Evaluating the connectivity, continuity and distance norm in mathematical models for community ecology, epidemiology and multicellular pathway prediction

Evaluating the connectivity, continuity and distance norm in mathematical models for community ecology, epidemiology and multicellular pathway prediction

W Allaerts[1]
Biological Publishing A&O and Immunology Department,
Erasmus MC, Rotterdam, The Netherlands

E-mail: w.allaerts@planet.nl

Abstract. The main global threats of the biosphere on our planet, such as a global biodiversity impairment, global health issues in the developing countries, associated with an environmental decay, unnoticed in previous eras, the rise of greenhouse gasses and global warming, urge for a new evaluation of the applicability of mathematical modelling in the physical sciences and its benefits for society. In this paper, we embark on a historical review of the mathematical models developed in the previous century, that were devoted to the study of the geographical spread of biological infections. The basic notions of connectivity, continuity and distance norm as applied by successive bio-mathematicians, starting with the names of Volterra, Turing and Kendall, are highlighted in order to demonstrate their usefulness in several new areas of bio-mathematical research. These new areas include the well-known fields of community ecology and epidemiology, but also the less well-known field of multicellular pathway prediction. The biological interpretation of these abstract mathematical notions, as well as the methodological criteria for these interpretative schemes and their corroboration with empirical evidence are discussed. In particular, we will focus on the boundedness norm in polynomial Lyapunov functions and its application in Markovian models for community assembly and in models for cellular pathways in multicellular systems. Finally, the usefulness of hybrid mathematical modelling in miscellaneous biological, environmental and public health issues will be discussed.

.

1. Introduction

The geographical spread of infectious diseases in an inhabited area or worldwide, has been a matter of primary concern for a long time, not only for practitioners and health policy makers but also for epidemiologists and mathematical biologists. Since the formulation of Kermack and McKendrick's

[1] To whom any correspondence should be addressed.

model [1] - for the evolution in time of a contagious disease in a closed population – many adaptations of their model have been put forward [2,3,4]. Kermack and McKendrick's deterministic model [1] forms the basis for analytical solutions for the spread of an epidemic, using a Volterra-type nonlinear integral equation set [5]. This model was also an inspiration for Turing's model for morphogenesis through a reaction-diffusion mechanism [6,7]. Kendall's [2] modification of this model resulted in a space-dependent analogue leading into a set of ordinary differential equations (ODE's). Kendall achieved this simplification by taking for the infectivity at each point a weighted spatial average of the density of infectives [4]. Kendall claims that a particular result of this model is that it predicts the conditions for an epidemic to develop into a pandemic, named as the 'pandemic threshold theorem' (PTT) [2]. According to this PTT, there will be a pandemic if and only if the population density σ exceeds the threshold density $\rho \equiv \gamma/\beta$, being the ratio of the intrinsic removal (γ) and infection (β) rates. In Kendall's model, the removal of 'infected-and-infectious' individuals occurs either by recovery of the disease (by acquiring immunity) or by death. Moreover, Kendall [2] suggests that the 'severity' of the pandemic, indicated by the parameter ζ, is obtained from the equation

$$\zeta = 1 - e^{-\sigma\zeta/\rho} \tag{1}$$

which always has the root $\zeta = 0$ and a unique positive root if and only if $\sigma > \rho$, which is the pandemic threshold condition ([2]). According to Diekmann [4], this interpretation is similar to the hair-trigger effect described by Aronson and Weinberger [8]. No matter how little infectivity is introduced in an arbitrarily small subset of the plane, eventually there will be a large effect at every point, provided the threshold is exceeded [4].

During the past decades, several alternative models were presented [9, 10], including models for the spreading of biological populations in spatially non-uniform conditions (e.g. [11]). Also, models have been worked out to encompass other aspects of the propagation of infectious diseases [12,13,14]. Finally, new paradigms were launched for the dispersion of individuals or particles, using topologically distinct approaches for defining the spatial connectedness of biological phenomena [15,16,17]. Diekmann and Heesterbeek [18] pointed to the difficulty of combining a deterministic description for the development of an epidemic outbreak (and especially the end of the outbreak) and a stochastic model that predicts new outbreaks. Models with combined deterministic and stochastic components are called 'hybrid models' [19,20]. At a demographic time scale, repeated epidemic outbreaks are well documented, e.g. in the case of measles and influenza [14,21,22]. Nevertheless, Diekmann and Heesterbeek know of only one paper in which a stochastic version of the Kermack-McKendrick ODE model is elaborated [20], which paper is an application of the asymptotic methods (i.e. approximate solutions) of the Fokker-Planck equation [18].

The impressive amount of mathematical papers found in literature, however, has not simplified the task of practitioners, policy makers or even biologists, in order to apply and interprete these mathematical tools for a given situation. And, although recent studies in mathematical and computational biology (e.g. [18]) have succesfully incorporated new biological, epidemiological and mathematical insights, for the average non-mathematical reader it is quite an elusive operation to grasp the ingredients of the mathematical argumentation. This is especially relevant when new outbreaks of known and unknown viruses are startling, like the epidemics of a recent past, like *foot-and-mouth-disease* (2000-2001), the *Severe Acute Respiratory Syndrome* (SARS) pandemic (2002-2003), the *Avian Influenza* epidemic (2004, 2014, …). Under these circumstances, comprehensive models to predict the geographical spread of infections may become matters of national priority.

([2]) This results from the monotone nondecreasing property of $e^{-\rho\zeta/\rho}$, provided that $\sigma > \rho$. The unique positive root may be found from the equation $\zeta = \frac{-\rho}{\sigma}\ln(1-\zeta)$.

The central notion in the geographical spread of contagious diseases is called the *contagion* of the disease (e.g. between a susceptible individual and an infectious agent). For predictive modeling, the geographical spread of this *contagion* has to be translated in mathematical terms. In order to deal with complex biological problems at an operational level which is easier to handle, the so-called *conjugacy principle* [23] constitutes a key approach. This conjugacy principle is synonymous to the *bypass principle*, which is a way of dealing with complexity by means of constructing a bypass which promotes a passage of the solution of a problem in a three-stage reduction process. The first and last stages are each other's inverses, so the bypass has the (abstract) form:

$$W = S\,T\,S^{-1} \qquad (2)$$

where W is called the conjugate of T under S

Although no perfect transformation exists in nature, when a conjugate is found for a given problem, the bypass principle requires the symmetry of the transformation process. In the present study, inverting the bypass is successful if the medical consequences of for instance the *contiguity* between susceptible and infectious agent can be read off from the predicted geographical spread of contagion. However, the complex notion of *contagion* in itself can be de-composed into few basic notions, like the notions of *norm (boundedness), continuity* and *connectivity*.

The most interesting aspect of the bypass principle is that it may be instructive when applying these notions from the field of epidemiology into the very different fields of biological science. Therefore, we will first analyse the use of these notions in mathematical models for the spread of infectious diseases (Section 2). Next, the biological interpretation of the theoretical outcome of some selected models will be explored (Section 3). In the last two sections, examples are given of the application of these notions in the fields of community ecology (Section 4) and multicellular pathway prediction (Section 5).

2. Mathematical background

2.1. Norm (boundedness)

One of the most basic notions in mathematical modelling with respect to spatial dispersion is the notion of 'norm'. Mathematically speaking, a norm $|\,V\,|$, for instance of a vector or a matrix (in $\Re_{n\,x\,n}$) is any mapping from $\Re_{n\,x\,n}$ into the real numbers $\Re$ which has the following properties:

i) $|\,V\,| = 0$ if and only if $V = 0$
ii) $|\,V + W\,| \leq |\,V\,| + |\,W\,|$ *(triangle inequality)*
iii) $|\,c\,V\,| = |\,c\,|\,.\,|\,V\,|$ for any V in $\Re_{n}$ and scalar c in $\Re$

(3)

Intuitively, it is understood that the norm of a vector represents its length, or, in Euclidean space $\Re_{n}$, the euclidean norm is simply the distance between two points. When enlarging our scope to all possible spaces, for instance when working with matrices in $\Re_{m\,x\,n}$, several norms like maximum (or box norm), euclidean and matrix norms can be used altogether. Some very useful results may come about. For instance, the euclidean norm of a matrix equals the square root of the *trace (tr)* of A* A (and: A* is the transpose of matrix A). When a non-symmetric matrix, say $A_{m\,x\,n}$, is multiplied by its transpose ($A_{n\,x\,m}$), a symmetric *(n x n)* matrix is the result. The *trace* of a symmetric matrix is the sum of the entries on its main diagonal. The euclidean norm, or $\sqrt{tr(A^{*}A)}$, is a current technique in scaling analysis like

Procrustes analysis [25,26]. According to Gower [26], scaling and rotation of matrices, in order to minimize the residual sum-of-squares (i.e. to find the best-fit configuration), can be seen as a multivariate form of analysis of variance.

Another important norm related to *m x n* matrices is the spectral norm ‖ A ‖. Let λ be the largest eigenvalue of A*A, then ‖ A ‖ is given by $\sqrt{\lambda}$ [24]. The spectral norm is used in the *convergence criterion* for infinite series of matrices.

For, if ‖ A ‖ ≤ a_k, and if $\sum_{k=0}^{\infty} a_k$ converges, then $\sum_{k=0}^{\infty} A_k$ also converges.

For *R* the radius of convergence of the real power series $\sum_{k=0}^{\infty} c_k x^k$, c_k being a sequence of real numbers, if ‖ A ‖ < *R*, then $\sum_{k=0}^{\infty} c_k A^k$ converges. Since the power series $1 + x + x^2 + \ldots$ converges for | x | < 1 and diverges for | x | > 1, the matrix power series $1 + A + A^2 + \ldots$ converges if ‖ A ‖ < 1 [24]. The latter rule is called the *convergence criterion* further on in this paper.

An example of the use of the supremum norm ([3]) for the analysis of the geographical spread of infections is found in Diekmann [4]. The Banach space on Ω ([4]), denoted as BC (Ω), is considered a convenient framework for this study when functions on this Banach space (C_T) are equipped with the supremum (sup) norm:

$$\| f \|_{C_T} = \sup_{0 \le t \le T} \| f[t] \|_{BC(\Omega)} \tag{4}$$

In section 3, we will further discuss the biological interpretation that can be given to this construction. The requirement of the sup norm is also named the *boundedness criterion*. Diekmann [4] uses the sup norm, together with the continuity and Lipschitz condition (see section 2.1. *"Continuity"*), to demonstrate the existence of a unique solution *U* to the nonlinear integral equation describing the geographical spread of an epidemic. The Lipschitz condition and sup norm enable to construct a series of mappings from C_T into C_T:

$$R_j : u \rightarrow R_j u : R_j u = Q u + f$$

([3]) The supremum norm is related to the max or Čebyšev norm: ‖ x ‖ = max ({ | x_1 |, | x_2 | ... | x_n |}). Compared with the euclidean norm, the following relationship exists [24]:

$$\max | v_i | \le \sqrt{\sum_{i=1}^{n} v_i^2} \qquad \text{for any } V \text{ in } \Re_n \text{ with } V = (v_1, \ldots v_n)$$

Also, the supremum, resp. infimum, is defined by the convergence theorem of monotone increasing, resp. decreasing, series.

([4]) Let us recall that a Banach space is a completely normed space, i.e. a space where every fundamental or Cauchy sequence converges. A sequence a_n is called a Cauchy sequence (or fundamental sequence) if for each ε > 0 a natural number n_δ exists, so that d (a_n,a_m) < ε for all m,n ≥ n_δ. The Banach space thus defined by Diekmann is the space of bounded continuous functions $f : [0,T] \rightarrow BC(\Omega)$, *f* equipped with the sup norm, and Ω being a closed subset of $\Re_n$ [4].

The rationale of these mappings is pointed out when they are used in an *iteration procedure*. If we put $u_0 = f$, and subsequently $u_{n+1} = Q\,u_n + f$, then the limit of u_n for $n \to \infty$ in C_T converges to the unique solution u [4]. This follows from the boundedness of the mapping as formulated in Banach's contraction mapping theorem (see below). It is important to note that in the terminology of Diekmann, the function $Q\,u\,(t)$ represents the Volterra part of the nonlinear differential equation:

$$u(t,x) = \int_0^t \int_\Omega g(u(t-\tau,\xi))S_0(\xi)A(\tau,x,\xi)d\xi d\tau + f(t,x) \tag{5}$$

[4]

with $u(t,x) = -\ln\frac{S(t,x)}{S_0(x)}$

and $g(y) = 1 - e^{-y}$. The variables *S* and *A* in Diekmann [4] represent the density of susceptible individuals *(S)* and the infectivity (A) at a certain position *x* due to an infection at a position ε (or the *contagion*, see below for mathematical interpretation). The parameter τ is the 'age of illness' that one infective has acquired at the moment of infection [4].
In equation (3) the Volterra part thus is of the form:

$$Qu[t] \sim \int_0^t \int_\Omega (1 - \frac{S(t-\tau,\xi)}{S_0})S_0(\xi)A(...)d\xi d\tau$$

The major advantage of Diekmann's approach is that once a solution at some time $T \in (0,T_0]$ is found, an iteration procedure can be formulated following the steps:

i) $u_0 = f$
ii) $u_{n+1} = Q\,u_n + f$

The spatial function describing the *contagion* - or the distance between x and ε - is the more difficult part of Diekmann's model. It has the nature of a radial function,

$$W(x,\xi) = V(x-\xi)$$

and is used by Diekmann (1978) in an iterated convolution procedure. According to Diekmann (1978) it can be demonstrated that the only non-divergent solution of the infinite series of convolutions of *V* with itself ([5]), namely:

$$v(x) = \int_{\Re_n} v(\xi)V(x-\xi)d\xi$$

implicates that $v(x) \equiv c$ (constant function), or, the function *v(x)* describing the minimum of *{w(x),p}* with *w(x)* a nonlinear convolution equation of the distance function V(x-ξ) and some threshold *p*, itself represents a constant function or $w(x) \equiv p$.

([5]) The convolution transformation on two functions $f_1(t), f_2(t)$ is an application of operational calculus that makes use of the product of the Laplace transforms or images (F_1,F_2) given by

$$F_1(p)F_2(p) = L\left\{\int_0^t f_1(\tau)f_2(t-\tau)d\tau\right\} = \int^\infty e^{-pt}\left[\int_0^t f_1(\tau)f_2(t-\tau)d\tau\right]dt$$

In this formula, the Laplace transform *(L)* relates a real valued function *f(t)* to the infinite integral of the complex function $e^{pt}f(t)$, where $p = a + bi$ (and $a > 0$) (e.g. [27], p. 464).

Diekmann [4] believes the infimum norm p of the iteration procedure :

$$p \leq u = inf\, u[\infty]\,(x)$$

corresponds to the threshold of the epidemic (see Kendall's *pandemic threshold theorem*, [2]).

Thieme [3] uses a quite different approach to discriminate the geographical spread of an epidemic due to the contacts between infective individuals at the one hand, and the spread of an epidemic due to the 'self-increase' of the infectious agents at the other hand. This distinction, needless to say, is biologically very important, for instance in view of the differences between bacterial and viral infections, or between viral and parasitic diseases requiring eukaryotic vector species for their distribution. However, also Thieme [3] uses the sup norm to demonstrate that a maximum bounded solution exists for the final size of an epidemic.

2.2. Continuity

The notion of continuity seems as difficult to grasp as it is fundamental in topology and analysis. The ambition to provide a rigorous foundation for the notion of continuity, as opposed to discontinuity, and which notion appears indispensable for infinitesimal analysis and differential calculus, dates back to the work of Richard Dedekind (1831-1916) and Georg Cantor (1845-1918)([6]). In abstract terms, a function is called (locally) continuous in x, if for every environment V of $f(x)$ an environment U of x exists, such that f^{-1} of $V\,(f[x])$ is contained in $U(x)$. The continuity of a function f thus requires the continuity of its inverse f^{-1}. Dedekind [28] is very concerned and rather dissatisfied with the lack of 'a purely arithmetic and perfectly rigorous foundation for the principles of infinitesimal analysis' and criticizes the use of sheer geometric evidence. Also the concept of a limiting value (shortly limit) is attributable to Dedekind's inheritance, and this concept forms the basis of differential and integral calculus. Continuously differentiable functions can be used in an iteration procedure, in order to find a numerical solution of an equation when analytical solutions fail. Therefore, however, the *convergence criterion* must be fulfilled (see section 2.1). Or, the iteration function $x \rightarrow g\,(x) : x_{n+1} = g\,(x_n)$ has a unique solution, if and only if $g(x) < 1$ (*convergence criterion*).

The pendant of convergence and continuity criteria for iteration procedures in topological spaces is found in Banach's *contraction mapping theorem*. For a Banach space (see footnote [3]), the theorem states that a mapping T of the metric space M to itself has a unique fixed point in M. Banach's contraction mapping theorem can be applied to a set of differential equations (of continuously differentiable functions) to demonstrate that iteration procedures converge to a unique solution, referring to a fixed point of the mapping (see also [4]). Essential to the contraction mapping T is the Lipschitz condition [30], namely that there exists $0 < k < 1$ such that $d(T_x,T_y) \leq k\, d(x,y)$ for all x,y in M [31]. Contraction mappings, i.e. fulfilling the Lipschitz condition, thus represent a class of mappings that do not increase distances [32]. In section 3.2. we will further comment on the biological interpretation of Diekmann's [4] use of Banach's contraction mapping theorem.

2.3. Connectivity

The physical or biological concept of connectivity is quite distinct from the topological concept of connectedness, although the latter may have some use to give a formal definition of the biological concept (see also [33]). In topological terms, connectedness of a space means that whenever it is

([6]) "If space has at all a real existence it is **not** necessary for it to be continuous; many of its properties would remain the same even were it discontinuous. And if we knew for certain that space was discontinuous there would be nothing to prevent us, in case we so desired, from filling up its gaps, in thought, and thus making it continuous (...)" ([28]; *fide* [29], p. 575).

decomposed as a union A ∪ B of two nonempty subsets, then A and B have some point in common or some point of A (resp. B) is a limit point of B (resp. A) [34]. The property of connectedness in general depends on the dimension and the homotopy type of a space [34]. For compact spaces, i.e. topological spaces that are both closed and bounded, *n*-dimensional structures can be simplicially approximated by an *n*-dimensional simplex. The advantage of a simplicial approximation s_f of a function *f* is that this s_f is continuous and homotopic to *f*. Consequently, the continuity of the polyhedral structure, from which the simplex is derived, is no longer required.

Simplicial approximation, like for instance used in *graph theory* [35], indeed may become a useful approach to study the connectivity of biological networks, when no complete topological characterization of the biological objects in these networks is possible [33]. Examples are easily found in the connectivity patterns of food chains, disease transmission (infectivity) or intra-species communication. However, the incorporation of the biological determinants related to the 'connectivity process' remains the most tedious part of mathematical modelling in these biological systems.

Recently, spatial models have been worked for biological phenomena that handle with individuals as discrete units and that abandon the space-averaged concentration notions that are typical of mean-field diffusion models [15,36,37]. These spatial individual-based models, also called *correlation models*, are kindred to probabilistic cellular automata and have some advantages compared to the more classical mean-field models [15]. The mathematical structure of these correlation models is based on the idea that two individuals are neighbours if they regularly interact with each other. This relation is the result of the geographical distribution of individuals (essentially two-dimensional) or "it may represent some more complex interaction structure such as that seen in childhood diseases like measles or sexually transmitted diseases like HIV/AIDS" ([15], p. 103). Therefore, Rand concludes, although the global structure is two-dimensional the local structure can be of a higher dimension. The correlation equations are derived in a number of steps, including some approximations of the space-averaged numbers of neighbours of a site (being another site or an edge) and measures to incorporate biases in the correction terms. The latter are important to ensure that the remaining variation can be modeled as random noise and a stochastic differential equation can be obtained ([15], p. 105).

A variation on the above model is found in the so-called *contact process* [38]. Applied to the dynamics of the spread of infection, the population is represented as a lattice of susceptible *(S)* or infected *(I)* individuals. Infection and recovery take place at a given rate, defining the transmissibility *(β)* and recovery process *(ν)* of the disease. At a critical transmissibility β_c long-range correlations become important and the pair approximation of the contact process is only poor. Below the critical β_c the infection dies out [38]. According to Rand [15], comparison of different models allows for estimating the value of β_c for a particular disease.

In other studies, the connectivity of the network is defined in terms of the probability of two nodes to be linked to each other [16,17]. In a scale-free network this connectivity distribution forms a continuum between a power law or an exponential distribution [17]. Albert and Barabási [17] found that in evolving networks the probability of a node *i* to increase its connectivity *k* was described by the following differential equation:

$$\frac{\partial k_i}{\partial t} = (p-q)m\frac{1}{N} + m\frac{k_{i+1}}{\sum_j (k_j + 1)} \tag{6}$$

where *p, q, m* represent the probability to add new links, the probability to rewire new links and the number of links connecting a new node to the system, respectively. The system size *N* and the total number of links $\sum_j k_j$ vary with time according to a simple linear relationship, namely:

$N(t) = m_0 + (1 - p - q)\,t$ (with m_0 the initial number of 'isolated nodes')
and $\sum_i k_i(t) = (1 - q)\,2mt - m$.

Assuming that k_i changes continuously, and defining the unit of time in the model as one event (attempt to growth/rewire or new link), Albert and Barabási (2000) derive the following *connectivity distribution*:
$P(k) \approx \left[k + \kappa(p,q,m)\right]^{-\gamma(p,q,m)}$
where $\kappa(p,q,m) = A(p,q,m) + 1$ and $\gamma(p,q,m) = B(p,q,m) + 1$.
The connectivity distribution thus follows a power law, when A and B meet certain restrictions. The meaning of A can be obtained from translating the restrictions set to the validity of the latter expression, namely that $A(p,q,m) + m + 1 > 0$. For fixed p and m this restriction translates into:

$$q < q_{\max} = \min\left\{1 - p, \frac{(1 - p + m)}{(1 + 2m)}\right\} \tag{7}$$

The connectivity distribution as expressed by the power law (see above) thus only holds for $q < q_{max}$, i.e. the connectivity does not equal or exceed the peak value of q. For $q > q_{max}$, the equation for $P(k)$ crosses from a scale-free (power law) regime to an exponential. The value of B is defined by the values chosen for p,q,m, namely:

$$B(p,q,m) = \frac{2m(1-q) + 1 - p - q}{m} \tag{8}$$

From the preceding analysis, we especially retain the property of an evolving network at a low connectivity state, corresponding to biological networks that are immature or not fully grown: in that case the probability distribution follows a power law instead of an exponential distribution.

3. Biological interpretation

3.1. Idealization

Application of mathematical principles and models to biological systems necessarily occurs through a process of abstraction and idealization of the biological system [33,39,40,41]. This also holds for mathematical models describing the geographical spread of an epidemic, or the geographical spread of an invading species or biological infection [42].
For instance, for Diekmann's [4] application of Banach's contraction mapping theorem, the uniform boundedness and continuity of the mapping functions (describing the biological infectivity) are crucial. To meet the requirement of boundedness (see section 2.2), the influence of boundaries has to be incorporated in the infectivity function. According to Diekmann [4] this means that the boundary has no active influence on the spread of the infection outside a certain habitat. Diekmann's model therefore especially relies on infections in a bounded environment, whereas anisotropic boundary effects are considered less important. The continuity with respect to the geographical spread of the susceptible species (denoted by x) is another idealization of the Diekmann model (for a discussion see [11]). The geographical spread of the infectious agent (ξ) in the Diekmann model [4] is normalized to unity, suggesting that no coupling exists between the spatial parameters x and ξ.
A distinct type of idealization is found in Thieme [3] in order to describe the habitat where spreading of the epidemic takes place. As the habitat for the susceptible population, Thieme [3] chooses a Borel measurable subset of $\Re_n$ (n= 1,2,3,...). Borel measurable subsets mean that an infinit, countable set of open or closed subsets can be found (with cardinal number ≤ the cardinal of the natural numbers). If the Borel measurable subsets are also bounded, they are called Lebesgue-measurable. Bounded functions are measurable, even when they are not continuous, because discontinuous measurable functions can be

approximated by continuous ones. The use of Borel measurable subsets, enables Thieme [3] to restrict solutions of the epidemic to the functions on an ordered topological space ($\aleph$), obtained by the union M_+ of bounded subsets provided with the pointwise convergence topology. Using monotone increasing (decreasing) sequences $\{X(u_n)\}$ that converge towards a sequence $X(u)$, and taking benefit of Banach's contraction theorem (see section 2.3), the fixed point properties of these sequences enable to imply the existence of an isotonic (resp. antitonic) compact operator on the topological space $\aleph$. The isotonic (resp. antitonic) compactness implies convergence towards a minimum (resp. maximum) solution on this topological space. According to Thieme [3], the extrapolation from the union of bounded measurable subsets M_+ to $\Re_n$ is biologically important in order to be able to analyse the intensity of the infectious influence of an epidemic when far removed from its origin. The slightly different framework used by Diekmann [4] is that of the Banach space of bounded continuous functions on a closed subset of $\Re_n$ (see section 2.1).

3.2. Functionality

In biological systems, material distances between two objects are usually the euclidean distance in E^3. When mapping functions that diminish euclidean distances are implicated (see section 2.2.), a 'biologically' functional interpretation ([7]) of these mappings is essential for biological relevance. A general concern is to avoid entering infinities (singularities) into the model. Similar to the *horror vacui*, there is also the scientific horror of infinite regress ([23], p. 15). In models for the spread of epidemics, we prefer not to ascribe massive invasions of infectious agents to an unknown, extra-terrestrial origin, neither do we explain full recovery from disease to some miraculous effect or *Deus ex machina*.
According to Diekmann [4], biological relevance for instance is found in the hair-trigger effect of a biological infection: "no matter how little infectivity is introduced in an arbitrarily small subset of Ω, eventually there will be a large effect at every point" ([4], p. 119). Both Thieme [3] and Diekmann [4] claim that their models demonstrate Kendall's pandemic threshold theorem: the biological threshold to an all or nothing geographical spread of an epidemic. It is important to note that in Diekmann's model both biological interpretations, the threshold phenomenon and the hair-trigger effect, are related to the same condition, namely that: $\gamma s_0 > 1$

in the equation $x_\infty = \gamma . s_0 (1 - e^{-x_\infty}) + f(\infty)$

Putting $x = \inf_{f(\infty)>0} x(\infty)$, then x satisfies the homogenous equation

$$y = \gamma . s_0 (1 - e^{-y})$$

which can only be positive for $\gamma s_0 > 1$ ([4], p. 118).

Knowing that $s_0 = \inf_{x \in \Omega} S_0(x)$ and $\gamma = \inf_{x \in \Omega} \int_{\Omega} \int_{0}^{\infty} A(\tau, x, \xi) d\tau d\xi$

indicating respectively the infimum of the density of infectives (S_0) at time 0 in x, and the infimum of the total infectivity at x due to a homogenously distributed density of infected individuals during the whole course of the disease. Diekmann's [4] model thus requires both infectivity and density of infectives (at homogeneity) to be non-zero and positive, in order to have a pandemic or general epidemic. If $\gamma s_0 \leq 1$, the epidemic will not spread.

Finally, the biological relevance of Diekmann's [4] and Thieme's [3] models can be inferred from the occurrence of traveling wave solutions. Traveling wave solutions are solutions of the form $u(t,x) = w(x$

([7]) 'Biologically' functional has quite a different connotation than the mathematical 'function' concept: it is rather comparable to the 'functor' notion in algebraic topology, without the bijective property (see also [43]).

+ ct) , thus depending on the linear combination of the independent variables *x* and *t*. Therefore, a substitution of these variables of the form $\xi = x+ct$
is recommended, which leads to the solution of the nonlinear convolution integral equation:

$$w(\xi) = \int_{-\infty}^{\infty} g(w(\eta)).V_c(\xi - \eta)d\eta \qquad \text{with } \xi = x+ct \text{ and where}$$

$$V_c(\xi) = \int_{0}^{\infty} H(\tau).V(\xi - c\tau)d\tau \qquad \text{[4] (see also footnote [4]).}$$

$H(\tau)$ denotes a non-negative distribution parameter on the time axis, for which $\int_0^{\infty} H(\tau)d\tau = 1$

The existence of a non-trivial solution for V_c depends on the value of *c* ([4], p. 125).

Before to proceed with the extrapolation of these functionalities to any biological system, it is however important to validate some of the assumptions made to derive the model. Among the assumptions formulated in Diekmann [4], we retain the following shortlist: *(a)* there are no changes in the susceptible population due to birth or migration (p. 110); *(b)* the disease induces permanent immunity, or susceptible individuals, once infected, cannot become susceptible again (p. 116); *(c)* the habitat space Ω does not consist of parts which are isolated with respect to infection (p. 119); *(d)* there is always a fraction of the susceptible population that escapes from being infected (p. 117). The combination of these 4 assumptions, taken together with the interpretation that all susceptible individuals "walk around during the day and return to their homes for the night" ([4], p. 110), yields a biological system which is neither realistic nor generally applicable (see e.g. [11,42] for more realistic approximations).

3.3. Extrapolation

Basically, extrapolation of results obtained from mathematical modelling is achieved in either one - or a combination - of the following approaches: generalizing the model by dropping one or more of the model assumptions (see examples in section 3.2), or extending the model by applying to a different, less restricted domain (or habitat, in the present case).

Similarly, in the posthumously published work of A.M. Turing on morphogenesis in plants [44], the functions describing the geometrical patterns of leafs *(phyllotaxis)* on a cylinder are considered instrumental to describe diffusion patterns in the plane. Or, according to Turing, functions in the plane having the symmetry of a lattice may be considered as obtained by unrolling the surface of a cylinder [45]. Applications are found in the geographical spread of an epidemic, Turing suggests. The situation where a lattice is changing with time may not only be interesting to describe the change of a phyllotactic pattern during plant growth, it may also apply to a changing lattice in the plane [46]. Turing uses the notion of 'flow matrices' that can be imagined to picture the change in the lattice as being due to "leafs being carried over the surface of a lattice by a fluid (whose velocity is a linear function of position)" ([44], p. 75).
However, the calculations of the functions describing these processes, according to Turing, were very difficult, and, moreover, before the advent of electronic computers in the following decades, they were practically unsolvable (see also [7]).

According to Diekmann and Heesterbeek [18], it is the combination of deterministic behaviour describing the fading out of an epidemic with the stochastic processes of new outbreaks, that causes the biggest methodological challenge. In other network models [16,17], random graphs are used to demonstrate that fading of an epidemic is not due to infections falling below a critical threshold. On the contrary, it is suggested that due to the occurrence of highly connected nodes, the latter may serve as

reservoirs from which new outbreaks eventually emerge. Scale-free networks are considered "extremely heterogeneous, their topology being dominated by a few highly connected nodes (hubs), which link the rest of the less connected nodes to the system" ([16], p. 651).

In the following sections, the application of the previous mathematical notions and methods into very different fields of the biological sciences are broached, namely the fields of community ecology and pathway prediction in multicellular organisms. A central problem in these new applications is the scarcity of available empirical data to corroborate the assumptions of the theoretical models, as for instance in the issues of biodiversity and the predictability of tumour metastasis [47,48].

4. Community Ecology

Two decades ago, it was already noted that "something was wrong in community ecology", that was linked with the gap between theory and empirical work (reviewed in [49]). Not only the lack of agreement about the meaning of basic terms and the confusion about the theoretical foundations seemed at stake. Also the quantitative analysis, aiming at the description of the processes of how the abundance of one species affects the abundance of another, appeared not suitable for describing how the species of a community change altogether [49], leaving aside the question how biodiversity of entire ecosystems changed and may further evolve in the present era [47]. Only a 'minimum' number of species, say a few predators and prey species, so far could be readily modelled. Moreover, many of the issues related to the turnover of species in natural communities, according to Law [49], are qualitative rather than quantitative. Also in the theories derived from Markovian models, the description of the community state is essentially qualitative and even subjective, namely, it may be based on the presence or absence of certain indicator species [49]. However, there exist an indirect way of measuring the effect of reduced biodiversity in models where the skewness of antigen exposure is essential for understanding the development of immunoglobulin-E mediated allergic diseases [50]. A possible way out of the impasse described above, concerning the qualitative changes in species composition, according to Law [49], may be found in considering only those subsets of species (from a regional pool), that have the property of **persistence.** This means that only the species are preserved that persist, i.e. that "have more than a transient existence". Taking the threat of a global declining biodiversity in mind, an alternative approach could envisage the interactions that are crucial for survival and to focus especially on the species that are threatened with extinction. It is one of our next project goals to pursue this approach to find an alternative method for the estimation of biodiversity or an alternative formulation of the mean species abundance (MSA) as an index of global biodiversity decline [47].

Following the persistence approach, Law [49] developed criteria for coexistence of n species, using a system of coupled ordinary differential equations (ODE) in an n-dimensional non-negative phase space. The criterion for coexistence of n species, according to common practice in theoretical ecology, is to evaluate the Jacobian matrix (J) in order to establish the existence of an equilibrium point $\hat{\mathbf{z}}$ with the property of *asymptotic Lyapunov stability* ([8]). The test for asymptotic stability consists of evaluating the eigenvalues of J at the equilibrium point, of which the real parts should all be strictly negative:

([8]) A solution of a differential equations used for describing dynamical systems is called **Lyapunov** stable if the solutions that start near an equilibrium point $\hat{\mathbf{z}}$, stay near $\hat{\mathbf{z}}$. If, moreover, a solution $\hat{\mathbf{z}}$ is Lyapunov stable and all solutions that start near $\hat{\mathbf{z}}$ converge to $\hat{\mathbf{z}}$, then $\hat{\mathbf{z}}$ is called asymptotically stable.

$$J = a_{ij} \qquad \text{with} \qquad a_{ij} = \frac{\partial}{\partial x_i}\left(x_i \cdot f_i(x)\right)\Big|_{x=\hat{z}} \tag{9}$$

A serious drawback of this method, according to Law [49], is that the property only applies for a small region around the equilibrium point, where the dynamic equations can be linearized. An alternative for the property of asymptotic stability (in the sense of Lyapunov) is to replace it by the more powerful alternative of *global asymptotic stability* ([51]; reviewed in [49]). In this approach, the property of global asymptotic stability implies that "no orbit can tend from the interior to the boundary", and the species therefore will coexist. *Permanence* of a species in this system is defined as "a 'skin' of thickness $\delta > 0$ around the boundary of the phase space". ODE's are then said to be permanent if "all orbits not initially in the boundary remain at least at a distance $d > \delta$ from the boundary" [49]. Following this approach, some aspects of community dynamics can be predicted, as permanence forms a criterion for persistence, although it still has its limitations ([49], p. 152). Nevertheless, it might help explaining why invader species might become permanent inhabitants, or why invasion resistance might prevent them from doing so. Also, the permanence algorithm for community assembly is supportive to the formulation of a final phase of (ecological) succession, consisting of two possible kind of end states, that are also corroborated by ecological findings. The first is that "a resident community is uninvadeble by any other species from the regional pool". The second is consistent with the "union of three or more subsets", which are communities that replace one another in a cyclic sequence (or when more complicated: a heterocyclic sequence) ([49], p. 164-165). It is obvious that these modelling studies may become helpful in describing ecologically stable communities, which show a diversity similar to the wide range of ecological habitats of the planet. The so far unwitnessed decline of global biodiversity however, probably with an anthropogenic signature, remains very hard to harmonize with what may be intrinsically called 'stability' analysis.

5. Pathway prediction in multicellular organisms

The analogy of transfer, from the assembly of biological species in an ecological community to a 'community' of cells in a multicellular organisms, seems logical but summons two important caveats. First, the stability of the multicellular organism is only a relative stability, because individuals are not only subject to a limited life cycle, but also diseases, dysfunctional behaviour of groups of cells or malignant cell proliferation may alter the survival of parts of the organism as well as of the individual as a whole 'micro-cosmos'. Moreover, the stability of the multicellular organism is not a goal in itself. Rather its stability reflects the qualitative appearance and performance of the organism in terms of an individual's well-being, and as such, it is the main purpose of medical care. One may argue that this approach is similar to the viewpoint that biodiversity is not solely a measure of numerical density of the total pool of species, but also a matter of functional performativity of ecosystem diversity, as it is perceived in various aspects of human culture, economy, and public health [47].

The question of a sustainable, healthy individual therefore surpasses the questions of morphogenesis (of a species and of the individual) as well as questions related to the various attempts to free the body of unwanted cells (of different kinds). Rather it poses the problem of sustainability of a heterogeneous, differentiated population of billions of cells in an functionally organized way, the healthy organism. The predictability of molecular and cellular transitions following distinct cellular pathways, indeed forms the main concern in cell reprogramming (stem cell regeneration) and the control of tumour metastasis [48].

In order to model cell lineage transformations in a multicellular organism, the problem of state transitions in theory can be approached by the use of Markovian models. In mathematical terms, Lyapunov functions can be useful, as suggested by Law [49]:

$P(x) = \prod_{i \in S} x_i^{h_i}$ reflecting the probability of a series of events, for some choice of $h_i > 0$.

In the multicellular environment of an organism, an event can be defined as either a molecular interaction between biomolecules or as a migration event of a cell, as for instance in circulating immune cells or metastatic tumour cells [48]. The probability of an event, drawn from a multitude of possible interactions or migration steps, has to exceed a threshold value to be a real possibility. Herein, clinical and/or metabolic data are essential to obtain information on the incidence of possible interactions (and diagnostic cues for visualization of these events).

In an *n*-dimensional Euclidean space this results in a set $\mathcal{M}$ of *n x n* matrices, for which a boundedness norm has to be defined, similar to the problem of asymptotic stability in community ecology (see ⁋ 4). In practical applications, however, finding a common Lyapunov function may become increasingly hard as the dimension *n* goes up [52]. The problem is also known as the approximation method of the *Joint Spectral Radius (JSR)*. The *JSR* represents the maximum growth rate obtained by taking arbitrary products of the matrices A_i from the iteration series $x_{k+1} = A_{\sigma(k)} x_k$ where the index $\sigma(k)$ results from a mapping from the integers to a finite set of indices {1,… m}. The *JSR* is formally defined as:

$$\rho(A_1, \ldots, A_m) := \lim_{k \to \infty} \max_{\sigma \in \{1, \ldots, m\}^k} \left\| A_{\sigma k} \ldots A_{\sigma 2} A_{\sigma 1} \right\|^{1/k} \tag{10}$$

According to Parrilo & Jadbabaie [52], a sum of squares (SOS) approximation shows the best results in order to obtain a class of bounds on the JSR that guarantee contractiveness properties for all the matrices in the set $\mathcal{M}$ (see also the notion of Banach contraction mapping, ⁋ 2.2). Namely, it seems possible to obtain upper bounds on the JSR by replacing norms with homogenous polynomials $p(x)$ (of degree $2d$), following the topological properties of compactness (of the unit ball in $\mathfrak{R}^n$) and continuity of a strictly positive homogeneous polynomial $p(x)$ (see also ⁋ 2.3).

Applications of this methodology are found in the metaheuristics of Ant Colony Optimization (ACO) algorithms, which have been proven to be useful in complex optimization problems, such as protein side-chain conformation and protein-protein interaction analysis [53]. It is suggested that with a number of adaptations of ACO to the context of multicellular networks, the heuristics of ACO could also become useful in the analysis of cellular pathway prediction [48].

6. Concluding Remarks

Starting from a historical survey of the modelling of infectious diseases, as originally formulated in the days of Volterra, Turing and Kendall, we analysed the abstract notions of norm/boundedness, continuity and connectivity throughout several areas of bio-mathematical research. These areas are not confined to the classical, mathematical roots of epidemiology, but also to some more recent applications such as community ecology and cellular network modelling.

On the one hand, it may be noted that the achievements and possibilities of mathematics and computation technology have generated the notion of computability of almost any complex problem. On the other hand, the limited availability of empirical data, from the field, or from the clinic or metabolic data bases, still hampers the practical usefulness of the mathematical models. As mentioned earlier, a thorough analysis of the biological issues at stake in for instance a declining biodiversity, environmental degradation or individual health impairment, may become instrumental for the fine-tuning of the bio-mathematical and computation techniques.

References

[1] Kermack W O and McKendrick A G 1927. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society (London), series A,* **115** 700-21.
[2] Kendall D G 1957. Discussion of 'Measles periodicity and community size' by M.S. Bartlett. *Journal of the Royal Statistical Society, series A,* **120** 64-7.
[3] Thieme H R 1977. A model for the spatial spread of an epidemic. *Journal of Mathematical Biology* **4** 337-51.
[4] Diekmann O 1978. Thresholds and travelling waves for the geographical spread of infection. *Journal of Mathematical Biology* **6** 109-130.
[5] Volterra V 1926. Variazione fluttuazioni del numero d'individui in specie animali conviventi. *Mem. Acad. Lincei* **2**: 3-113, translation in: ed R N Chapman (1931), *Animal Ecology* pp 409-48 (New York: McGraw Hill).
[6] Turing A M 1952. The chemical basis of morphogenesis. *Phil. Trans. R. Soc. London (B)* **237** 37-72.
[7] Allaerts W 2003. Fifty years after Alan M. Turing. An extraordinary theory of morphogenesis. Belgian Journal of Zoology **133** 3-14.
[8] Aronson D G and Weinberger H F 1975. Nonlinear diffusion in population genetics, combustion, and nerve pulse propagation. In: ed J A Goldstein, *Partial differential equations and related topics, Lecture notes in Math.* vol 446 pp 5-49 (Berlin: Springer).
[9] Van den Bosch F, Metz J A J and Diekmann O 1990. The velocity of spatial population expansion. *Journal of Mathematical Biology* **28** 529-65.
[10] Van den Bosch F, Hengeveld R and Metz J A J 1992. Analysing the velocity of animal range expansion. *Journal of Biogeography* **19** 135-50.
[11] Hengeveld R and Van den Bosch F 1997. Invading into an ecologically non-uniform area. In: B Huntley, et al. (eds.), *Past and future rapid environmental changes: the spatial and evolutionary responses of terrestrial biota. NATO AS1 Series,* vol 147 pp 217-25 (Berlin, Heidelberg: Springer).
[12] Diekmann O 1979. Run for your life. A note on the asymptotic speed of propagation of an epidemic. *Journal of Differential Equations* **33** 58-73.
[13] Cliff A D, Haggett P, Ord J D and Versey G R 1981. *Spatial Diffusion* (Cambridge: Cambridge University Press).
[14] Murray J D 1989. Geographical Spread of Epidemics, chapter 20. In: ed J D Murray, *Mathematical Biology* (Berlin, Heidelberg: Springer).
[15] Rand D A 1999. Correlation equations and pair approximations for spatial ecologies. In: ed J McGlade, *Advanced Theoretical Ecology: Principles and Applications* pp. 100-42 (London: Blackwell Science).
[16] Jeong H, Tombor B, Albert R, Ottval Z N and Barabási A-L 2000. The large-scale organization of metabolic networks. *Nature* **407** 651-54.
[17] Albert R and Barabási A-L 2000. Topology of Evolving Networks: Local Events and Universality. *Physical Review Letters* 85 5234-37.
[18] Diekmann O and Heesterbeek J A P 2000. *Mathematical Epidemiology of Infectious Diseases. Model building, analysis and interpretation* (New York: John Wiley and Son).
[19] Nåsell I 1985. *Hybrid Models of Tropical Infections* (Berlin: Springer).

[20] Nåsell I 1995. The threshold concept in stochastic epidemic and endemic models. In: ed D Mollison, *Epidemic models: their structure and relation to data* (Cambridge: Cambridge University Press).
[21] Bartlett M S 1957. Measles periodicity and community size. *Journal of the Royal Statistical Society, series A,* **120** 48-60.
[22] Cliff AD, Haggett P and Smallman-Raynor M 1993. *Measles: an historical geography of a major human viral disease from global expansion to local retreat, 1840-1990* (London: Blackwell).
[23] Melzak Z A 1983. *Bypasses. A simple approach to complexity* (New York: John Wiley and Sons).
[24] Brinkmann H W and Klotz E A 1971. *Linear algebra and analytic geometry* (London, Reading, Menlo Park: Addison-Wesley Publishing Company).
[25] Schönemann P H and Carroll R M 1970. Fitting one matrix to another under choice of a central dilation and a rigid motion. *Psychometrika* **35** (2) 245-55.
[26] Gower J C 1975. Generalized Procrustes analysis. *Psychometrika* **40** (1) 33-51.
[27] Piskounov N 1980. *Calcul différentiel et intégral,* tome II (Moscow: Éditions Mir).
[28] Dedekind R 1872. Stetigkeit und irrationale Zahlen (Braunschweig, 1872); Eg. translation by W W Beman, Continuity and irrational numbers, in: R Dedekind (1963), *Essays on the theory of numbers* (Dover).
[29] Fauvel J and Gray J (1987)(eds.). *The History of Mathematics* (London, Milton Keynes: Mac Millan Press & The Open University).
[30] Lipschitz R 1877. *Lehrbuch der Analyse* (Bonn).
[31] Aksoy A G and Khamsi M A (1990). *Nonstandard methods in Fixed Point Theory* (Berlin, New York: Springer).
[32] Kirk W A 1965. A fixed point theorem for mappings which do not increase distances. *American Mathematics Monthly* **72**: 1004-6.
[33] Allaerts W 1999. Local and global patterns during morphogenesis of the retinotectal topographical mapping in the vertebrate brain. *Acta Biotheoretica* **47** 99-122.
[34] Armstrong M A 1979. *Basic Topology* (Berlin: Springer).
[35] Wilson R J 1975 (2nd ed.). *Introduction to graph theory* (London: Longman).
[36] Keeling M J and Grenfell B T 1997. Disease extinction and community size: modeling the persistance of measles. *Science* **275** 65-7.
[37] Keeling M J, Rand D A and Morris A J 1997. Correlation models for childhood epidemics. *Proceedings of the Royal Society of London, series B* **264** 1149-56.
[38] Bezuidenhout C and Grimmett G 1990. The critical contact process dies out. *Annals of probability* **18** 1462-82.
[39] Stewart I and Golubitsky M 1992. *Fearful Symmetry. Is God a Geometer?* (London: Blackwell Publishers, Penguin Books).
[40] Presnov E V and Isaeva V V 1990. Local and global aspects of biological morphogenesis. *Speculations in Science and Technology* **13** 68-75.
[41] Allaerts W and Roelants H 1993. Positional information limits the self-explaining endeavour in morphogenetic theory (in the sense of Turing). Towards the understanding of the functioning of biological forms. *Belgian Journal of Zoology* **123** 263-82.
[42] Hengeveld R 1989. *Dynamics of biological invasions* (London, New York: Chapman & Hall).
[43] Allaerts W 1999. The biological function paradigm applied to the immunological self-non-self discrimination: critique of Tauber's phenomenological analysis. *Journal for General Philosophy of Science* **30** 155-71.
[44] Saunders P T 1992. *Collected works of A.M. Turing*. vol. 3: *Morphogenesis* (Amsterdam, London: Elsevier Science Publishers).
[45] Turing A M 1992 (posthum.). Morphogen theory of phyllotaxis. In: ed P T Saunders, *Collected works of A.M. Turing*. vol. 3: *Morphogenesis*, pp. 49-123 (Amsterdam, London: Elsevier Science Publishers).

[46] Swinton J 2013. Turing, Morphogenesis, and Fibonacci Phyllotaxis: Life in pictures. In: eds S B Cooper and J van Leeuwen, *Alan Turing: His Work and Impact*, pp. 834-849 (Amsterdam, Boston, London: Elsevier).
[47] Allaerts W 2018. Why is biodiversity of cardinal importance for public health? *International Journal of Environment & Agricultural Science* **2** (1): 013.
[48] Allaerts W 2018. Annotation and predictability of cellular pathways: III. Computability and potential use of parallel Ant Colony Optimization algorithms. *Functional and Structural Genomics and Medicine* **1** (1): 102-9.
[49] Law R 1999. Theoretical aspects of community assembly. In: ed J McGlade, *Advanced Theoretical Ecology: Principles and Applications* pp. 143-171 (London: Blackwell Science).
[50] Allaerts W and Chang T W 2017. Skewed exposure to environmental antigens complements hygiene hypothesis in explaining the rise of allergy. *Acta Biotheoretica* **65** (2) 117-34.
[51] Goh B S 1977. Global stability in many-species systems. *American Naturalist* **111** 135-42.
[52] Parrilo P A and Jadbabaie A 2008. Approximation of the joint spectral radius using sum of squares. *Linear Algebra and its Applications* **428** 2385-402.
[53] Lü Q, Xia X Y, Chen R, Miao D J, Chen S S, Quan L J, et al. 2012. When the lowest energy does not induce native structures: Parallel minimization of multi-energy values by hybridizing searching intelligences. *PLOS One* **7** (9): e44967.