GBIO 2060

Lectures on Mathematical Modelling of Biological Systems

G. Bastin

August 22, 2018

Contents

Lecture 1

Dynamical Modelling of Infectious Diseases

1.1 Introduction

The aim of this lecture is to give an elementary introduction to mathematical models that are used to explain epidemiologic phenomena and to assess vaccination strategies. We focus on *infectious diseases*, i.e. diseases where individuals are infected by pathogen micro-organisms (like, for instance, viruses, bacteria, fungi or other microparasites). Some well known examples of such infectious diseases are :

- Viral infectious diseases : AIDS, Chickenpox (Varicella), Common cold, Cytomegalovirus Infection, Dengue fever, Ebola hemorrhagic fever, Hepatitis, Influenza (Flu), Measles, Mononucleosis, Mumps, Poliomyelitis, Rubella, SARS, Smallpox (Variola), Viral meningitis, Viral pneumonia, West Nile disease, Yellow fever.
- Bacterial infectious diseases : Cholera, Diphtheria, Legionellosis, Leprosy, Lyme disease, Pertussis (Whooping Cough), Plague, Pneumococcal pneumonia, Salmonellosis, Scarlet Fever, Syphilis, Tetanus, Tuberculosis, Typhus.
- Parasitic infectious diseases : Malaria, Taeniasis, Toxoplasmosis.
- Prion infectious diseases : Creutzfeldt-Jakob disease.

1.2 The basic SIR model

A first fundamental mathematical model for epidemic diseases was formulated by Kermack and McKendrick in 1927 (see the fac-simile of their paper in Appendix). This model applies for epidemics having a relatively short duration (compared to life duration) that take the form of "a sudden outbreak of a disease that infects (and possibly kills) a substantial portion of the population in a region before it disappears" (Brauer, 2005). In this model, the population is classified into three groups : (i) the group of individuals who are

uninfected and susceptible (S) of catching the disease, (ii) the group of individuals who are infected (I) by the concerned pathogen, (iii) the group of recovered (R) individuals who have acquired a permanent immunity to the disease. The propagation of the disease is represented by a compartmental diagram shown in Fig.1.1. The model is derived under

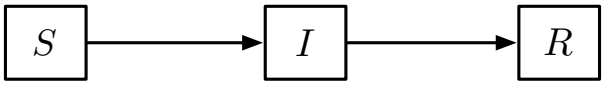

Figure 1.1: Compartmental diagram

three main assumptions : (i) a closed population (no births, no deaths, no migration), (ii) spatial homogeneity, (iii) disease transmission by contact between susceptible and infected individuals. The model is a system of three differential equations:

$$\frac{dS}{dt} = -\beta IS, \tag{1.1}$$

$$\frac{dI}{dt} = \beta IS - \gamma I, \tag{1.2}$$

$$\frac{dR}{dt} = \gamma I. \tag{1.3}$$

In these equations, S denotes the number of susceptibles, I the number of infected individuals and R the number of immune individuals at time t. The total population

$$N = S + I + R.$$

is constant by assumption since we have

$$\frac{dN}{dt} = \frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0$$

from the model equations.

In the first and second model equations (1.1)-(1.2), the term βIS represents the disease transmission rate by contact between susceptible and infected individuals. This rate is assumed to be proportional to the sizes of both groups with a proportionality coefficient β. In the second and third equations (1.2)-(1.3), the parameter γ is the specific rate at which infected individuals recover from the disease.

Let us consider an epidemic outbreak in a population where, at the initial time, only a few individuals are infected. The initial conditions are

$$S(0) \approx N, \;\; I(0) = N - S(0) \approx 0, \;\; R(0) = 0.$$

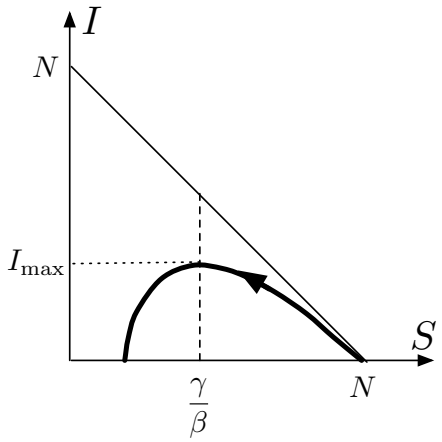

Figure 1.2: Epidemic trajectory

A typical trajectory of the system solution in the I-S phase plane is given in Fig.1.2. From this curve, a fundamental observation is the existence of a Threshold Effect. The maximum value of the curve occurs at $S = \gamma/\beta$. This means that an epidemic will start and amplify only if $S(0) \approx N$ is larger than γ/β or equivalently if

$$R_0 = \frac{N\beta}{\gamma} > 1.$$

Under this condition, the number of infectives will increase until the number of susceptibles is reduced to γ/β and will decrease thereafter. Thus the number R_0 represents a threshold for an epidemic to occur. In the literature, this number is also called *Basic Reproduction Ratio* because it represents the average number of susceptibles which are contaminated by one infective.

Dividing equation (1.2) by equation (1.1), we obtain

$$\frac{dI}{dS} = \left(\frac{\gamma}{\beta S} - 1\right).$$

Integrating this equation, we have

$$I = \frac{\gamma}{\beta}\log S - S + C \quad \text{with} \quad C \approx N - \frac{\gamma}{\beta}\log N. \tag{1.4}$$

From this equation, we can compute the instantaneous maximum number of infectives (see Fig.1.2) as

$$I_{\max} = N\left(1 - \frac{1 + \log R_0}{R_0}\right)$$

Equation (1.4) also implies that I must vanish at some positive value of S. This means that the trajectory terminates on the S-axis at a positive value as shown in Fig.1.2. Therefore,

the epidemic terminates before all susceptibles have become infected and some individuals escape the disease entirely. We can now determine how many susceptibles remain or equivalently the final value $R(\infty)$ of the immune population size.

Dividing equation (1.2) by equation (1.3), we have

$$\frac{dS}{dR} = -\frac{\beta}{\gamma}S \quad \Rightarrow \quad S(R) = S(0)e^{-\frac{\beta}{\gamma}R} \approx Ne^{-\frac{\beta}{\gamma}R}$$

Then

$$\frac{dR}{dt} = \gamma I = \gamma(N - S - R) = \gamma\left(N - Ne^{-\frac{\beta}{\gamma}R}\right).$$

Therefore

$$t \to \infty \quad \Longrightarrow \quad I \to 0 \quad \Longrightarrow \quad \frac{dR}{dt} = 0$$

$$\Longrightarrow \quad N\left[1 - e^{-\frac{\beta}{\gamma}R(\infty)}\right] = R(\infty) \tag{1.5}$$

Equation (1.5) has a unique solution $R(\infty)$ between 0 and N as long as $R_0 > 1$. We denote $x = R(\infty))/N$ the fraction the population that has contracted the disease before the epidemic collapses. By solving (1.5), we have the following relation between R_0 and x:

$$R_0 = -\frac{\log(1-x)}{x}.$$

An interesting application of the SIR model is reported in the book of Murray [1] on the basis of influenza epidemic data in an English boarding school published in 1978 by The British Medical Journal [2]. The epidemic lasted 22nd January to 4th February 1978. A large fraction of the N = 763 boys in the school were infected and are represented by dots in Fig.1.3. The curves in the figure represent solutions of the SIR model fitted to the data using least squares. The estimated parameter values are

$$\beta N = 1.66 \text{ per day}, \quad \frac{1}{\gamma} = 2.2 \text{ days}, \quad R_0 = 3.65.$$

1.3 Temporary immunity : the SIRS model for endemic diseases

In this section, we describe how the basic Kermack-McKendrick model is modified in order to describe how a disease in a population can persist when the immunity of recovered individuals is temporary (and not permanent as we have assumed in the previous section).

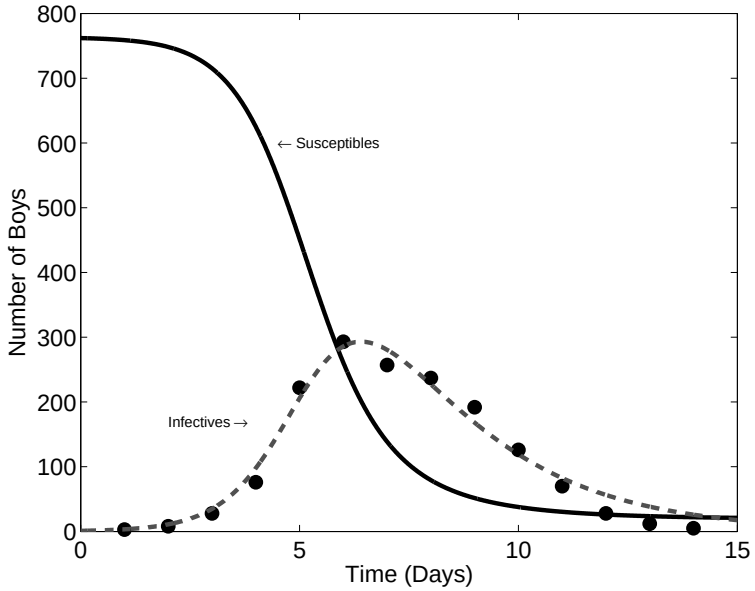

Figure 1.3: An influenza epidemic in an English boarding school in 1978 (reprinted from [3], see also http://www.modelinginfectiousdiseases.org/)

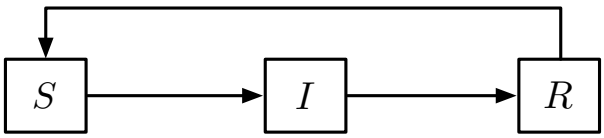

Figure 1.4: SIRS compartmental diagram

This is illustrated with the SIRS compartmental diagram of Fig.1.4. An additional parameter δ is introduced in order to represent the specific rate of immunity loss. The SIRS model is as follows.

$$\frac{dS}{dt} = -\beta IS + \delta R, \tag{1.6}$$

$$\frac{dI}{dt} = \beta IS - \gamma I, \tag{1.7}$$

$$\frac{dR}{dt} = \gamma I - \delta R. \tag{1.8}$$

As above the total population $N = S + I + R$ is constant ($dN/dt = 0$).

This system has two equilibria which are the two possible constant solutions of equations (1.6)-(1.7)-(1.8). The first one is the **disease free equilibrium** :

$$S^* = N, \quad I^* = 0, \quad R^* = 0.$$

The second one is the **endemic equilibrium** :

$$S^* = \frac{\gamma}{\beta}, \quad I^* = \frac{N - \left(\frac{\gamma}{\beta}\right)}{1 + \left(\frac{\gamma}{\delta}\right)}, \quad R^* = \frac{N - \left(\frac{\gamma}{\beta}\right)}{1 + \left(\frac{\delta}{\gamma}\right)}.$$

Obviously, the endemic equilibrium exists only if the numerators are strictly positive. This implies that the following condition must hold :

$$R_0 = \frac{N\beta}{\gamma} > 1.$$

Hence the condition for the existence of an endemic equilibrium in the SIRS model is in agreement with the condition for an epidemic to occur in the SIR model.

In order to analyse the system trajectories and the equilibrium stability, we consider the second order system obtained from equations (1.6)-(1.7) by substituting $R = N - S - I$:

$$\frac{dS}{dt} = -\beta IS + \delta(N - S - I),$$

$$\frac{dI}{dt} = \beta IS - \gamma I.$$

The Jacobian matrix of this system, evaluated at an equilibrium point (S^*, I^*) is

$$J = \begin{pmatrix} -(\beta I^* + \delta) & -(\beta S^* + \delta) \\ \beta I^* & (\beta S^* - \gamma) \end{pmatrix}. \tag{1.9}$$

An equilibrium is asymptotically stable if $\text{trace}(J) < 0$ and $\det(J) > 0$. Otherwise it is unstable.

For the *disease free equilibrium* $(S^* = N, I^* = 0)$ we have

$$\text{trace}(J) = \beta N - \gamma - \delta, \quad \det(J) = -\delta(\beta N - \gamma)$$

Consequently

if $R_0 < 1$, then $\text{trace}(J) < 0$, $\det(J) > 0$ and the disease free equilibrium is **stable**,

if $R_0 > 1$, then $\det(J) < 0$ and the disease free equilibrium is **unstable**.

For the *endemic equilibrium* which exists only if $R_0 > 1$, we have

$$\text{trace}(J) = -(\beta I^* + \delta) < 0, \quad \det(J) = \beta I^*(\beta S^* + \delta) = \beta I^*(\gamma + \delta) > 0$$

Consequently the endemic equilibrium is necessarily **stable** when it exists.

1.4 The SIR model with demography

We now reconsider the basic SIR model of Section 1.2 in the case where demographic effects are taken into account. Therefore, as it is illustrated with the compartmental diagram of Fig.1.5, births (or immigration) at the rate ν as well as deaths (or emigration) at the rate μ are introduced in the model:

$$\frac{dS}{dt} = \nu N - \beta IS - \mu S, \tag{1.10}$$

$$\frac{dI}{dt} = \beta IS - \gamma I - \mu I, \tag{1.11}$$

$$\frac{dR}{dt} = \gamma I - \mu R. \tag{1.12}$$

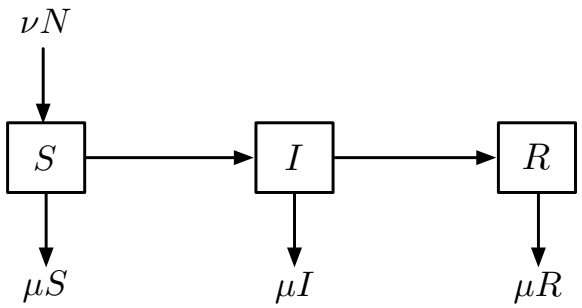

Figure 1.5: Compartmental diagram of the SIR model with birth (or immigration) and mortality (or emigration).

In order to have a constant total population $N = S + I + R$ ($dN/dt = 0$), we assume that $\mu = \nu$. The system has two equilibria and we shall see that the analysis is quite similar to the previous case. The first equilibrium is the **disease free equilibrium** :

$$S^* = N, \quad I^* = 0, \quad R^* = 0.$$

The second equilibrium is the **endemic equilibrium** :

$$S^* = \frac{\gamma + \mu}{\beta} \Rightarrow \frac{S^*}{N} = \frac{\gamma + \mu}{\beta N} = \frac{1}{R_0}, \quad I^* = \frac{\mu(N - S^*)}{\beta S^*} = \frac{\mu(R_0 - 1)}{\beta}.$$

Obviously, the endemic equilibrium exists only if $S^* < N$ and $I^* > 0$ which means that, as in the previous cases, the *Basic Reproduction Ratio* R_0 must be greater than 1:

$$R_0 = \frac{N\beta}{\gamma + \mu} > 1.$$

The Jacobian matrix of the sub-system (1.10)-(1.11), evaluated at an equilibrium point (S^*, I^*) is

$$J = \begin{pmatrix} -(\beta I^* + \mu) & -\beta S^* \\ \beta I^* & \beta S^* - \gamma - \mu \end{pmatrix}. \tag{1.13}$$

It is readily checked that the disease free equilibrium ($S^* = N, I^* = 0$) is stable if $R_0 \leqslant 1$ and unstable if $R_0 > 1$.

For the *endemic equilibrium* which exists only if $R_0 > 1$, we have

$$\text{trace}(J) = -(\beta I^* + \mu) = -\mu R_0 < 0, \quad \det(J) = \beta^2 I^* S^* = \mu(\gamma + \mu)(R_0 - 1) > 0$$

Consequently the endemic equilibrium is necessarily **stable** when it exists. Moreover, the endemic equilibrium is a *focus* if the following inequality holds:

$$4\det(J) > (\text{trace}(J))^2 \quad \Rightarrow \quad 4\mu(\gamma + \mu)(R_0 - 1) > \mu^2 R_0^2.$$

and the trajectories exhibit oscillations as shown in Fig.

1.5 Variants and generalisations

1.5.1 SIR model with vaccination

In this section, we explore how the SIR model of the previous section can be modified in order to explain how epidemic diseases can be eradicated by vaccination. A new parameter σ is introduced in the model which represents the specific vaccination rate of the newborns. The model is now written as follows.

$$\frac{dS}{dt} = (\mu - \sigma)N - \beta IS - \mu S, \tag{1.14}$$

$$\frac{dI}{dt} = \beta IS - \gamma I - \mu I, \tag{1.15}$$

$$\frac{dR}{dt} = \gamma I - \mu R. \tag{1.16}$$

The disease free equilibrium is:

$$S^* = (1 - \frac{\sigma}{\mu})N = (1 - p)N, \quad I^* = 0$$

where $p = \sigma/\mu$ is the fraction of the newborn population which is vaccinated. The endemic equilibrium is

$$S^* = \frac{\gamma + \mu}{\beta}, \quad I^* = \frac{\mu(N - S^*) - \sigma N}{\beta S^*}.$$

Table 1.1: Estimates of R_0 and of the corresponding vaccinated fraction p of the population to achieve eradication.

Disease	Estimate of Threshold R_0	Minimal p (%) for eradication
Smallpox	4	75
Measles	13	92
Whooping cough	13	92
Chicken pox	10	90
Diphteria	5	80
Poliomyelitis	6	83

The disease is eradicated if the disease free equilibrium is the only possible stable equilibrium and if the endemic equilibrium does not exist. This is achieved if the following condition holds:

$$p = \frac{\sigma}{\mu} > (1 - \frac{1}{R_0}).$$

It is remarkable that not everybody has to be vaccinated in order to prevent an endemic disease. This is called *herd immunity*. Typical examples of R_0 values and the corresponding critical level of vaccination are given in Table 1.1. It can be seen that the critical level of vaccination is about 80% only for severe diseases like smallpox and polyomelithis which have been eradicated in developped countries, while it is much higher for childhood diseases like measles and whooping cough (about 92%) which are fortunately much less severe.

1.5.2 SEIR model with a latent period

A latent period is a phase of the disease where individuals are already infected by the pathogens but not yet infectious (i.e. cannot yet transmit the disease to other people). The compartmental diagram is extended with an additional E compartment representing the "exposed" fraction of the population during the latent period as shown in Fig.1.6. An additionnal parameter η is introduced which represents the specific transfer rate from E to I. The derivation of the model equations is left as an exercise. For this model, the threshold parameter is

$$R_0 = \frac{\beta N}{\gamma + \mu} \frac{\eta}{\eta + \mu}.$$

1.5.3 Age structured SIR model

A very relevant issue is obviously to extend the basic SIR model to heterogeneous populations. There are infectious diseases (e.g. sexually transmitted diseases) where considering

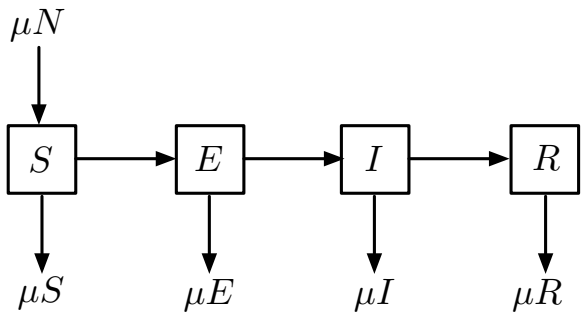

Figure 1.6: Compartmental diagram of the SEIR model with a latent period.

all susceptible and infected individuals as a single homogeneous group cannot sufficiently capture the dynamics of disease propagation. Individuals may differ in characteristics that are epidemiologically relevant. Traits as age, sex or genetic composition can influence susceptibility and infectivity of individuals. In order to account for such individual characteristics, it is useful to extend the dimension of the system, either by considering higher order systems of ordinary differential equations, or, in a more radical way, systems of partial differential equations (PDEs). As a matter of illustration, we present here the example of a PDE model which is derived under the assumption that individuals aging matters.

In this type of models, the dependent variables S, I and R are functions of both time and age. More precisely, the variable $S(t,a)$ represents the age distribution of the population of susceptible individuals at time t. This means that

$$\int_{a_1}^{a_2} S(t,a)da$$

is the number of susceptible individuals with ages between a_1 and a_2. Similar definitions are introduced for the age distributions $I(t,a)$ of Infected individuals and $R(t,a)$ of recovered individuals. Then a natural extension of equation (1.10) is to describe the dynamics of $S(t,a)$ by the following hyperbolic partial differential equation:

$$\frac{\partial S}{\partial t} = -\frac{\partial S}{\partial a} - \mu(a)S - \phi(t,a)S$$

$$\phi(t,a) = \int_0^{\infty} \beta(a,b)I(t,b)db$$

where $\beta(a,b)$ denotes the transmission coefficient by contact between a susceptible having age a and an infected having age b.

1.5.4 SIR model with time-varying parameters and periodic forcing

Various childhood diseases exhibit sustained periodic oscillations with pluriennal epidemic cycles which are not explained by the models that we have presented so far. Typical

examples are shown in Fig.1.7 for measles in Rekyavik (with a 4 year cycle) and chicken pox in New York (with an annual cycle). During the last twenty years, numerous scientific

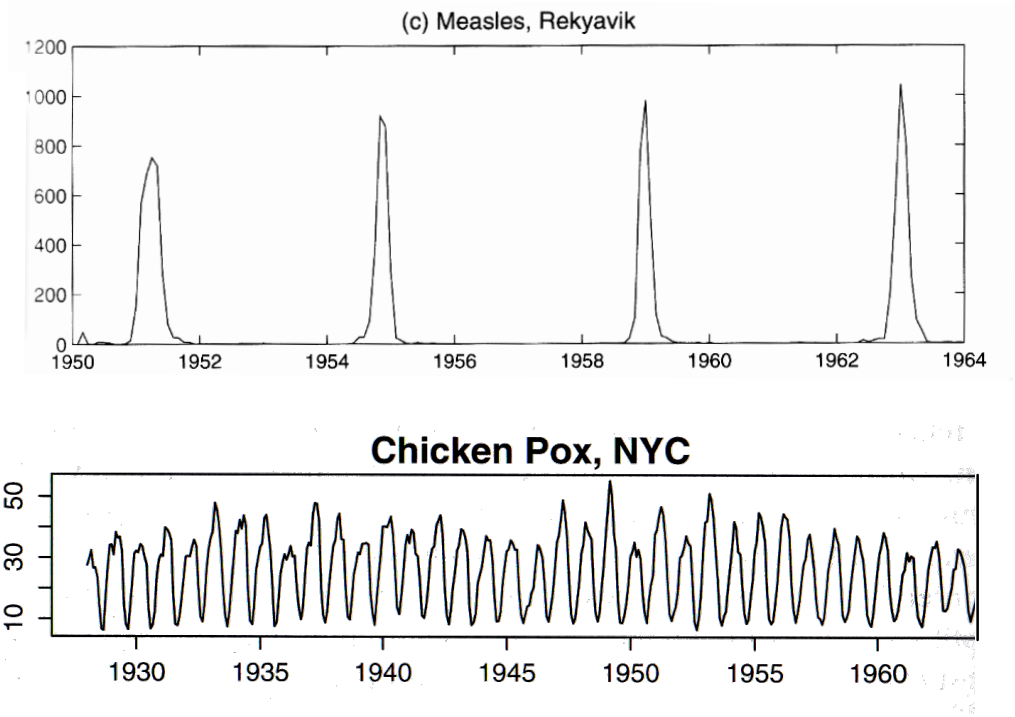

Figure 1.7: Observed notifications of measles in Rekyavik and chicken-pox in New-York before the vaccination era (Reprinted from [6])

studies have demonstrated that seasonal variations in disease transmission rates appear to be a major factor to explain sustained epidemic cycles. This effect of seasonality is illustrated in Fig.1.8 with the trajectory of a simple SIRD model where the transmission coefficient β is an annual sinusoidal forcing function

$$\beta(t) = \beta_0\big[1 + \beta_1 \cos(2\pi t)\big]$$

with the time t in years and $\beta_1 = 0.2$, (i.e. a 20% seasonal variation). It can be seen, in this example, that the annual forcing induces biennal epidemics. In fact, by changing the value of the birth-mortality rate quadriennal epidemics that are reminiscent to the observed data for Rekyavik in Fig.1.7 could be induced as well.

Now, if in addition to seasonal variations of transmission rate, time varying birth rates $\mu(t)$ are introduced in the model according to the real-life demography, then very realistic simulations are achieved as illustrated in Fig.1.9. After 1950, we see sustained biennal

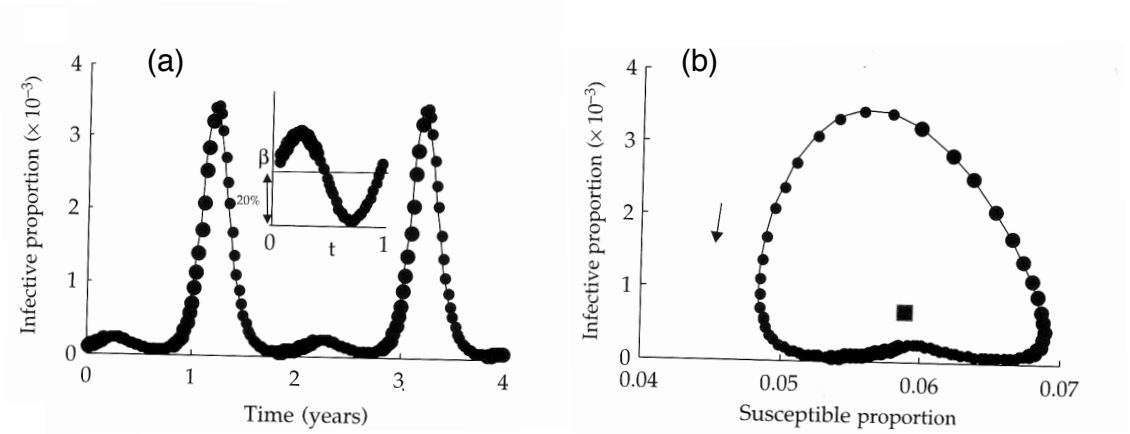

Figure 1.8: Impact of seasonal forcing on the SIRD model. (a) time evolution. (b) phase plane trajectory (the square dot is the unforced equilibrium) (reprinted from [5]).

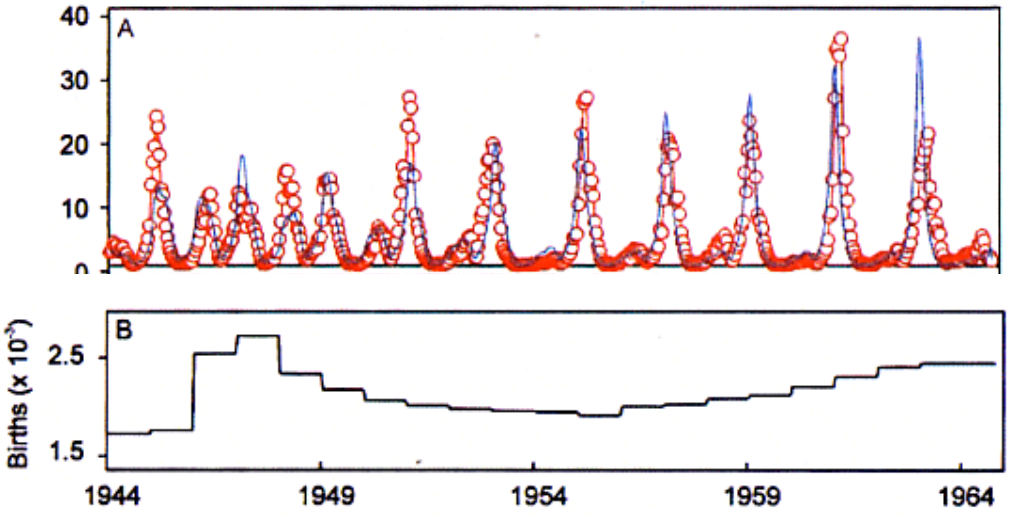

Figure 1.9: Observed measles data in London (circles) and corresponding simulation of an SIR model (solid line) with annual forcing of $\beta(t)$ and time varying birth rate $\mu(t)$ (reprinted from [4]).

cycles. In contrast, the baby boom in the period 1945-1950 increases the recruitment of susceptibles and induces annual epidemics (see [4] for more details).

1.5.5 Modelling of vector-borne diseases

Vector-borne diseases constitute another clas of diseases where it is relevant to extend the basic SIR model. A typical example is mosquito-borne diseases such as malaria, yellow fever, dengue or chikungunya. In such diseases, susceptible humans are infected by the parasite when they are bitten by infectious mosquitoes while the susceptible mosquitoes become themselves infected when they bite infectious humans. Furthermore, the immunity of recovered humans is generally temporary. It is therefore natural to consider two coupled SIR-type models for the humans and for the mosquitoes as shown in Fig.1.10 where the dotted lines represent the reciprocal infections between humans and mosquitoes. The derivation of the corresponding equations is left as an exercice (an interesting reference is [5]).

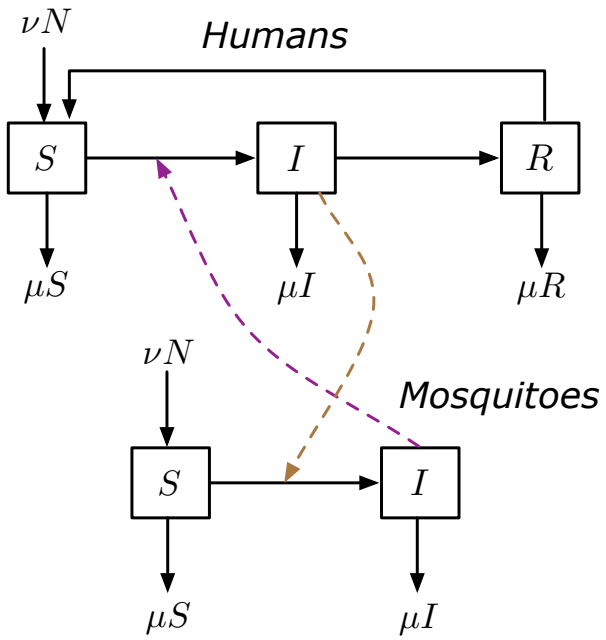

Figure 1.10: Compartmental diagram for mosquito-borne diseases.

1.6 References

[1] J.D. Murray, *Mathematical Biology*, 3rd edition, Springer, 2007 (Volume 1, Chapter 10).

[2] Anonymous contribution to the British Medical Journal *The Lancet*, March 1978, p. 587.

[3] M. J. Keeling and P. Rohani, *Modeling Infectious Diseases in Humans and Animals*, Princeton University Press, 2007 (voir aussi : http://www.modelinginfectiousdiseases.org/)

[4] B.T. Grenfell, O.N. Bjornstad and B.F. Finkenstadt, *Dynamics of measles epidemics: scaling noise, determinism and predictability with the SIR model*, Ecological Monographs, Vol. 72(2), pp. 185-202, 2002.

[5] G.A. Ngwa, *Modelling the dynamics of endemic malaria in growing populations*, Discrete and Continuous Dynamical Systems - Series B, Vol. 4(4), pp. 11731202 ,2004

[6] S.P. Ellner and J. Guckenheimer, *Dynamic Models in Biology*, Princeton University Press, 2006 (Chapter 6).

Lecture 2

Quantitative Modelling of Metabolic Systems

2.1 Metabolic networks

The intracellular metabolism of living cells is usually represented by a metabolic network under the form of a directed hypergraph[1] that encodes a set of elementary biochemical reactions taking place within the cell. In this hypergraph, the nodes represent the involved metabolites and the edges represent the metabolic fluxes. A typical example of metabolic network is shown in Fig.2.2. The metabolic network involves two groups of nodes: boundary nodes and internal nodes. Boundary nodes have only either incoming or outgoing edges, but not both together. Boundary nodes are further separated into input (or initial) and output (or terminal) nodes. Input nodes correspond to substrates that are only consumed but not produced. Output nodes correspond to final products that are only produced but not consumed. In contrast, the internal (or intermediary) nodes are the nodes that have necessarily both incoming and outgoing incident edges. They correspond to intracellular metabolites that are produced by some of the metabolic reactions and consumed by other reactions inside the cell.

It is assumed that the cells are cultivated in batch mode in a stirred tank reactor. The dynamics of substrates and products in the bioreactor are represented by the following basic differential equations:

(2.1a)
$$\frac{d\mathbf{s}(t)}{dt} = -\mathbf{v}_s(t)X(t)$$

(2.1b)
$$\frac{d\mathbf{p}(t)}{dt} = \mathbf{v}_p(t)X(t)$$

where $X(t)$ is the biomass concentration in the culture medium, $\mathbf{s}(t)$ is the vector of substrate concentrations, $\mathbf{p}(t)$ the vector of product concentrations, $\mathbf{v}_s(t)$ the vector of

[1] A *hypergraph* is a generalization of a graph, where edges can connect any number of vertices.

specific uptake rates and $\mathbf{v}_p(t)$ the vector of specific production rates. (From now on, the time index "t" will be omitted).

Obviously, the specific rates $\mathbf{v}_s$ and $\mathbf{v}_p$ are not independent. They are quantitatively related through the intracellular metabolism represented by the metabolic network. In order to explicit this relation, the quasi steady-state paradigm of metabolic flux analysis (MFA) is adopted (e.g. [17]). This means that for each internal metabolite of the network, it is assumed that the net sum of production and consumption fluxes, weighted by their stoichiometric coefficients, is zero. This is expressed by the algebraic relation:

$$\mathbf{N}\mathbf{v} = 0 \qquad \mathbf{v} \geqslant 0 \tag{2.2}$$

where $\mathbf{v} = (v_1, v_2, \ldots, v_m)^T$ is the m-dimensional vector of fluxes and $\mathbf{N} = [n_{ij}]$ is the $n \times m$ stoichiometric matrix of the metabolic network (m is the number of fluxes and n the number of internal nodes of the network). More precisely, a flux v_j denotes the rate of reaction j and a non-zero n_{ij} is the stoichiometric coefficient of metabolite i in reaction j.

2.2 Elementary pathways and input-output bioreactions

For a given metabolic network, the set $\mathcal{S}$ of admissible flux distributions is the set of vectors $\mathbf{v}$ that satisfy the finite set (2.2) of homogeneous linear equalities and inequalities. Each admissible $\mathbf{v}$ must necessarily be non-negative and belong to the kernel of the matrix $\mathbf{N}$. Hence the set $\mathcal{S}$ is the pointed polyhedral cone which is the intersection of the kernel of $\mathbf{N}$ and the nonnegative orthant. This implies that any flux distribution $\mathbf{v}$ can be expressed as a non-negative linear combination of a set of vectors $\mathbf{e}_i$ which are the edges (or extreme rays) of the polyhedral cone and form therefore a *unique* convex basis (see e.g. [19]) of the flux space:

$$\mathbf{v} = w_1\mathbf{e}_1 + w_2\mathbf{e}_2 + \cdots + w_p\mathbf{e}_p \quad w_i \geqslant 0. \tag{2.3}$$

The $m \times p$ non-negative matrix $\mathbf{E}$ with column vectors $\mathbf{e}_i$ obviously satisfies $\mathbf{NE} = 0$ and (2.3) is written in matrix form as

$$\mathbf{v} = \mathbf{E}\mathbf{w} \quad \text{with} \quad \mathbf{w} \triangleq (w_1, w_2, \ldots, w_p)^T. \tag{2.4}$$

From a metabolic viewpoint, the vectors $\mathbf{e}_i$ of the convex basis encode the simplest metabolic paths that connect the substrates (input nodes) to the products (output nodes). More precisely, the non-zero entries of a basis vector $\mathbf{e}_i$ enumerate the fluxes of a sequence of biochemical reactions starting at one or several substrates and ending at one or several products. These simple pathways between substrates and products are called extreme pathways (ExPa) or elementary (flux) modes (EM) of the network ([16] and [9]). Since the intermediary reactions are assumed to be at quasi steady-state, a single macroscopic bioreaction is then readily defined from an elementary flux mode by considering only the involved initial substrates and final products.

Let us now come back to the basic model (2.1) in order to elucidate the relation between the specific consumption and production rates $\mathbf{v}_s$ and $\mathbf{v}_p$ induced by the metabolic

network. Obviously $\mathbf{v}_s$ and $\mathbf{v}_p$ are linear combinations of some of the metabolic fluxes. This is expressed by defining appropriate matrices $\mathbf{N}_s$ and $\mathbf{N}_p$ such that

$$\mathbf{v}_s = \mathbf{N}_s\mathbf{v} \qquad \mathbf{v}_p = \mathbf{N}_p\mathbf{v}. \tag{2.5}$$

From (2.4) and (2.5), it follows that the model (2.1) is rewritten as:

$$\frac{d}{dt}\begin{pmatrix}\mathbf{s}\\ \mathbf{p}\end{pmatrix} = \begin{pmatrix}-\mathbf{N}_s\\ \mathbf{N}_p\end{pmatrix}\mathbf{E}\mathbf{w}X = \mathbf{K_e}\mathbf{w}X \tag{2.6}$$

where

$$\mathbf{K_e} \triangleq \begin{pmatrix}-\mathbf{N}_s\\ \mathbf{N}_p\end{pmatrix}\mathbf{E}. \tag{2.7}$$

is the stoichiometric matrix of the set of bioreactions encoded by the EFMs. Equation (2.6) can be regarded as the dynamic model of a bioprocess governed by the bioreactions with stoichiometry $\mathbf{K_e}$ and specific reaction rates $\mathbf{w}$. In other terms, each weighting coefficient w_i in (2.3) can equally be interpreted as the specific reaction rate of the bioreaction encoded by the EFM $\mathbf{e}_i$: the flux vector $\mathbf{v}$ is thus a linear combination of EFMs whose non-negative weights are the macroscopic bioreaction rates w_i.

However an important issue concerns the number of distinct bioreactions that are generated when computing the EFMs. It may become very large because it combinatorially increases with the size of the underlying metabolic network [2]. Furthermore, even when the number of EFMs is rather limited, it appears that the resulting set of bioreactions can be significantly redundant for the design of a dynamic model that fully explains the available experimental data. There is therefore clearly a need for reducing the model size as much as possible and trying to determine a minimal subset of bioreactions that are able to fully describe the available experimental data.

2.3 Metabolic flux analysis

Metabolic flux analysis (MFA) is the exercise of calculating the admissible flux distributions $\mathbf{v}$ that satisfy the steady state balance equation $\mathbf{N}\mathbf{v} = \mathbf{0}$ together with an additional set of linear constraints added by using experimental measurements. Here we consider the case where the measurements are collected in a vector $\mathbf{v}_m$ which is a linear function of the unknown flux distribution $\mathbf{v}$ and is expressed as

$$\mathbf{v}_m = \mathbf{N}_m\mathbf{v} \tag{2.8}$$

[2]The Double Description (DD) method ([8]) is the simplest known algorithm for computing the convex basis of the solution space (see [3] for a review). In the context of metabolic networks various refinements have been proposed that differ from the original DD algorithm mainly by their initialization. A first specific algorithm was presented by [15]. Recently, the implementation of the DD method for metabolic networks has received various further improvements (e.g. [4] and [6]).

with $\mathbf{N}_m$ being a known $\dim\{\mathbf{v}_m\} \times n$ matrix. Then, from equations (2.2)-(2.8), we have the following fundamental equation of metabolic flux analysis

$$\mathbf{\Sigma}\begin{pmatrix}\mathbf{v}\\ 1\end{pmatrix} = \mathbf{0} \quad \text{with} \quad \mathbf{\Sigma} \triangleq \begin{pmatrix}\mathbf{N} & \mathbf{0}\\ \mathbf{N}_m & -\mathbf{v}_m\end{pmatrix} \quad \text{and} \quad \mathbf{v} \geqslant \mathbf{0}. \tag{2.9}$$

For a given metabolic network and a given set of measurements, the solution of the MFA problem is defined as the space $\mathcal{F}$ of admissible flux distributions. i.e. the set of non-negative vectors $\mathbf{v}$ that satisfy the finite set (2.9) of homogeneous linear equalities and inequalities. Each admissible $\mathbf{v}$ must be such that the non-negative vector $(\mathbf{v}^T\ 1)$ belongs to the kernel of the matrix $\mathbf{\Sigma}$. Hence, as emphasized in [11, Chapter 4]-[13], the set $\mathcal{F}$ is a polytope in the positive orthant $\mathbb{R}^m_+$. This means that any admissible flux distribution $\mathbf{v}$ can be expressed as a convex combination of a set of q non-negative basis vectors $\mathbf{f}_i$ which are the edges (or extreme rays) of this polyhedral cone and form therefore a *unique* convex basis of the flux space $\mathcal{F}$. In other words, the solution of the MFA problem is the *admissible flux space* $\mathcal{F}$ defined as

$$\mathcal{F} \triangleq \big\{\mathbf{v} : \mathbf{v} = \sum_{i=1}^{q} \omega_i \mathbf{f}_i,\ \ \omega_i \geqslant 0,\ \ \sum_{i=1}^{q} \omega_i = 1\big\}. \tag{2.10}$$

The smallest "hyper-rectangular" set that encloses $\mathcal{F}$ in $\mathbb{R}^m$ is called the *flux spectrum* (e.g. [7]) and is defined as the set

$$\mathcal{F}_o = \big\{\mathbf{v} : v_i^{\min} \leqslant v_i \leqslant v_i^{\max}\big\}$$

where the bounds $v_i^{\min}$ and $v_i^{\max}$ are defined from the convex basis vectors as follows:

$$v_i^{\min} \triangleq \min\big\{f_{ki},\ k = 1, \ldots, p\big\}, \quad v_i^{\max} \triangleq \max\big\{f_{ki},\ k = 1, \ldots, p\big\},$$

where f_{ki} denotes the i-th element of the basis vector $\mathbf{f}_k$.

2.4 Case study: Application to CHO cells

As a matter of illustration and motivation to the methodology presented above, we consider the example of chinese hamster ovary (CHO) cells cultivated in batch mode in stirred flasks in a serum-free medium ([1]). During the growth phase, the cell metabolism is described by the metabolic network presented in Fig.2.2. This network describes only the part of the metabolism concerned with the utilisation of the two main energetic nutrients (glucose and glutamine). The metabolism of the amino acids provided by the culture medium is not considered. The network involves the Glycolysis pathway, the Pentose-Phosphate pathway and the Krebs cycle. Moreover it is assumed that a part of the glutamine is used for the making of nucleotides which are lumped into a single species with equal shares of purines and pyrimidines (see [11] and [14] for further motivation and details).

In this network, there are

- two input substrates : Glucose and Glutamine;

- five output products : Lactate, CO_2, NH_4, Alanine and Nucleotides;
- $n = 18$ internal metabolites : Glucose-6-phosphate, Fructose-6-Phosphate, Dihydroxy-acetone-phosphate, Glyceraldehyde-3 phosphate, Pyruvate, Acetyl-coA, Citrate, α-ketoglutarate, Fumarate, Malate, Oxaloacetate, Aspartate, Glutamate, CO_2, Ribose-5-Phosphate, Ribulose-5-Phosphate, Xylose-5-Phosphate, Erythrose-4-Phosphate;
- $m = 24$ metabolic fluxes denoted v_1 to v_{24} in Fig. 2.2.

Without loss of generality, all the intermediate metabolites that are not located at branch points are omitted from the network. The stoichiometric matrix $\mathbf{N}$ is given in Table 2.2. The matrices $\mathbf{N}_s$ and $\mathbf{N}_p$ are given in Tables 2.3 and 2.4 respectively.

This network has eleven elementary modes given in Table 2.6 (computed with METATOOL[3]) and from which the following set of input/output bioreactions can be derived:

($\mathbf{e}_1$) Glucose $\rightarrow$ 2 Lactate
($\mathbf{e}_2$) 2 Glucose + 3 Glutamine $\rightarrow$ Alanine + Nucleotide + 9 CO_2
($\mathbf{e}_3$) Glutamine $\rightarrow$ Lactate + 2 NH_4 + 2 CO_2
($\mathbf{e}_4$) Glutamine $\rightarrow$ 2 NH_4 + 5 CO_2
($\mathbf{e}_5$) Glutamine $\rightarrow$ Alanine + NH_4 + 2 CO_2
($\mathbf{e}_6$) 2 Glucose + 3 Glutamine $\rightarrow$ Lactate + Alanine + Nucleotide + 6 CO_2
($\mathbf{e}_7$) 3 Glucose $\rightarrow$ 5 Lactate + 3 CO_2
($\mathbf{e}_8$) 2 Glucose + 3 Glutamine $\rightarrow$ 2 Lactate + Nucleotide + NH_4 + 6CO_2
($\mathbf{e}_9$) Glucose $\rightarrow$ 6 CO_2
($\mathbf{e}_{10}$) Glucose $\rightarrow$ 6 CO_2
($\mathbf{e}_{11}$) 2 Glucose + 3 Glutamine $\rightarrow$ Nucleotide + NH_4 + 12 CO_2

We observe that the two bioreactions corresponding to elementary modes $\mathbf{e}_9$ and $\mathbf{e}_{10}$ are identical (Glucose $\rightarrow$ 6 CO_2) although the two concerned elementary pathways are different.

It can be checked that the stoichiometric matrix of this set of bioreactions is given by the matrix product

$$\mathbf{K_e} \triangleq \begin{pmatrix} -\mathbf{N}_s \\ \mathbf{N}_p \end{pmatrix} \mathbf{E}.$$

Moreover, there are five measured extra-cellular species : the two substrates (Glucose and Glutamine) and three excreted products (Lactate, Ammonia, Alanine). The values of the average specific uptake and excretion rates (vector $\mathbf{v_m}$), computed by linear regression during the growth phase (see [12]), are given in Table 2.1. The corresponding matrix $\mathbf{N}_m$ is given in Table 2.5. We observe that in this case:

$$\text{the matrix } \mathbf{N}_m \quad \text{is a sub-matrix of} \quad \begin{pmatrix} \mathbf{N}_s \\ \mathbf{N}_p \end{pmatrix}.$$

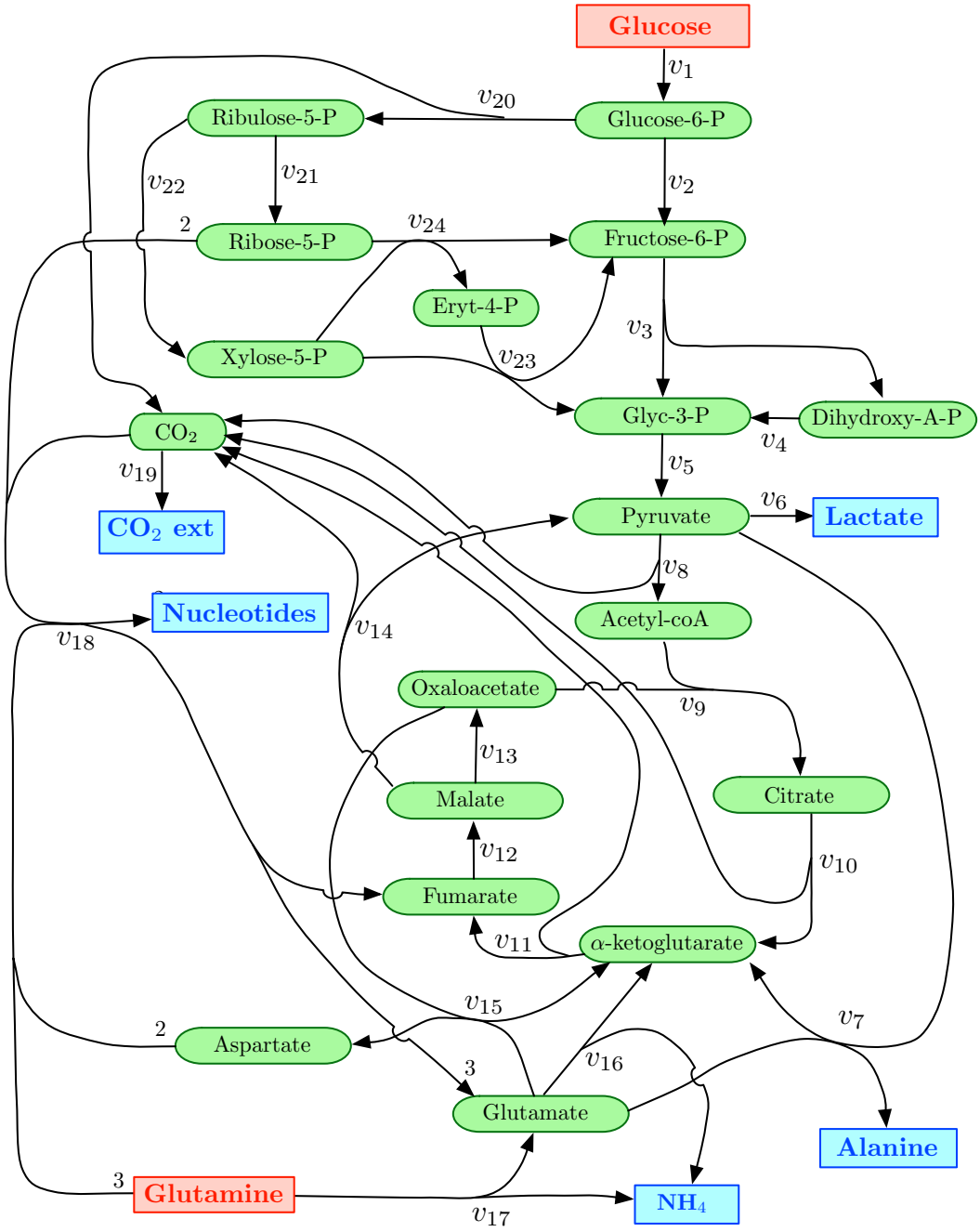

Figure 2.1: Metabolic network : rectangular boxes represent input/ouput nodes, elliptic boxes represent internal nodes. (The numbers along some arrows indicate stoichiometric coefficients).

Glucose	Glutamine	Lactate	NH_4	Alanine
4.0546	1.1860	7.3949	0.9617	0.2686

Table 2.1: Specific uptake and excretion rates (mM/(d$\times 10^9$cells)).

The admissible flux space $\mathcal{F}$ is generated by a convex basis that includes $p = 2$ basis vectors that are given in Table 2.7 (computed with METATOOL). Obviously the values given in this Table are also the limiting values $v_i^{\min}$ and $v_i^{\max}$ of the flux spectrum. It is remarkable that, although the MFA problem is here underdetermined, the values of the fluxes v_1, v_6, v_7, v_{14}, v_{15}, v_{16}, v_{17}, v_{18} and v_{19} are exactly given without uncertainty. This is obviously normal for the three fluxes v_1 (Glucose), v_6 (Lactate), v_7 (Alanine) that are constrained to be equal to their measured values. But we observe that it is also the case for other fluxes like for instance the production fluxes of Nucleotides v_{18} and CO_2 v_{19} that are not measured at all, and also for some intracellular fluxes like for instance the anaplerotic flux v_{14}.

2.5 Minimal dynamical bioreaction models

Here we focus on the special case where the MFA system (2.9) is exactly determined and and has a single well-defined solution which can obviously be decomposed in the convex basis as expressed by (2.3). But even if the flux vector $\mathbf{v}$ satisfying equation (2.9) is unique, it must be emphasized that *the decomposition of* $\mathbf{v}$ *in the convex basis* $\{\mathbf{e}_i\}$ *is* **not** *unique* which is the algebraic expression of the fact that the set of bioreactions used in the dynamical model (2.6) is redundant. Using (2.4), system (2.9) is equivalent to the system:

$$\left(\begin{array}{c} \mathbf{NE} \\ \mathbf{N}_m\mathbf{E} \end{array} \right) \mathbf{w} = \left(\begin{array}{c} \mathbf{0} \\ \mathbf{v}_m \end{array} \right) \quad \mathbf{w} \geqslant 0. \tag{2.11}$$

We observe that the first equation $\mathbf{NEw} = \mathbf{0}$ is trivially satisfied independently of $\mathbf{w}$ since $\mathbf{NE} = 0$ by definition. Hence, system (2.11) may be reduced to the second equation:

$$\mathbf{N}_m\mathbf{E}\mathbf{w} = \mathbf{v}_m \quad \mathbf{w} \geqslant 0.$$

or equivalently:

$$\left(\begin{array}{cc} \mathbf{N}_m\mathbf{E} & -\mathbf{v}_m \end{array} \right) \left(\begin{array}{c} \mathbf{w} \\ 1 \end{array} \right) = 0 \quad \mathbf{w} \geqslant 0. \tag{2.12}$$

In this form, it is clear that the set of admissible reaction rate vectors $\mathbf{w}$ that satisfy (2.12) also constitutes a convex polytope. Therefore there exists a set of appropriate edge vectors $\mathbf{h}_i$ such that any arbitrary convex combination of the form:

$$\mathbf{w} = \sum_i z_i \mathbf{h}_i \quad z_i \geqslant 0 \quad \sum_i z_i = 1 \tag{2.13}$$

[3] http://pinguin.biologie.uni-jena.de/bioinformatik/networks/ (see also [10]).

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{10}	v_{11}	v_{12}
Glucose-6-P	1	-1	0	0	0	0	0	0	0	0	0	0
Fructose-6-P	0	1	-1	0	0	0	0	0	0	0	0	0
Glyc-3-P	0	0	1	1	-1	0	0	0	0	0	0	0
Dihydroxy-A-P	0	0	1	-1	0	0	0	0	0	0	0	0
Pyruvate	0	0	0	0	1	-1	-1	-1	0	0	0	0
Acetyl-coA	0	0	0	0	0	0	0	1	-1	0	0	0
Citrate	0	0	0	0	0	0	0	0	1	-1	0	0
α-ketoglutarate	0	0	0	0	0	0	1	0	0	1	-1	0
Fumarate	0	0	0	0	0	0	0	0	0	0	1	-1
Malate	0	0	0	0	0	0	0	0	0	0	0	1
Oxaloacetate	0	0	0	0	0	0	0	0	-1	0	0	0
Glutamate	0	0	0	0	0	0	-1	0	0	0	0	0
Aspartate	0	0	0	0	0	0	0	0	0	0	0	0
Ribulose-5-P	0	0	0	0	0	0	0	0	0	0	0	0
Ribose-5-P	0	0	0	0	0	0	0	0	0	0	0	0
Xylose-5-P	0	0	0	0	0	0	0	0	0	0	0	0
Erythrose-4-P	0	0	0	0	0	0	0	0	0	0	0	0
CO_2	0	0	0	0	0	0	0	1	0	1	1	0

	v_{13}	v_{14}	v_{15}	v_{16}	v_{17}	v_{18}	v_{19}	v_{20}	v_{21}	v_{22}	v_{23}	v_{24}
Glucose-6-P	0	0	0	0	0	0	0	-1	0	0	0	0
Fructose-6-P	0	0	0	0	0	0	0	0	0	0	1	1
Glyc-3-P	0	0	0	0	0	0	0	0	0	0	1	0
Dihydroxy-A-P	0	0	0	0	0	0	0	0	0	0	0	0
Pyruvate	0	1	0	0	0	0	0	0	0	0	0	0
Acetyl-coA	0	0	0	0	0	0	0	0	0	0	0	0
Citrate	0	0	0	0	0	0	0	0	0	0	0	0
α-ketoglutarate	0	0	1	1	0	0	0	0	0	0	0	0
Fumarate	0	0	0	0	0	1	0	0	0	0	0	0
Malate	-1	-1	0	0	0	0	0	0	0	0	0	0
Oxaloacetate	1	0	-1	0	0	0	0	0	0	0	0	0
Glutamate	0	0	-1	-1	1	3	0	0	0	0	0	0
Aspartate	0	0	1	0	0	-2	0	0	0	0	0	0
Ribulose-5-P	0	0	0	0	0	0	0	1	-1	-1	0	0
Ribose-5-P	0	0	0	0	0	-2	0	0	1	0	0	-1
Xylose-5-P	0	0	0	0	0	0	0	0	0	1	-1	-1
Erythrose-4-P	0	0	0	0	0	0	0	0	0	0	-1	1
CO_2	0	1	0	0	0	0	-1	1	0	0	0	0

Table 2.2: Stoichiometric Matrix **N**

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{10}	v_{11}	v_{12}
Glucose	1	0	0	0	0	0	0	0	0	0	0	0
Glutamine	0	0	0	0	0	0	0	0	0	0	0	0

	v_{13}	v_{14}	v_{15}	v_{16}	v_{17}	v_{18}	v_{19}	v_{20}	v_{21}	v_{22}	v_{23}	v_{24}
Glucose	0	0	0	0	0	0	0	0	0	0	0	0
Glutamine	0	0	0	0	1	3	0	0	0	0	0	0

Table 2.3: Matrix $\mathbf{N}_s$

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{10}	v_{11}	v_{12}
Lactate	0	0	0	0	0	1	0	0	0	0	0	0
NH_4	0	0	0	0	0	0	0	0	0	0	0	0
Alanine	0	0	0	0	0	0	1	0	0	0	0	0
CO2	0	0	0	0	0	0	0	0	0	0	0	0
Nucleotides	0	0	0	0	0	0	0	0	0	0	0	0

	v_{13}	v_{14}	v_{15}	v_{16}	v_{17}	v_{18}	v_{19}	v_{20}	v_{21}	v_{22}	v_{23}	v_{24}
Lactate	0	0	0	0	0	0	0	0	0	0	0	0
NH_4	0	0	0	1	1	0	0	0	0	0	0	0
Alanine	0	0	0	0	0	0	0	0	0	0	0	0
CO2ext	0	0	0	0	0	0	1	0	0	0	0	0
Nucleotides	0	0	0	0	0	1	0	0	0	0	0	0

Table 2.4: Matrix $\mathbf{N}_p$

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{10}	v_{11}	v_{12}
Glucose	1	0	0	0	0	0	0	0	0	0	0	0
Glutamine	0	0	0	0	0	0	0	0	0	0	0	0
Lactate	0	0	0	0	0	1	0	0	0	0	0	0
NH_4	0	0	0	0	0	0	0	0	0	0	0	0
Alanine	0	0	0	0	0	0	1	0	0	0	0	0

	v_{13}	v_{14}	v_{15}	v_{16}	v_{17}	v_{18}	v_{19}	v_{20}	v_{21}	v_{22}	v_{23}	v_{24}
Glucose	0	0	0	0	0	0	0	0	0	0	0	0
Glutamine	0	0	0	0	1	3	0	0	0	0	0	0
Lactate	0	0	0	0	0	0	0	0	0	0	0	0
NH_4	0	0	0	1	1	0	0	0	0	0	0	0
Alanine	0	0	0	0	0	0	0	0	0	0	0	0

Table 2.5: Measurement matrix $\mathbf{N}_m$

	$\mathbf{e_1}$	$\mathbf{e_2}$	$\mathbf{e_3}$	$\mathbf{e_4}$	$\mathbf{e_5}$	$\mathbf{e_6}$	$\mathbf{e_7}$	$\mathbf{e_8}$	$\mathbf{e_9}$	$\mathbf{e_{10}}$	$\mathbf{e_{11}}$
v_1	1	2	0	0	0	2	3	2	1	3	2
v_2	1	0	0	0	0	0	0	0	1	0	0
v_3	1	0	0	0	0	0	2	0	1	2	0
v_4	1	0	0	0	0	0	2	0	1	2	0
v_5	2	0	0	0	0	0	5	0	2	5	0
v_6	2	0	1	0	0	1	5	2	0	0	0
v_7	0	1	0	0	1	1	0	0	0	0	0
v_8	0	1	0	1	0	0	0	0	2	5	2
v_9	0	1	0	1	0	0	0	0	2	5	2
v_{10}	0	1	0	1	0	0	0	0	2	5	2
v_{11}	0	4	1	2	1	3	0	3	2	5	5
v_{12}	0	5	1	2	1	4	0	4	2	5	6
v_{13}	0	3	0	1	0	2	0	2	2	5	4
v_{14}	0	2	1	1	1	2	0	2	0	0	2
v_{15}	0	2	0	0	0	2	0	2	0	0	2
v_{16}	0	0	1	1	0	0	0	1	0	0	1
v_{17}	0	0	1	1	1	0	0	0	0	0	0
v_{18}	0	1	0	0	0	1	0	1	0	0	1
v_{19}	0	9	2	5	2	6	3	6	6	18	12
v_{20}	0	2	0	0	0	2	3	2	0	3	2
v_{21}	0	2	0	0	0	2	1	2	0	1	2
v_{22}	0	0	0	0	0	0	2	0	0	2	0
v_{23}	0	0	0	0	0	0	1	0	0	1	0
v_{24}	0	0	0	0	0	0	1	0	0	1	0

Table 2.6: Matrix **E** of elementary modes.

is necessarily an admissible $\mathbf{w}$ satisfying (2.12). The convex basis vectors $\mathbf{h}_i$ have an important and critical property : the number of non-zero entries is **at most** equal to the number of measured uptake and excretion rate i.e. the size of the vector $\mathbf{v}_m$ (see [3] and Section 3.5 in [11]). From a metabolic viewpoint, each vector $\mathbf{h}_i$ is a particular solution $\mathbf{w}$ of (2.12), or equivalently a particular way (among an infinity) of computing the flux distribution $\mathbf{v}$ that satisfies (2.11):

$$\mathbf{v} = \mathbf{E}\mathbf{h}_i \quad \forall i. \tag{2.14}$$

Of course in this expression, the non-zero entries of the vector $\mathbf{h}_i$ are interpreted as the weights of the respective contributions of the corresponding EFMs in the computation of the flux distribution $\mathbf{v}$. But, at the same time, they can also be interpreted as being the specific rates of the bioreactions that are encoded by the EFMs and are involved in the dynamic model (2.6).

Hence each convex basis vector $\mathbf{h}_i$ brings two different pieces of information. First it tells which EFMs and consequently which bioreactions are sufficient to explain the

	$\mathbf{f}_1$	$\mathbf{f}_2$
v_1	4.0546	4.0546
v_2	3.5979	2.1279
v_3	3.5979	3.1079
v_4	3.5979	3.1079
v_5	7.1958	6.7058
v_6	7.3949	7.3949
v_7	0.2686	0.2686
v_8	0.4900	0.0
v_9	0.4900	0.0
v_{10}	0.4900	0.0
v_{11}	1.6760	1.1860
v_{12}	1.9043	1.4143
v_{13}	0.9467	0.4567
v_{14}	0.9577	0.9577
v_{15}	0.4567	0.4567
v_{16}	0.4607	0.4607
v_{17}	0.5010	0.5010
v_{18}	0.2283	0.2283
v_{19}	3.8420	3.8420
v_{20}	0.4567	1.9267
v_{21}	0.4567	0.9467
v_{22}	0.0	0.9800
v_{23}	0.0	0.4900
v_{24}	0.0	0.4900

Table 2.7: Convex basis of the flux space *(mM/(d×10^9cells))*

measured uptake and excretion rates $\mathbf{v_m}$. These EFMs are designated by the position of the non-zero entries of $\mathbf{h_i}$. Secondly, the value of each non-zero entry of $\mathbf{h}_i$ is the value of the reaction rate of the corresponding bioreaction.

For each basis vector $\mathbf{h}_i$, we can then define a selection matrix $\mathbf{S_i}$ that encodes the corresponding selection of bioreactions. Then the dynamical model (2.6) is reduced to a minimal form:

$$\frac{d}{dt}\begin{pmatrix}\mathbf{s}\\ \mathbf{p}\end{pmatrix} = \mathbf{K_i}\mathbf{r}_i X \tag{2.15}$$

where $\mathbf{K_i} \triangleq \mathbf{K_e}\mathbf{S_i}$ and $\mathbf{r}_i \triangleq (\mathbf{S_i})^T\mathbf{h}_i$ respectively denote the stoichiometric matrix and the vector of the specific reaction rates of the selected minimal set of bioreactions.

Therefore, we see that the computation of the convex basis vectors $\mathbf{h}_i$ provides the tool for determining all the minimal dynamical models that are both compatible with the metabolic network and the available measurements. Furthermore, it is worth to clearly

	$\mathbf{e_1}$	$\mathbf{e_2}$	$\mathbf{e_3}$	$\mathbf{e_4}$	$\mathbf{e_5}$	$\mathbf{e_6}$	$\mathbf{e_7}$	$\mathbf{e_8}$	$\mathbf{e_9}$
v_1	1	1	0	0	0	2	2	2	2
v_2	1	1	0	0	0	0	0	0	0
v_3	0	0	0	0	0	2	2	2	2
v_4	1	1	0	0	0	0	0	0	0
v_5	2	2	0	0	0	0	0	0	0
v_6	2	0	0	1	0	2	1	0	0
v_7	0	0	1	0	0	0	1	1	0
v_8	0	2	0	0	1	0	0	1	2
v_9	0	2	0	0	1	0	0	1	2
v_{10}	0	2	0	0	1	0	0	1	2
v_{11}	0	2	1	1	2	3	3	4	5
v_{12}	0	2	1	1	2	4	4	5	6
v_{13}	0	2	0	0	1	2	2	3	4
v_{14}	0	0	1	1	1	2	2	2	2
v_{15}	0	0	0	0	0	2	2	2	2
v_{16}	0	0	0	1	1	1	0	0	1
v_{17}	0	0	1	1	1	0	0	0	0
v_{18}	0	0	0	0	0	1	1	1	1
v_{19}	0	6	2	2	5	6	6	9	12

Table 2.8: Matrix **E** of elementary flux modes.

understand that all of these minimal models are totally **equivalent** because **they all provide exactly the same internal flux distribution and the same dynamical simulation results**.

2.5.1 Illustration with the case-study of CHO cells

In this example, we consider a slightly simplified metabolic network where the PPE pathway is neglected. The network is shown in Fig. 2.2. It has nine EFMs that are collected in matrix **E** (see Table 2.8) and from which the following set of input/output bioreactions is readily derived:
(b1) Glc $\rightarrow$ 2 Lac
(b2) Glc $\rightarrow$ 6 CO_2
(b3) Gln $\rightarrow$ Ala + NH_4 + 2 CO_2
(b4) Gln $\rightarrow$ Lac + 2 NH_4 + 2 CO_2
(b5) Gln $\rightarrow$ 2 NH_4 + 5 CO_2
(b6) 2 Glc + 3 Gln $\rightarrow$ 2 Lac + Nucl + NH_4 + 6CO_2
(b7) 2 Glc + 3 Gln $\rightarrow$ Lac + Ala + Nucl + 6 CO_2
(b8) 2 Glc + 3 Gln $\rightarrow$ Ala + Nucl + 9 CO_2

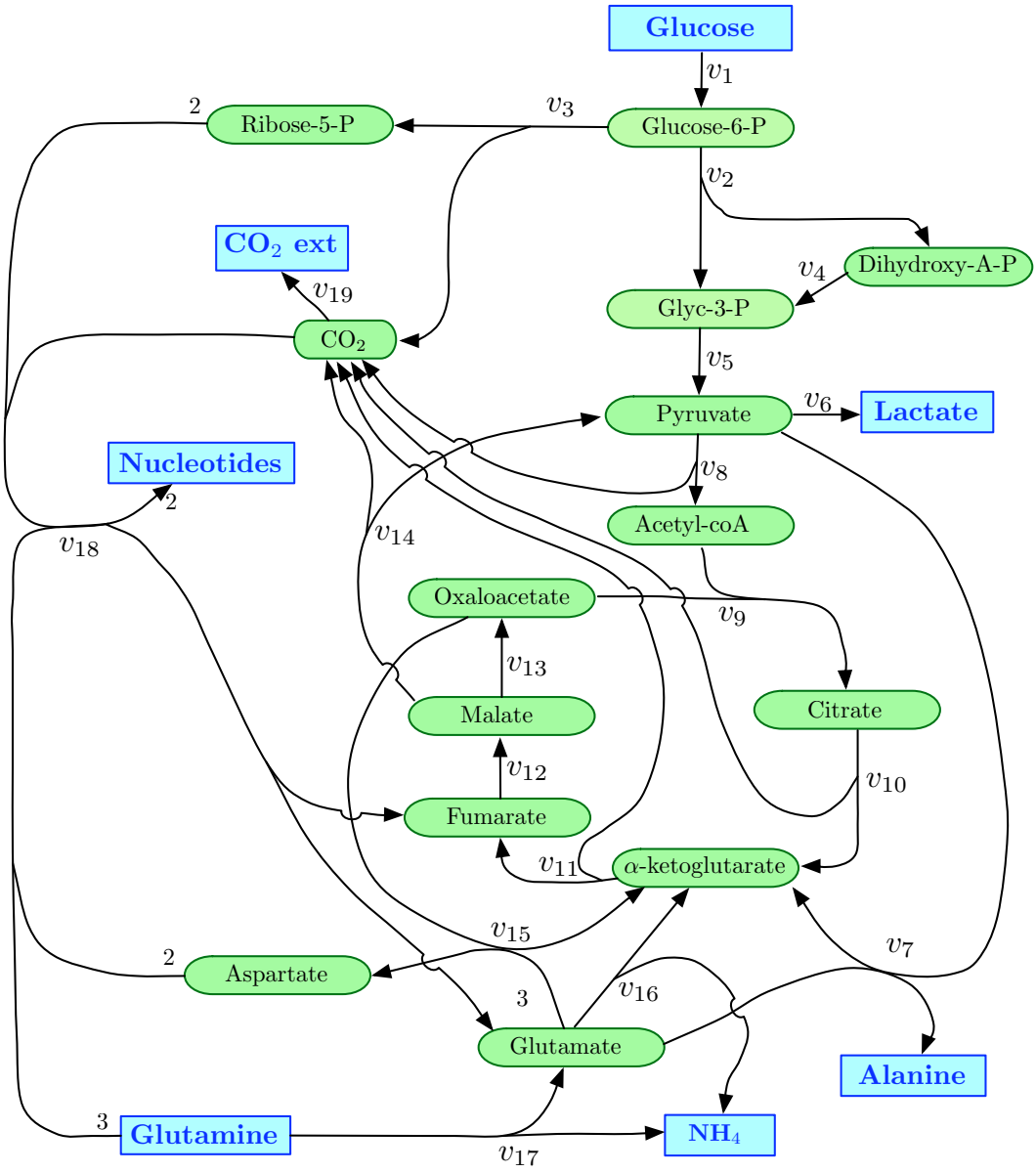

Figure 2.2: Metabolic network for the special case where the PPE pathway is neglected.

(b9) 2 Glc + 3 Gln $\rightarrow$ Nucl + NH_4 + 12 CO_2

In this special case, the system (2.9) is a system of 19 equations with 19 unknowns which is uniquely determined. The solution is given in Table 2.9 and corresponds to the first column of Table 2.7. We are then in a position to compute the set of vectors $\mathbf{h}_i$ and the result is shown in Table 2.10. We first observe that there are 12 different vectors $\mathbf{h_i}$ in this Table. They all produce exactly the same value of the flux distribution $\mathbf{v}$ of Table 2.9 when premultiplied by the matrix $\mathbf{E}$ as expected according to (2.14). Furthermore, as

v_1	v_2	v_3	v_4	v_5
4.0546	3.5979	0.4567	3.5979	7.1958
v_6	v_7	v_8	v_9	v_{10}
7.3949	0.2686	0.4900	0.4900	0.4900
v_{11}	v_{12}	v_{13}	v_{14}	v_{15}
1.6760	1.9043	0.9467	0.9577	0.4567
v_{16}	v_{17}	v_{18}	v_{19}	
0.4607	0.5010	0.2283	3.8420	

Table 2.9: Metabolic fluxes.

predicted by the theory, we also observe that there are exactly 5 non-zero entries in each vector $\mathbf{h_i}$. From these observations, we can conclude that there are 12 different equivalent minimal dynamical models of the form (2.15) for the considered process. For each of these models, Table 2.10 tells us which 5 bioreactions (among the set (b1)-(b9)) are used and the value of their reaction rates.

The design of a particular dynamic bioreaction model is finally completed by chosing arbitrarily any vector $\mathbf{h_i}$ in Table 2.10 and assuming that the selected bioreactions have Michaelis-Menten kinetics with maximum specific rates μ_i given by the non-zero entries of $\mathbf{h_i}$. This automatically gives by construction a model which necessarily produces simulations that fit the experimental data with a high accuracy as shown in Fig. 2.3.

	$\mathbf{h_1}$	$\mathbf{h_2}$	$\mathbf{h_3}$	$\mathbf{h_4}$	$\mathbf{h_5}$	$\mathbf{h_6}$
(b1)	3.5833	3.4671	3.3529	3.5979	3.4691	3.5979
(b2)	0.0146	0.1308	0.2450	0.0	0.1288	0.0
(b3)	0.0403	0.0403	0.0403	0.0403	0.2686	0.2686
(b4)	0.0	0.4607	0.4607	0.0	0.0	0.0
(b5)	0.4607	0.0	0.0	0.4607	0.2324	0.2324
(b6)	0.0	0.0	0.0	0.0	0.2283	0.0995
(b7)	0.2283	0.0	0.2283	0.1991	0.0	0.0
(b8)	0.0	0.2283	0.0	0.0293	0.0	0.0
(b9)	0.0	0.0	0.0	0.0	0.0	0.1288

	$\mathbf{h_7}$	$\mathbf{h_8}$	$\mathbf{h_9}$	$\mathbf{h_{10}}$	$\mathbf{h_{11}}$	$\mathbf{h_{12}}$
(b1)	3.5979	3.5979	3.3529	3.5979	3.5813	3.5979
(b2)	0.0	0.0	0.2450	0.0	0.0167	0.0
(b3)	0.1398	0.0403	0.2686	0.0695	0.2686	0.2686
(b4)	0.0	0.1991	0.2324	0.0	0.2324	0.1991
(b5)	0.3612	0.2617	0.0	0.4314	0.0	0.0333
(b6)	0.0995	0.0	0.2283	0.0	0.0	0.0
(b7)	0.0	0.0	0.0	0.1991	0.0	0.0
(b8)	0.1288	0.2283	0.0	0.0	0.0	0.0
(b9)	0.0	0.0	0.0	0.0293	0.2283	0.2283

Table 2.10: Specific reaction rates for the 12 equivalent minimal dynamic models (mM/(d $\times 10^9$cells)

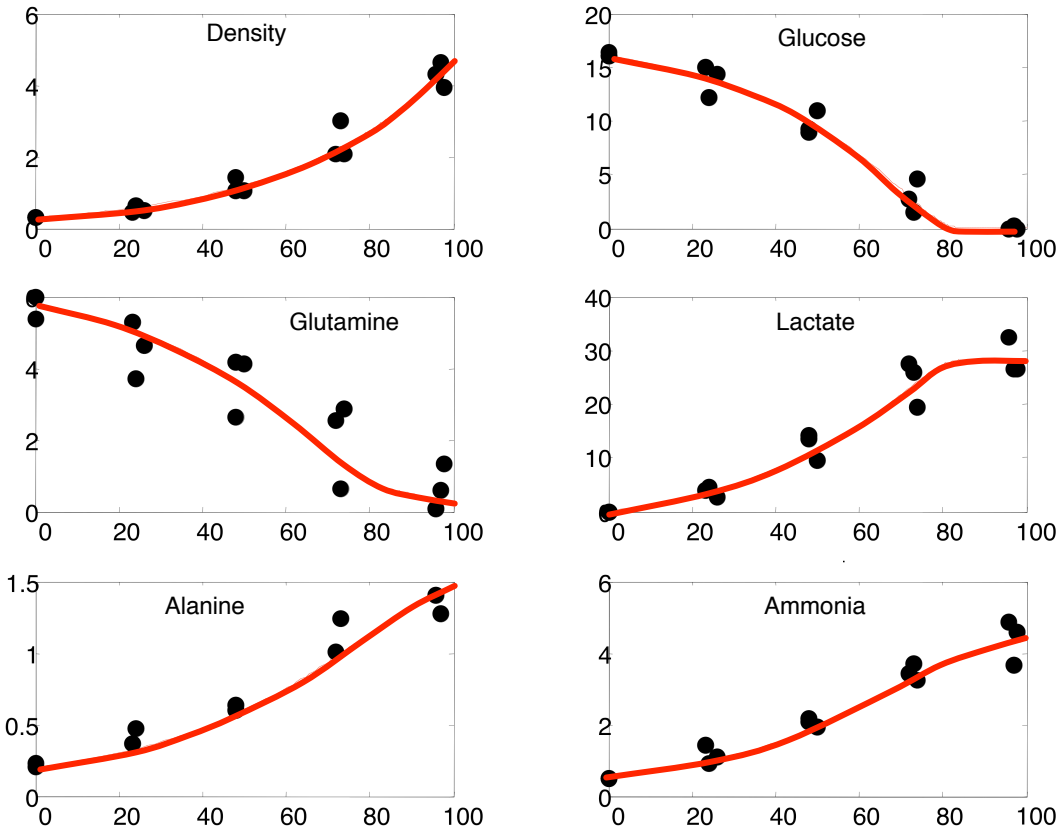

Figure 2.3: Simulation results with a minimal dynamical model. The black dots are the experimental data. The solid lines are the solutions of the simulation

.

Bibliography

[1] J.S. Ballez, J. Mols, J. Burteau, S.N. Agathos, and Y-J. Schneider. Plant protein hydrolysates support CHO-320 cells proliferation and recombinant ifn-gamma production in suspension and inside microcarriers in protein-free media. Cytotechnology, 44:103-114, 2004.

[2] S.L. Bell and B.O. Palsson. EXPA: a program for calculating extreme pathways in biochemical reaction networks. Bioinformatics, 21(8):1739-1740, 2005.

[3] K. Fukuda and A. Prodon. The double description method revisited. In R. Euler and M. E. Deza I. Manoussakis, editors, Combinatorics and Computer Science, volume 1120 of Lecture Notes in Computation Sciences, pages 91-111. Springer-Verlag, 1996.

[4] J. Gagneur and S. Klamt. Computation of elementary modes : a unifying framework and the new binary approach. BMC Bioinformatics, 5:175, 2004.

[5] J.E. Haag, A. Vande Wouwer, and P. Bogaerts. Dynamic modeling of complex biological systems : a link between metabolic and macroscopic descriptions. Mathematical Biosciences, 193:25-49, 2005.

[6] S. Klamt, J. Gagneur, and A. von Kamp. Algorithmic approaches for computing elementary modes in large biochemical networks. IEE Proceedings Systems Biology, 152:249-255, 2005.

[7] F. Llaneras and J. Pico. An interval approach for dealing with flux distributions and elementary modes activity patterns. Journal of Theoretical Biology, 246:290-308, 2007.

[8] T.S. Motzkin, H. Raiffa, G.L. Thompson, and R.M. Thrall. The double description method. In H.W. Kuhn and A.W. Tucker, editors, Contribution to the Theory of Games Vol. II, volume 28 of Annals of Mathematical Studies, pages 51-73, Princeton, New Jersey, 1953. Princeton University Press.

[9] J. Nielsen, J. Villadsen, and G. Liden. Bioreaction Engineering Principles. Kluwer Academic, 2002.

[10] T. Pfeiffer, I. Sanchez-Valdenebro, J.C. Nu, F. Montero, and S. Schuster. Metatool : for studying metabolic networks. Bioinformatics, 15(3):251-257, March 1999.

[11] A. Provost. Metabolic design of dynamic bioreaction models. PhD thesis, Faculty of Engineering, Université catholique de Louvain, November 2006.

[12] A. Provost and G. Bastin. Dynamical metabolic modelling under the balanced growth condition. Journal of Process Control, 14(7):717-728, 2004.

[13] A. Provost and G. Bastin. Metabolic flux analysis: an approach for solving non-stationary underdetermined systems. In CD-Rom Proceedings 5th MATHMOD Conference, Paper 207 in Session SP33, Vienna, Austria, 2006.

[14] A. Provost, G. Bastin, S.N. Agathos, and Y-J. Schneider. Metabolic design of macroscopic bioreaction models : Application to Chinese hamster ovary cells. Bioprocess and Biosystems Engineering, 29(5-6):349-366, 2006.

[15] R. Schuster and S. Schuster. Refined algorithm and computer program for calculating all non-negative fluxes admissible in steady states of biochemical reaction systems with or without some flux rates fixed. Computer Applications in the Biosciences, 9(1):79-85, February 1993.

[16] S. Schuster, D.A. Fell, and T. Dandekar. Detection of elementary flux modes in biochemical networks : a promising tool for pathway analysis and metabolic engineering. Trends in Biotechnology, 17(2):53-60, February 1999.

[17] G. Stephanopoulos, J. Nielsen, and A. Aristidou. Metabolic Engineering : Principles and Methodologies. Academic Press, San Diego, 1998.

[18] R. Urbanczik. Enumerating constrained elementary flux vectors of metabolic networks. IET Systems Biology, 1(5):274-279, 2007.

[19] H. Weyl. The elementary theory of convex polyhedra. In Contributions to the Theory of Games Vol. I, Annals of Mathematical Studies, pages 3-18, Princeton, New Jersey, 1950. Princeton University Press.

[20] F. Zhou, J-X. Bi, A-P. Zeng, and J-Q. Yuan. A macrokinetic and regulator model for myeloma cell culture based on metabolic balance of pathways. Process Biochemistry, 41:2207-2217, 2006.

Lecture 3

Mathematical Models in Population Genetics

3.1 Introduction

This lecture gives an elementary introduction to mathematical models of population genetics. "Population genetics is the science of genotypic variation in interbreeding populations. It is a quantitative theory which is based on models that translate precise assumptions about mating patterns and the action of selection into mathematical equations for the evolution of genotype frequencies[1]".

3.2 Mendelian genetics

In living organisms, reproduction involves the passing of a genetic code from generation to generation. The code is carried on **chromosomes**. Many animal and plant cells are **diploid** : they have chromosomes arranged in matched pairs, each member of the pair being a version of the same chromosome (with the possible exception of the chromosome determining sex). **Genes** are segments of chromosomes that code for specific heritable characters (like the texture and the colour of the peas in Mendel's experiments). The genes may appear in variant forms that are called **alleles** (thus the gene for pea color has yellow and green alleles). In sexual reproduction, the two chromosomes of each pair are inherited from the two parents (one from each parent). In **monoceious** populations, individuals house both male and female organs, and any individual can mate with any other or even with itself: plants with flowers that both contain an ovum and produce pollen provide a common example. In **dioceious** populations, such as humans, individuals are either male or female, but not both.

Here we consider the case of a monoceious diploid population and we denote two possible alleles of a certain gene as **A** and **B**. The **genotype** of an individual with respect to this gene is the set of the two alleles it carries. There are thus three possible genotypes:

[1] From [1], Chapter 1, Introduction

AA, **AB**, **BB** (note that **AB** and **BA** are indistinguishable). Assuming perfect random mating in an infinite population, one can determine the probability of a given match to give an offspring with a given genotype. The result is given in Table 3.1.

	Offspring		
Parents	**AA**	**AB**	**BB**
AA x AA	1	0	0
AA x AB	1/2	1/2	0
AA x BB	0	1	0
AB x AB	1/4	1/2	1/4
AB x BB	0	1/2	1/2
BB x BB	0	0	1

Table 3.1: Probabilities of offspring of a given genotype resulting from the mating of two given parent genotypes. For example, the mating of an **AA** with an **AB** will give, on the average, 50% of **AA** and 50% of **AB**.

3.3 Hardy-Weinberg equilibrium

Our concern is now to describe the change, in a population, of the genotype frequencies from generation to generation. In order to avoid useless complications, we assume no overlapping of the generations (as it is the case, for example, with annual plants) : matings occur only between individuals of the same generation. We introduce the following notations:

X = frequency of genotype **AA**,
$2Y$ = frequency of genotype **AB**,
Z = frequency of genotype **BB**.

Obviously we have $X + 2Y + Z = 1$.

Let X_n, $2Y_n$ and Z_n denote these frequencies at the n-th generation. Then, using the Mendelian probabilities of Table 3.1, we have at the $n+1$-th generation:

$$
\begin{aligned}
X_{n+1} &= X_n^2 + \frac{1}{2}(4X_nY_n) + \frac{1}{4}(4Y_n^2) = (X_n + Y_n)^2 \\
2Y_{n+1} &= \frac{1}{2}(4X_nY_n) + 2X_nZ_n + \frac{1}{2}(4Y_n^2) + \frac{1}{2}(4Y_nZ_n) = 2(X_n + Y_n)(Y_n + Z_n) \\
Z_{n+1} &= Z_n^2 + \frac{1}{2}(4Z_nY_n) + \frac{1}{4}(4Y_n^2) = (Z_n + Y_n)^2.
\end{aligned}
$$

Then at the next $n+2$-th generation, we have:

$$\begin{aligned} X_{n+2} &= \left[X_{n+1} + Y_{n+1}\right]^2 \\ &= \left[(X_n + Y_n)^2 + (X_n + Y_n)(Y_n + Z_n)\right]^2 \\ &= \left[(X_n + Y_n)^2 + (X_n + Y_n)(Y_n + 1 - X_n - 2Y_n)\right]^2 \\ &= (X_n + Y_n)^2 \\ &= X_{n+1} \end{aligned}$$

$Z_{n+2} = Z_{n+1}$ by symmetry.

$Y_{n+2} = Y_{n+1}$ because $X + 2Y + Z = 1$

We conclude readily that, starting from any initial condition, an equilibrium is reached in exactly one generation and remain equal for all further generations. This is the so-called *Hardy-Weinberg equilibrium* which satisfies the following two relations:

(3.1)

Hardy-Weinberg equilibrim

$$Y^2 = XZ$$
$$X + 2Y + Z = 1$$

This result is obviously of great importance for the Darwinian theory since it explains why random mating may indeed preserve the genetic diversity in a population. We remark also that since the two equations (3.1) hold together, only one frequency is independent and must be known (e.g. from experimental observations) in order to fully charaterize the proportions of the three genotypes in the population.

The Hardy-Weinberg equilibrium can be generalized in various ways:

(1) to the case of more than two alleles of a given gene (like in the ABO blood system) where the frequencies at the equilibrium satisfy the relations $f_{ij}^2 = 4f_{ii}f_{jj}$;

(2) to the case of one gene with two alleles in a diocieous population where it can be shown that the HW equilibrium is achieved in two generations (instead of one);

(3) to the case of overlapping generations where it can be shown that the HW equilibrium is achieved asymptotically (not in finite time) though with a rather fast convergence.

3.4 Evolution dynamics

Evolution occurs when some of the underlying assumptions of the Hardy-Weinberg analysis are violated, for instance under selection, mutations or population migrations. Our purpose, in this paragraph, is to analyse the evolution of the genotype frequencies in case of selective fecondation, i.e. when the genotypes have different reproductive fitness. We

introduce the following notations for the allele frequencies in the population.

$p = X + Y =$ frequency of alleles **A** in the population,
$q = 1 - p = Y + Z =$ frequency of alleles **B** in the population.

It follows that the HW genotype frequencies can be expressed as well as follows:

$$X = p^2 \quad 2Y = 2pq = 2p(1-p) \quad Z = q^2 = (1-p)^2.$$

Let us compute the recurrence of p from generation to generation. We have

$$\begin{aligned} p_{n+1} &= (X_n + Y_n)^2 + (X_n + Y_n)(Y_n + Z_n), \\ &= p_n^2 + p_n q_n, \\ q_{n+1} &= q_n^2 + p_n q_n. \end{aligned}$$

It follows that

$$p_{n+1} + q_{n+1} = p_n^2 + 2p_n q_n + q_n^2 = (p_n + q_n)^2 = 1.$$

Now, let us assume that the genotypes have different reproductive fitness. Then the total gene pool at the next generation is proportional to

$$(1+\rho)p_n^2 + 2(1+\sigma)p_n q_n + (1+\tau)q_n^2. \tag{3.2}$$

where the small parameters ρ, σ, τ represent the deviation with respect to the ideal situation (i.e. in case of selective fecondation when reproductive fitness differs between genotypes). Similarly, the pool of alleles **A** is proportional to

$$(1+\rho)p_n^2 + (1+\sigma)p_n q_n. \tag{3.3}$$

Hence, we can compute the freqency p_{n+1} at the next generation under selective fecondation as the ratio of (3.3) over (3.2):

$$p_{n+1} = \frac{(1+\rho)p_n^2 + (1+\sigma)p_n(1-p_n)}{(1+\rho)p_n^2 + 2(1+\sigma)p_n(1-p_n) + (1+\tau)(1-p_n)^2},$$

which leads to the iteration

$$p_{n+1} = \frac{p_n + \left[\rho p_n^2 + \sigma p_n(1-p_n)\right]}{1 + \rho p_n^2 + 2\sigma p_n(1-p_n) + \tau(1-p_n)^2}. \tag{3.4}$$

We observe that, as expected, we have $p_{n+1} = p_n$ if $\rho = \sigma = \tau = 0$ (no selection).

In order to find the relevant fixed points (or equilibria) of the iteration (3.4), we have to find the values of $p \in [0,1]$ such that

$$p = \frac{p + \left[\rho p^2 + \sigma p(1-p)\right]}{1 + \rho p^2 + 2\sigma p(1-p) + \tau(1-p)^2} = f(p).$$

We can easily see that $p = 0$ and $p = 1$ are fixed points. If $p \neq 0$ we have

$$\begin{aligned} & \rho p^2 + 2\sigma p(1-p) + \tau(1-p)^2 - \rho p - \sigma(1-p) = 0, \\ \Rightarrow \quad & \sigma(1-p)(2p-1) + \tau(1-p)^2 - \rho p(1-p) = 0, \\ \Rightarrow \quad & p(\rho - 2\sigma + \tau) + \sigma - \tau = 0. \end{aligned}$$

We conclude that

$$p = \frac{\sigma - \tau}{(\sigma - \tau) + (\sigma - \rho)}$$

is a third fixed point of the iteration (3.4) which belongs to $[0, 1]$ if and only if

$$\sigma > \max\{\rho, \tau\} \quad \text{or} \quad \sigma < \min\{\rho, \tau\}.$$

The stability of the fixed points depends on the value of $f'(p)$. There are three possibilities.

- Case 1: $\rho < \sigma < \tau$. In this case we have $f'(0) < 1$ and $f'(1) > 1$. This means that $p = 0$ is stable and $p = 1$ is unstable. *Allele* **A** *is favored by the natural selection.*

- Case 2: $\tau < \sigma < \rho$. Then $f'(0) > 1$ and $f'(1) < 1$, $p = 0$ is unstable and $p = 1$ is stable. *Allele* **B** *is favored by the natural selection.*

- Case 3: $\sigma > \max\{\rho, \tau\}$. Then $f'(0) > 1$ and $f'(1) > 1$, $p = 0$ and $p = 1$ are both unstable. The third fixed-point is stable. *Both alleles coexist in the population.*

- Case 4: $\sigma < \min\{\rho, \tau\}$. Then $f'(0) < 1$ and $f'(1) < 1$, $p = 0$ and $p = 1$ are both stable. The third fixed-point is unstable. *Depending of the initial conditions, either* **A** *or* **B** *is favored by the natural selection.*

3.5 References

[1] D. Ocone, Course Notes on *Discrete and Probabilistic Models in Biology*, http://www.math.rutgers.edu/courses/338/coursenotes/coursetext.html

[2] W.J. Ewens, *Population Genetics*, Methuen, London (1969)

Lecture 4

Modelling of within-host HIV dynamics

4.1 The basic model

In the basic model of HIV dynamics, three species are taken into account: the uninfected target cells (mainly a fraction of CD4+T cells), the infected cells and the blood free viruses. The model is as follows (see e.g. [2], [5], [6]):

$$\frac{dT}{dt} = \lambda - \delta_T T - \beta V T, \tag{4.1}$$

$$\frac{dI}{dt} = \beta V T - \delta_I I, \tag{4.2}$$

$$\frac{dV}{dt} = pI - \delta_V V. \tag{4.3}$$

In these equations, T denotes the density of susceptible target cells, I the density of infected cells and V the viral concentration (also called viral load) at time t.

The target cells are assumed to be produced at the exogenous rate λ and to be infected by the virus according to a simple mass-action principle with a proportionality coefficient β. The free viruses are supposed to be produced by infected cells with a specific rate p. The parameters δ_T, δ_I and δ_V are natural decay rates (death and/or degradation).

Remark : Simplified model. In some publications (e.g. [7]), a simplified model is obtained by assuming that the virus concentration is just proportional to the infected cell density, which corresponds to a quasi steady-state approximation of equation (4.3):

$$pI - \delta_V V \simeq 0.$$

Then the basic model reduces to

$$\frac{dT}{dt} = \lambda - \delta_T T - \tilde{\beta} IT,$$
$$\frac{dI}{dt} = \tilde{\beta} IT - \delta_I I,$$

with $\tilde{\beta} \triangleq \beta p/\delta_V$.

4.2 Immune response

Cell-mediated immune response refers to the killing of infected CD4+T-cells by cytotoxic T-cells (such as CD8T-cells). The simplest extension of the basic model is as follows (e.g. [1], [9]):

$$\frac{dT}{dt} = \lambda - \delta_T T - \beta VT,$$
$$\frac{dI}{dt} = \beta VT - kIC - \delta_I I,$$
$$\frac{dV}{dt} = pI - \delta_V V,$$
$$\frac{dC}{dt} = \gamma I - \delta_C C.$$

where C denotes the density of cytotoxic cells, kIC represents the infected cells killing rate and γI the proliferation rate of cytotoxic cells induced by the presence of infected cells.

A simplfied 2nd-order model is used by Nowak in the book *Evolutionary Dynamics* [4], and also in [7], under the quasi steady-state assumptions that T is constant and V is proportional to I. In such case, the model reduces to:

$$\frac{dI}{dt} = \alpha I - kIC - \delta_I I,$$
$$\frac{dC}{dt} = \gamma I - \delta_C C.$$

Remark that in the recent survey paper [6] (2013) it is mentioned that the immune response models have not been validated from experimental data and the relevance of the immune response modelling is still a subject of research and debate.

4.3 Mutants

For simplicity, as in [5], we consider the very simplest case where there is one wild-type virus (with concentration V_w) and one mutant (with concentration V_m).

The basic model is then expanded as follows:

$$\frac{dT}{dt} = \lambda - \delta_T T - \beta_w V_w T - \beta_m V_m T,$$
$$\frac{dI_w}{dt} = \beta_w(1-\varepsilon_w)V_w T + \beta_m \varepsilon_m V_m T - \delta_I I_w,$$
$$\frac{dI_m}{dt} = \beta_w \varepsilon_w V_w T + \beta_m(1-\varepsilon_m)V_m T - \delta_I I_m,$$
$$\frac{dV_w}{dt} = p_w I_w - \delta_V V_w,$$
$$\frac{dV_m}{dt} = p_m I_m - \delta_V V_m.$$

In these equations, I_w and I_m represent the densities of cells infected by wild and mutant viruses respectively, ε_w is the probability of mutation from wild-type to mutant and ε_m the inverse mutation probability. In absence of therapy, it can be shown that the selection favors the wild-type virus if $\varepsilon_w > \varepsilon_m$ and $\beta_w p_w > \beta_m p_m$.

4.4 Latency

Latency refers to the fact that the infection of CD4+T-cells can give latent cells that carry the virus genome and reproduce but do not produce new viruses. To take this situation into account, the basic model can be expanded as follows (e.g. [5], [8]):

$$\frac{dT}{dt} = \lambda - \delta_T T - \beta V T,$$
$$\frac{dI}{dt} = q\beta V T - \delta_I I,$$
$$\frac{dL}{dt} = (1-q)\beta V T - \delta_L L,$$
$$\frac{dV}{dt} = p_I I + p_L L - \delta_V V.$$

In this model L is the density of latent cells and q is the probability that, upon infection, a cell will enter active viral replication. In [5], $p_L << p_I$ is a small parameter representing the virus production at a much slower rate than normal replication. In contrast, in [8], $p_L(t)$ is a pulsed signal representing random intermittent viral "blips".

4.5 Antiretroviral therapies

Antiretroviral therapies are based on two fundamental mechanisms:

(1) using drugs containing *reverse transcriptase inhibitors* which reduce the virus production by infected cells; this is modelled by multiplying the infection term by $(1-u_1)$ where $0 \leqslant u_1 \leqslant 1$ is the control input that represents the effectiveness of the drug;

(2) using drugs containing *protease inhibitors* which partially prevent the produced viruses to be infectious; the effectiveness of these drugs is denoted u_2 with $0 \leqslant u_2 \leqslant 1$.

The basic model is then expanded as follows (see e.g. [6]):

$$\begin{aligned}
\frac{dT}{dt} &= \lambda - \delta_T T - (1-u_1)\beta V T,\\
\frac{dI}{dt} &= (1-u_1)\beta V T - \delta_I I,\\
\frac{dV}{dt} &= (1-u_2) pI - \delta_V V.
\end{aligned}$$

Remark that this nonlinear control system is a flat system !

4.6 Resistance to antiretroviral therapies

Antiretroviral therapies may fail because of virus mutation. In order to deal with this issue, the equations must combine both the presence of mutations and the use of therapies. We consider the general case where there are n virus mutants and m antiretroviral therapies which can be used in combination. In such case, the models proposed above can be extended in the following way:

$$\frac{dT}{dt} = \lambda - \delta_T T - \sum_{j=1,n} \left[\beta_j V_j T \Big(\prod_{k=1,m} (1-\eta_{jk}u_k) \Big) \right],$$

$$\frac{dI_i}{dt} = \beta_i V_i T \Big(1 - \sum_{\substack{j=1,n\\ j\neq i}} \varepsilon_{ij} \Big) \Big(\prod_{k=1,m} (1-\eta_{ik}u_k) \Big) + \sum_{\substack{j=1,n\\ j\neq i}} \left[\beta_j V_j T \varepsilon_{ji} \Big(\prod_{k=1,m} (1-\eta_{jk}u_k) \Big) \right] - \delta_I I_i,$$

$$\frac{dV_i}{dt} = p_i I_i - \delta_V V_i, \qquad i = 1, \ldots, n.$$

In these equations, ε_{ij} is the probability of mutation from species i to species j, u_i is the intensity of the i-th drug administration, η_{ik} is the effectiveness of the i-th drug against mutant k. The other notations are obvious. This model can also be written in a more

compact matrix form:

$$\frac{dT}{dt} = \lambda - T(\mathbf{V}^T\mathbf{b}(\mathbf{u})),$$
$$\frac{dI_i}{dt} = -\delta_I I_i + T(\mathbf{e}_i^T\mathbf{V})\phi_i(\mathbf{u}), \qquad i = 1,\ldots,n,$$
$$\frac{dV_i}{dt} = p_i I_i - \delta_V V_i,$$

with the definitions

$$\mathbf{u} \triangleq \mathrm{diag}\{u_1,\ldots,u_m\}, \quad \phi_j(\mathbf{u}) \triangleq \prod_{k=1,m}(1-\eta_{jk}u_k),$$
$$\mathbf{b}(\mathbf{u}) \triangleq (\beta_1\phi_1(\mathbf{u}),\ldots,\beta_n\phi(\mathbf{u}))^T, \quad \mathbf{V} \triangleq (V_1,\ldots,V_n)^T,$$
$$\mathbf{e}_i \text{ is a column vector with i-th entry} = \beta_i\Big(1 - \sum_{\substack{j=1,n\\ j\neq i}}\varepsilon_{ij}\Big)$$
$$\text{and j-th entry} = \beta_j\varepsilon_{ji} \text{ for } j \neq i.$$

Using quasi-steady state approximations (T constant and V_i proportional to I_i) as in [3], we get a linear model for the dynamics of the infected cells:

$$\frac{dI_i}{dt} = (\mathbf{e}_i^T\mathbf{D}\mathbf{I})\phi(\mathbf{u}) - \delta_I I_i, \qquad i = 1,\ldots,n,$$

with the definitions

$$\mathbf{I} \triangleq (I_1,\ldots,I_n)^T, \qquad \mathbf{D} \triangleq \mathrm{diag}\{p_1T/\delta_V,\ldots,p_nT/\delta_V\}.$$

4.7 Resistance to immune response

In the book *Evolutionary Dynamics* [4], Nowak proposes a model to analyze the resistance against immune response induced by virus mutations. Using the quasi-steady state approximations T constant and V_i proportional to I_i, the model is as follows:

$$\frac{dV_i}{dt} = \alpha V_i - k_1V_iC_i - k_2V_iC_0, \qquad i = 1,\ldots,n,$$
$$\frac{dC_i}{dt} = \gamma V_i - \delta_C C_i - k_3C_i\sum_{i=1,n}V_i \qquad i = 1,\ldots,n,$$
$$, \quad \frac{dC_0}{dt} = k_4\sum_{i=1,n}V_i - \delta_C C_0 - k_3C_0\sum_{i=1,n}V_i.$$

In this model, V_i denotes the viral load of mutant i, C_i denotes the immune response specifically directed against virus strain$i$ and C_0 denotes a global immune response against

all virus strains. Furthermore, it is assumed that "the virus can impair immune responses" and this effect is represented by the terms $-k_3 C_i V$ and $-k_3 C_0 V$.

This model has been received with skepticism and does not seem to have been experimentally validated. Furthermore, it is not fully consistent with the previous models presented above.

Bibliography

[1] R.J. De Boer and A.S. Perelson. Target cell limited and immune control models of HIV infection: a comparison. *Journal of Theoretical Biology*, 190:201–214, 1998.

[2] P. De Leenheer and H.L. Smith. Virus dynamics: a global analysis. *SIAM Journal of Applied Mathematics*, 63(4):1313–1327, 2003.

[3] E. Hernandez-Vargas, P. Colaneri, R. Middleton, and F. Blanchini. Discrete-time control for switched positive systems with application to mitigating viral escape. *International Journal of Robuste and Nonlinear Control*, 2010.

[4] M. A. Nowak. *Evolutionary Dynamics - Exploring the Equations of Live.* Harvard University Press, 2006.

[5] M. A. Nowak, S. Bonhoeffer, G.M. Shaw, and R.M. May. Anti-viral drug treatment: dynamics of resistance in free virus and infected cell populations. *Journal of Theoretical Biology*, 184:203–217, 1997.

[6] A.S. Perelson and R.M. Ribeiro. Modeling the within-host dynamics of HIV infection. *BMC Biology*, 1741-7007/11/96, 2013.

[7] R.M. Ribeiro and S. Bonhoeffer. Production of resistant HIV mutants during antiretroviral therapy. *PNAS*, 97:7681–7686, 2000.

[8] L. Rong and A.S. Perelson. Modeling latently infected cell activation: viral and latent reservoir persistence, and viral blips in HIV-infected patients on potent therapy. *PLOS Computational Biology*, 5(10):e1000533, October 2009.

[9] D. Wodarz, R.A. Arnaout, M.A. Nowak, and J.D. Lifson. Transient antiretroviral treatment during acute simian immunodeficiency virus infection facilitates long-term control of the virus. *Phil. Trans. Royal Society of London B*, 355:1021–1029, 2000.