [19]). This study involved a sample of US persons aged 75 years and under. In our studies (see above subsection 7.6 and Ref. [17]), we showed the possible existence of discrete components of demographic mortality. These two approaches - conducted in completely different ways - show a correlation of results.

## **REFERENCES**

### Author Note on References:

*The literature on demographic mortality is enormous and covers a broad spectrum of scientific and social aspects. It is beyond the scope and capacity of this article to attempt to review and make reference with these studies. We will therefore only refer to works strictly instrumental to the contents of this work, leaving the correlation with other studies on the topic to the interest and experience of the reader. This choice can also be justified - in the author's opinion - by the fact that the approach illustrated here is original and not directly related to other works and research models*

- [1] L. A. Gavrilov and N. S. Gavrilova, “The Biology of Life Span: a Quantitative Approach” Center on Aging NORC and The University of Chicago, Illinois, USA, slide presentation available at: “ <https://slidetodoc.com/>”
- [2] D. Makowiec, D. Stauffer and M.Zieliński, “Gompertz law in simple computer model of ageing of biological population”,<http://arxiv.org/abs/cond-mat/0107107v1>
- [3] A. Racco, M. Argollo de Menezes and T.J. Penna “Search for an unitary mortality law through a theoretical model for biological ageing”, <http://arxiv.org/abs/adap-org/9709002v1>
- [4] M. D. Pascariu, A. Lenart & V.Canudas-Romo (2019) “The maximum entropy mortality model: forecasting mortality using statistical moments”, Scandinavian Actuarial Journal, 2019:8, 661-685, DOI: 10.1080/03461238.2019.1596974
- [5] A. Boulougari, K.Lundengård, M. Rančić, S.Silvestrov, S. Suleiman & B. Strass (2019) “Application of a power-exponential function-based model to mortality rates forecasting”, Communications in Statistics: Case Studies, Data Analysis and Applications, 5:1, 3-10, DOI: 10.1080/23737484.2019.1578705
- [6] S.J. Clark “A General Age-Specific Mortality Model With an Example Indexed by Child Mortality or Both Child and Adult Mortality”,Demography. 2019 June ; 56(3): 1131–1159. doi:10.1007/s13524-019-00785-3.



- [7] ISTAT Web site ,<http://dati.istat.it>
- [8] SQU Systems, <https://squ-systems.eu> , Web site managed by the author and devoted to diophantine STC equation systems
- [9] E. Fermi, “Molecules, Crystals and Quantum Statistics”, W.A.Benjamin,1966, pag. 264 and subseq.
- [10] P. Richmond and B. M. Roehner “Predictive implications of Gompertz’s law.”, <http://arxiv.org/abs/1509.07271v1>
- [11] L. A. Gavrilov and N. S. Gavrilova, “Mortality Measurement at Advanced Ages: a Study of the Social Security Administration Death Master File”, North American Actuarial Journal. 15 (3): 432–447. doi:10.1080/10920277.2011.10597629. PMC 3269912.PMID 22308064.
- [12] Nicolas Bacaer, *Histoires des mathématiques et de populations* ed. Cassini, Paris, 2008.
- [13] S. Méléard, “Modèles aléatoires en Ecologie et Evolution”, Springer
- [14] K., Henderson| M. Loreau “An ecological theory of changing human population dynamics”, *People and Nature*. 2019;1:31–43.DOI: 10.1002/pan3.8
- [15] M. Ausloos, “Gompertz and Verhulst frameworks for growth AND decay description”, arXiv:1109.1269v2
- [16] G. Alberti “A conjecture on demographic mortality at high ages.” <https://arxiv.org/abs/2307.10279>
- [17] G. Alberti “More on the mortality conjecture: the components of demographic mortality” <https://doi.org/10.13140/rg.2.2.14304.46086>
- [18] G. Alberti "On two asymptotic limits for demographic mortality Life Tables data" <https://doi.org/10.13140/rg.2.2.27428.86405>
- [19] Xiaotao Shen, Chuchu Wang, Xin Zhou, Wenyu Zhou, Daniel Hornburg, Si Wu & Michael P. Snyder “Nonlinear dynamics of multi-omics profiles during human aging ”, *Nature Aging*, <http://dx.doi.org/10.1038/s43587-024-00692-2>