Individual-based models for stage structured populations: formulation of development equations

Giuseppe Buffoni
Via del Canaletto 152, 19126 La Spezia - Italy
e-mail: giuseppe.buffoni@santateresa.enea.it

Sara Pasquali
CNR-IMATI
Via Bassini 15, 20133 Milano - Italy
e-mail: sara@mi.imati.cnr.it

Abstract

Individual-based models for the growth dynamics of a population have to formalize mathematically the basic processes defining the life history of an individual of the population. In particular, the process of development, from birth to death, of a generic individual. The model equations for this process are stochastic equations and describe the time evolution of the status of an individual, in terms of a physiological age. In this paper we are addressed to the formulation of development models, when regression effects on the status of an individual are forbidden; this is the case when the physiological age is defined in terms of an abstract indicator, i.e. the percentage of development. Different stochastic models of the development process are presented, and their responses are compared. Their behaviours are analyzed by varying the stochasticity level, which takes into account the degree of intraspecific variability. A multi-stage system, representing a copepod population, is used to illustrate the results of the analysis, by means of numerical simulations. Furthermore, some computational aspects, related to the implementation of an individual-based model, are illustrated: choice of time step, roundoff errors, averages over the realizations.

Keywords: Individual Based Models, stage structured populations, stochastic development equations, “no regression” models.

1 Introduction

The individual-based models (IBMs) track all the individuals in a population. We quote from Grimm et al. (1999): “Individual-based modelling refers in the following to simulation models that treat individuals as unique and discrete entities which have at least one property in addition to age that changes during the life cycle, e.g. weight, rank in a social hierarchy, etc.”. In this modelling approach, the dynamics of the overall population is obtained by performing numerical simulations of the life histories of the individuals of the initial population, and those of the recruitment yielded over time. The status of an individual (the individual state variable) is individuated by a physiological age (Metz and Diekmann, 1986; Curry and Feldman, 1987;
Munholland and Dennis, 1992) which can be defined in terms of biometric descriptors (such as a characteristic length, a weight) or by an abstract indicator of the maturity of an individual, i.e. the percentage of development. The model equations describe the time evolution of the status of an individual, i.e. its life history, which is assumed completely determined by the main biological processes of development (growth of an individual), reproduction, mortality, and in case by some individual characteristics concerning the specific problem under study. Individual-based modelling is a Lagrangian approach to modelling. Some IBMs are mainly addressed to spatial dynamic problems, neglecting the development of the individuals, and then the changes of population abundance (Gómez-Mourelo, 2005; Goodwin et al., 2006). On the other side, some IBMs simulate both spatial and growth dynamics (Choi et al., 2006; Nehrbass and Winkler, 2007), and the effects of environment temporal variability on population dynamics (Mazaris and Matsinos, 2006).

On the rise, features, identification of peculiarities, differences with Eulerian models based on either differential or discrete balance equations, and applications of IBMs one can refer to the book of DeAngelis and Gross (1992), and to the papers, dealing with and about IBMs, by Huston et al. (1988), DeAngelis et al. (1990, 1994), Judson (1994), Grimm (1999), Buffoni and Pasquali (2003, 2007).

IBMs are applied to populations characterized either by a continuous-size structure, such as fish and coral populations (DeAngelis et al., 1979; DeAngelis et al., 1990; Kirby et al., 2004), or by a discontinuous-stage structure, such as insect and crustacean copepod populations (Batchelder et al., 2002; Zadereev et al., 2003; Buffoni et al., 2004; Mazzocchi et al., 2006; Buffoni et al., 2007; Gilioli and Pasquali, 2007).

In the applications, when the system shows strong heterogeneity at the scale of individuals (spatial, physiological, behavioural heterogeneity), the implementation of IBMs allows to take into account for details relevant to the dynamics of the system (Bolker et al., 1997; Grimm, 1999): the Lagrangian approach overcomes the Eulerian approach (Buffoni and Pasquali, 2003). Stochastic Lagrangian methods are used in transport and diffusion problems of particles in both solid and fluid media (Cashwell and Everett, 1959; Haji-Sheikh and Sparrow, 1967; Thomson, 1987; Buffoni et al., 1997; Graham and Moyeed, 2002). For example, the computation of the neutron distribution function in highly heterogeneous nuclear reactors is based on the random-walk method. Since successive events are regulated by accidents or by chance, the method was referred to as Monte Carlo method, originated by Ulam and Von Neumann (Metropolis and Ulam, 1949; Cashwell and Everett, 1959).

In this paper we assume that the individuals belong to a stage-structured population, where the stages are defined by sharp biological events (eggs eclosion, moult, adult emergence, beginning and end of oviposition, death). Moreover, the individual state variable, i.e. the physiological age of an individual considered as a stochastic process, is defined as the percentage of development in a stage (Lee et al., 1976; Curry and Feldman, 1987; Munholland and Dennis, 1992). We address our analysis to the processes of development of an individual: the main objective is the formulation of the development equations of an individual and their dynamics responses, especially when linked together with mortality and production processes. Different formulations can be found in literature. In some IBMs the development of an individual is described in terms of deterministic (differential) equations (DeAngelis et al., 1979; Batchelder et al., 2002). Here we follow a computational approach in the formulation of the process equations. Thus, the development process is described in terms of stochastic difference equations, rather than stochastic differential equations. The development of an individual is regarded as an accumulation of small increments of physiological age over time. These increments are given by the
contributions of a deterministic term, due to the stage-specific mean development rate, and of a stochastic term, due to the variability of development time among the individuals. The rates of the biological processes, introduced in the model equations, are assumed dependent on the environmental variables, mainly on temperature.

We start with development processes which allow regression of the physiological age (e.g., when the process is described in terms of particular biometric descriptors, such as a weight); they are essentially diffusion processes with drift. Then, we will introduce development processes in which regression episodes are not allowable (e.g., when the process is described in terms of the percentage of development). We present different formulations for this class of development processes. The behaviours of three development schemes have been analyzed, in particular by varying the stochasticity level, which takes into account for the degree of intraspecific variability. First we compare the development schemes in a single stage system, and then in a multi-stage system, relative to a copepod population. The behaviours of the development schemes are analyzed and illustrated in some idealized situations, in order to separate the effects of the various processes: development, mortality and production. Some remarks on the computational aspects of IBMs can be found in the last section of the paper.

2 The history of an individual

To introduce the argument and for future reference, the steps identifying the history of an individual are briefly illustrated (Buffoni et al., 2004; Buffoni et al., 2007). Assume $M$ stages in the life of an individual: egg stage, and a number of juvenile and adult stages. For simplicity, only adult females in stage $M$ are assumed reproductive individuals. Let

$$X_i^t = \text{percentage of development at time } t \text{ of an individual in stage } i$$

and let $t_i^{-1}$ be the initial time instant of the stay of an individual in stage $i$, so that $X_i^{t_i-1} = 0$. Here, the development of an individual is regarded as a discrete time stochastic process characterized by a fixed time step $\Delta t$. The steps describing the life history of an individual are listed below.

- At the beginning of the life of an individual, i.e. at time $t^0$, the sex of the individual is determined by means of a choice process based on the sex ratio.
- Then, for $i = 1, 2, ..., M$, we follow the development of an individual through each stage by means of the development equations

$$X_{i+\Delta t}^t = X_i^t + Y_i^t(\Delta t), \quad t > t_i^{-1}, \quad X_i^{t_i-1} = 0, \quad (1)$$

where the increment function $Y_i^t(\Delta t)$ will be critically discussed in the next section. The development process in stage $i$ ends when at a time $t_i$ we register $X_i^{t_i} \geq 1$.
- At each time $t$, the survival or death of an individual in the time interval $(t, t + \Delta t)$ is determined by means of a choice process based on the survival probability in $(t, t + \Delta t)$.
- For an adult female in stage $M$, at each time $t$ the number of eggs produced in the time interval $(t, t + \Delta t)$ is computed on the basis of the fecundity profile.
- At each time $t$, behavioural, hierarchical, ... characteristics of the individuals, which have to be specified by the problem under study, are determined.
Thus, the life history of an individual is depicted by the time evolution of its status, from birth, through stages until death. Some processes are governed by deterministic and stochastic components (development and production processes), while some events are regulated only by chance (the events survival-death in terms of the survival probability, and female-male in terms of the sex ratio of the population). The overall population dynamics is obtained by putting together the information from the life histories of all the individuals. A number of realizations of the population dynamics should be carried out, in order to obtain stable distributions by means of averaging procedures.

3 Development process: formulation of stochastic difference equations

To simplify the notation the stage index \(i\) is omitted in this section. Thus, let now \(X_t\) represents the percentage of development at time \(t\) of an individual in a generic stage. Equation (1) is now written as

\[
X_{t+\Delta t} = X_t + Y_t(\Delta t), \quad t > t_0; \quad X_{t_0} = 0,
\]

(2)

where \(t_0\) is the initial time instant of the stay of an individual in a generic stage. The increment function \(Y_t(\Delta t) = \Delta X_t\) takes into account of both deterministic and stochastic effects on the development process. We may write

\[
Y_t = E(Y_t) + Y'_t,
\]

where the expected value \(E(Y_t)\), assumed positive, represents the deterministic component, and \(Y'_t\), with \(E(Y'_t) = 0\), the stochastic component. Moreover, the random variables \(Y'_t, Y'_s\), for \(t \neq s\), are assumed independent.

Let \(\mu_{X_t}, \mu_{Y_t}, \sigma^2_{X_t}, \sigma^2_{Y_t}\) be the expected values and variances of \(X_t\) and \(Y_t\). Since the development rate depends on the environmental variables, mainly on temperature, \(\mu_{Y_t}\) and \(\sigma^2_{Y_t}\) may depend explicitly on time \(t\). From (2) we have

\[
X_{tn} = \sum_{j=0}^{n-1} Y_{tj},
\]

(3)

and then

\[
\mu_{X_{tn}} = \sum_{j=0}^{n-1} \mu_{Y_{tj}}, \quad \sigma^2_{X_{tn}} = \sum_{j=0}^{n-1} \sigma^2_{Y_{tj}},
\]

(4)

where \(t_j = j\Delta t + t_0\). While the numerical characteristics mean and variance of \(X_t\) can be easily determined from those of \(Y_t\), in general this is not the case for its distribution.

The computation of the time evolution of the individual development in a stage by means of (2) ends when at a time \(t_N = N\Delta t + t_0\) we register \(X_{t_N} \geq 1\) (or when the individual is eliminated due to its death). \(N\) is an integer-valued random variable, and it is said to be a “stopping time” for the sequence \(Y_{t_1}, Y_{t_2}, ...\) in (3) (Ross, 1983, p.59). \(t_N\) is the exit time from the stage of an individual, and the random variable \(t_N - t_0\) is

\[
\tau = t_N - t_0 = N\Delta t = \text{time spent in a stage by an individual},
\]
called the residence time of an individual in a stage. As for $X_t$, in general the distribution of $N$ cannot be determined analytically. However, under the assumption that the $Y_{t_j}$ in (3) are independent and identically distributed random variables having finite expectations

$$E(Y_{t_j}) = \mu_Y < +\infty,$$

(assumption of constant temperature) the Wald’s equation (Ross, 1983, p. 59) holds

$$\mu_{XtN} = E(N) \mu_Y.$$  

The stochastic process $X_t$ is generally described in terms of a diffusion process with a deterministic drift, given by the average development rate. For this type of processes, when the stochastic component is dominant with respect to the deterministic drift, regression effects in the development process can be produced. The outcome $Y_t < 0$, and then $X_{t+\Delta t} < X_t$, for some $t$, can be obtained (Figure 1(a)). Now, we first introduce the basic formalism of the diffusion processes, and how the numerical characteristics $\mu_{Y_t}$ and $\sigma_{Y_t}^2$ are related to experimental data. Then we will perform an analysis of some processes characterized by different choices of the probability distribution of $Y_t$, for which regression effects are not allowed (Figure 1(b)). Moreover, we will compare the outcomes of the models.

![Figure 1: Possible outcomes of $X_{t+\Delta t}$ (circles) in a development process with regression (a) and when regression effects are forbidden (b).](image)

### 3.1 Diffusion process

The development as a diffusion process with a deterministic drift is defined by

$$Y_t = v\Delta t + \sqrt{2k} \Delta W_t,$$

(7)

where $v > 0$ is the drift velocity, $\sqrt{2k}$ ($k \geq 0$) is a measure of the stochasticity level, and $\Delta W_t$ are independent increments of a Wiener process (Gardiner, 1994, p. 69) satisfying

$$E(\Delta W_t) = 0, \quad E((\Delta W_t)^2) = \Delta t.$$  

(8)

For computational purpose, the increments $\Delta W_t$ are expressed as

$$\Delta W_t = \xi_t \sqrt{\Delta t},$$

(9)
with $\xi_t$ standard normal random numbers.

Equation (7) may be written in the form $Y_t = V_t \Delta t$, where $V_t$ is the random variable representing the individual development rate (or the individual growth velocity) given by

$$V_t = \frac{X_{t+\Delta t} - X_t}{\Delta t} = v + \sqrt{2k\Delta W_t}, \quad (10)$$

The precise meaning of the two parameters $v$ and $k$ introduced in (7) follows from (10). In fact, taking into account (8), we have that

$$E(V_t) = v, \quad E((V_t - v)^2) = 2k \Delta t. \quad (11)$$

Thus, $v = \text{average of } V_t = \text{average of the individual development rate}$, while $k = \text{var}(V_t) \cdot \Delta t/2$. Assuming a finite value of the variance of $V_t$, the noise time scale is $\Delta t$, and $k$ becomes infinitesimal as $\Delta t \to 0$. Other formulations of the development process with finite time scale of noise (Buffoni et al., 2007) are suggested by models of particle dispersion in water (Thomson, 1987).

Let $f(s; \mu, \sigma^2)$ denote the normal density of a random variable with mean $\mu$ and variance $\sigma^2$. The density of $Y_t$ defined in (7) is $f(y; \mu_{Y_t}, \sigma_{Y_t}^2)$ where

$$\mu_{Y_t} = v \Delta t, \quad \sigma_{Y_t}^2 = 2k \Delta t. \quad (12)$$

The process $X_t$ defined in (2), with $Y_t$ given by (7), is a normal diffusion process with drift; whose density is $f(x; \mu_{X_t}, \sigma_{X_t}^2)$; the equation for $\mu_{X_t}$ and $\sigma_{X_t}^2$ are

$$\Delta \mu_{X_t} = v \Delta t, \quad \Delta \sigma_{X_t}^2 = 2k \Delta t. \quad (13)$$

In the case of continuous time, the density $\tilde{f}(x,t) = f(x; \mu_{X_t}, \sigma_{X_t}^2)$ satisfies an advection-diffusion equation (Karlin and Taylor, 1981, p. 159) with drift $v$ and diffusion coefficient $k$.

### 3.2 Drift velocity and diffusion coefficient

Up to this point we have summarized the basic formalism of an advection-diffusion process. Now the task is the estimation of the parameters $v$ and $k$ from the available experimental data. Let $D$ and $S^2$ be the average and variance, respectively, of the experimental residence time in a stage of an individual. Again, assume a situation at constant temperature, so that $D$ and $S^2$ are constant.

For a pure deterministic development process, i.e. when $k = 0$ in (7), we should have $Y_t = \mu_{Y_t} = v \Delta t = \Delta t / D$. Thus, with a suitable time step $\Delta t$, the process (7), leads to $\tau = t_N - t_0 = D$ and $X_{t_N} = 1$.

The same choice is generally adopted also for a stochastic development process, i.e. when $k > 0$ in (7) (Curry and Feldman, 1987). Thus we let

$$\mu_Y = \frac{\Delta t}{D} \quad \text{so that from (12)} \quad v = \frac{1}{D}. \quad (14)$$

Since we stop the development process at step $N$ when $X_{t_N} \geq 1$, we have $\mu_{X_{t_N}} \geq 1$, and then (6) implies that $1 \leq E(N) \mu_Y$. Thus, from (14) it follows that for any $k > 0$

$$D \leq E(N) \Delta t = E(\tau). \quad (15)$$
However, it will be shown from the results of numerical simulations in subsection 3.6 and section 4, that for $k$ not sufficiently high $D \simeq E(\tau)$, so that the choice (14) can be adopted. On the other side, for $k$ high we obtain a sharp inequality $D < E(\tau)$; in this situation we would take $v = (1 + \epsilon)/D$, with $\epsilon > 0$, to try to maintain $D \simeq E(\tau)$.

The level of the stochastic component is measured here by the ratio

$$\eta = \frac{S}{D}. \quad (16)$$

Moreover, we assume $\sigma^2_{Y_t}$ proportional to $\eta^2$, and we let

$$\sigma^2_{Y_t} = \frac{\Delta t}{D} \eta^2 \quad \text{so that from (12)} \quad k = \frac{1}{2D} \eta^2. \quad (17)$$

Numerical simulations of development processes in a single stage (subsections 3.5 and 3.6) and in a multi-stage system (section 4) will be performed, under constant environmental conditions (stationary daily cycles of temperature and solar radiation, availability of food, absence of disease and pest), by varying the parameter $\eta$ to illustrate the behaviour of the system “population” versus the degree of intraspecific variability.

### 3.3 Forward dispersion with normal distribution

When regression effects are not allowable in the development process (Figure 1(b)), in a first approach (Buffoni et al., 2004) we can modify equation (7) by assuming

$$Y_t = \max\{0, Z_t\} = \frac{1}{2}(Z_t + |Z_t|), \quad (18)$$

where $Z_t$ is defined as $Y_t$ in (7)

$$Z_t = p\Delta t + \sqrt{2q} \Delta W_t; \quad (19)$$

$p\Delta t$ and $2q\Delta t$ are mean and variance of $Z_t$, respectively. From (18) and (19) it follows that all the negative values assumed by $Z_t$ shrink into the null value of $Y_t$. Thus, the cumulative distribution function of $Y_t$ (Figure 2) is given by

$$P(Y_t \leq y) = 0 \quad \text{for } y < 0,$$

$$P(Y_t = 0) = F(0; p\Delta t, 2q\Delta t), \quad (20)$$

$$P(Y_t \leq y) = F(y; p\Delta t, 2q\Delta t) \quad \text{for } y > 0,$$

where $F(s; \mu, \sigma^2) = \int_{-\infty}^{s} f(s'; \mu, \sigma^2) ds'$.

**Theorem 1** Let $Y_t$ be defined by (18)-(19). Then, its mean and variance, given by

$$\mu_{Y_t} = \int_{0}^{+\infty} s f(s; p\Delta t, 2q\Delta t) ds,$$

$$\sigma^2_{Y_t} = \mu^2_{Y_t} + \int_{0}^{+\infty} s^2 f(s; p\Delta t, 2q\Delta t) ds,$$
Figure 2: Cumulative distribution and density for the forward dispersion with normal distribution. Continuous line: \( p < 0 \); dashed line: \( p > 0 \).

satisfy the following properties

\[
\frac{\partial \mu_{Y_i}}{\partial \Delta t} > 0, \quad \mu_{Y_i} \simeq \sqrt{\frac{q \Delta t}{\pi}} + \frac{1}{2} p \Delta t \quad \text{as} \quad \Delta t \to 0, \quad \mu_{Y_i} \simeq p \Delta t \quad \text{as} \quad \Delta t \to \infty,
\]

(21)

\[
\frac{\partial \sigma_{Y_i}^2}{\partial \Delta t} > 0, \quad \sigma_{Y_i}^2 \simeq q \Delta t \left( \frac{\pi - 1}{\pi} + p \sqrt{\frac{\Delta t}{q \pi}} \right) \quad \text{as} \quad \Delta t \to 0, \quad \sigma_{Y_i}^2 \simeq 2q \Delta t \quad \text{as} \quad \Delta t \to \infty.
\]

(22)

Proof. See Appendix.

The choice \( p = v \) and \( q = k \) implies that \( \mu_{Y_i} \) is greater than \( v \Delta t \) and \( \sigma_{Y_i}^2 \) is lesser than \( 2k \Delta t \), as specified in (12), while we have to choose the parameters \( p \) and \( q \) in such a way that mean and variance of \( \mu_{Y_i} \) equal the values in (12).

Theorem 2 Let

\[
\theta_0 = \frac{v \Delta t}{\sqrt{2k \Delta t}} = v \sqrt{\frac{\Delta t}{2k}},
\]

(23)

where \( v \) and \( k \) are estimated from experiments by means of (14) and (17). Let \( H_1(\theta) \) and \( H_2(\theta) \) be defined as in the proof of Theorem 1 in Appendix. Then, we can assign \( \mu_{Y_i} = v \Delta t \) and \( \sigma_{Y_i}^2 = 2k \Delta t \) as in (12) by taking

\[
p = p^* = \frac{v}{H_1(\theta^*)}, \quad q = q^* = \frac{k}{H_2(\theta^*)},
\]

(24)

where \( \theta^* \) is the unique solution to the equation

\[
\chi(\theta) = \theta_0^2 \quad \text{with} \quad \chi(\theta) = \theta^2 \frac{H_1^2(\theta)}{H_2(\theta)}.
\]

(25)
3.4 Forward dispersion with gamma and beta distributions

An alternative approach is to give up the splitting of the increment $Y_t$ in a deterministic component with positive mean and a stochastic component with zero mean and defined in terms of independent increments of a Wiener process. We may assume directly for $Y_t$ a probability distribution null on $(-\infty, 0)$. A suitable choice, for example, seems to be either gamma distribution $g(y; \alpha, \beta)$ or beta distribution $b(y; \alpha, \beta)$

$$g(y; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta y} y^{\alpha-1} I_{(0,\infty)}(y),$$

$$b(y; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1} I_{(0,1)}(y),$$

where $\alpha > 0$, $\beta > 0$ and $I_J(y)$ is the indicator function of the set $J$. The parameters $\alpha$ and $\beta$ are expressed in terms of the desired mean and variance of $Y_t$. We put as in (12), for gamma distribution,

$$\mu_{Y_t} = \frac{\alpha}{\beta} = v\Delta t, \quad \sigma^2_{Y_t} = \frac{\alpha}{\beta^2} = 2k\Delta t,$$

obtaining

$$\alpha = \frac{v^2\Delta t}{2k}, \quad \beta = \frac{v}{2k}. \quad (29)$$

Similarly, for beta distribution, we put

$$\mu_{Y_t} = \frac{\alpha}{\alpha + \beta} = v\Delta t, \quad \sigma^2_{Y_t} = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} = 2k\Delta t,$$

obtaining

$$\alpha = v\Delta t \left[-1 + \frac{v}{2k}(1-v\Delta t)\right], \quad \beta = (1-v\Delta t) \left[-1 + \frac{v}{2k}(1-v\Delta t)\right]. \quad (31)$$

In the case of gamma distribution, from definition (29), $\alpha$ and $\beta$ are always positive. Moreover, only $\alpha$ depends on $\Delta t$ and it is proportional to $\Delta t$. When $\alpha$ increases, the gamma distribution moves from a decreasing shape (for $\alpha \leq 1$) to an unimodal function (for $\alpha > 1$) with increasing mode $(\alpha - 1)/\beta$. The comparison between cumulative distribution functions of the forward dispersion with normal distribution and the gamma distribution is given in Figure 3.

In the case of a beta distribution, from definition (31), in order to have positive $\alpha$ and $\beta$, the inequality $v\Delta t + 2k/v < 1$ must be necessarily verified. This inequality implies that $2k/v < 1$, i.e. the process is advective dominant, and furthermore the limitation $\Delta t < 1/v$ on the time step must hold. As for the gamma distribution, to have a decreasing distribution we must have that $\alpha/\beta = v\Delta t/(1-v\Delta t) < 1$. Thus, the limitation $\Delta t < 1/(2v)$ on the time step must hold.

Other distributions can be chosen for $Y_t$, as truncated normal, Weibull, Snedecor-Fisher distributions (Johnson et al., 1994, 1995), however we do not consider them here because in all these cases it is not easy to express mean and variance of the distribution in a closed form as function of $v$ and $k$. 

Proof. See Appendix.
Figure 3: Cumulative distributions for the forward dispersion with normal distribution (continuous line) and gamma distribution (dashed line) in the case $\Delta t < 2k/[(\pi - 1)v^2]$ (see formulae (39) and (40)).

### 3.5 Distribution of $X_t$

The distribution of $X_t$ can be theoretically obtained only in special cases. For the diffusion process, outlined in section 3.1, the density of $X_t$ can be derived in the general case of time dependent $v$ and $k$. In fact, from (4) and (12) we have that the density of $X_t$ is $f(x; \bar{v}t, 2\bar{k}t)$, where $t = n\Delta t + t_0$, and $\bar{v}$ and $\bar{k}$ are average values over $(0, t)$ of $v$ and $k$, respectively. When the density of $Y_t$ is the gamma distribution $g(y; \alpha, \beta)$, described in section 3.4, the density of $X_t$ is the gamma distribution $g(x; n\alpha, \beta)$ in the case of time independent $v$ and $k$ (Johnson et al., 1994, p. 349). In case of time dependent $\alpha$, $X_t$ is again gamma distributed with parameters $\sum_{j=0}^{n-1} \alpha_tj$ and $\beta$. When both $\alpha$ and $\beta$ are time dependent, $X_t$ has a distribution (Johnson et al., 1994, p. 384) with a complicated analytical expression that we omit here. The distribution of $X_t$ cannot be theoretically obtained in the other cases, but it can be numerically approximated. In particular, in case of $v$ and $k$ time independent, the central limit theorem assures that the sum $X_t = \sum_{j=0}^{n-1} Y_{tj}$ is normally distributed with mean $vt$ and variance $2kt$ for $n$ sufficiently large. The condition $n$ large is satisfied taking $\Delta t$ sufficiently small. If $v$ and $k$ are time dependent and $n$ is large, $X_t$ is approximately normal with mean $\sum_{j=0}^{n-1} v_{tj} \Delta t$ and variance $2\Delta t \sum_{j=0}^{n-1} k_{tj}$ under suitable conditions on the first three moments of the variables $Y_{tj}$ (Laha and Rohatgi, 1979, p. 280).

In the following we shall denote by

- scheme S1 = normal process (7)
- scheme S2 = forward normal process with $p = p^*$ and $q = q^*$ defined by (24)
- scheme S3 = gamma process with $\alpha$ and $\beta$ given by (29).

All the three schemes have $\mu_{Y_t}$ and $\sigma_{Y_t}^2$ given by (12). Numerical simulations have been performed also with the forward normal process taking $p = v$ and $q = k$. This scheme is not calibrated as the schemes S1, S2, S3, having $\mu_{Y_t} > v\Delta t$ and $\sigma_{Y_t}^2 < 2k\Delta t$. It follows that this scheme
shows a higher mean $\mu_{X_t}$ and a lower variance $\sigma_{X_t}^2$. Increasing $\Delta t$, in accordance with the results of theorem 1, $\mu_{Y_t}$ and $\sigma_{Y_t}^2$ tend to the values in (12); thus, if this scheme is used to avoid regression effects in the development process, then a time step $\Delta t$ sufficiently high should be used to maintain the prescribed average and variance. However, this could be in contrast with an higher temporal resolution requested by some processes. For small $\eta$ (i.e. for small $k$) the outcomes of these schemes approach those of scheme S1.

Comparisons between the schemes S1-S3 are obtained by performing realizations of the process $X_t$ by using the data relative to the last stage (adult stage) of the copepod population described in the next section (Table 3), for which the experimental mean residence time is $D = 20\, d$:

$$v = \frac{1}{D} = 0.05\, d^{-1}, \quad k = \frac{1}{2D}\, \eta^2 = 0.025\, \eta^2\, d^{-1}. \quad (32)$$

The numerical simulations have been carried out for various ratios $\eta$ defined in (16). Values of $\Delta t = 0.1\, d$ and $0.05\, d$ produce nearly the same results. The process $X_t$ is followed, starting from $t_0 = 0$, up to a final fixed time $t_f = 25\, d$, allowing that the variable $X_t$ may exceed the value 1.

The numerical characteristics $\mu_{X_t}$ and $\sigma_{X_t}^2$ show a linear trend versus $t$ (see (4), (14), (17)) for all the development schemes. Obviously, by construction very similar results are obtained for all the three schemes, and for arbitrary $\eta \leq 0.4$ (for $\eta = 0.4$ differences of $\simeq 2\%$ between the schemes at $20\, d$ are founded).

The histograms representing the distributions of $X_t$ have been constructed (the histograms for $\eta = 0.4$ and $\Delta t = 0.1\, d$ are reported in Figure 4). The scheme S1 shows a tail of negative value of $X_t$, while for S2, S3 the range of $X_t$ is positive. For times $t$ sufficiently small and arbitrary $\eta$, the normal approximation for the distributions relative to S2, S3 is not valid. However, for large $t$, depending on the level of $\eta$, the normal approximation holds.

### 3.6 Residence time in a stage

Now we follow, starting from $t_0 = 0$, the process $X_t$ up to the first time $t_N$ such that $X_{t_N} \geq 1$, for 10 000 individuals, by using $\Delta t = 0.1\, d$. Thus, we can compute the residence times $\tau = t_N - t_0$ of the individuals in the stage characterized by the data in (32), for which $D = 20\, d$ (Table 1). All the three schemes show similar mean residence times for small $\eta$ ($\eta \leq 0.2$). Increasing $\eta$ ($\eta = 0.4$) again the mean residence times for the schemes S2, S3 are similar, but are $\simeq 1.5$ days longer than the mean residence time relative to S1, which is a good approximation of the experimental residence time. The percentages of the individuals with residence time in the stage less than $15\, d$ and greater than $25\, d$ increase with $\eta$, i.e. the range of values for the residence time in the stage increases with the variability among the individuals.

In the case of $\eta$ large, the difference between the estimated and the experimental mean residence time will lead to a substantial difference between the estimated and experimental total life time, as will be illustrated in the next section.

The distribution of $\tau$ can be theoretically obtained in the case of the scheme S3, for which the increments $Y_{t_j}$ in (3) are independent and identically distributed nonnegative random variables having finite expectation. The distribution of the stopping time $N$ of $X_{t_n}$ defined as

$$N = \sup \{ n : X_{t_n} \leq 1 \}$$

is obtained from the distribution of $X_{t_n}$. In fact (Ross, 1983, p.56),

$$P(N = n) = P(X_{t_n} \leq 1) - P(X_{t_{n+1}} \leq 1),$$

11
Figure 4: Distributions of $X_t$ for the dispersion schemes S1, S2, S3, with $\Delta t = 0.1 \, d$, at times 5 $d$, 10 $d$, 20 $d$, 25 $d$. 

12
\begin{tabular}{|c|c|c|c|}
\hline
 & $\eta = 0.1$ & $\eta = 0.2$ & $\eta = 0.4$ \\
\hline
S1 & 20.22 $d$ & 20.36 $d$ & 20.31 $d$ \\
& 0.14\% - 1.30\% & 7.02\% - 12.40\% & 27.20\% - 23.18\% \\
\hline
S2 & 20.25 $d$ & 20.51 $d$ & 21.63 $d$ \\
& 0.17\% - 1.21\% & 7.34\% - 13.22\% & 21.70\% - 30.84\% \\
\hline
S3 & 20.24 $d$ & 20.55 $d$ & 21.78 $d$ \\
& 0.23\% - 0.94\% & 7.95\% - 13.07\% & 20.25\% - 32.18\% \\
\hline
\end{tabular}

Table 1: Mean residence times in the adult stage of an individual and percentages of individuals with residence time less than 15 $d$ (left) and greater than 25 $d$ (right), for different levels of stochasticity $\eta$. 10000 realizations performed with $\Delta t = 0.1$ $d$.

where

$$P (X_{t_n} \leq x) = \int_{0}^{x} g(x'; n\alpha, \beta) dx'.$$

The mean residence time is then given by

$$E(\tau) = \sum_{n=1}^{\infty} n \ P(N = n) \ \Delta t. \quad (33)$$

In the case of scheme S3, the computed residence time in Table 1 are in good agreement with the theoretical values obtained from (33) and shown in Table 2 (differences of about 1%).

\begin{tabular}{|c|c|c|c|}
\hline
 & $\eta = 0.1$ & $\eta = 0.2$ & $\eta = 0.4$ \\
\hline
S3 & 20.05 $d$ & 20.35 $d$ & 21.55 $d$ \\
\hline
\end{tabular}

Table 2: Theoretical values of $E(\tau)$ from (33) in the case of scheme S3 for the data in (32) (adult stage) and $\Delta t = 0.1$ $d$.

## 4 The multi-stage system: application to a copepod population

Now, we simulate the development of an individual through a system of stages: from the egg stage, through a number of juvenile stages, to the reproductive adult stage. The effects of different development schemes and stochasticity levels, taking into account the degree of intraspecific variability, on the temporal dynamics of the copepod *Temora stylifera*, a common and abundant species in coastal areas in the Western Mediterranean (Estrada et al., 1985), are here illustrated. The annual cycle of this species has been depicted in a long-term study conducted in the Gulf of Naples (Mazzocchi and Ribera d’Alcala’, 1995). It is characterized by a slow numerical increase of the population in early summer, a peak of abundance in late summer/autumn, and a rapid decline in winter. This population has been previously studied with an IBM under different food conditions in Mazzocchi et al., 2006. A simplified structure of the population (Mazzocchi et al., 2006) is defined by grouping the individuals in $M = 6$ stages: eggs in stage 1, immature individuals in stages 2-5 and reproductive individuals in stage 6. The detailed description of the IBM used in the numerical simulations presented in this section, together with remarks and references relative to the data introduced in the model equations (reported in Table 3 for a temperature of 20\(^0\)C) can be found in Mazzocchi et al., 2006, and Buffoni et al., 2007.
Table 3: Values of the mean experimental residence times in the stage $D^i$ in $d$, mortality rates $\mu^i$ in $d^{-1}$, and mean eggs production $f_0$ in $eggs\ female^{-1}\ d^{-1}$. The time $D^6$ is the average adult female life span. The rates $\mu^4$ and $\mu^5$ take into account both natural and predation mortalities.

Some idealized situations will be considered, trying to separate the effects of the various processes. First, the stochastic development process from egg to juvenile and adult stages, and the consequent stage distributions of the individuals is illustrated. Then, the effects of the stochastic mortality process on the stage distribution is shown. Finally, the population growth with a deterministic production of individuals is estimated; a deterministic mechanism is used to simplify the interpretation of the dynamic process, because here we address mainly to the behaviour of the development processes. As stochasticity level for the development process we consider the ratio

$$\eta = \frac{S^i}{D^i},$$

where $S^i$ and $D^i$ have been defined in subsection 3.2, assumed independent of $i$.

### 4.1 Development of individuals through stages

We consider here the development stochastic process, not affected by mortality events. Only the adults are assumed to leave the population system at the end of their life. Let us define the random variable

$$\hat{X}_t = (i - 1) + X^i_t, \quad \text{for } t^{i-1} \leq t < t^i \text{ and } i = 1, 2, ... M,$$

where $i$ is the stage index and $X^i_t$, $t^i$ have been defined in section 2. The times $t^0$ and $t^M$ correspond to the beginning and to the end of the life of an individual, respectively.

The trajectory $(t, \hat{X}_t)$, $t^0 \leq t \leq t^M$, is a representation of the life history of an individual (Figure 5). For the scheme S1, $\hat{X}$ is not a monotone function of $t$, because regression events are inherent of this type of advection-diffusion processes. On the contrary, the scheme S2 shows non decreasing trajectories, while S3 increasing trajectories. The increase of $\eta$ produces a scattering of the trajectories (Figure 5).

The residence times $\tau^i$ of an individual in the stages during the development process from egg to adult, and its life span

$$\Lambda = \sum_{i=1}^{M} \tau^i$$

(34)

are random variables and depend both on the development scheme and on the stochasticity level (Table 4). All the three schemes produce nearly the same mean stage residence times, and consequently the life span of an individual, for small $\eta$ ($\eta \leq 0.2$). For $\eta = 0.4$ again the schemes S2, S3 produce similar results, but all the stage residence times are longer than those relative to S1. We obtain that the mean life span of an individual for S2, S3 is $\simeq 2.6\ d$ longer than that relative to S1.

Let

$$\psi^i(t) = \text{average number of individuals in stage } i \text{ at time } t,$$
Figure 5: Trajectories $\hat{X}_t$ representing the life histories of three individuals for $\eta = 0.2$ and $\eta = 0.4$. The symbols triangle, circle and asterisk, denote the ends of the stages.

<table>
<thead>
<tr>
<th>STAGE</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Life span</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta = 0.2$</td>
<td>1.07</td>
<td>1.07</td>
<td>8.16</td>
<td>3.10</td>
<td>10.11</td>
<td>20.18</td>
<td>43.69</td>
</tr>
<tr>
<td>$\eta = 0.4$</td>
<td>1.11</td>
<td>1.10</td>
<td>8.30</td>
<td>3.14</td>
<td>10.20</td>
<td>20.31</td>
<td>44.16</td>
</tr>
<tr>
<td>S2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta = 0.2$</td>
<td>1.07</td>
<td>1.07</td>
<td>8.22</td>
<td>2.82</td>
<td>10.19</td>
<td>20.47</td>
<td>43.84</td>
</tr>
<tr>
<td>$\eta = 0.4$</td>
<td>1.12</td>
<td>1.13</td>
<td>8.18</td>
<td>3.29</td>
<td>10.79</td>
<td>21.64</td>
<td>46.77</td>
</tr>
<tr>
<td>S3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta = 0.2$</td>
<td>1.07</td>
<td>1.07</td>
<td>8.21</td>
<td>3.12</td>
<td>10.25</td>
<td>20.49</td>
<td>44.19</td>
</tr>
<tr>
<td>$\eta = 0.4$</td>
<td>1.13</td>
<td>1.13</td>
<td>8.72</td>
<td>3.30</td>
<td>10.80</td>
<td>21.63</td>
<td>46.71</td>
</tr>
</tbody>
</table>

Table 4: Mean residence times $E(\tau^i) \ (d)$ of an individual in the stage and mean life span $\Lambda \ (d)$ for different levels $\eta$. 10000 realizations performed with $\Delta t = 0.1 \ d$. 15
i.e. the average stage distribution of the individuals during the development process of a single generation, and let \([t_{min}^i, t_{max}^i]\) the time interval where \(\psi^i(t) > 0\). The estimation of these distributions (Figure 6) has been performed considering the histories of 10 000 individuals. The distributions \(\psi^i(t)\) for S2, S3 are practically identical for \(0 < \eta \leq 0.4\), and show a time delay with respect to those relative to S1 only for high ratios \(\eta\) and for the stages \(i = 4, 5, 6\). This fact is due to the difference in mean residence times between S1 and S2, S3, in the case \(\eta = 0.4\): in scheme S3, individuals are “slower” than in S1. The delay is evidenced by plotting the total population distribution

\[
\Psi(t) = \sum_{i=1}^{M} \psi^i(t)
\]

versus time (Figure 7).

Figure 6: Stage distributions \(\psi^i(t), i = 1, 2, \ldots, 6\), for \(\eta = 0.2\) and \(\eta = 0.4\), when mortality events are not taken into account. 10000 realizations performed with \(\Delta t = 0.1\) d. Scheme S1: continuous line, scheme S3: dashed line.

The quantity

\[
\Phi(t) = \frac{\Psi(t)}{\Psi(0)}
\]
is monotone non increasing with $t$ (Figure 7) and is a measure of the probability of finding an individual at time $t$ in the “system population”. In other words

$$\Phi(t) = P(\Lambda > t - t_0)$$

where $\Lambda$ is the life span of an individual defined in (34), or

$$1 - \Phi(t) = P(\Lambda \leq t - t_0).$$

Thus,

$$\varphi(t) = -\frac{d\Phi}{dt}$$

is the probability density of $\Lambda$. Moreover, taking into account that $\Phi(t) = 0$ for $t >> t_0$, we have that

$$E(\Lambda) = \int_{t_0}^{+\infty} (t - t_0)\varphi(t)dt = \int_{t_0}^{+\infty} \Phi(t)dt.$$  \hspace{1cm} (35)

The computation of $E(\Lambda)$ from (35) agree with the results reported in Table 4 (differences of about 0.1%).

In the case of the scheme S3, $E(\Lambda)$ can also be obtained through formula (34). The values of $E(\Lambda)$ obtained using (35) are in good agreement with the sum of $E(\tau^i)$ given by (33).
4.2 Effects of the mortality process

Let us now consider the effect due to the stochastic mortality process during the development of the individuals. The estimation of the distributions $\psi^i(t)$ (Figure 8) has been performed by using the mortality rates in Table 3, and considering the histories of 10,000 individuals. Again, the distributions for S2, S3 are very similar for $0 < \eta \leq 0.4$, and show a time delay with respect to those relative to S1. Moreover, their maxima are lesser than those of the distributions produced by S1. This last effect is particularly marked for the adult distributions (Figure 9): the schemes S2, S3 show adult distributions under the adult distribution for S1, during the large part of the adult life. This fact is due to the differences in mean residence times for the various schemes. In scheme S3 the mean residence time for adults is 1.5 days longer than that of scheme S1, that is scheme S3 has about 15 steps more than S1. At each step a control is made to check if the individual die or survive. In the case of scheme S3 there are more controls than in S1, so the probability of death is higher. Consequently, there are less individuals in adult stage for scheme S3.

We note that, starting from 10,000 eggs, the mortality process reduces the individuals from thousands in the first 3 stages to hundreds in the last 3 stages. Thus, the distributions relative to the last stages are not so smoothed as those of the first stages.

4.3 Population growth

The numerical simulations of the growth of the population have been performed by assuming both a stochastic mortality process (as in subsection 4.2) and a deterministic egg production process given by $0.5 f_0 \text{ eggs adult}^{-1} d^{-1}$, where 0.5 is the sex ratio and the value of $f_0$ is reported in Table 3.

Numerical simulations of the Malthusian growth of the population have been carried out starting with 100 initial eggs, for a time of 180 days, corresponding to about four generations. Asymptotically the increase is exponential for the three schemes. For $\eta \leq 0.2$ the three schemes increase with nearly the same growth rate. For $\eta = 0.4$ the growth rates are different; at 180 days the number of adult individuals for S2, S3 are about 25% and 35%, respectively, of the adults relative to S1 (Figure 10). In fact, as described in subsection 4.2, for $\eta = 0.4$ the mortality events influence the development of the individuals so that the schemes S2, S3 underestimate the number of adults with respect to the adults obtained with the scheme S1. Let

$$Q = 0.5 f_0 \int_{t_{\min}}^{t_{\max}} \psi^6(t) \, dt$$

be the potential production of eggs by adults in a single generation. For the distributions in Figure 9 we have

<table>
<thead>
<tr>
<th></th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q</td>
<td>25145</td>
<td>21399</td>
<td>20563</td>
</tr>
</tbody>
</table>

Table 5: Potential production of eggs in a single generation for $\eta = 0.4$

It follows that the productions of eggs in a single generation obtained with the schemes S2, S3 are 85% and 82%, respectively, of the production relative to S1. Thus, we would expect major differences during a growth process consisting of more generations.
Figure 8: Stage distributions $\psi^i(t)$, $i = 1, 2, \ldots, 6$, for $\eta = 0.4$, obtained taking into account the mortality process. 10000 realizations performed with $\Delta t = 0.1 \, d$. Scheme S1: continuous line, scheme S3: dashed line.
Figure 9: Adult distributions $\psi^\delta(t)$ for $\eta = 0.4$, obtained taking into account the mortality process. Scheme S1: continuous line, scheme S2: dotted line, scheme S3: dashed line.
Figure 10: Malthusian growth: total population and adults versus time, taking into account mortality and production processes, starting from 100 initial eggs. $\Delta t = 0.1$. Scheme S1: continuous line, scheme S2: dotted line, scheme S3: dashed line.

5 Computational aspects

The choice and the implementation of the development scheme should be supported by sound bi-ecological arguments, in particular when regression effects are not allowable and the stochasticity level is high. Here we report some remarks on computational aspects regarding the solution of the development equations, and in general the implementation of IBMs.

5.1 Choice of time step

Let us now assume that there are $M$ stages in the life of an individual characterized by parameters $D^i$, $S^i$ and $v^i$, $k^i$ as defined in subsection 3.1. The quantity $D^i$ is the average duration of an individual in the stage $i$, thus we should have

$$\Delta t \ll \min_i D^i. \quad (36)$$

Since $D^i$ depends on environmental variables, $\Delta t$ is also determined by the characteristic times of the fluctuations of these variables. Furthermore, the time step $\Delta t$ should be taken to approximate the chosen shape (either decreasing or unimodal) of the probability density of $Y^i_t(\Delta t)$ and to maintain the prescribed average and variance of the development rate.

For the forward dispersion process (18)-(19), if we take in (19) $p = v^i$ and $q = k^i$, then the distribution of $Y^i_t$ is of the type shown in Figure 2 by the dashed line. The average $\mu_{Y^i_t}$ is greater than $v^i \Delta t$, and the variance $\sigma_{Y^i_t}^2$ is smaller than $2k^i \Delta t$

$$\mu_{Y^i_t} = v^i \Delta t H_1(\theta^i_0) > v^i \Delta t, \quad \sigma_{Y^i_t}^2 = 2k^i \Delta t H_2(\theta^i_0) < 2k^i \Delta t,$$
where
\[ \theta_0^i = v^i \sqrt{\frac{\Delta t}{2k^i}}. \] (37)

Only for sufficiently high values of \( \Delta t \), \( \mu_{Y_i} \) and \( \sigma_{Y_i}^2 \) approximate \( v^i \Delta t \) and \( 2k^i \Delta t \). This could be in contrast with an higher temporal resolution requested by some processes.

Again for the forward dispersion (18)-(19), let us now assume in (19) \( p = p^{*i} \) and \( q = q^{*i} \), where \( p^{*i} \) and \( q^{*i} \) are obtained from (24),
\[ p^{*i} = \frac{v^i}{H_1(\theta^{*i})}, \quad q^{*i} = \frac{k^i}{H_2(\theta^{*i})}; \]
here \( \theta^{*i} \) is solution to (25), written as
\[ \chi(\theta) = (\theta_0^i)^2. \] (38)

From (47)-(24) it follows that
\[ p^{*i} < v^i, \quad q^{*i} > k^i. \]
Moreover, \( p^{*i} \) can be either negative (when \( \theta^{*i} < 0 \)) or positive (when \( \theta^{*i} > 0 \)); in any case, the average value \( \mu_{Y_i} = p^{*i} \Delta t H_1(\theta^{*i}) \) is always positive. The sign of \( \theta^{*i} \) and \( p^{*i} \) comes to depend on \( \Delta t \). In fact, as \( \chi(0) = 1/(\pi - 1) \), from (38) it follows that
\[ (\theta_0^i)^2 < \frac{1}{\pi - 1}, \quad \text{i.e.} \quad \Delta t < \delta^i, \implies \theta^{*i} < 0, \quad p^{*i} < 0, \]
\[ (\theta_0^i)^2 > \frac{1}{\pi - 1}, \quad \text{i.e.} \quad \Delta t > \delta^i, \implies \theta^{*i} > 0, \quad p^{*i} > 0, \] (39)
where\[ \delta^i = \frac{2k^i}{(\pi - 1)(v^i)^2} = \frac{(S^i)^2}{(\pi - 1)D^i}. \] (40)
The last equality in (40) follows from (14) and (17). The cumulative distribution and density for the two situations are illustrated in Figure 2. When \( (\theta_0^i)^2 = 1/(\pi - 1) \), i.e. when \( \Delta t = \delta^i \), then \( \theta^{*i} = 0 \) and \( p^{*i} = \lim_{\theta^{*i} \rightarrow 0} v^i/H_1(\theta^{*i}) = 0 \).

The parameter \( (\theta_0^i)^2 \) in (37) has a meaningful interpretation. Let \( (\Delta x)^i = v^i \Delta t \) be the deterministic increment in the time interval \([t, t + \Delta t] \). Then \( (\theta_0^i)^2 \) may be written as
\[ (\theta_0^i)^2 = \frac{v^i(\Delta x)^i}{2k^i}, \]
which represents the Peclet number associated to the development process, and it is increasing with \( \Delta t \). It is a measure of the contribution to the process of the deterministic and stochastic components. Thus, when this Peclet number is sufficiently small, the probability density of \( Y_i^i(\Delta t) \) is decreasing; this happens for any \( i \) when
\[ \Delta t < \min_i \delta^i. \] (41)
We note that if \( S^i/D^i < 1 \), then from (40) it follows \( \delta^i < D^i \); thus, inequality (41) implies \( \Delta t < \min_i D^i \), a weak form of (36).
For the forward dispersion with a gamma distribution the role of Peclet number associated to the development process is played by the parameter $\alpha$ defined in (29), here written as $\alpha = (\theta^2_0)$. For $\alpha \leq 1$, i.e. for

$$\Delta t \leq \frac{2v^3_0}{(v^3)^2} = \frac{(S^3)^2}{D^3},$$

(42)

the probability density of $Y^i_t(\Delta t)$ is decreasing; this happens for any $i$ when

$$\Delta t < \min_i \frac{(S^3)^2}{D^3}.$$  

(43)

Again, when $S^i/D^i < 1$, inequality (43) implies $\Delta t < \min_j D^j$.

The cumulative distributions for the forward dispersion with normal distribution and gamma distribution show different trends only for small values of the independent variable (Figure 3).

5.2 Effects of roundoff errors

Let $X^i_t$ be defined as in section 2. $X^i_t$ satisfies equation (1) where

$Y^i_t(\Delta t)$ is a random variable whose probability distribution depends on $v^3$ and $k^3$. The stochastic difference equation (1) can be considered as the Euler approximation of a stochastic differential equation. This numerical scheme has an order of convergence of $\sqrt{\Delta t}$. Moreover, in the literature it is retained that it gives a good approximation of a stochastic differential equation when the drift and diffusion coefficients are nearly constant (Kloeden and Platen, 1992). However, a critical situation occurs when the uncertainties of the development rates are very small with respect to the average values $v^3$. Consider the limit case of the deterministic process

$$X^i_{t+\Delta t} = X^i_t + v^i \Delta t, \quad t > t^{i-1}; \quad X^i_{t^{i-1}} = 0.$$  

The time $t^i$ is numerically obtained by the equation $X^i_{t^i} \geq 1$, while it should be given by $t^i = t^{i-1} + 1/v^i = t^{i-1} + m^i \Delta t$, with $m^i = number of time steps in stage i$. It is well known that the Euler scheme is not an appropriate numerical approximation for deterministic equations. In fact, in this case roundoff errors may produce errors in the duration $t^i - t^{i-1}$ of an individual in stage $i$, and then in its life span; in some cases this error results of 1 or 2 days. In general the roundoff errors tend to increase the life span of an individual and affect the birth date of the offsprings and consequently the time evolution of the overall population. A simple trick to avoid this shortcoming is to replace the exit test $X^i_{t^i} \geq 1$ with the following inequality $X^i_{t^i} \geq 1 - \varepsilon v^i \Delta t$, where $\varepsilon$ is a sufficiently small number.

5.3 Averages over the realizations

In a recent paper (Gómez-Mourelo, 2005) it is asserted that “Individual-based modeling was expected to transform ecology; however this transformation has not happened yet. There can be different reasons for this failure (Grimm, 1999; Grimm et al., 1999), and we believe that an important reason can be the lack of rigorous verification of individual-based models”. This procedure can be performed through comparisons of the outcomes from “equivalent” Lagrangian and Eulerian models, based on PDE (Fokker-Planck equations, advection-diffusion equations) for the average number of individuals (Buffoni et al., 2004 and 2007; Gómez-Mourelo, 2005; Buffoni and Pasquali, 2007). Since stochasticity is in general involved in movement and development processes, and in individual attributes, different runs of the same IBM will produce different
outcomes. Thus, the results of a suitable number $R$ of realizations of the IBM are needed to obtain stable average distributions to be compared with the Eulerian results. Building an Eulerian model equivalent to an IBM, and moreover computationally efficient, is feasible certainly when only some basic processes, such as movement, development, birth and death processes, are considered. For example, modelling only movements of passive particles or individuals leads to either linear or nonlinear (Gómez-Mourelo, 2005) advection-diffusion problems. Otherwise, when some behavioural, hierarchical, ... characteristics of an individual have to be modelled (for a complete review of such IBMs see DeAngelis and Mooij, 2005), this is not always the case, or it is a hard task. Therefore, to support the results of sophisticated IBMs, comparisons with other models at least in the basic situations are desirable.

Dispersion processes of passive particles (passive tracers) in water basins can be easily simulated by means of either Eulerian or Lagrangian models (Buffoni et al., 1996 and 1997). In a two dimensional physical system the Eulerian formalism produces a particle distribution (tracer concentration) $\phi(t, x, y)$ in the basin; on the other side, the Lagrangian formalism produces a set of positions $(X_{rt}, Y_{rt}), r = 1, 2, ..., R,$ of the $R$ released particles, where $(X_{rt}, Y_{rt})$ are solutions to the stochastic equation of motion. For fixed $t,$ the isolines $\phi(t, x, y) = \text{constant}$ are compared with the space distribution of the $R$ Lagrangian particles. Since the particles are all identical, the number $R$ represents the number of independent realizations of the dispersion process carried out; stable average distributions are obtained with values of $R$ which depend on the space and time scales of the physical problem ($R = 500$ and $R = 10000$ in two problems reported in Buffoni et al., 1996 and 1997). In the paper by Gómez-Mourelo (2005) the Lagrangian movements of $R = 5000$ eels in a river are simulated, and their “empirical distributions” are compared with “theoretical distributions”, solutions to very special advection-diffusion problems. The comparison is confirmed by means of statistical tests. In Graham and Moyeed (2002) “a strategy is proposed to determine in an efficient way how large the sample size should be to produce results with given confidence limits. The main feature is the need to perform repeated calculations with samples of a given size.”

Let us now consider IBMs for simulating the growth of a population. In the pioneer work by DeAngelis et al. (1979) an IBM was formulated for the dynamics of a population whose individuals where the young-of-the-year largemouth bass, which were structured in size classes, taking the fish length as size. The model was used to explain the differences in the observed behaviour (dependent on the rate of cannibalism) of populations in laboratory experiments. For each situation, seven different simulation runs were performed, and the experimental data were compared with the range of predictions of the seven simulations, without any explicit justification of the choice of seven realizations.

The problem of simulating the life history through all the stages of a single individual, as sketched in section 2, may be strongly dependent on the individual, when the uncertainties $k^i$ of the development rates are high with respect to the average values $v^i$. Thus, a sufficiently large number of life histories of individuals should be performed to estimate the average time evolution $\mu_{X^i_t}$. In the numerical simulations of section 3, 1000-2000 realizations were needed to estimate correctly the variance $\sigma_{X^i_t}^2$. Going on, the dynamics of the overall population is obtained by performing numerical simulations of the life histories of the individuals of the initial population, and those of the recruitment yielded over time generation after generation. Two realizations of the overall population dynamics may present quite different results in the number of individuals, strictly determined by the birth and death processes. The random contribution to the development process is less relevant to the various outcomes. An average procedure over the realizations is necessary to obtain an “average dynamics”. We perform $R$ realizations of the
dynamics by assuming the same initial conditions. The realizations are identified by different sequences of pseudo-random numbers, needed in the development, birth and death processes. Let $\psi_i^r(t)$ be the number of individuals at time $t$ in stage $i$ for the $r$-th realization. The average of the number of individuals at time $t$ in stage $i$ is then obtained by

$$\Psi_i^R(t) = \frac{1}{R} \sum_{r=1}^{R} \psi_i^r(t).$$

In a specific problem relative to the dynamics of a zooplankton population, Buffoni et al. (2007) found that a number of realizations $R = 20 - 30$ was sufficient to produce a stable average distribution of the individuals. The number of realizations to estimate the overall population dynamics is very small with respect to that of the individual life history. In fact, one realization of the dynamics of the overall population, consisting of a very high number $I_r$ of individuals (thousands), takes implicitly into account the variability among the individuals, which is proportional to $\sum_{r=1}^{R} I_r \approx R \times I_{\text{average}}$ (Graham and Moyeed, 2002). Note that $I_r$ depends on both the initial number of individuals and the final time of simulation, determined by the type of dynamics: exponential growth of the population, convergence to either an attractor (equilibrium state, limit cycle) or a chaotic evolution. Reliable results are obtained with large values of $R$, however with high computational cost; thus, the choice of $R$ is a compromise between stability of the average distribution and computational cost.

**Appendix**

**Proof of Theorem 1**

The direct computation of $\mu_{Y_t}$ and $\sigma_{Y_t}^2$ leads to the following expressions

$$\mu_{Y_t} = p \Delta t H_1(\theta), \quad \sigma_{Y_t}^2 = 2q \Delta t H_2(\theta),$$

where

$$\theta = \frac{p \Delta t}{\sqrt{2q \Delta t}} = p \sqrt{\frac{\Delta t}{2q}},$$

$$H_1(\theta) = \frac{1}{\theta} f(\theta; 0, 1) + F(\theta; 0, 1), \quad H_2(\theta) = \theta^2 H_1(\theta) (1 - H_1(\theta)) + F(\theta; 0, 1).$$

For the functions $H_1(\theta)$ and $H_2(\theta)$ (Figure 11), recalling that $\theta f + \partial f / \partial \theta = 0$, we obtain

$$\frac{\partial H_1(\theta)}{\partial \theta} = \frac{-f(\theta; 0, 1)}{\theta^2} < 0, \quad H_1(\theta) \simeq \frac{1}{2} + \frac{1}{\sqrt{2\pi \theta}} \text{ as } \theta \to 0,$$

$$\lim_{\theta \to -\infty} H_1(\theta) = 0, \quad \lim_{\theta \to +\infty} H_1(\theta) = 1,$$

and, observing that $\theta H_1(\theta) > 0 \forall \theta$,

$$\frac{\partial H_2(\theta)}{\partial \theta} = 2\theta H_1(\theta) (1 - F(\theta; 0, 1)) > 0, \quad H_2(\theta) \simeq \frac{\pi - 1}{2\pi} + \frac{\theta^2}{4} \text{ as } \theta \to 0,$$

$$\lim_{\theta \to -\infty} H_2(\theta) = 0, \quad \lim_{\theta \to +\infty} H_2(\theta) = 1.$$
Properties (21) and (22) follow directly from (44)-(48).

Proof of Theorem 2

From (12), (44), (45) we obtain equation (25) for $\theta$. The function $\chi(\theta)$ (Figure 12) is an increasing function of $\theta$, it is defined in zero (because $\theta H_1(\theta)$ is regular in zero) and satisfies the limits

$$\lim_{\theta \to -\infty} \chi(\theta) = 0, \quad \lim_{\theta \to \infty} \chi(\theta) = \infty.$$ 

Thus, equation (25) has a unique solution for any given $\theta_0^2$. 

26
References


