Dynamics in Bio-Mathematical Perspective

Odo Diekmann
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands
and
Institute of Theoretical Biology, University of Leiden
Groenhovenstraat 5, 2311 BT Leiden, The Netherlands

1. Introduction

1.1. Biology and mathematics: a continual interaction

In mathematics and science it is nowadays almost compulsory to follow the narrowing road of specialization. In a period in which the would-be universal scientist is forced to read night and day (and even while doing so is confronted with an ever increasing back-log), intense co-operation between specialists in different fields seems to be a designated way to escape from the various pitfalls (the Scylla of narrowmindedness and the Charybdis of unproductivity). This negative argument in favour of interdisciplinary projects is easily supplemented with more positive ones, such as: co-operation between people having different backgrounds increases the chances of discovering unexpected but enlightening connections and, last but not least, may enhance working pleasure considerably.

The interplay of mathematics and the sciences is not an instantaneous one-way process but rather a process of repeated cross-fertilization. Foggy notions and questions about real world phenomena have to be clarified when one tries to reformulate them in terms of a mathematical model. The incorporation of specific models (and the problems they pose) within a mathematical framework of some generality serves as a test for the mathematical structure itself and may lead to the creation of a new, extended and improved structure based on a deeper understanding. The outcome of a mathematical analysis may trigger renewed investigations, with different eyes, of the natural phenomena which one is trying to describe and understand.

In this lecture I intend to illustrate the general statements above by means of a few selected examples. These examples have in common that they are concerned with dynamics, the time-evolution of states, in the context of biological

(more precisely, population dynamical and epidemiological) models. This characteristic provides a first justification and interpretation of the title. A second interpretation derives from the fact that the interaction between biology and mathematics is itself a dynamical process. I will try to describe the examples in such a way that at least part of this process becomes visible. I will stress the mutual influence by paying special attention to the way things have developed to what they are now (and by speculating a little bit about future developments). Of course there are many cases in which by now well-known mathematical techniques are used to answer by now well-defined biological questions but, however useful that may be, this is not the kind of applications of mathematics in biology I want to describe. Instead I will concentrate on situations in which the mathematical and the biological aspects coevolve towards a state in which they are adapted to each other at the benefit of both. Inevitably the composition of the audience and my own background create some bias to the effect that the mathematical aspects will be overemphasized.

Many interesting and important recent results and developments of dynamical systems theory are not touched upon in this lecture (no chaos, for instance). Most of the work (even of that with a biological flavour) in which the Department of Applied Mathematics of the Centre for Mathematics and Computer Science (and its predecessor, the Mathematical Centre) was involved during the last 40 years, will not be described. I concentrate on two problems which, I feel, are well suited to illustrate some general features of the coevolution of mathematics and science, which are more or less representative of the work done at the Department of Applied Mathematics, and which are interesting by themselves. The solution of the first problem requires hard nonlinear analysis (up to six or seven constants have to be chosen suitably to get the estimates right). The solution of the second problem is based on soft linear functional analysis (an abstract framework has to be defined to make things easy and straightforward).

Chapter 2 deals with the first problem, the description and analysis of the geographical spread of an infectious disease. In Section 1.2 I give a preview of the main questions and answers while emphasizing the conceptual aspects and neglecting the technical ones.

Finding an appropriate mathematical framework for models of physiologically structured populations is the main issue of Chapter 3. Although biologically not the most interesting case, I concentrate on age structured populations for didactic reasons (to understand the equations of age dependent population dynamics requires comparatively little energy of the uninitiated reader; see [64,51] for a systematic exposition of models and equations in the general case and for a snapshot of the state-of-the-art of the rapidly growing mathematical theory). An introductory preview of the basic ideas and problems is given in Section 1.3.

1.2. The speed of propagation and intermediate asymptotics

In Chapter 2 we consider a situation with very simple dynamics. A steady state, called 0, is unstable and any biologically realizable perturbation, no matter how small, gives rise to a sequence of events (an orbit) which ends in a stable steady state, called ∞. Real world examples range from fires (combustion theory), over the development of an infectious disease to the taking over by a favourable mutant gene. Despite the dynamical simplicity one can ask a difficult question: how fast will the transition $0 \to \infty$ effectively take place. The sting is in the adverb 'effectively', which makes the answer 'It will take an infinite time' inappropriate. The mathematical theory of dynamical systems centers around the asymptotic behaviour of trajectories for large time and, in particular, the classification of limit sets. Transients are the Cinderellas which do the hard and dirty work, but which are hardly ever regarded as interesting by themselves.

Our question can be rephrased in terms of the physical notion of 'time scale' (see, for instance, LIN, SEGEL [47]),but in a nonlinear problem several time scales can be involved (in the present case one has at least three phases: an initial phase governed by the linearization near 0, an intermediate phase governed by the nonlinearity and a final phase governed by the linearization near ∞). So do we have to take recourse to numerical calculations, taking for granted the inherent imperfection that variation of parameters may lead to large amounts of numbers from which it is hard to deduce the essential information?

Let us first indulge in our basic question, while concentrating, for the sake of exposition, on the case of an infectious disease affecting some agricultural crop. A farmer finding his wheat-field invaded by a certain rust wants to estimate how much of the field will be unaffected at harvest time (note that the upper limit for the time window accentuates that the problem does not fit into the standard large time asymptotic realm). It appears that the problem has a spatial dimension too. At first sight this only seems to complicate the matter but, as we will see, it actually enables us to bring asymptotics back into the play.

Assume, as an 'idealization', that the field extends infinitely far in all directions. Then we can look for travelling plane waves, a special kind of self-similar solutions. The rationale for our interest in these special solutions lies in the idea that an observer moving with the right speed might be able to study the transients. Or, in other words, in a moving coordinate system the transients may look like 'frozen' spatial transitions.

A robust conclusion obtains: travelling plane waves exist for all speeds $c \geqslant c_0$ for some c_0 and this minimal wave speed c_0 is the asymptotic speed of propagation of disturbances in a sense which is on the one hand excellently adapted to the biological connotation and, on the other hand, mathematically precise. By 'robust' we mean that the conclusion is valid for a large class of models which are quite different from a mathematical point of view, yet describe biologically similar phenomena. The equations corresponding to these models take divergent forms as is manifest from the adjectives: reaction-diffusion, integro-differential, integro-difference, Volterra-Hammerstein.

Comparison theorems and the construction of suitable lower- and upper-solutions are indispensable tools for their analysis.

It is an experimental fact, derived from simulation studies, that the quantity c_0 is highly relevant for a description of propagation in *finite* fields during *finite* time intervals. In the interesting book *Similarity, Self-Similarity and Intermediate Asymptotics* [7] G.I. BARENBLATT writes:

> "Self-similar solutions also describe the 'intermediate asymptotic' behaviour of solutions of wider classes of problems in the range where these solutions no longer depend on the details of the initial and/or boundary conditions, yet the system is still far from being in a limiting state"

(and he stresses the importance of self-similar solutions as an aid in interpreting large amounts of data obtained from computer simulations). Unfortunately it appears to be rather hard to prove (or even formulate) precise mathematical statements about intermediate asymptotics (and I cannot resist the temptation of writing a commonplace: this subject deserves to be more widely and deeply studied!). However, even though the theoretical basis is perhaps not as solid as it should be, we arrive at a clear-cut conclusion: the transition $0 \to \infty$ takes place with a well-defined speed c_0.

Once such a strong result is available, it becomes worth-while to embark upon a more detailed modelling exercise dealing with such questions as: how do the ingredients of the model relate to measurable biological quantities? Moreover, the computation of c_0 from the ingredients is a point of concern and, finally, the prediction of c_0 found from the model should be tested against the speed found in the field (measurements usually indeed display a constant rate of expansion!).

1.3. About states and state-spaces

In order to give a realistic description of disease propagation it does not suffice to classify an individual plant as either healthy or infected. The production of infectious agents (say spores) is determined by the state of the particular plant, where 'state' should incorporate everything relevant for determining the spore production now and in the future, given the course of the environment (the weather, for instance). This is not an unusual situation. Individuals are not really the 'atoms' of population dynamics, simply because they differ in traits as age, size, energy reserves etc., which are of great influence on their population dynamical behaviour (giving birth, dying, consumption of limiting nutrients, occupying territoria etc.). An obvious idea is to introduce a (finite dimensional) *individual state space* Ω and to conceive of the population as a frequency distribution (sometimes called the population density) n over Ω. The dynamics of the individuals (their ageing, growing, metabolism, etc.) are described by ordinary differential equations and simple bookkeeping arguments at the population level lead to a first order partial differential equation for n. These partial differential equations may exhibit several unusual features:

birth terms are non-local and the support of n may concentrate on a lower dimensional manifold in Ω.

A convenient conceptual framework for the description of dynamical phenomena can be build from the notions of *state, next-state operators* and *generator* (and, in addition, *input* and *output* but these are not essential for our purposes now). In the present context the notion of state figures at two levels. At the individual level the state corresponds to the finitely many characteristics, say summarized in a vector x, which uniquely fix the population dynamical 'status' of an individual. The variable x takes values in Ω, a subset of $\mathbb{R}^k$. At the population level the state is given by the frequency distribution n and we have still to specify to which space X of functions on Ω $n(t)$ is assumed to belong.

Operators $T(t,t_0)$ map the population state at t_0 onto the population state at time t, thereby providing a complete description of the dynamics. Even though the collection of operators $T(t,t_0)$ is just a mathematical incarnation of its real world counterpart it is usually impossible to give a direct mathematical definition. They have a clear and well-defined interpretation but, as a rule, it is impossible to calculate explicitly how they act on the basis of nothing but modelling assumptions. Instead we usually first derive the (infinitesimal) *generator* $A(t_0)$ by calculating changes of the state in small time intervals h up to first order in h and, after dividing by h, taking the limit $h\downarrow 0$. Hence $A(t_0)$ is, at least formally, the derivative of $T(t,t_0)$ with respect to t evaluated at $t=t_0$. The advantage of the 'infinitesimal' formulation is that the different contributions to the dynamics from the various 'forces' are uncoupled in the limit $h\downarrow 0$ whereas, in contrast, they are strongly intermingled in finite time intervals (an individual which has died cannot give birth!). The 'local' differential equation $\frac{dn}{dt}=A(t)n$ is much easier derived from a verbal description of a model then the 'global' solution operators $T(t,t_0)$. This is, of course, one of the main reasons for the omnipresence of differential equations in (applied) mathematics.

Part of the bookkeeping arguments alluded to above are formal $h\downarrow 0$ calculations which yield the equation $\frac{dn}{dt}=A(t)n$ in the form of a partial differential equation supplemented with appropriate boundary conditions. So here $A(t)$ is a differential (or integro-differential or differential-difference) operator acting on functions of the variable x. In this derivation we don't bother about the precise definition of the population state space X or about the sense of convergence as $h\downarrow 0$. In the partial differential equation formulation we think of n as a function of two variables, $n(t,x)=n(t)(x)$, and neither X nor the sense in which the equation should hold is specified during a derivation by formal calculus.

Partly for the sake of exposition and partly because more general population models are not elaborated in detail yet, we assume from now on that the environmental circumstances are constant in time. So experiments starting from the same initial state are identical, whether we perform them now or two weeks from now. Time translations don't matter then and, slightly abusing

notation, we may write $T(t,t_0)=T(t-t_0)$ and assume that A is independent of t. Moreover, let us assume that density dependence may be neglected such that, as a consequence, all our operators will be linear.

In any book on the functional analytic theory of semigroups (HILLE and PHILLIPS [42], BUTZER and BERENS [14], DAVIES [18], PAZY [56], GOLDSTEIN [33], VAN CASTEREN [15], NAGEL [53]) one finds the following definitions. Let X be a Banach space with norm $\|\cdot\|$, and let for each $t\geqslant 0$, $T(t)$ be a bounded linear operator on X. Assume that:

(i) $T(0)=I$, where I denotes the identity operator on X,

(ii) $T(t+s)=T(t)T(s)$, $t,s\geqslant 0$,

(iii) $\lim_{t\downarrow 0}\|T(t)\phi-\phi\|=0$, for all $\phi\in X$.

Then $\{T(t)\}$ is called a *strongly continuous semigroup* (of bounded linear operators) on X.

The prefix 'semi' reflects the restriction $t\geqslant 0$. Note that (i) and (ii) yield a mathematical formulation of intuitive ideas about next-state operators. The condition (iii) is, as one can easily verify by exploiting (i) and (ii), equivalent with the condition that orbits are continuous, i.e. for each $\phi\in X$ the map $t\mapsto T(t)\phi$ is continuous from $\mathbb{R}_+$ to X.

The infinitesimal generator A of $\{T(t)\}$ is the, in general unbounded, operator defined by

$$A\phi=\lim_{h\downarrow 0}\frac{1}{h}(T(h)\phi-\phi)$$

whenever the limit exists. So $D(A)$, the domain of A, is by definition the set of $\phi\in X$ for which this limit exists.

Although we use the same symbols and terminology, we are at the moment dealing with two different 'worlds'. In one lives a formally derived partial differential equation, in the other an unspecified semigroup and generator acting on an unspecified Banach space X. It seems conceivable to make the connection by removing the largely conceptual difference between $n(t,x)$, a function of two variables, and $n(t)(x)$, a function of t with values in a space X of functions of x. But is this worth the effort? Does an abstract approach make life easy? A controversial question to which different people may give opposite answers.

One of the high-lights of semigroup theory is the Theorem of Hille and Yosida which gives a precise characterization of the generators of strongly continuous semigroups. So if we make a choice for the function space X and define, on the basis of the appearance of the partial differential equation, the operator A, in particular its domain, we may try to verify the necessary and sufficient conditions of the Hille-Yosida Theorem. If we are successful this yields an existence and uniqueness result for solutions of the time evolution problem. So here we first reinterpret our partial differential equation as an equation of the form $\frac{dn}{dt}=An$, then associate with A the semigroup $T(t)$ and finally define $n(t,x,\phi)=(T(t)\phi)(x)$, where $\phi(x)=n(0,x)$ is the initial condition

at $t=0$ which is (assumed to be) given. This is a usual procedure for dealing with parabolic equations, where A is an elliptic operator for which a large body of results about spectrum and resolvent estimates, the key ingredients of a verification of the Hille-Yosida conditions, is available (see HENRY [39] or FRIEDMAN [32]).

When dealing with physiologically structured population models (or with delay equations, i.e. differential equations which do incorporate some influence of the past on the future, see HALE [36]) we proceed differently. The solution $n(t,x,\phi)$ of the initial value problem is rather easily defined constructively (see section 3.1 for an example). Next we define $T(t)\phi=n(t,\cdot,\phi)$ and calculate from this definition the generator A. So here we obtain only a posteriori a reinterpretation of the partial differential equation as the abstract ordinary differential equation $\frac{dn}{dt}=An$ and the profit is far from self-evident.

In the linear case a basic advantage of the semigroup approach derives from available results concerning the connection between the spectrum of A and the asymptotic behaviour of $T(t)$ (some of the more recent results in this area were motivated by models from age dependent population dynamics! See PRÜSS [57-59] and WEBB [69]). In the case of ordinary differential equations in $\mathbb{R}^k$ this is just the connection between the eigenvalues of the matrix A and the asymptotic behaviour of solutions. But in an infinite dimensional situation there may exist spectral values which are not eigenvalues and a careful analysis is needed. I don't review this interesting theory here, but confine myself to remarking that it serves as a mayor motive for putting specific evolution problems in the semigroup framework. The very recent and highly interesting lecture notes *One-Parameter Semigroups of Positive Operators* [53], edited by R. NAGEL, gives a wealth of results culminating in an extensive study of the special (but rather important also from an 'applied' point of view) case of positive operators. Also see HEIJMANS [40,41]. DIEKMANN, METZ, KOOIJMAN and HEIJMANS [25] or WEBB [69] for an exposition directed towards applications in population dynamics.

Bypassing a vast literature on the generation of nonlinear semigroups (e.g. BARBU [6], BREZIS [12], CRANDALL [17]), we recall that in local stability and bifurcation theory one deals with perturbations of linear problems. Many results in this area can be obtained from simple estimates and the implicit function theorem once has formulated the appropriate variant of the *variation-of-constants formula*

$$T(t)=T_0(t)+\int_0^t T_0(t-\tau)BT(\tau)d\tau.$$

Here $T_0(t)$ is a semigroup generated by A_0, B is a bounded perturbation and $T(t)$ is the semigroup generated by A_0+B. In stability and bifurcation problems B is small in an appropriate sense but not necessarily linear. The variation-of-constants formula enables us to estimate how the smallness of B affects the solution operators $T(t)$ and to prove the principle of linearized stability, the center manifold theorem etc. in completely the same way as one

does in the case of ordinary differential equations. As a side-remark we mention that an appropriate form of *relative* boundedness of B is sufficient for this purpose (see, for instance, HENRY [39]).

We conclude that a basic advantage of the semigroup approach is that one can prove many results once and for all in the general setting such that subsequently one can draw conclusions about solutions of specific evolution equations by showing that the general results apply.

Following this approach in the case of physiologically structured population models (and in the case of delay equations as well) we run into some disappointment: the general abstract framework does not fit as good as one feels it ought to fit! The problem that arises is explained in Section 3.2 by means of an example. Rather than concluding that the 'basic advantage' is not so big after all and sitting down under it, we take up the challenge, analyse the difficulty and find that the equations do fit excellently within a somewhat extended general framework. In retrospect the extension is quite natural from a mathematical point of view as well and one can easily explain the framework in mathematical terms, without any reference to models from population dynamics or any other application. We emphasize, however, that the tension between general theory and specific applications (as exemplified in feelings of irritation and frustation: why are these damned problems so resistant against an abstract approach which intends to make them easy instead of difficult!?) serves as a catalyser for finding the key ideas.

The work on physiologically structured population models has only just begun and much remains to be done. At the end of the paper I will stress the need for young talented people to carry out the program.

2. THE GEOGRAPHICAL SPREAD OF AN INFECTIOUS DISEASE

2.1. A mathematical prototype: linear diffusion

In this section I will present some rather simple explicit calculations which, I hope, illuminate the main concepts and results. The simplest differential equation

$$\dot{u}=ku \tag{2.1}$$

states that the rate of production of 'particles' (say genes or spores) is proportional, with constant k, to their density u. Assume $k>0$. Then $u=0$ is an unstable steady state and, in some sense, $u=\infty$ is a stable steady state. Next suppose our particles are subject to random spatial migration in a plane and replace (2.1) by the diffusion equation

$$\frac{\partial u}{\partial t}=D\Delta u+ku \tag{2.2}$$

where

$$\Delta u=\frac{\partial^2 u}{\partial x_1^2}+\frac{\partial^2 u}{\partial x_2^2}$$

and where the diffusion coefficient D is a measure for the variance of the motion. The fundamental solution

$$u(t,x)=\frac{1}{4\pi Dt}e^{-\frac{|x|^2}{4Dt}+kt}=\frac{1}{4\pi Dt}e^{kt(1-\frac{|x|^2}{4kDt^2})} \tag{2.3}$$

describes what happens when we start at $t=0$ with one particle located at $x=0$. From this explicit expression it follows immediately that for any fixed $\epsilon>0$

$$u(t,x)\overset{t\to\infty}{\longrightarrow}\begin{cases}0 & \text{if } |x|^2>(4Dk+\epsilon)t^2\\ \infty & \text{if } |x|^2<(4Dk-\epsilon)t^2\end{cases} \tag{2.4}$$

So, asymptotically for $t\to\infty$, nothing has happenend yet outside growing circles of radius $t\sqrt{4Dk+\epsilon}$ and everything has happened already inside growing circles of radius $t\sqrt{4Dk-\epsilon}$. Therefore we call

$$c_0=2\sqrt{Dk} \tag{2.5}$$

the *asymptotic speed of propagation* of disturbances (the need to provide c with an index will become evident soon).

Two questions arise:

(i) can we obtain more information about the structure of the transition $0\to\infty$ in the vicinity of the boundary of the growing circles?

(ii) is it possible to derive (or at least guess) the speed $c_0=2\sqrt{Dk}$ a priori, i.e. without solving equation (2.2) explicitly?

It will appear that the answer to (i) provides a first step towards the answer of (ii).

So far we have exploited the radial symmetry of the fundamental solution (2.3) by concentrating at circles, i.e. using $|x|^2$ as our basic variable. But let us now choose some arbitrary unit vector ζ and look explicitly in the direction of ζ by taking for x a representation

$$x=\alpha(t,\theta)\zeta+y, \quad \text{with } y\cdot\zeta=0, \tag{2.6}$$

where θ represents a 'local' one-dimensional coordinate and the scalar function α has still to be determined. Upon substitution in (2.3) we find

$$u(t,x)=\frac{1}{4\pi Dt}e^{kt(1-\frac{\alpha^2}{4kDt^2})}e^{-\frac{|y|^2}{4Dt}} \tag{2.7}$$

which is bounded away from 0 and ∞ for $t\to\infty$ provided we make sure that

$$\alpha^2(t,\theta)=4kDt^2(1-\frac{\ln t}{kt}+\frac{h(t,\theta)}{kt})$$

for some bounded function h. For the special choice

$$\alpha^2(t,\theta)=4kDt^2(1-\frac{\ln t}{kt}-\frac{\ln 4\pi D}{kt}+\frac{\theta}{\sqrt{kDt}}) \tag{2.8}$$

we find that for $t \to \infty$

$$u(t,x) \to e^{-\sqrt{\frac{k}{D}}\theta} \tag{2.9}$$

uniformly for y and θ in compact subsets (note that the t-dependent constraint on the range of θ 'dissolves' in the limit $t \to \infty$). The formula (2.8) implies that

$$\alpha(t,\theta) = m(t) + \theta + O(\frac{\ln^2 t}{t}), \quad t \to \infty, \tag{2.10}$$

where

$$m(t) = 2\sqrt{Dk}\, t - \sqrt{\frac{D}{k}} \ln t - \sqrt{\frac{D}{k}} \ln 4\pi D \tag{2.11}$$

So asymptotically for $t \to \infty$ the solution behaves in the direction ζ like a plane wave (no dependence on y!) of the form $\exp(-\sqrt{\frac{k}{D}}\theta)$ which travels approximately with speed

$$\dot{m}(t) = 2\sqrt{Dk} - \sqrt{\frac{D}{k}} \frac{1}{t}. \tag{2.12}$$

Since ζ is arbitrary we conclude that the solution u 'decomposes' into plane waves travelling in all directions with speed $2\sqrt{Dk}$ and that these waves describe the transition between the inside of the circles ($\theta \to -\infty$) and the outside ($\theta \to +\infty$).

We could as well search for travelling plane wave solutions of the diffusion equation (2.2) directly. Substituting

$$u(t,x) = w(x \cdot \zeta - ct) \tag{2.13}$$

we find for w the ordinary differential equation

$$Dw'' + cw' + kw = 0 \tag{2.14}$$

where a prime denotes the derivative with respect to the variable

$$\theta = x \cdot \zeta - ct. \tag{2.15}$$

The solutions of (2.14) are of the form $w(\theta) = C\exp(\lambda\theta)$ with

$$\lambda = \frac{-c \pm \sqrt{c^2 - 4Dk}}{2D} \tag{2.16}$$

and C an arbitrary constant. The biological interpretation requires that w is non-negative. Consequently we are forced to adopt a lower bound for the speed c:

$$c^2 \geq 4Dk \tag{2.17}$$

So $c_0 = 2\sqrt{Dk}$ is the *minimal* wave speed (and $e^{-\sqrt{\frac{k}{D}}\theta}$ is the corresponding travelling plane wave solution) and we have found a characterization of the

asymptotic speed c_0 which allows for its determination without demanding a prohibitive effort

The following argument due to J.A.J. Metz makes the result intuitively understandable. By manipulating the initial condition suitably we can produce travelling waves in much the same way as one can create the illusion of steady movement in an array of electric lights by turning them on and off appropriately. Only one thing can spoil this game: if we try to make the speed too low the inherent 'infection' mechanism of our excitable medium takes over. Therefore this inherent infection speed is exactly the lowest possible wave speed!

2.2. Host -pathogen systems

Let $S(t,x)$ denote the density of unaffected host plants. For the domain of x (the habitat or field) we simply take $\mathbb{R}^2$. Let $A(\tau,x,y)$ describe the infectivity at x caused by the pathogen on a plant at y which was infected τ time units ago, then, by the law of mass action,

$$\frac{\partial S}{\partial t}(t,x)=S(t,x)\int_0^\infty\int_{\mathbb{R}^2}\frac{\partial S}{\partial t}(t-\tau,\, y)A(\tau,x,y)dyd\tau. \tag{2.18}$$

If, in the infinite past, S was S_0 (a given function) one obtains upon integrating (2.18) from $-\infty$ to t:

$$u(t,x)=\int_0^\infty\int_{\mathbb{R}^2}g(u(t-\tau,\, y))S_0(y)A(\tau,x,y)dyd\tau \tag{2.19}$$

where

$$u(t,x):=-\ln\frac{S(t,x)}{S_o(x)} \tag{2.20}$$

and

$$g(u)=1-e^{-u} \tag{2.21}$$

Similarly the equation

$$u(t,x)=\int_0^t\int_{\mathbb{R}^2}g(u(t-\tau,\, y))S_0(y)A(\tau,x,y)dyd\tau+f(t,x) \tag{2.22}$$

corresponds to an initial value problem in which at $t=0$ S is given by S_0 and the (given) function f describes the infectivity due to the pathogen already present at $t=0$.

Note that in this model the hosts don't move but the pathogen does by *non-local* interaction (for instance realized by spore dissemination), that an incubation period (time delay between infection and spore production) is incorporated and that the diminution of unaffected hosts makes the problem nonlinear. These features create as many striking differences with the diffusion equation of the foregoing section, but nevertheless the description of Section 1.2 reduces both to the same denominator. So let's see whether similar conclusions can be obtained.

We first make two simplifying assumptions:

$$S_0(x)=S_0, \quad \text{a constant,} \tag{2.23}$$

$$A(\tau,x,y)=H(\tau)V(|x-y|). \tag{2.24}$$

The first means that initially the density of unaffected hosts is everywhere the same, the second that the medium for the interaction is homogeneous and isotropic (only the distance between x and y matters; so no prevailing wind) and that the dispersal of infectious agents is so fast relative to the time scale of the incubation and infectivity period that the processes of creation and transport of infectious agents are effectively uncoupled.

If $u(t,x)=w(x\cdot\zeta-ct)$ is to be a solution of (2.19), under the assumption (2.23)-(2.24), the function w has to be a solution of the nonlinear convolution equation on the line

$$w(\theta)=S_0\int_{-\infty}^{\infty} g(w(\eta))\tilde{V}_c(\theta-\eta)d\eta, \quad -\infty<\theta<+\infty, \tag{2.25}$$

where

$$\tilde{V}_c(\eta):=\int_0^{\infty} H(\tau)\tilde{V}(\eta-c\tau)d\tau \tag{2.26}$$

with $\tilde{V}$ the so-called marginal infectivity kernel defined by

$$\tilde{V}(\eta):=\int_{-\infty}^{\infty} V(\sqrt{\eta^2+\sigma^2})d\sigma. \tag{2.27}$$

In the analysis of (2.25) an important role is played by the characteristic equation

$$L_c(\lambda)=1 \tag{2.28}$$

where

$$L_c(\lambda):=S_0\int_{-\infty}^{\infty}\tilde{V}_c(\eta)e^{-\lambda\eta}d\eta=S_0\int_0^{\infty}e^{-\lambda c\tau}H(\tau)d\tau\int_{\mathbf{R}^2}V(|x|)e^{-\lambda x_1}dx. \tag{2.29}$$

This characteristic equation is obtained by linearizing (2.25) around the constant solution $w\equiv 0$ followed by substitution of an exponential function. Let us assume that both H and V are nonnegative and integrable and that V decreases faster than exponentially for $|x|\to\infty$. Then some straightforward arguments show that the definition

$$c_0:=\inf\{c>0 | L_c(\lambda)=1 \text{ for some } \lambda>0\} \tag{2.30}$$

makes sense (and that $0<c_0<\infty$), provided

$$L_c(0) = S_0\int_0^{\infty}H(\tau)d\tau\int_{\mathbf{R}^2}V(|x|)dx>1. \tag{2.31}$$

The condition (2.31) is the famous threshold condition of mathematical epidemiology which has the following interpretation: the number of secundary infections produced by a single newly infected individual placed in a hypothetical population (of density S_0) consisting permanently of susceptibles only should exceed one. Clearly any epidemic will peter out immediately if this condition is not satisfied! From now on we assume that (2.31) holds.

THEOREM 1. *For any $c \geqslant c_0$ there exists a nonincreasing solution w of (2.25) with $w(-\infty)=p$ and $w(+\infty)=0$ where p is the unique positive root of the scalar equation*

$$p=\gamma S_0 g(p) \quad \text{where} \quad \gamma := \int_0^\infty H(\tau)d\tau \int_{\mathbf{R}^2} V(|x|)dx.$$

For $c>c_0$, the basic idea of the proof in [20,70] is to use the information obtained from $L_c(\lambda)$ and the properties of g in the construction of two functions ϕ and ψ such that $\phi \leqslant \psi$, $T\phi \geqslant \phi$, $T\psi \leqslant \psi$, where T denotes the (monotone!) integral operator that is associated with the right-hand side of (2.25). For $c=c_0$ one can either follow the same procedure, but the construction is a little bit more complicated, see [70], or one can resort to a limiting argument which shows that the set of speeds is closed, see [13].

The characterization of the set of speeds is completed by the following complementary result.

THEOREM 2. *For $0 \leqslant c < c_0$ there are no nonconstant solutions of equation (2.25) with $0 \leqslant w(\theta) \leqslant p$.*

One can prove Theorem 2 in at least two different ways. In one approach one has to construct a compactly supported function ψ such that, for δ positive and sufficiently small, $T(\delta\psi) \geqslant \delta\psi$ and $\liminf_{n\to\infty} T^{(n)}(\delta\psi) \geqslant p$. Subsequently one shows that for an arbitrary nontrivial solution w of (2.25) there exists a positive δ such that $w \geqslant \delta\psi$ and the result $w \geqslant p$ follows from the monotonicity of T; see [70].

In the second approach one uses Tauberian theorems (notably Pitt's form of Wiener's Tauberian Theorem) to deduce that an arbitrary solution of (2.25) with $0 \leqslant w(\theta) \leqslant p$ has to decrease exponentially to zero for $\theta \to +\infty$. Furthermore, by manipulating a bit with Laplace transforms, one can show that the exponent has to be a real root of the characteristic equation (2.28) and consequently the nonexistence of such roots implies the nonexistence of solutions of (2.25) between 0 and p; see [24] for the details.

The advantage of the second approach is that the same method is suitable for obtaining results about uniqueness modulo translation:

THEOREM 3. *For fixed $c \geqslant c_0$ equation (2.25) admits modulo translation one and only one nonconstant solution between 0 and p.*

The case $c>c_0$ is dealt with in [24] but LUI [48] has extended the proof to the case $c=c_0$; BARBOUR [5] has given a different uniqueness proof based on probabilistic arguments.

In conclusion of this section we state two results which together define the sense in which c_0 is the asymptotic speed of propagation of disturbances.

THEOREM 4. *Let f be a nonnegative bounded continuous function from $\mathbb{R}_+ \times \mathbb{R}^2$ into $\mathbb{R}$ such that the projection of the support of f on $\mathbb{R}^2$ is compact. Then*

$$\lim_{t\to\infty}(\sup\{u(t,x)|\ |x|\geqslant ct\})=0$$

for any $c>c_0$, where u is the solution of equation (2.22).

THEOREM 5. *Let f be a nonnegative continuous function then*

$$\liminf_{t\to\infty}(\min\{u(t,x)|\ |x|\leqslant ct\})\geqslant p$$

for any $c\in(0,c_0)$, provided f is not identically zero.

The proofs are based on a comparison principle and the construction of suitable upper- and lower-solutions [21,67,68]. An understanding of the way in which Volterra convolution equations generate dynamical systems [23] is very helpful.

So, with the part of ∞ assigned to p, a dynamical picture emerges that is identical to the one of the linear diffusion equation.

2.3. Into the field

As presented in Section 2.2 the results have hardly any appeal to researchers in plant pathology. The functions H and V are introduced in the abstract and the theorems are completely unreadable. In an attempt to bridge the communication gulf J.A.J. Metz asked F. van den Bosch, at that time a student in theoretical biology at the University of Leiden, to learn both languages and act as an interpreter. This is a far from easy job but several recent preprints witness that the attempt was quite succesful [9,10]. In joint work with J.C. Zadoks of the Laboratory for Phytopathology of the Agricultural University of Wageningen they developed several mechanistic submodels for spore dispersal from which V can be derived, they introduced flexible yet parameter sparse kernels H that fit published data on spore production well, they developed approximation formulae and numerical procedures to calculate c_0 from the defining equations

$$L_c(\lambda)=1,\quad \frac{\partial I_c}{\partial\lambda}(\lambda)=0$$

with a pocket calculator in negligible time, they expressed both the 'input' quantities S_0, H and V and the 'output' quantity c_0 in standard phytopathological terminology and, finally they showed that the model predictions match up to simulation studies [72] and agree reasonably with the speed measured in a field experiment. They built the connection between some parts of the

biological and the mathematical world by making biologically palpable what is mathematically so easily introduced ('Let H and V denote ...').

So far their work deals with the expansion of a connected area of infested plants within a field (a focus or hot-spot). But, as HEESTERBEEK [38] has described and classified in detail, one can consider the spread of an infectious disease in a crop at different geographical scales. One can concentrate on focus expansion, on changes in the number and size of foci within one field or on a large number of fields in different phases of disease development. In the first two cases the temporal scale is the growing season but in the last case one may have to pay attention to overwintering. This last case is particularly relevant in view of so-called quarantine-diseases (pests which are accidentally introduced in countries or continents in which they were unknown before). Although from a mathematical point of view the phenomena are almost identical on all these scales, it is a far from trivial modelling problem to make the available results applicable to the various situations and to figure out what additional results are needed. Work on these problems is in progress.

2.4. Some history and other things worth knowing

The subject of a wave-like transition from an unstable state to a stable one seems to be born in 1937 with the publication of two highly influential papers.

In his paper 'The wave of advance of advantageous genes' [30] FISHER discusses the nonlinear diffusion equation on the line

$$\frac{\partial u}{\partial t} = D\frac{\partial^2 u}{\partial x^2} + f(u)$$

with $f(u) = ku(1-u)$ and he finds that travelling waves exist for all $c \geqslant c_0 = 2\sqrt{Dk}$. A little puzzled by the indeterminacy of velocity he examines the behaviour of a finite aggregate of discrete particles, subject to random scattering and increase in number, and concludes from this study that c_0 has to be the 'true' speed. In a celebrated paper of the same year 1937 KOLMOGOROFF, PETROVSKY and PISCOUNOFF [46] prove that the solution corresponding to the special discontinuous initial condition

$$u(0,x) = \begin{cases} 0 & x<0 \\ 1 & x \geqslant 0 \end{cases}$$

converges to the travelling wave w of minimal velocity c_0 in the sense that $u(t,\ x+m(t)) \to w(x)$, uniformly in x, for $t \to \infty$ and for appropriate choice of $m(t)$, and that $\dot{m}(t) \to c_0$. Already in 1948 KENDALL [43] observes that this result cannot hold for all initial conditions, but that it is likely that for compactly supported initial data the solution develops into two diverging travelling waves of minimal velocity. Since that time several important contributions to the solution of the convergence problem have been made by various authors, culminating in a complete solution by M. BRAMSON [11] which, remarkably, uses the Feyman-Kac integral formula in conjunction with sample path estimates for Brownian motion as the basic technical device. No results about

convergence to travelling waves in higher dimensional spatial domains seem to be known.

But the speed-ambiguity which annoyed Fisher was fully resolved in 1975 when ARONSON and WEINBERGER [2,3] introduced the notion of the *asymptotic speed* of propagation of disturbances and showed that, even in higher space dimension, this speed coincides with the minimal velocity of travelling plane waves. The papers by Aronson and Weinberger mark the beginning of an explosive increase in published papers on nonlinear reaction-diffusion equations with biological applications, see for instance FIFE [27,28], OKUBO [55] and DIEKMANN and TEMME [26].

As the title of his paper indicates, Fisher was interested in the speed at which an advantageous mutant gene would spread in a spatially distributed population. In a similar spirit SKELLAM [63] investigated the regional spread of oak trees in the post-glacial period and the dispersal of the muskrat after its escape from 'prison' in Europe, and AMMERMAN and CAVALLI-SFORZA [1] analysed the neolithic transition in Europe (the shift from hunting and gathering to early farming as a new way of life). KENDALL [44] initiated the modelling of the spatial spread of epidemics (his work has been continued by MOLLISON [52]). As a rather sinister example NOBLE [54] has treated the propagation of the Black Death in medieval Europe. A much studied wildlife disease is rabies [4].

The model of Section 2.2 is a space-dependent analogue of the basic model of KERMACK and MCKENDRICK (which was introduced as early as 1927 [45]; also see [50]). It was developed and analysed independently by THIEME [65] and DIEKMANN [20] and later extended to vector-borne and other multi-type diseases by RADCLIFFE and RASS [60]. A remarkable feature of both the epidemic equation and the nonlinear diffusion equation with $f(u)=ku(1-u)$ is that c_0 is determined by the linearization at the unstable state. This is true for a large class of nonlinearities but not for all (in this connection one discriminates between pulled waves, the ones we have met, and pushed waves which are more strongly determined by the nonlinearity; see, for instance, ROTHE [61], HADELER and ROTHE [35]).

If in the genetics model heterozygotes are inferior one has *two* steady states which are 'seperated' from one another by an unstable steady state. In this case there exists usually a unique (modulo translation) wave travelling at an exactly determined velocity. In order to bring about a transition from one stable state to the other perturbations now have to be sufficiently large over a sufficiently large domain (super-threshold, as it is called) but once this is so the transition takes effectively place with the wave velocity, see FIFE and MCLEOD [29] and [27,28].

WEINBERGER [71] has introduced and analysed a discrete time equation which is sufficiently general to cover both discrete and continuous spatial domains and which allows for seasonal influences and spatial anisotropy (prevailing winds!). As a consequence the speed may depend on the direction. Let $c_0(\zeta)$ be the minimal speed of travelling plane waves in the direction ζ then Weinberger shows that the (convex) set

$$S = \{x \in \mathbb{R}^2 | x \cdot \zeta \leqslant c_0(\zeta) \text{ for all unit vectors } \zeta\}$$

replaces the circles in the results that characterize the asymptotic speed of propagation. Many other results for this class of equations were obtained by LUI in an interesting series of papers [48,49].

Aronson and Weinberger have achieved a major conceptual break-through by introducing the notion of 'asymptotic speed of propagation'. This notion combines practical relevance with mathematical elegance. Analysis of a multitude of models has by now made clear that it provides a robust link between observed spatial expansion of many different substances and the behaviour of solutions of mathematical equations. The characterization as the minimal wave speed makes it computable and hence applicable.

It is not always easy to apply applied mathematics. The spirit of the papers by Fisher and by Aronson and Weinberger is quite different and so is the jargon. The style of the papers by Thieme and Diekmann puts off many potentially interested people. We need chains of communicating people with overlapping knowledge and interests to let the stream of scientific information and inspiration flow freely back and forth between scientists and mathematicians. In Section 2.3 I briefly described such a chain and indicated its highly valuable products.

The early papers (FISHER [30], SKELLAM [63]) are quite explicitly concerned with natural phenomena. Next comes a period in which 'applicability' is still a motivation, but nevertheless mathematical analysis is the principal thing. The right concept is created and strong results are obtained. It requires additional energy to come full circle and let the mathematical results bear upon the original scientific questions. Most likely new questions arise in this 'final' phase and the process repeats indefinitely ('the march of science along a spiral staircase').

3. MATHEMATICAL MODELS OF STRUCTURED POPULATIONS AND PERTURBED DUAL SEMIGROUPS

3.1. The background

The first impulse to a general theory of physiologically structured population models was given in 1967, a year which showed a remarkable outburst of innovative papers [8,31,62]. But, perhaps due to the lack of a cut and dried mathematical framework, the subsequent development was disappointing in view of the very promising start. In the first half of 1983 a colloquium on the Dynamics of Structured Populations was held at the Centre for Mathematics and Computer Science attempting to revive the spirit of the pioneering papers and, at the same time, to start building the required mathematical framework. The colloquium served as a starting point for intense interdisciplinary interaction of the core participants. The fruits of this interaction obtained so far have been documented extensively elsewhere [51]. Here I want to concentrate on one particular mathematical aspect while refering to [51] for a general survey and many concrete examples displaying various amounts of biological complexity and realism.

3.2. Age-dependent population growth

Let the individuals of a population be characterized by their age a. Let $n(t,a)$ denote the age distribution at time t, i.e.

$$\int_{a_1}^{a_2} n(t,\alpha)d\alpha = \text{number of individuals with age between } a_1 \text{ and } a_2 \text{ at time } t.$$

The individuals age, may give birth or die. The first process is described by the differential equation $\frac{da}{dt}=1$, the second by the age-specific per capita birth rate $\beta(a)$ and the third by the age-specific per capita death rate $\mu(a)$. Since

$$n(t+h,a+h)=n(t,a)-h\mu(a)n(t,a)+O(h^2)$$

we derive for n the balance law

$$\frac{\partial n}{\partial t} = -\frac{\partial n}{\partial a} - \mu n \tag{3.1}$$

which we supplement with the boundary condition

$$n(t,0) = \int_0^\infty \beta(\alpha)n(t,\alpha)d\alpha \tag{3.2}$$

to express that the influx at the boundary $a=0$ equals the total birth rate. Finally we assume that at $t=0$ the age distribution equals a given function ϕ:

$$n(0,a)=\phi(a). \tag{3.3}$$

In order to minimize inessential (for the present purpose) technical and notational detail we take μ to be identically zero throughout this paper. To get a feel for the problem we begin by taking $\beta(a)\equiv 0$ as well. In the absence of births and deaths the solution of (3.1) - (3.3) is evidently

$$n(t,a,\phi)= \begin{cases} \phi(a-t) & ,a\geq t \\ 0 & ,a<t \end{cases} \tag{3.4}$$

as follows also directly from the interpretation.

A reasonable choice of population state space is $L_1(\mathbb{R}_+)$. Putting

$$T_0(t)\phi = n(t,\cdot,\phi) \tag{3.5}$$

we obtain a strongly continuous semigroup of bounded linear operators on $L_1(\mathbb{R}_+)$ with infinitesimal generator

$$\begin{cases} A_0\phi=-\phi' \\ D(A_0)=\{\phi|\phi(a)=\int_0^a \phi'(\alpha)d\alpha \text{ with } \phi'\in L_1(\mathbb{R}_+)\} \end{cases} \tag{3.6}$$

(recalling that one out of several equivalent definitions of an absolutely continuous function is 'a function which is, locally, the integral of an L_1-function', we can also write $D(A_0)=\{\phi|\phi$ is absolutely continuous, $\phi(0)=0$ and ϕ' is

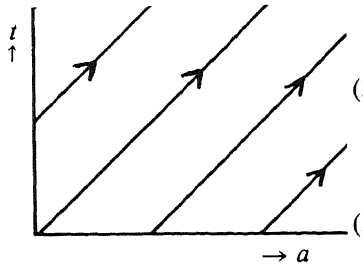
integrable over $\mathbb{R}_+$}; in the following we abbreviate 'absolutely continuous' to AC).

The standard solution procedure in case of non-zero birth rate is the following. First consider the birth rate

$$b(t) = \int_0^\infty \beta(\alpha)n(t,\alpha)d\alpha \tag{3.7}$$

as known. Then

$$n(t,a,\phi) = \begin{cases} \phi(a-t) & ,a \geqslant t \\ b(t-a) & ,a<t \end{cases} \tag{3.8}$$



where, although we haven't expressed this in our notation, b depends on ϕ. Substituting (3.8) into (3.2) we obtain the linear renewal (i.e. Volterra convolution) equation

$$b(t) = \int_0^t \beta(\alpha)b(t-\alpha)d\alpha + f(t) \tag{3.9}$$

with

$$f(t) = \int_t^\infty \beta(\alpha)\phi(\alpha-t)d\alpha = \int_0^\infty \beta(\alpha+t)\phi(\alpha)d\alpha. \tag{3.10}$$

Assume $\beta \in L_\infty(\mathbb{R}_+)$. Standard contraction mapping arguments imply that (3.9) has a unique solution represented by

$$b = \sum_{n=o}^{\infty} \beta^{n*} * f \tag{3.11}$$

where the star denotes the convolution product, $\beta^{0*} * f := f$, $\beta^{1*} := \beta$, $\beta^{n} * := \beta^{(n-1)*} * \beta$, $n \geqslant 2$. Substituting (3.11) into (3.8) we finally arrive at a series expansion for the age distribution n, which has the following interpretation. Let's call those individuals which were present at time $t = 0$ the zero'th generation. Then f describes the offspring of the zero'th generation and the corresponding term in the expansion of n is, for this reason, called the first generation. Similarly the n-th term describes the n-th biological generation and the expansion is called the generation expansion.

The semigroup

$$T(t)\phi = n(t,\cdot,\phi) \tag{3.12}$$

is generated by

$$\begin{cases} A\phi = -\phi' \\ D(A) = \{\phi | \phi(a) = \int_0^\infty \beta(\alpha)\phi(\alpha)d\alpha + \int_0^a \phi'(\alpha)d\alpha \text{ with } \phi' \in L_1(\mathbb{R}_+)\} \end{cases} \tag{3.13}$$

(or, equivalently, $D(A)=\{\phi|\phi$ is AC, $\phi(0)=\int_0^\infty \beta(\alpha)\phi(\alpha)d\alpha$ and ϕ' is integrable over $\mathbb{R}_+\}$).

A striking point is that *all* information about the birth rate enters in the domain of A and that the action of A is independent of β. This is highly unpleasant for several reasons:

(i) within the present functional analytic framework there is no analogue of the renewal equation (3.9) which we can solve iteratively; a puzzling and somewhat irritating phenomenon.

(ii) if we deal with nonlinear birthrates (describing density dependence) we don't have at our disposal a variation-of-constants formula. The lack of this important tool forms an obstacle for the development of the local stability and bifurcation theory and as a consequence ad hoc approaches dominate the field [69,59,19].

So, once again, is an abstract approach beneficial? It looks as though we made life more complicated, instead of simpler, by introducing a semigroup.

A little reflection reveals that the difficulty is due to the fact that all newborns have (by definition) one and the same age $a=0$. The range of the birth operator is spanned by the (Dirac) measure concentrated at $a=0$ which is not an element of $L_1(\mathbb{R}_+)$. So the 'perturbation' of the generator maps out of the state space into some bigger space but, as we have seen, solving the differential equation we come back into the smaller space. An analogous phenomenon occurs with delay equations [36,22].

Should we enlarge the state space and let our age distributions live in the space of regular Borel measures? This is a natural and sensible action (in fact one can argue right from the start that this is the appropriate state space) but we have to pay a technical price: the semigroup is no longer strongly continuous (indeed, translation of a concentrated measure is not continuous).

In Section 3.4 it is shown that we need not choose the least of two evils but that, instead, we can make great play with the good things of two spaces neither of which is ideal by itself. It appears that duality provides us with a systematic procedure to create the appropriate 'bigger' space and that a general theory can be built which encompasses both age-dependent population models and delay equations. The key Section 3.4 is essentially a summary of the preprint [16] by CLÉMENT, DIEKMANN, GYLLENBERG, HEIJMANS and THIEME.

3.3. Dual semigroups

Let $\{T(t)\}$ be a strongly continuous semigroup of bounded linear operators on a Banach space X generated by A. The adjoint operators $T^*(t)$ form a semigroup on the dual space X^*. $\{T^*(t)\}$ is weak * continuous but need not be strongly continuous if we equip X^* with the norm topology (unless X is reflexive). A^*, the adjoint of A, is the weak * generator of $\{T^*(t)\}$. Note that A^* need not be densely defined.

In their classic treatise [42] HILLE and PHILLIPS showed that the dialogue of a space and a semigroup demands a duality theory which is made to measure.

We need a special star, called sun and represented by the symbol $\odot$. Let $X^{\odot}$ denote the maximal invariant subspace on which $\{T^*(t)\}$ is strongly continuous. Then

$$X^{\odot} = \{\phi^* \in X^* | \lim_{t \downarrow 0} \| T^*(t)\phi^* - \phi^* \| = 0\}, \tag{3.14}$$

$X^{\odot}$ is norm-closed and $\overline{D(A^*)} = X^{\odot}$. Let $\{T^{\odot}(t)\}$ denote the strongly continuous semigroup on $X^{\odot}$ which is obtained by restriction of $\{T^*(t)\}$ and let $A^{\odot}$ denote its generator. Then $A^{\odot}$ is the part of A^* in $X^{\odot}$, i.e. the largest restriction of A^* with both domain and range in $X^{\odot}$.

On $X^{\odot *}$, the dual space of $X^{\odot}$, we obtain a weak * continuous semigroup $\{T^{\odot *}(t)\}$ with weak * generator $A^{\odot *}$. Let

$$X^{\odot\odot} = \{\phi^{\odot *} \in X^{\odot *} | \lim_{t \downarrow 0} \| T^{\odot *}(t)\phi^{\odot *} - \phi^{\odot *} \| = 0\}. \tag{3.15}$$

It follows rather easily that X can be embedded into $X^{\odot *}$ and henceforth we identify X and its embedding. Then X becomes a subspace of $X^{\odot\odot}$.

DEFINITION. *X is called $\odot$-reflexive with respect to A iff $X = X^{\odot\odot}$.*

It is known that X is $\odot$-reflexive with respect to A iff $(\lambda I - A)^{-1}$ is $X^{\odot}$-weakly compact. Moreover, X is $\odot$-reflexive with respect to A iff $X^{\odot}$ is $\odot$-reflexive with respect to $A^{\odot}$.

3.4. Perturbation theory for dual semigroups

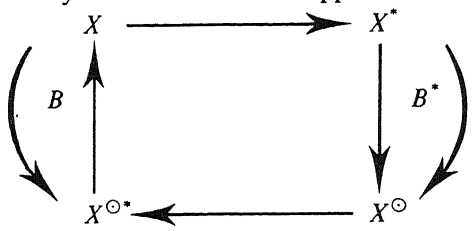
Let $\{T_0(t)\}$ be a strongly continuous semigroup on X generated by A_0 and assume that X is $\odot$-reflexive with respect to A_0. Let $B: X \to X^{\odot *}$ be a bounded linear operator. The variation-of-constants equation

$$T(t)\phi = T_0(t)\phi + \int_0^t T_0^{\odot *}(t-\tau) B T(\tau)\phi \, d\tau \tag{3.16}$$

can be shown to make sense and to admit a unique solution $\{T(t)\}$ (which can be represented by a 'generation' series). Here the integral is a weak * integral which in principle takes values in $X^{\odot *}$ but in fact takes values in the closed subspace $X^{\odot\odot} = X$. By duality and restriction we obtain semigroups $\{T^*(t)\}$, $\{T^{\odot}(t)\}$ and $\{T^{\odot *}(t)\}$ on $X^*, X^{\odot}$ and $X^{\odot *}$ respectively, since it can be shown that the spaces of strong continuity do *not* depend on B. Similarly the domains of the weak * generators on the 'big' spaces are independent of B. The following theorem summarizes part of the results.

THEOREM. *The operator $A\phi = A_0^{\odot *}\phi + B\phi$ with $D(A) = \{\phi \in D(A_0^{\odot *}) | A_0^{\odot *}\phi + B\phi \in X\}$ is the generator of a strongly continuous semigroup $\{T(t)\}$ on X and the variation-of-constants formula (3.16) holds.*

The symmetry of the framework is apparent from the diagram



When X is not $\odot$-reflexive with respect to A_0 this symmetry is disturbed. Nevertheless similar results hold. A canonical embedding of $X^{\odot\odot}$ into X^{**} seems to play a leading part, but it is not yet precisely clear how the most elegant and efficient argumentation proceeds, so we refrain from further discussion here.

3.5. Age-dependent population dynamics revisited

If we consider age-distributions over the non-compact domain $\mathbb{R}_+$ we don't get $\odot$-reflexivity. However, if β has compact support (or, in other words, if very old individuals don't produce offspring) we may limit our bookkeeping of individuals without losing relevant information. For the purpose of the present exposition, we therefore replace $L_1(\mathbb{R}_+)$ by $L_1(0,a_{\max})$ for some constant $a_{\max}$. So $X=L_1(0,a_{\max})$.

Let $A_0\phi=-\phi'$ with $D(A_0)=\{\phi|\phi$ is AC and $\phi(0)=0\}$ be, as before, the generator of the semigroup

$$(T_0(t)\phi)(a) = \begin{cases}\phi(a-t) & ,a\geqslant t\\ 0 & ,a<t.\end{cases}$$

On the dual space $X^*=L_\infty(0,a_{\max})$ we have the semigroup

$$(T_0^*(t)\psi)(a) = \begin{cases}\psi(a+t) & ,a+t\leqslant a_{\max}\\ 0 & ,a+t>a_{\max}\end{cases}$$

with weak * generator

$A_0^*\psi=\psi'$

$D(A_0^*)=\{\psi|\psi$ has a Lipschitz continuous representative which is zero at $a=a_{\max}\}$.

Clearly $X^\odot=C_0(0,a_{\max})=\{\psi|\psi$ has a continuous representative which is zero at $a=a_{\max}\}$ and $X^{\odot*}=M[0,a_{\max})$, the space of all complex regular Borel measures on $[0,a_{\max})$. It is well-known [14] that the subspace of $M[0,a_{\max})$ on which translation is continuous, i.e. $X^{\odot\odot}$, is exactly the closed subspace of all absolutely continuous measures. The mapping which associates with any ϕ in X the measure μ in $X^{\odot*}$ defined by $\mu(\omega)=\int_\omega\phi(\alpha)d\alpha$ describes the canonical identification of X and $X^{\odot\odot}$.

Let $B:X\to X^{\odot*}$ be defined by

$$B\phi = \int_0^{a_{\max}}\beta(\alpha)\phi(\alpha)d\alpha\cdot\delta = <\beta,\phi>\delta$$

where δ is the (Dirac) measure concentrated at $a=0$. Then the results of Section 3.4 apply and we conclude that the part of $A_0^{\odot *}+B$ in X generates a semigroup $T(t)$ which satisfies the variation-of-constants equation

$$T(t)\phi = T_0(t)\phi + \int_0^t T_0^{\odot *}(t-\tau)BT(\tau)\phi d\tau. \tag{3.17}$$

Since B has one-dimensional range we can go a little further. Define $b(t)=<\beta,T(t)\phi>$ then, applying the functional β to (3.17), we find after a little calculation that b has to satisfy the scalar equation

$$b(t) = f(t) + \int_0^t \beta(t-\tau)b(\tau)d\tau \tag{3.18}$$

where $f(t):=<\beta,T_0(t)\phi>=\int_0^{a_{\max}}\beta(\alpha)\phi(\alpha-t)d\alpha$. Thus we recover the renewal equation (3.9).

If, conversely, b is a solution of (3.18) with f of the form $f(t)=<\beta,T_0(t)\phi>$ for some $\phi\in X$ then $T(t)\phi$ is obtained by a simple substitution into the now explicit expression (3.17):

$$T(t)\phi = T_0(t)\phi + \int_0^t T_0^{\odot *}(t-\tau)\delta b(\tau)d\tau \tag{3.19}$$

Thus we obtain a reformulation of (3.8).

We conclude that the abstract framework of Sections 3.3 and 3.4 is rich enough for the (re)formulation of the (quasi-) explicit formulas of the direct approach via the renewal equation.

Now we can also ease those attentive readers who worried about the fact that the epidemic model of Section 2.2 was formulated as an integral equation (notably with respect to the time variable) and not as a (abstract) differential equation. When we think of 'age' as 'time elapsed since infection' and adopt a nonlinear 'birth = infection' condition one can make the connection between the nonlinear renewal equation via the variation-of-constants formula (3.17) exactly as in the present linear case.

3.6. Physiologically structured population models: a challenge for the future

The biological motivation for studying physiologically structured population models is described at length in the lecture notes [51] and the survey paper [64]. The mathematical form taken by these models is:

$$\frac{\partial n}{\partial t} + \text{divergence (velocity } n) = \text{sources} - \text{sinks}, \quad x\in\Omega,$$

$$\nu\cdot \text{velocity}|_{\partial\Omega_+} = \text{source}$$

where the individual 'velocity' $\frac{dx}{dt}$ and the sources and sinks are specified according to the specific situation at hand. Here ν denotes the inward normal to $\partial\Omega$, the boundary of the individual state space Ω, and $\partial\Omega_+$ is the part of $\partial\Omega$

at which $\nu\cdot$ velocity >0, i.e. characteristics enter Ω. The solution concept is based on integration along characteristics.

In a recent survey on *Infinite Dimensional Dynamics* [37] J.K. HALE writes:

> 'For the successful development and application of dynamical systems in infinite dimensions, we need intensive interaction between two special groups of researchers. The first group consists of mathematicians who are well trained in dynamical systems and know both the analytic and the geometric theory of differential equations in finite dimensions. They should also know well the classical and modern theory of partial differential and functional differential equations and have a strong background in applications-especially physics and engineering. The other group of researchers should be primarily concerned with applications, but should be well trained in ordinary and partial differential equations. It does not take much reflection to see that there are very few people with these qualifications. More resources need to be allocated for training young people to carry out this program'.

Then Hale goes on to describe functional differential equations and parabolic systems as special cases in which the type of interaction he has in mind has led to considerable success (and to make some remarks about hyperbolic systems and chaotic dynamics). It seems quite conceivable that the equations of physiologically structured population dynamics will be at home in a similar survey written many years from now. But whether this will happen or not, only time will tell.

REFERENCES

1. A.J. AMMERMAN, L.L. CAVALLI-SFORZA (1984). *The Neolithic Transition and the Genetics of Populations in Europe,* Princeton Univ. Press.
2. D.G. ARONSON, H.F. WEINBERGER (1975). Nonlinear diffusion in population genetics, combustion and nerve pulse propagation. J.A. GOLDSTEIN (ed.). *Partial Differential Equations and Related Topics, Springer Lect. Notes in Math. 446,* 5-49.
3. D.G. ARONSON, H.F. WEINBERGER (1978). Multidimensional nonlinear diffusion arising in population genetics. *Adv. in Math. 30,* 33-76.
4. P.J. BACON (ed.). (1985). *Population Dynamics of Rabies in Wildlife,* Academic Press.
5. A.D. BARBOUR (1977). The uniqueness of Atkinson and Reuter's epidemic waves. *Math. Proc. Camb. Phil. Soc. 82,* 127-130.
6. V. BARBU (1976). *Nonlinear Semigroups and Differential Equations in Banach Spaces,* Noordhoff, Leiden.
7. G.I. BARENBLATT (1979). *Similarity, Self-Similarity and Intermediate Asymptotics,* Plenum.

8. G.I. BELL, E.C. ANDERSON (1967). Cell growth and division. I. A mathematical model with applications to cell volume distributions in mammalion suspension cultures. *Biophys. J. 7,* 329-351.
9. F. VAN DEN BOSCH, J.A.J. METZ, J.C. ZADOKS. *The Asymptotic Speed of Travelling Epidemic Waves,* preprint
10. F. VAN DEN BOSCH, J.C. ZADOKS, J.A.J. METZ. *Focus Formation in Plant Diseases.* I. The constant rate of focus expansion. II. Realistic parameter-sparse models. Preprints.
11. M. BRAMSON (1983). Convergence of solutions of the Kolmogorov equation to travelling waves. *Memoir of the AMS 285.*
12. H. BREZIS (1977). *Opérateurs Maximaux Monotones et Semi-Groupes de Contractions dans les Espaces de Hilbert,* North Holland, Amsterdam.
13. K.J. BROWN, J. CARR (1977). Deterministic epidemic waves of critical velocity. *Math. Proc. Camb. Phil. Soc. 81,* 431-436.
14. P.L. BUTZER, H. BERENS (1967). *Semi-groups of Operators and Approximation,* Springer, Berlin.
15. J. VAN CASTEREN (1985). *Generators of Strongly Continuous Semigroups,* Pitman, Boston.
16. PH. CLÉMENT, O. DIEKMANN, M. GYLLENBERG, H.J.A.M. HEIJMANS, H.R. THIEME. *Perturbation Theory for Dual Semigroups. I. The Sun-Reflexive Case,* preprint.
17. M.G. CRANDALL (1986). Nonlinear semigroups and evolution governed by accretive operators. *Proc. Symp. Pure Math. AMS 45* Part 1, 305-337.
18. E.B. DAVIES (1980). *One-Parameter Semigroups,* Academic Press, London.
19. W. DESCH, W. SCHAPPACHER (1985). Spectral properties of finite-dimensional perturbed linear semigroups. *J. Diff. Equ. 59,* 80-102.
20. O. DIEKMANN (1978). Thresholds and travelling waves for the geographical spread of infection. *J. Math. Biol. 6,* 109-130.
21. O. DIEKMANN (1979). Run for your life. A note on the asymptotic speed of propagation of an epidemic. *J. Diff. Equ. 33,* 58-73.
22. O. DIEKMANN. *Perturbed Dual Semigroups and Delay Equations,* preprint.
23. O. DIEKMANN, S.A. VAN GILS (1984). Invariant manifolds for Volterra integral equations of convolution type. *J. Diff. Equ. 54,* 139-180.
24. O. DIEKMANN, H.G. KAPER (1978). On the bounded solutions of a nonlinear convolution equation. *Nonl. Anal. Th. Math. Appl. 2,* 721-737.
25. O . DIEKMANN, J.A.J. METZ, S.A.L.M. KOOIJMAN, H.J.A.M. HEIJMANS (1984). Continuum population dynamics with an application to *Daphnia magna. Nieuw Archief voor Wiskunde 4,* 82-109.
26. O. DIEKMANN, N.M. TEMME (1976). *Nonlinear Diffusion Problems, MC Syllabus 28,* Math. Centrum, Amsterdam.
27. P.C. FIFE (1978). Asymptotic states for equations of reaction and diffusion. *Bull. AMS 84,* 693-726.
28. P.C. FIFE (1979). *Mathematical Aspects of Reacting Diffusing Systems, Springer Lect. Notes in Biomath. 28.*
29. P.C. FIFE, J.B. MCLEOD (1977). The approach of solutions of nonlinear diffusion equations to travelling front solutions. *Arch. Rat. Mech. Anal.*

65, 335-361.
30. R.A. FISHER (1937). The wave of advance of advantageous genes. *Ann. of Eugenics 7,* 355-369.
31. A.G. FREDRICKSON, D. RAMKRISHNA, H.M. TSUCHIYA (1967). Statistics and dynamics of procaryotic cell populations. *Math. Biosc. 1,* 327-374.
32. A. FRIEDMAN (1969). *Partial Differential Equations,* Holt-Rinehart & Winston.
33. J.A. GOLDSTEIN (1985). *Semigroups of Operators and Applications,* Oxford University Press.
34. M.E. GURTIN, R.C. MACCAMY (1974). Nonlinear age-dependent population dynamics. *Arch. Rat. Mech. Anal. 54,* 281-300.
35. K.P. HALDELER, F. ROTHE (1975). Travelling fronts in nonlinear diffusion equations. *J. Math. Biol. 2,* 251-263.
36. J.K. HALE (1977). *Theory of Functional Differential Equations,* Springer.
37. J.K. HALE (1985). *Infinite Dimensional Dynamics,* Report, Brown Univ. *Providence. R.I.*
38. H. HEESTERBEEK (1985). *Over Modellering van Continentale Epidemieën,* Laboratory of Phytopathology, Agricultural University Wageningen.
39. D. HENRY (1981). *Geometric Theory of Semilinear Parabolic Equations, Springer Lect. Notes. in Math. 840.*
40. H.J.A.M. HEIJMANS (1985). *Dynamics of Structured Populations,* Thesis, Univ. of Amsterdam.
41. H.J.A.M. HEIJMANS (1986). Structured populations, linear semigroups and positivity. *Math. Z. 191,* 599-617.
42. E. HILLE, R.S. PHILLIPS (1957). *Functional Analysis and Semi-Groups,* Amer. Math. Soc., Providence R.I.
43. D.G. KENDALL (1948). A form of wave propagation associated with the equation of heat conduction. *Proc. Camb. Phil. Soc 44,* 591-593.
44. D.G. KENDALL (1965). Mathematical models of the spread of infection. *Mathematics and Computer Science in Biology and Medicine,* Medical Research Councel, London 213-224.
45. W.O. KERMACK, A.G. MCKENDRICK (1927). A contribution to the mathematical theory of epidemics. *Proc. Roy. Soc. A 115,* 700-721.
46. A. KOLMOGOROFF, I. PETROVSKY, N. PISCOUNOFF (1937). Étude de l'equation de la diffusion avec croissance de la quantité de matière et son application a un problème biologique. *Bull. Univ. Etat Moscou Ser. Int. A. Math. Méc. 1 # 6,* 1-25 *(Bjal. Moskovskovo Gos. Univ. 17,* 1-72).
47. C.C.LIN, L.A. SEGEL (1974). *Mathematics Applied to Deterministic Problems in the Natural Sciences,* Macmillan, New York.
48. R.LUI. A nonlinear integral operator arising from a model in population genetics. *SIAM J. Math. Anal. I* Monotone initial data. *13* (1982), 913-937. II Initial data with compact support. *13* (1982), 938-953. III Heterozygote inferior case. *16* (1985), 1180-1206. IV Clines. *17* (1986), 152-168.
49. R. LUI (1983). Existence and stability of travelling wave solutions of a

nonlinear integral operator. *J. Math. Biol. 16,* 199-220.
50. J.A.J. METZ (1978). The epidemic in a closed population with all susceptibles equally vulnerable; some results for large susceptible populations and small initial infections. *Acta Biotheor. 27,* 75-123.
51. J.A.J. METZ, O. DIEKMANN (eds.). *Dynamics of Physiologically Structured Populations, Springer Lecture Notes in Biomathematics,* to appear in 1986.
52. D. MOLLISON (1977). Spatial contact models for ecological and epidemic spread. *J. Roy. Stat. Soc. B 39,* 283-326.
53. R. NAGEL (ed.) (1986). *One-Parameter Semigroups of Positive Operators, Springer Lect. Notes in Math. 1184.*
54. J.V. NOBLE (1974). Geographic and temporal development of plagues. *Nature 250,* 726-729.
55. A. OKUBO (1980). *Diffusion and Ecological Problems: Mathematical Models, Biomathematics Vol. 10,* Springer.
56. A. PAZY (1983). *Semigroups of Linear Operators and Applications to Partial Differential Equations,* Springer, New York.
57. J. PRÜSS (1981). Equilibrium solutions of age-specific population dynamics of several species. *J. Math. Biol. 11,* 65-84.
58. J. PRÜSS (1983). On the qualitative behaviour of populations with age-specific interactions. *Comp. & Maths. with Appls. 9,* 327-339.
59. J. PRÜSS (1983). Stability analysis for equilibria in age-specific population dynamics. *Nonl. Anal. Th. Math. Appl. 7,* 1291-1313.
60. J. RADCLIFF, L. RASS (1984). The spatial spread and final size of the deterministic non-reducible *n*-type epidemic. *J. Math. Biol. 19,* 309-327.
61. F. ROTHE (1981). Convergence to pushed fronts. *Rocky Mountain J. Math. 11,* 617-633.
62. J.W. SINKO, W. STREIFER (1967). A new model for age-size structure of a population. *Ecology 48,* 910-918.
63. J.G. SKELLAM (1951). Random dispersal in theoretical populations. *Biometrika 38,* 196-218.
64. W.STREIFER (1974). Realistic models in population ecology. A. MAC FADYEN (ed.). *Advances in Ecological Research 8,* 199-266.
65. H.R. THIEME (1977). A model for the spatial spread of an epidemic. *J. Math. Biol. 4,* 337-351.
66. H.R. THIEME (1977). The asymptotic behaviour of solutions of nonlinear integral equations. *Math. Z. 157,* 141-154.
67. H.R. THIEME (1979). Asymptotic estimates of the solutions of nonlinear integral equations and asymptotic speeds for the spread of populations. *J. Reine Angew. Math. 306,* 94-121.
68. H.R. THIEME (1979). Density-dependent regulation of spatially distributed populations and their asymptotic speed of spread. *J. Math. Biol. 8,* 173-187.
69. G.F. WEBB (1985). *Theory of Nonlinear Age-Dependent Population Dynamics,* Marcel Dekker.
70. H.F. WEINBERGER (1978). Asymptotic behaviour of a model in population genetics. J.M. CHADAM (ed.). *Nonlinear Partial Differential Equations and*

Applications, Springer Lect. Notes in Math. 648, 47-98.

71. H.F. WEINBERGER (1982). Long-time behaviour of a class of biological models. *SIAM J. Math. Anal. 13,* 353-396.
72. J.C. ZADOKS, P. KAMPMEIJER (1977). *Epimul, a Simulator of Foci and Epidemics in Mixtures of Resistant and Susceptible Plants, Mosaics and Multilines,* simulation monograph, PUDOC, Wageningen.