Mathematical Modelling and Analysis
Volume 23, Issue 3, 473–491, 2018
https://doi.org/10.3846/mma.2018.028

http://mma.vgtu.lt
ISSN: 1392-6292
eISSN: 1648-3510

Supply Chain Modeled as a Metabolic Pathway*

Almaz Mustafin and Aliya Kantarbayeva

Satbayev University
22 Satbayev St., Almaty 050013, Kazakhstan
E-mail(*corresp.*): `butsuri123@gmail.com`

Received October 14, 2017; revised June 20, 2018; accepted June 20, 2018

Abstract. A new model of economic production process is proposed (in the form of a set of ODEs) based on an idea that nonconsumable factors of production facilitate the conversion of inputs to output in much the same catalytic way as do enzymes in living cells when transforming substrates into different chemical compounds. The output of a converging, multi-resource, single-product supply chain network is shown to depend on the minimum of its inputs in the form of the Leontief–Liebig production function, providing the validity of the clearing function approximation. In turn use of the clearing function is legitimate when the machine processing time is much shorter than the machine loading time.

Keywords: supply chain, production function, limiting factor, clearing function.

AMS Subject Classification: 90B30.

1 Problem description

The neoclassical production function is a mathematical statement relating the rate of production of a certain finished, or partly finished, intermediate, commodity (output), Y, to a set of the involved factors of production, $\phi_1, \phi_2, \ldots, \phi_n$ (e.g., [33]):

$$Y = F(\phi_1, \phi_2, \ldots, \phi_n). \tag{1.1}$$

In its broad meaning, factor of production is any entity which can lead to increased output as its availability is increased. Factors may be of material, energy, human and financial nature. In what follows we restrict ourselves to material factors of production and the case of a single product.

* This is an extended and revised version of a preliminary conference report that was presented at the MMA2017 [25].

The factors of production may be of two types: consumable and nonconsumable. By convention, consumable factors are referred to as "inputs". We will use term "resource" interchangeably with "input", in spite of economists quite often identify resources with factors of both types. By consumption we understand irreversible conversion, physical embodiment, of a resource into a material product. As the production process takes place, the resource is certainly consumed, used up. Consuming the resource means tending to reduce its availability.

The nonconsumable factors of production, commonly known as primary factors, to which belong land, capital and labor, are often lumped together as "funds". They are not resources by the definition in use. This is not to imply that funds are less important, but that they must be treated in a different way from resources. Funds are not materially transformed into an output they produce. They are transforming tools that turn the involved inputs into a product, but are not themselves embodied physically in the product. Although funds are not used up, their amount can change and they are subject to wear-and-tear.

In terms of dimensions, output Y in formula (1.1), being the quantity of the commodity produced in a unit of time, is a flow variable. Factors of production that represent resources, most commonly are flows, although in some cases they may be stocks. Funds always are stock variables.

Cybernetically, a manufacturing technology for the single-product case may be considered as a converter of the resources $R_1, R_2, \ldots, R_n$ into a product P by means of the funds $\varphi_1, \varphi_2, \ldots, \varphi_m$:

$$\underset{\text{inputs}}{R_1, R_2, \ldots, R_n} \longrightarrow \boxed{\begin{array}{c}\text{TECHNOLOGY}\\ \underset{\text{funds}}{\varphi_1, \varphi_2, \ldots, \varphi_m}\end{array}} \longrightarrow \underset{\text{product}}{P}$$

Resources are fed to the converter from the outside, while funds act inside of the black box of the technology. Funds are not spent on the output, however they function to make the transformation of inputs to product feasible and efficient, and to enable control of that transformation.

A special case of production function implying zero substitutability between inputs is the Leontief technology [18, p. 38], first proposed for the mathematical apparatus of the input–output (IO) analysis which arose to deal with the problem of interindustry demand. It has the form

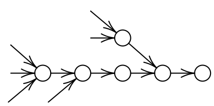
$$Y = Y_0 \min\left(x_1/x_{10},\, x_2/x_{20}, \ldots,\, x_n/x_{n0}\right), \tag{1.2}$$

where x_i are inputs, x_{i0} are the constant per unit input requirements, Y is output, and Y_0 is the scale factor having the dimension of Y. The Leontief production function (1.2) assumes there is only one technique for producing output and requires combination of the inputs in a fixed ratio; the elasticity of substitution is zero. In the basic version of the Leontief technology, intended for static analysis, inputs are meant to be resource influxes.

In such sectors as agriculture, forestry and fishery the use of production function with stock arguments proved to be more appropriate because it is dictated by the very specifics of the relationship between yield (harvest, crop, catch, etc.) and nutrients. This is expressed by the Liebig production function [27], in its simplest version formally identical with the Leontief function (1.2). The difference is that arguments x_i in the Liebig function have the meaning of concentrations. As a matter of fact, the Liebig production function is a mathematical formulation of the famous "law of the minimum" [5]. This law states that the rate of growth of a plant or crop is affected not by the most abundant mineral resource, but by the most deficient one. Essentially, a plant will only yield as much as the least available nutrient allows.

Neither the Leontief, nor the Liebig production functions are derived; they are merely postulated. Formally, the Leontief function can be inferred from the production function with constant elasticity of substitution (CES), as is done, e.g., in [35, p. 20]. But still the CES production function per se is nothing more than a formal mathematical construction.

In an industry sector, a path from primary resources to a finished product may run through a complex supply chain network of "elementary" production units—IO converters. In such a network, the product of one unit, or workstation, serves as the resource for the other. We restrict our consideration to the case of converging structure of supply chain, which allows for multiple initial resources and a single final product. The following commutative diagram shows an example of one-product supply chain with six initial suppliers:



Here circles stand for IO production units, while arrows indicate material flows.

As empirical studies evidence, in some occasions, at least in the short run, the production function of a whole sector can be effectively represented by the Leontief function [34]. The resulting lumped, or aggregated, production function depends not on the total number of the factors of production involved, but on the scarcest one (limiting factor). The behavior of the entire system turns out to be governed by only very few degrees of freedom. The questions arise:

- What are the reasons for such an enormous reduction of the description? Namely, what peculiar properties of the supply chain generate the production function of the Leontief type?
- Does the production system have the property of scale invariance?
- How are the formalisms of Leontief and Liebig interrelated?

These and other questions are addressed in the present paper.

In this paper we suggest what may be called a reverse engineering of the black box of the fixed proportions production function. In effect, we construct a hypothetical open box that behaves in the same way in the hope to deduce design features of the black box. Our approach differs radically from the conventional and is based on an idea that the transformation of resources into products in a manufacturing process occurs similar to the conversion of substrates into new substances in enzyme-catalyzed biochemical reactions. As is known, in a living cell a substrate molecule binds to an enzyme to form a short-lived substrate-enzyme complex. The complex then breaks up into a product and the original enzyme, which can then catalyze a new reaction (e.g., [4]). In any production unit of a supply chain, funds play role of enzymes. A set of raw materials, or parts, or semi-finished goods is jointly transformed into a complete product by means of machinery and equipment, and qualified and trained workforce. As we are going to demonstrate, the rate of production is given by the minimum of its input supplies and funds, and this property does not depend on the length of the chain and how much it is branched.

It is worth noting that the models we study do not pursue any optimal economic solutions under imposed constraints, such as cost minimizing or profit maximizing. We rather accentuate the emergent self-regulation in the system being considered, which is not due to external interference of any rational agents.

2 Brief literature review

At present, queueing theory, a major branch of operations research that deals with mathematical models of waiting lines using the apparatus of probability theory, is held to be the principal mathematical tool to describe production processes (e.g., [14]). A comprehensive review of the literature on applications of queueing theory in manufacturing is clearly beyond the scope of this paper, especially since we employ a different approach. Instead, we briefly review the results of this theory most relevant to our study.

A queueing phenomenon in manufacturing is characterized by three main elements: (i) jobs (production lots or parts); (ii) the queue (buffer); and (iii) the server (machine), the purpose of which is to handle the jobs. The lots arrive from outside the system according to a statistical distribution of their interarrival times. Any lot joins a queue in the buffer and waits until the machine is available. At various times, lots are selected for processing by the machine. The basis on which the jobs are selected is determined by the queue discipline. Typically, jobs are served in order of arrival.

Important to us are implications of the most basic and tractable queueing model M/M/1. In standard Kendall's notation, letters "M" and "M" in the descriptor designate exponential distribution of interarrival and processing times respectively ("M" comes from Markovian), and "1" indicates that the number of parallel servers is just one. Let λ^{-1} be the mean interarrival time and μ^{-1} the mean processing time. Define the utilization, $u = \lambda/\mu$, as the fraction of time the machine is not idle. Utilization has no dimension and can never exceed 1, otherwise, the queue length will explode. Let W be the current work in progress (WIP), i.e. the total number of jobs in the manufacturing system, or the total length of the queue. Then it can be shown [ibid., ch. 8] that in steady-state conditions, the utilization is related to the WIP by the formula $W = u/(1-u)$. The throughput, Y, of any queueing system is the rate at which jobs successfully leave the system. For the M/M/1 infinite buffer case, $Y = \lambda$ if the system is stable. (Everything that arrives must eventually depart.) Therefore, one may obtain $Y = \mu W/(1 + W)$. As is seen, there is a trade-off between the throughput and the WIP in the M/M/1 model. If a high throughput is required, the machine should always be busy and the inventory level needs to be high, thereby lengthening the time a job is in the manufacturing system (from entering the buffer in front of the machine until leaving the machine). Conversely, if the WIP level is low, the machine is not processing for most of the time, yielding a small throughput.

The deterministic relation between the WIP and the throughput of a steady-state production process bears a name of "clearing function". Dealing with stochastic model of real production process involves severe computational load. Instead, there is a general consensus to employ proper aggregate models intended to represent average behavior in some sense. Nowadays the formalism of nonlinear clearing functions devised by U. Karmarkar [16] shows considerable promise in production planning. The Karmarkar clearing function has the following form:

$$Y = Y_m W/(K + W), \tag{2.1}$$

where Y is the throughput and W is the WIP. Constant Y_m represents the maximum yield achieved by the production system at saturating WIP level; constant K is the WIP at which the throughput is half of the maximum. The alternative forms of clearing function sharing the common properties of monotonicity, concavity and saturation have been proposed as well [21].

Liebig's law of the minimum, or the effect of limiting (constraining) factor, is known popularly in production networks as the immanence of bottlenecks. A bottleneck is usually defined as a process in a production chain, such that its limited capacity reduces the performance of the whole chain. The first systematical treatment of methods for identifying bottlenecks in production was given by E. Goldratt [9]. If bottlenecks are static, conventional techniques for their detection may be applied, such as the utilization method, by which the machine with the highest utilization is considered the bottleneck, and the queue length method, by which the machine with the longest queue length or waiting time is judged to be the bottleneck.

However, it is quite challenging to detect a so-called wandering, or shifting, bottleneck, which tends to evolve over time [30]. Wandering bottlenecks appear because of unexpected random events such as machine failures that disturb the smooth flow of jobs. As a result, in the course of manufacturing, the bottleneck machines might shift temporarily. The shifting bottleneck distinguishes between a momentary bottleneck, describing the bottleneck at a given time, and an average bottleneck describing rather the bottleneck behavior over a selected period of time. A more robust method for identifying shifting bottlenecks has been suggested [ibid.], based on measuring the period of time a machine is active without any interruption. By active time is meant time when the machine is causing the following machine to wait. This method allows to find several bottlenecks and sort them according to their size. Other effective methods have been developed as well. For instance, the shifting bottleneck heuristic [1, 22], that employs disjunctive graphs to model the dependency of job processing on different machines. Based on the calculation of longest paths within the disjunctive graphs, the overall scheduling problem is decomposed into smaller scheduling problems for single or parallel machines.

Production lines have multiple steps that collectively perform the required task. It is desired that jobs are processed and requested dynamically by every machine in the system as to maintain a steady flow of jobs leading to the bottleneck machines. The effective lot flow control is possible only through coordinated operations of the machines. Supply chain coordination is the process of managing dependencies between supply chain entities in order to achieve mutually defined goals. A number of decentralized-coordination models are developed to improve the tradeoff between throughput and lead time [6, 31].

Presumably, N. Georgescu-Roegen was the first to come up with an idea that a production factor such as labor is like a catalyst [7, p. 319]. Virtually at the same time, I. Poletaev proposed a general mathematical theory of systems with limiting factors, in which production yield is given by a switching function of the Leontief–Liebig type with such arguments as flows of input components and stocks of fund components [17]. Independently, D. Chernavskii deepened the analogy between biosynthesis and industrial production. In particular, he pointed out: "Both by spirit and by methods of research, modeling the economic and production processes is closely related to the subject we expounded above. There is nothing surprising in that biological systems with their basic variables—concentrations of substances—are similar to economic ones, where variables are the quantities of products or commodities, and the role of enzyme concentration is played by the number of machines in a shop or automatic line. In this regard, both the kinetic models of biophysics and biochemistry, and the economic models belong to the common branch of cybernetics, the so-called theory of complex systems" [29, p. 134, own translation].

Though the above mentioned inspiring insights remained barely noticed over the years, the interest to the problem has been revived after appearance of the concept of "industrial metabolism" proposed by R. Ayres [3]. The word "metabolism" in its original biological meaning characterizes the totality of internal biochemical processes in living organism. An individual cell or the whole organism consumes energy-rich, low-entropy substances to maintain its basic functions, as well as for growth and reproduction. This process is necessarily accompanied by the release of high-entropy waste. Industrial metabolism is an integrated set of physical processes aimed at transforming raw materials, energy, labor and capital into goods and associated waste. The analogy between biological and industrial metabolism is about the fact that in both types of systems takes place the conversion of material substances driven by a flow of free energy. These ideas stimulated the appearance of works by one of the authors [24, 26]. Treating the act of resource-to-product conversion as a sort of enzyme-catalyzed reaction was shown to result in hyperbolic dependence of the output on the inventory, similar to the Michaelis–Menten saturation curve. The throughput of a cascade of such converters is shown to be determined by a single limiting production factor.

In recent years, sophisticated tools borrowed from the arsenal of queueing theory and supply chain management become more and more relevant in the studies of intracellular metabolic networks [12, 15, 19]. Research work in this direction is spoken of as "biologistics" [10]. In spite of the growing awareness that the biochemical activity of a living cell is similar to the operation of an industrial factory where products of one machine are used by other machines for manufacturing of their own products, biologistics, however, is lacking in attempts to recognize enzyme catalysis mechanism underlying a manmade production system, not the other way around. In other words, nobody attempts to consider machine as an enzyme, rather than protein as a molecular machine. To the best of our knowledge, the origins of the Leontief production function in supply chains has not been tackled so far to any noticeable extent. The present work is focused on that interesting problem.

"The Mecca of the economist lies in economic biology rather than in economic dynamics" wrote A. Marshall, one of the founders of neoclassical economics, in the preface to his famous *Principles of Economics* [20, p. xxv]. Marshall's holy place may yet be attained as mechanistic postulates so long dominated economics will give way to the holistic approach offered by the modern theory of complex systems.

3 The generic model of a production unit

Consider a simple production unit—an IO processor, which converts a single resource into a product. The unit consists of a fixed number of identical machines. It is precisely the machinery that represents a nonconsumable factor of production, or fund, in the manufacturing unit under consideration. Each of the machines can process any of the arriving lots of the resource and we assume here that they do so one at a time. The resource arrives at the production unit from the outside, and if all machines are busy processing jobs, the arriving portion of the resource has to wait. In our model the terms "inventory", "work in progress" (WIP) and "buffer stock" are regarded as synonyms and mean the current stock of resource in the unit. Lots waiting for service pile up in a common buffer which feeds all machines. When a machine finishes the processing of its current job, it grabs another portion of the resource from the buffer.

We can write the scheme of this event in the form of pseudochemical equations:

$$\begin{tikzcd} {} \arrow[r, "r"] & x \arrow[u, "q"] \arrow[r, "a \quad \alpha"] & v \arrow[r, "\beta \quad b"] \arrow[d, bend left, "\beta"] & y \\ & & u \arrow[u, bend left, "\alpha"] & \end{tikzcd} \tag{3.1}$$

Here r is the supply rate of the resource (say, in lots per time unit), x is the inventory, q is the specific rate with which the resource is being lost (or dispatched to a storage, or exported elsewhere), u is the number of idle machines, v is the number of busy (operative) machines, and y is the quantity of successfully produced commodity. The constants a, b, α and β depict the various rates with which these processes proceed.

A flow diagram such as the one given by equations (3.1) can be translated into a set of differential equations that describe rates of change of stock quantities of the participating material agents. The diagram (3.1) encodes both the sequence of steps and the rates with which these steps occur. To write corresponding equations, we naturally can choose to use what chemists call the "law of mass action", which states that when two or more agents are involved in a conversion step, the rate of conversion is proportional to the product of their quantities. By convention, the mass-action rate constants are the proportionality constants. Unlike standard chemical reaction schemes, in (3.1) more than one rate constant—by the number of input and output agents—may correspond to one transformation step, because dimensions of quantities may not coincide. Namely, a is the capture rate of a unit of the resource by a machine, α stands for how many idle machines get involved in work in a unit of time per unit of resource, β^{-1} is the mean processing time of a machine (service time of a job at a machine), and b is the output per a machine, or the number of units of product that a single machine will deliver at one unit of time. They are indicated in the diagram as arrow labels.

Keeping track of each participant allows us to derive the following set of equations:

$$\mathrm{d}x/\mathrm{d}t = r - aux - qx, \tag{3.2a}$$
$$\mathrm{d}u/\mathrm{d}t = \beta v - \alpha ux, \tag{3.2b}$$
$$\mathrm{d}v/\mathrm{d}t = \alpha ux - \beta v, \tag{3.2c}$$
$$\mathrm{d}y/\mathrm{d}t = bv, \tag{3.2d}$$

where t is time. All parameters in the model are nonnegative.

Adding equations (3.2b) and (3.2c) reveals a conserved quantity u_0, the total number of machines, idle and busy: $u + v = u_0$. This is not at all surprising, since fund is neither formed nor destroyed in the process of manufacturing.

With the aid of this conservation law the system (3.2) can be simplified by eliminating either u, or v. We arbitrarily choose to eliminate u. Furthermore, we see that (3.2d) is just a slave equation with respect to (3.2a), (3.2b) and (3.2c); it can always be solved later on, once solutions for x, u and v are known. These steps lead to the following:

$$\mathrm{d}x/\mathrm{d}t = r - ax(u_0 - v) - qx, \tag{3.3a}$$
$$\mathrm{d}v/\mathrm{d}t = \alpha x(u_0 - v) - \beta v. \tag{3.3b}$$

Introduce new dimensionless parameters: the loss rate constant $\gamma = q/(au_0)$ and the influx $\varrho = \alpha r/(a\beta u_0)$, such that $\gamma \ll 1$ and $|\varrho - 1| \gg \gamma$. In terms of γ and ϱ, to $\mathcal{O}(\gamma)$ the steady-state solutions of (3.3) are

$$\overline{x}_\pm = \begin{cases} \frac{\beta}{\alpha}\left(\frac{\varrho-1}{\gamma} + \frac{1}{\varrho-1} - \frac{\gamma\varrho}{(\varrho-1)^3}\right), & \text{for } \pm(\varrho-1) > 0; \\ \frac{\beta\varrho}{\alpha}\left(\frac{1}{1-\varrho} - \frac{\gamma}{(1-\varrho)^3}\right), & \text{for } \pm(1-\varrho) > 0; \end{cases} \tag{3.4a}$$

$$\overline{v}_\pm = \begin{cases} u_0\left(1 - \gamma/(\varrho-1)\right), & \text{for } \pm(\varrho-1) > 0; \\ \varrho u_0\left(1 - \gamma/(1-\varrho)\right), & \text{for } \pm(1-\varrho) > 0. \end{cases} \tag{3.4b}$$

In (3.4), fixed point $(\overline{x}_+, \overline{v}_+)$ is physically feasible because (i) it is always positive, and (ii) asymptotically stable: both eigenvalues of the Jacobian matrix of the system (3.3) evaluated at $(\overline{x}_+, \overline{v}_+)$ have negative real parts. The solution $(\overline{x}_-, \overline{v}_-)$ is nonphysical because it yields $\overline{x}_- < 0$ for all legitimate values of ϱ.

Moreover, it is always saddle-type unstable. From now on we drop subscript at the feasible fixed point. Inserting $\overline{v}$ into (3.2d) gives the steady-state output of the IO unit under consideration, i.e. the production function. To zeroth order in γ,

$$\mathrm{d}y/\mathrm{d}t = b \min\left(\alpha r/(a\beta),\, u_0\right). \tag{3.5}$$

(Here we reverted to the dimensional parameters.) Quite apparently, the equation (3.5) is a Leontief–Liebig production function for two factors, r and u_0. The former is flow and the latter is stock. They enter the function on equal terms. In our approach, flows and stocks are "equalized in rights". Thus, the steady-state output of an elementary resource-product converter is determined either by the resource supply rate or by the given installed capacity, whichever is in shortest availability. At subcritical arrival rates, when $\varrho < 1$, the fraction of busy machines is of order ϱ, so the machinery is not a limiting factor. At supercritical arrival rates, however, when $\varrho > 1$, all the machines are engaged barely coping with the huge WIP that becomes inversely related to γ.

In steady state, one obtains from (3.3b) how $\overline{x}$ and $\overline{v}$ are related: $\overline{v} = \alpha u_0 \overline{x}/(\beta + \alpha\overline{x})$. Substituting this in (3.2d) yields

$$\mathrm{d}y/\mathrm{d}t = b u_0 \overline{x}/\left[(\beta/\alpha) + \overline{x}\right]. \tag{3.6}$$

Being the dependency of the throughput (i.e. the number of lots per unit of time that leave the manufacturing system) on the current WIP, (3.6) is nothing but the clearing function of the production unit (3.1) under consideration. In more exact terms, we derived the clearing function of the Karmarkar type [16] (cf. (2.1)).

One can recognize (3.6) as another version of the Michaelis–Menten equation of enzyme kinetics [4]. The hyperbolic IO relationship of the type (3.6) is not uncommon in biology: it describes the sigmoidal oxygen-binding curve of haemoglobin and the fraction of a macromolecule saturated by ligand as a function of the ligand concentration (Hill equation [8]), growth rate of microorganisms in a nutrient solution (Monod equation [23]), numerical response of predator to prey population density (Holling type II response [13]), and the like. A distinguishing characteristic of the equation (3.6) is saturated response of the output to the inventory. For low levels of $\overline{x}$, the output is roughly proportional to $\overline{x}$. At high $\overline{x}$ levels, though, the rate of production approaches a constant value, bu_0.

Recall that in deriving (3.6) we considered the production unit in a steady-state mode of operation. Now we are going to show that under certain additional assumptions the relationship (3.6) remains valid even in nonsteady-state conditions.

To begin with, we nondimensionalize the equations (3.3) by introducing the following scaled variables and parameters (in addition to already defined γ and ϱ): $\xi = \alpha x/\beta$, $\eta = v/u_0$, $\tau = tau_0$, and $\varepsilon = au_0/\beta$.

The quantity $(au_0)^{-1}$ is chosen to be a new unit of time. It is a characteristic time a job spends waiting (in the buffer) before beginning service. In other words, it is the resource lifetime in the production unit. This time is seen to be inversely proportional to the total number of installed machines, u_0.

The dimensionless equations then become

$$\mathrm{d}\xi/\mathrm{d}\tau = \varrho + (\eta - 1 - \gamma)\xi, \quad \varepsilon\,\mathrm{d}\eta/\mathrm{d}\tau = \xi - \eta(1+\xi). \tag{3.7}$$

According to the chosen scaling, ε is the ratio of the processing time to the characteristic waiting time. Hereinafter we assume this ratio to be small: $\varepsilon \ll 1$. This is a necessary condition for most of the subsequent reasoning to be valid, although we are aware that the assumption made can not be ensured for every existing production process. We will turn back to this topic in Section 6.

Inasmuch as $\varepsilon \ll 1$, the system (3.7) is singularly perturbed. The slow variable is resource, ξ, and the fast variable is the number of machines in service, η. The standard practice of reducing such systems is multiple-scale analysis (e.g., [28]) whereby fast variable is adiabatically eliminated. One has to establish the validity of the adiabatic elimination in each specific case. In particular, Fenichel–Tikhonov theorem requires, among other things, (i) quasi-steady state of the fast equation to be an isolated root of the algebraic equation $\mathrm{d}\eta/\mathrm{d}\tau = 0$ and to retain stability at all allowed values of the slow variable, and (ii) initial conditions of the fast equation to fall within the domain of influence of that quasi-steady state [ibid.]. It is worthy of note that from the chemists' side M. Bodenstein pioneered the quasi-steady-state approximation as far back as in 1913. The influential work to clarify the applicability of the technique to enzymatic reactions have been carried out by L. Segel and M. Slemrod [32].

To decompose system (3.7) into fast and slow parts, introduce fast time variable $\vartheta = \tau/\varepsilon$. Now rescale (3.7) by replacing τ with $\vartheta\varepsilon$ and, after taking $\varepsilon = 0$, it becomes

$$\mathrm{d}\xi/\mathrm{d}\vartheta = 0, \quad \mathrm{d}\eta/\mathrm{d}\vartheta = \xi - \eta(1+\xi). \tag{3.8}$$

This is the fast subsystem, where ξ is replaced by its initial value and treated as parameter. It yields the inner solution, valid for $\tau = \mathcal{O}(\varepsilon)$.

Setting $\varepsilon = 0$ in (3.7) leads to the slow subsystem

$$\mathrm{d}\xi/\mathrm{d}\tau = \varrho + (\eta - 1 - \gamma)\xi, \tag{3.9a}$$
$$0 = \xi - \eta(1+\xi), \tag{3.9b}$$

which produces the outer solution, valid for $\tau = \mathcal{O}(1)$. In this singular limit as $\varepsilon \to 0$, the subsystem defines a slow flow along the curve (slow manifold) given by (3.9b). Outer solution is valid for those values of ξ, for which the quasi-steady states of the fast subsystem (3.8) are stable.

The quasi-equilibrium for the fast subsystem (3.8) is given by

$$\eta = \xi/\,(1+\xi) \tag{3.10}$$

and it is asymptotically stable for any positive ξ. Quantity η is the simultaneous fraction of busy machines, v/u_0. As long as the momentary WIP keeps small, i.e. $\xi \ll 1$ (or, in dimensional form, $x \ll \beta/\alpha$), this fraction remains adequately small, meaning the equipment base is strongly underloaded. However at high levels of WIP, for $\xi \gg 1$ (or $x \gg \beta/\alpha$), all the installed machines become busy. Note that the fraction of busy machines does not depend on ε.

Hence, it follows from (3.9) that for time scales on the order of $\tau = \mathcal{O}(1)$ the process of resource-to-product conversion is given by the equations

$$\mathrm{d}\xi/\mathrm{d}\tau = \varrho - \xi/(1+\xi) - \gamma\xi, \tag{3.11a}$$

$$\mathrm{d}\zeta/\mathrm{d}\tau = \mu\,\xi/(1+\xi). \tag{3.11b}$$

Here we have written down the slave equation (3.2d) in dimensionless form by having introduced a normalized product quantity $\zeta = y/y_0$ and the combined parameter $\mu = b/(ay_0)$, where y_0 is a proper unit for y.

The equation (3.11b) is the clearing function in nondimensional form. It looks formally identical with (3.6), however as opposed to (3.6), the argument ξ standing for WIP does not have to be constant in time. In deriving (3.11b) we did not require the production unit to operate in a steady-state mode. And yet, the number of busy machines, η, being the fast variable, after a short transient of order $\mathcal{O}(\varepsilon)$ keeps in a quasi-steady state with respect to the current inventory, ξ. In (3.11a), the resource supply rate, ϱ, may not be necessarily constant, but if the timescale of its typical variations is much longer than the machine processing time, then the equation for the clearing function (3.11b) will remain valid.

4 Linear supply chain

Now we pass on to two serially connected production units operating by the generic mechanism as discussed in the preceding section:

$$
\begin{tikzcd}
 & q_0 & & q_1 & & \\
{} \arrow[r, "r_0"] & x_0 \arrow[u] \arrow[r, "a_1 \quad \alpha_1"] & v_1 \arrow[r, "\beta_1 \quad b_1"] \arrow[d, bend left, "\beta_1"] & x_1 \arrow[u] \arrow[r, "a_2 \quad \alpha_2"] & v_2 \arrow[r, "\beta_2 \quad b_2"] \arrow[d, bend left, "\beta_2"] & x_2 \\
 & & u_1 \arrow[u, bend left, "\alpha_1"] & & u_2 \arrow[u, bend left, "\alpha_2"] &
\end{tikzcd}
\tag{4.1}
$$

The first (upstream) unit converts the resource x_0 to the product x_1, which, in turn, serves as a resource to the second (downstream) unit. The second unit uptakes x_1 and converts it to the product x_2. The two units may represent a fragment of a sequential supply chain, or a linear production line.

Upon adiabatical exclusion of the fast (fund) variables u_1, v_1, u_2 and v_2 the corresponding balance equations for the slow variables x_0, x_1 and x_2 become:

$$\frac{\mathrm{d}x_0}{\mathrm{d}t} = r_0 - \frac{a_1\beta_1 u_{10} x_0}{\beta_1 + \alpha_1 x_0} - q_0 x_0, \tag{4.2a}$$

$$\frac{\mathrm{d}x_1}{\mathrm{d}t} = \frac{\alpha_1 b_1 u_{10} x_0}{\beta_1 + \alpha_1 x_0} - \frac{a_2\beta_2 u_{20} x_1}{\beta_2 + \alpha_2 x_1} - q_1 x_1, \tag{4.2b}$$

$$\frac{\mathrm{d}x_2}{\mathrm{d}t} = \frac{\alpha_2 b_2 u_{20} x_1}{\beta_2 + \alpha_2 x_1}. \tag{4.2c}$$

Here $u_{10} = u_1 + v_1$ and $u_{20} = u_2 + v_2$ are the respective installed machinery of units 1 and 2.

Defining the dimensionless quantities $\xi_0 = x_0\alpha_1/\beta_1$, $\xi_1 = x_1\alpha_2/\beta_2$, $\xi_2 = x_2/x_{20}$, $\tau = a_1 u_{10} t$, $\gamma_0 = q_0/(a_1 u_{10})$, $\gamma_1 = q_1/(a_2 u_{20})$, $\varrho_0 = r_0\alpha_1/(a_1\beta_1 u_{10})$,

$\varrho_1 = b_1\alpha_2 u_{10}/(a_2\beta_2 u_{20})$, $\mu_1 = a_2 u_{20}/(a_1 u_{10})$, and $\mu_2 = b_2 u_{20}/(a_1 u_{10} x_{20})$, where x_{20} is an appropriate unit for x_2, we rescale system (4.2) to

$$\mathrm{d}\xi_0/\mathrm{d}\tau = \varrho_0 - \xi_0/(1+\xi_0) - \gamma_0\xi_0, \tag{4.3a}$$

$$\mathrm{d}\xi_1/\mathrm{d}\tau = \mu_1\left[\varrho_1\xi_0/(1+\xi_0) - \xi_1/(1+\xi_1) - \gamma_1\xi_1\right], \tag{4.3b}$$

$$\mathrm{d}\xi_2/\mathrm{d}\tau = \mu_2\xi_1/(1+\xi_1). \tag{4.3c}$$

As a matter of convenience, introduce an auxiliary quantity $w_1 = \varrho_1\xi_0/(1+\xi_0)$ such that $\mu_1 w_1$ is the dimensionless output by the first unit. The values of ξ_0 and hence, w_1, do not depend on parameters of the second unit. The equation (4.3a) has the following steady-state solutions:

$$\bar{\xi}_0 = \begin{cases} \varrho_0/(1-\varrho_0) + \mathcal{O}(\gamma_0), & \text{for } \varrho_0 < 1; \\ (\varrho_0 - 1)/\gamma_0 + 1/(\varrho_0 - 1) + \mathcal{O}(\gamma_0), & \text{for } \varrho_0 > 1. \end{cases} \tag{4.4}$$

Accordingly, the steady states of w_1 turn out to be

$$\overline{w}_1 = \begin{cases} \varrho_0\varrho_1 + \mathcal{O}(\gamma_0), & \text{for } \varrho_0 < 1; \\ \varrho_1 + \mathcal{O}(\gamma_0), & \text{for } \varrho_0 > 1. \end{cases} \tag{4.5}$$

To zeroth order in γ_0, this is equivalent to

$$\overline{w}_1 = \varrho_1 \min\left(\varrho_0,\, 1\right). \tag{4.6}$$

To within a constant factor, formula (4.6) is the production function of the first unit, as is found above (cf. (3.5)).

For the steady-state values of ξ_1 we obtain from (4.3b):

$$\bar{\xi}_1 = \begin{cases} \overline{w}_1/(1-\overline{w}_1) + \mathcal{O}(\gamma_1), & \text{for } \overline{w}_1 < 1; \\ (\overline{w}_1 - 1)/\gamma_1 + 1/(\overline{w}_1 - 1) + \mathcal{O}(\gamma_1), & \text{for } \overline{w}_1 > 1. \end{cases} \tag{4.7}$$

It is a matter of direct verification to prove that positive steady states of the system of equations (4.3a) and (4.3b) are stable.

Substituting the steady-state values of ξ_1 in equation (4.3c) yields, to $\mathcal{O}(1)$ in γ_0 and γ_1,

$$\mathrm{d}\xi_2/\mathrm{d}\tau = \mu_2 \min\left(\overline{w}_1,\, 1\right) = \mu_2 \min\left(\varrho_0\varrho_1,\, \varrho_1,\, 1\right), \tag{4.8}$$

where we used the equation (4.6). This is the dimensionless production function of the unbranched two-link supply chain (4.1). Turning back to the dimensional quantities, we get

$$\mathrm{d}x_2/\mathrm{d}t = b_2 \min\left(\alpha_1\alpha_2 b_1 r_0/(a_1 a_2\beta_1\beta_2),\, \alpha_2 b_1 u_{10}/(a_2\beta_2),\, u_{20}\right). \tag{4.9}$$

This is another Leontief–Liebig production function for the arguments r_0, u_{10}, and u_{20}. Again, the overall output is controlled either by the resource supply rate or by an installed capacity of one of the two production units, whichever is more deficient.

Suppose, of two IO units placed in series in (4.1), the second unit happens to control the overall output, while parameters of the first unit do not affect the operation of the chain. According to equation (4.9), this situation corresponds to $u_{20} < \alpha_2 b_1 u_{10}/(a_2\beta_2)$ and $u_{20} < \alpha_1\alpha_2 b_1 r_0/(a_1 a_2 \beta_1\beta_2)$. As this takes place, there is a substantial level of the inventory, x_1, in the buffer of the second production unit—to the extent that the smallness of q_1 warrants. Indeed, by formula (4.7),

$$\overline{x}_1 \approx \begin{cases} \dfrac{\alpha_1\alpha_2 b_1 r_0 - a_1 a_2 \beta_1\beta_2 u_{20}}{a_1\alpha_2\beta_1 q_1}, & \text{for } \alpha_1 r_0 < a_1\beta_1 u_{10}; \\ \dfrac{\alpha_2 b_1 u_{10} - a_2\beta_2 u_{20}}{\alpha_2 q_1}, & \text{for } \alpha_1 r_0 > a_1\beta_1 u_{10}. \end{cases} \tag{4.10}$$

In other words, the WIP piles up in the second unit to such a level, that makes the rate of production of x_2 practically insensitive to the variations in x_1:

$$\mathrm{d}x_2/\mathrm{d}t = b_2 u_{20} \tag{4.11}$$

in conformity with the equation (4.2c).

Now consider the case of $u_{10} < a_2\beta_2 u_{20}/(\alpha_2 b_1)$ and $u_{10} < \alpha_1 r_0/(a_1\beta_1)$ for which the first unit operates relatively slow. Then, in view of (4.9), the output of the second unit is given by

$$\mathrm{d}x_2/\mathrm{d}t = \alpha_2 b_1 b_2 u_{10}/(a_2\beta_2). \tag{4.12}$$

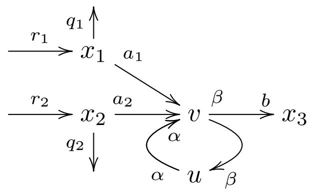
Clearly, the rate of production of x_2 is completely determined by the throughput of the first unit and does not depend on the machinery of the second unit. In this case the steady-state WIP in the second unit can be estimated using the equation (4.7):

$$\overline{x}_1 \approx \begin{cases} \dfrac{\alpha_1 b_1\beta_2 r_0}{a_1 a_2\beta_1\beta_2 u_{20} - \alpha_1\alpha_2 b_1 r_0}, & \text{for } \alpha_1 r_0 < a_1\beta_1 u_{10}; \\ \dfrac{b_1\beta_2 u_{10}}{a_2\beta_2 u_{20} - \alpha_2 b_1 u_{10}}, & \text{for } \alpha_1 r_0 > a_1\beta_1 u_{10}. \end{cases} \tag{4.13}$$

The intermediate product x_1 does not pile up in the buffer of the second unit and undergoes conversion into the product x_2 without delay.

5 The converging branch

Finally, consider a production node in a chain that has at most one successor, but is supposed to have two predecessors. In this converging branch, characterized by the vertex with in-degree 2 and out-degree 1, two independent suppliers provide components x_1 and x_2 to the downstream manufacturer which then yields product x_3:



(5.1)

This structure can represent, for example, the fragment of a modular assembly supply chain, which nowadays has found applications in many manufacturing industries. In modular supply chain, product modules are being apportioned to intermediate sub-producers. As a result, only a few assembled modules will be delivered to the final producer, which reduces the complexity of the final assembly process.

It is a straightforward matter to draw the balance equations for the scheme (5.1). According to our assumption, the rate of uptake of either of two resources under enzymatic facilitation of the machinery would be proportional to ux_1x_2, where u is the number of idle machines. Thus the equations describing the process will be

$$\mathrm{d}x_1/\mathrm{d}t = r_1 - a_1ux_1x_2 - q_1x_1, \tag{5.2a}$$
$$\mathrm{d}x_2/\mathrm{d}t = r_2 - a_2ux_1x_2 - q_2x_2, \tag{5.2b}$$
$$\mathrm{d}u/\mathrm{d}t = \beta v - \alpha ux_1x_2, \tag{5.2c}$$
$$\mathrm{d}v/\mathrm{d}t = \alpha ux_1x_2 - \beta v, \tag{5.2d}$$
$$\mathrm{d}x_3/\mathrm{d}t = bv. \tag{5.2e}$$

Noting from (5.2c) and (5.2d) that $u + v = u_0 = \text{const}$ and introducing dimensionless variables and parameters $\xi_1{=}x_1\sqrt{\alpha r_2/(\beta r_1)}$, $\xi_2{=}x_2\sqrt{\alpha r_1/(\beta r_2)}$, $\eta = v/u_0$, $\xi_3 = x_3/x_{30}$, $\tau = ta_1u_0\sqrt{\beta r_2/(\alpha r_1)}$, $\varrho_1 = r_1\alpha/(a_1\beta u_0)$, $\varrho_2 = r_2\alpha/(a_2\beta u_0)$, $\gamma_1 = q_1\sqrt{\alpha r_1/(\beta r_2)}/\,(a_1u_0)$, and $\gamma_2 = q_2\sqrt{\alpha r_2/(\beta r_1)}/\,(a_2u_0)$, we rewrite the system (5.2) in a nondimensional form

$$\mathrm{d}\xi_1/\mathrm{d}\tau = \varrho_1 - (1-\eta)\xi_1\xi_2 - \gamma_1\xi_1, \tag{5.3a}$$
$$\mathrm{d}\xi_2/\mathrm{d}\tau = \mu_2\left[\varrho_2 - (1-\eta)\xi_1\xi_2 - \gamma_2\xi_2\right], \tag{5.3b}$$
$$\varepsilon\,\mathrm{d}\eta/\mathrm{d}\tau = (1-\eta)\xi_1\xi_2 - \eta, \tag{5.3c}$$
$$\mathrm{d}\xi_3/\mathrm{d}\tau = \mu_3\,\eta, \tag{5.3d}$$

where $\varepsilon = a_1u_0\sqrt{r_2/(\alpha\beta r_1)}$, $\mu_2 = a_2r_1/(a_1r_2)$, $\mu_3 = b\sqrt{\alpha r_1/(\beta r_2)}/\,(a_1x_{30})$, and x_{30} is a proper unit for x_3. Note that (5.3d) is slave equation.

Just as in the cases considered above, parameters ε and μ_2^{-1} characterize by how much the dynamics of the respective variables η and ξ_2 is faster than that of ξ_1. Taking ε to be small while μ_2 to remain within $\mathcal{O}(1)$, the variable η can be replaced by its quasi-steady-state value

$$\eta = \xi_1\xi_2/(1+\xi_1\xi_2). \tag{5.4}$$

Plugging this in the system (5.3) we obtain slow equations

$$\mathrm{d}\xi_1/\mathrm{d}\tau = \varrho_1 - \xi_1\xi_2/(1+\xi_1\xi_2) - \gamma_1\xi_1, \tag{5.5a}$$
$$\mathrm{d}\xi_2/\mathrm{d}\tau = \mu_2\left[\varrho_2 - \xi_1\xi_2/(1+\xi_1\xi_2) - \gamma_1\xi_2\right], \tag{5.5b}$$
$$\mathrm{d}\xi_3/\mathrm{d}\tau = \mu_3\,\eta. \tag{5.5c}$$

The validity of the reduction of (5.3) to (5.5) is ensured, in conformity with Fenichel–Tikhonov theorem, by stability of quasi-steady state (5.4) of the fast equation (5.3c) at all positive ξ_1 and ξ_2.

For small loss parameters γ_1 and γ_2, steady-state solutions of the pair of equations (5.5a) and (5.5b) are as follows:

$$(\overline{\xi}_1,\overline{\xi}_2) = \begin{cases} \left(\frac{\gamma_2\varrho_1}{(1-\varrho_1)(\varrho_2-\varrho_1)} + \mathcal{O}(\gamma_1\gamma_2^2),\ \frac{\varrho_2-\varrho_1}{\gamma_2} + \mathcal{O}(\gamma_1)\right), \\ \qquad\qquad \text{for } \varrho_1 < 1 \wedge \varrho_1 < \varrho_2; \\ \left(\frac{\varrho_1-1}{\gamma_1} + \mathcal{O}(\gamma_2),\ \frac{\varrho_2-1}{\gamma_2} + \mathcal{O}(\gamma_1)\right), \\ \qquad\qquad \text{for } \varrho_1 > 1 \wedge \varrho_2 > 1; \\ \left(\frac{\varrho_1-\varrho_2}{\gamma_1} + \mathcal{O}(\gamma_2),\ \frac{\gamma_1\varrho_2}{(1-\varrho_2)(\varrho_1-\varrho_2)} + \mathcal{O}(\gamma_1^2\gamma_2)\right), \\ \qquad\qquad \text{for } \varrho_1 > \varrho_2 \wedge \varrho_2 < 1. \end{cases} \tag{5.6}$$

Besides, it can be shown that the positive fixed points of the pair of equations (5.5a) and (5.5b) are stable.

Substituting the steady-state values of ξ_1 and ξ_2 from (5.6) into (5.4) gives $\overline{\eta}$. Inserting the latter into equation (5.5c) yields, to $\mathcal{O}(1)$ in small γ_1 and γ_2, the dimensionless production function of the converging branch (5.1):

$$\mathrm{d}\xi_3/\mathrm{d}\tau = \mu_3 \min\left(\varrho_1,\ \varrho_2,\ 1\right). \tag{5.7}$$

In its dimensional form, this will look like

$$\mathrm{d}x_3/\mathrm{d}t = b \min\left(\alpha r_1/(a_1\beta),\ \alpha r_2/(a_2\beta),\ u_0\right). \tag{5.8}$$

Clearly, the result (5.8) is a Leontief–Liebig production function of three factors of production: r_1, r_2, and u_0. This can be extended to multiple inputs. We have focused so far on models with just a few inputs where the concept of modelling with low-order “chemical reactions” is perhaps most natural. However, it is important to recognize that we use the notation of chemical reactions simply to describe things that combine and the things that they produce, and that this framework can be used to model higher-order phenomena in a similar way.

It is easily comprehended from the above analysis, that kinetics of any one-product supply chain of arbitrary length with multiple resources would lead to the overall production function of the Leontief type, providing individual production nodes of the chain follow the generic mechanism similar to that of the enzyme catalysis.

6 Discussion and conclusions

Going over to comment our results, we would like to emphasize that emergence of the bottleneck effect in a supply chain is stipulated by two key features of the suggested construction of the Leontief's black box: (1) two strongly varying timescales involved in the production process—longer, for the inventory level, and shorter, for the number of machines engaged in processing, and (2) weak outflux of the inventory from the buffer.

Presence of time hierarchy makes possible the saturated response of the output to the WIP in the form of the Karmarkar clearing function (in fact the Michaelis–Menten equation), whereas the side buffer leakage secures finiteness of the steady-state inventory level.

The formalism of clearing functions is widely used for production planning. But how good is such an approximation? When is it expected to hold, and under what conditions would it fail? These questions are seldom if ever addressed in the current literature on operations research. The rare exception seems to be the review [2] recognizing the quasi-steady-state nature of the clearing function and relative slowness of the WIP dynamics. However the fast variable, which is supposed to stay in quasi-steady state towards the WIP, remains unspecified in the mentioned work. In terms of our bio-inspired model, with the background given above—especially with the concept of two timescales, we are able to suggest a more sound justification of the clearing function: the momentary number of machines in the operating state would be in a quasi-steady state with respect to the WIP provided the processing time is much shorter than the characteristic waiting time (the typical time it takes to load a machine with resource). Consequently, the condition $\varepsilon \ll 1$ is expected to be sufficient to assure the validity of the clearing function in nonsteady supply chains.

Thus it is shown that the Leontief production function naturally appears in supply chains where output of each individual production node is universally characterized by a saturated response to the WIP. To ensure this type of response it is suffices to assume that conversion of inputs to outputs in material production occurs similarly to the conversion of substrates into different substances in enzyme-catalyzed biochemical reactions. The part of enzyme is played by machinery. In the general case this may be any nonconsumable, or primary, factor of production (fund), such as capital, land, or labor.

The production line consisting of units of such a type, has the property of scale invariance: the production function of the whole chain is similar to the production function of any constituent unit. As we found out, a more correct form of the Leontief function is not its conventional flow–flow notation, but the Leontief–Liebig form, where resources and funds intermingle.

It turns out that the output of a one-product supply chain (possibly with multiple inputs and converging branches) is solely controlled by the minimum of its input supplies and funds. The dependence of the output only on the properties of the bottleneck allows the production system to effectively simplify the control, acting only on the bottleneck unit. The considered self-regulation principle is useful for understanding the functioning of complex production networks.

Just as the deterministic approach fails to capture the discrete and stochastic nature of chemical reactions at low concentrations, so does the continuous mass-action treatment of production process at small quantities of factors of production, whether resources or funds. As many manufacturing processes involve IO conversions at extremely small quantities, such discrete stochastic effects are well relevant for our bio-inspired model. For some supply chains, large fluctuations in the WIP may be dangerous. The evolution of the number of parts of a given type due to interactions with machines–catalysts can be described by Markov processes, which can be formalized, for example, in terms of the chemical master equation [11]. Exploring these possibilities will constitute a future direction for work on the model.

Acknowledgements

The results presented in this paper are part of the self-funded research projects 0116RK00250 and 0116RK00275 registered at the Ministry of Education and Science of the Republic of Kazakhstan.

References

[1] J. Adams, E. Balas and D. Zawack. The shifting bottleneck procedure for job shop scheduling. *Management Science*, **34**(3):391–401, 1988. https://doi.org/10.1287/mnsc.34.3.391.

[2] D. Armbruster. The production planning problem: clearing functions, variable lead times, delay equations and partial differential equations. In D. Armbruster and K. G. Kempf(Eds.), *Decision Policies for Production Networks*, pp. 289–302. Springer, London, 2012. `http://dx.doi.org/10.1007/978-0-85729-644-3_12`

[3] R.U. Ayres and U.E. Simonis(Eds.). *Industrial Metabolism: Restructuring for Sustainable Development.* United Nations University Press, Tokyo, 1994. ISBN 9789280808414.

[4] A. Cornish-Bowden. *Fundamentals of Enzyme Kinetics.* Wiley-Blackwell, Weinheim, 4th edition, 2012. ISBN 9783527330744.

[5] S.A. Ebelhar, W. Chesworth and Q. Paris. Law of the Minimum. In W. Chesworth(Ed.), *Encyclopedia of Soil Science*, pp. 431–437. Springer Netherlands, Dordrecht, 2008. ISBN 9781402039959

[6] J.W. Fowler, G.L. Hogg and S.J. Mason. Workload control in the semiconductor industry. *Production Planning & Control: The Management of Operations*, **13**(7):568–578, 2002. https://doi.org/10.1080/09537280210000026294.

[7] N. Georgescu-Roegen. *Analytical Economics: Issues and Problems.* Harvard University Press, Cambridge, MA, 1966. ISBN 9780674281639. https://doi.org/10.4159/harvard.9780674281639.

[8] R. Gesztelyi, J. Zsuga, A. Kemeny-Beke, B. Varga, B. Juhasz and A. Tosaki. The Hill equation and the origin of quantitative pharmacology. *Archive for History of Exact Sciences*, **66**(4):427–438, 2012. https://doi.org/10.1007/s00407-012-0098-5.

[9] E.M. Goldratt and J. Cox. *The Goal: A Process of Ongoing Improvement.* North River Press, Great Barrington, MA, 4th edition, 2014. ISBN 9780884271956.

[10] D. Helbing, D. Armbruster, A.S. Mikhailov and E. Lefeber. Information and material flows in complex networks. *Physica A: Statistical Mechanics and its Applications*, **363**(1):xi–xvi, 2006. https://doi.org/10.1016/j.physa.2006.01.042.

[11] D.J. Higham. Modeling and simulating chemical reactions. *SIAM Review*, **50**(2):347–368, 2008. https://doi.org/10.1137/060666457.

[12] P. Hochendoner, C. Ogle and W.H. Mather. A queueing approach to multi-site enzyme kinetics. *Interface Focus*, **4**(20130077):1–11, 2014. https://doi.org/10.1098/rsfs.2013.0077.

[13] C.S. Holling. The functional response of predators to prey density and its role in mimicry and population regulation. *Memoirs of the Entomological Society of Canada*, **97**(S45):5–60, 1965. https://doi.org/10.4039/entm9745fv.

[14] W.J. Hopp and M.L. Spearman. *Factory Physics: Foundations of Manufacturing Management.* McGraw-Hill, New York, NY, 3rd edition, 2008. ISBN 9780072824032.

[15] C. Kandemir-Cavas, L. Cavas, M.B. Yokes, M. Hlynka, R. Schell and K. Yurdakoc. A novel application of queueing theory on the Caulerpenyne secreted by invasive Caulerpa taxifolia (Vahl) C.Agardh (Ulvophyceae, Caulerpales): a preliminary study. *Mediterranean Marine Science*, **9**(1):67–76, 2008. https://doi.org/10.12681/mms.144.

[16] U.S. Karmarkar. Manufacturing lead times, order release and capacity loading. In S. C. Graves, A. H. G. Rinnooy Kan and P. H. Zipkin(Eds.), *Logistics of Production and Inventory*, volume 4 of *Handbook in Operations Research and Management Science*, chapter 6, pp. 287–329. North Holland, Amsterdam, 1993. https://doi.org/10.1016/S0927-0507(05)80186-0. ISBN 0444874720

[17] G.I. Kolesova and I.A. Poletaev. Nekotorye voprosy issledovaniia sistem s limitiruiushchimi faktorami [Selected problems in research of the systems with limiting factors]. In *Upravliaemye sistemy [Controllable systems]*, number 3, pp. 71–80. Institute of Mathematics, Siberian Branch of the USSR Academy of Sciences, Novosibirsk, 1969. (In Russian)

[18] W. Leontief. *The Structure of American Economy, 1919–1939; An Empirical Application of Equilibrium Analysis.* Oxford University Press, New York, NY, 2nd edition, 1951.

[19] E. Levine and T. Hwa. Stochastic fluctuations in metabolic pathways. *Proceedings of the National Academy of Sciences*, **104**(22):9224–9229, 2007. https://doi.org/10.1073/pnas.0610987104.

[20] A. Marshall. *Principles of Economics.* Palgrave Classics in Economics. Palgrave Macmillan UK, London, 8 edition, 2013. ISBN 9780230249295.

[21] H. Missbauer and R. Uzsoy. Optimization models of production planning problems. In K.G. Kempf, P. Keskinocak and R. Uzsoy(Eds.), *Planning Production and Inventories in the Extended Enterprise: A State of the Art Handbook, Volume 1*, volume 151 of *International Series in Operations Research and Management Science*, chapter 16, pp. 437–507. Springer, New York, NY, 2011. http://dx.doi.org/10.1007/978-1-4419-6485-4_16

[22] L. Mönch, J.W. Fowler and S.J. Mason. *Production Planning and Control for Semiconductor Wafer Fabrication Facilities: Modeling, Analysis, and Systems*, volume 52 of *Operations Research/Computer Science Interfaces.* Springer, New York, NY, 2013. ISBN 9781461444718.

[23] J. Monod. The growth of bacterial cultures. *Annual Review of Microbiology*, **3**(1):371–394, 1949. https://doi.org/10.1146/annurev.mi.03.100149.002103.

[24] A. Mustafin. K printsipam promyshlennogo metabolizma [A contribution to the principles of industrial metabolism]. *AlPari*, (4–5):80–83, 1999. (In Russian)

[25] A. Mustafin and A. Kantarbayeva. The Leontief's black box reverse engineered. In *22nd International Conference "Mathematical Modelling and Analysis", May 30 – June 2, 2017, Druskininkai, Lithuania. Abstracts*, p. 46, VGTU Technika, Vilnius, 2017. ISBN 9786094760228.

[26] A.T. Mustafin. A bottleneck principle for techno-metabolic chains. Discussion paper 504, Institute of Economic Research, Kyoto University, Kyoto, Japan, September 1999.

[27] G.O. Nijland, J. Schouls and J. Goudriaan. Integrating the production functions of Liebig, Michaelis–Menten, Mitscherlich and Liebscher into one system dynamics model. *NJAS – Wageningen Journal of Life Sciences*, **55**(2):199–224, 2008. https://doi.org/10.1016/s1573-5214(08)80037-1.

[28] R.E. O'Malley. *Singular Perturbation Methods for Ordinary Differential Equations*, volume 89 of *Applied mathematical sciences*. Springer-Verlag, New York, NY, 1991. ISBN 038797556X. https://doi.org/10.1007/978-1-4612-0977-5.

[29] Yu.M. Romanovskii, N.M. Stepanova and D.S. Chernavskii. *Chto takoe matematicheskaia biofizika: Kineticheskie modeli v biofizike [What is mathematical biophysics: Kinetic models in biophysics]*. Prosveshchenie, Moscow, 1971. (In Russian)

[30] C. Roser, M. Nakano and M. Tanaka. Shifting bottleneck detection. In E. Yücesan, C.-H. Chen, J. L. Snowdon and J. M. Charnes(Eds.), *Proceedings of the 2002 Winter Simulation Conference, 8–11 Dec. 2002, San Diego, CA*, pp. 1079–1086. IEEE, 2002. https://doi.org/10.1109/wsc.2002.1166360.

[31] S.S. Sana. Optimal production lot size and reorder point of a two-stage supply chain while random demand is sensitive with sales teams' initiatives. *International Journal of Systems Science*, **47**(2):450–465, 2014. https://doi.org/10.1080/00207721.2014.886748.

[32] L.A. Segel and M. Slemrod. The quasi-steady-state assumption: a case study in perturbation. *SIAM Review*, **31**(3):446–477, 1989. https://doi.org/10.1137/1031091.

[33] R.W. Shephard. *Theory of Cost and Production Functions*. Princeton University Press, Princeton, NJ, 2016. ISBN 9780691647524.

[34] A. Tipper. Capital-labour substitution elasticities in New Zealand: one for all industries? Working Paper 12-01, Statistics New Zealand, Wellington, New Zealand, July 2012.

[35] H.R. Varian. *Microeconomic Analysis*. Norton, New York, NY, 3rd edition, 1992. ISBN 9780393957358.