MODEL BASED UNDERSTANDING OF BIOLOGICAL SYSTEMS: APPROACHES AND TECHNIQUES

S. Ondimu
School of Biosystems and Environmental Engineering, Agricultural and Biosystems Engineering Department, Jomo Kenyatta Univeristy of Agriculture and Technology, Nairobi
Email: sondimu@jkuat.ac.ke

Abstract
Modeling can enable system-level understanding of biosystems. However, biological systems exhibit a number of complex properties and characteristics which are not amenable to standard scientific and engineering approaches towards study, understanding and manipulation of systems. This paper presents various approaches and techniques which can be used in modeling biosystems. Through a systematic review of literature it was found that there are a number of novel approaches and techniques that can be utilized in modeling biosystems.

Key words: *biosystems,* modeling approaches, modeling techniques

I.0 Introduction

A biosystem is made up of dynamic and complex set of interdependent or temporally interacting elements structured as a functional unit. Parts of a biosystem are in most cases themselves systems composed of other parts. Energy, material and information flow among the different elements of a biosystem as well as from and to the surrounding of the system via semi-permeable membranes or boundaries. Though equilibrium seeking, elements of a biosystem can exhibit oscillating, chaotic, or exponential behavior. Modelling can enable system-level understanding of biological systems. However, biological systems exhibit a number of complex properties and characteristics which present a challenge to the study and manipulation of biological systems. Standard scientific approaches towards study, understanding and manipulation of biosystems are bottom up and often involve formulation and solving of many complex equations. In addition, the data involved is often very noisy and limited. It is very tricky to estimate a vast number of equation parameters based on such data. Furthermore, it is difficult to identify the really functions of systems when using so many equations together. Engineering approaches can provide a solution to this bottleneck. However, traditional engineering approaches are reductionist in nature; often involve "brute force" (mechanistic) methodologies; are application-based; often focus on the objective at hand and often assume time-invariant, linear as well as predictable system dynamics. Thus, there is need for appropriate approaches and techniques for modeling biosystems. The aim of this paper is to present various approaches and technologies which can be utilized in modeling biosystems. The study was conducted by a systematic review of literature. It was found that biosystems can be described using white models (where adequate information is available); black models (where scant or no information is available) or grey models (where only part of the system information is available).

1.1 Properties of Biosystems

Biological systems are complex composites which display a number of properties:

(i). **Multifunctional and Self-healing:** Biological systems accumulate functions e.g. bones support the body and manufacture red blood cells. Leaves of plant leaves synthesize food (photosynthesis) and carry out gaseous exchange. Most bio-systems can self-repair themselves. In plants injured bark self-heals almost perfectly to its initial state. In humans many organs and systems do self-heal almost completely (exceptions are teeth and brains which do not self-repair:- but one part of the brain can take the lost functions).

(ii). **Hierarchical Organization Structure:** Hierarchical organization based on building blocks is inherent in the design of bio-systems. For instance cells are the building blocks of tissues; tissues the building blocks of organs; organs the building blocks of systems and systems building blocks of organisms. The organisms form the building blocks of populations, populations the building blocks of ecosystems and lastly ecosystems constitute the building blocks of the biosphere.

(iii). **Self-Organization and Self-Assembly**: Biosystems are self-organizing and self-assembling. For instance, essential amino acids organize and assemble themselves to form proteins. Simple sugars self-organize themselves into complex carbohydrates. Cells assemble and organize themselves into tissues. Many tissues come together to form functional organs and so on. However most modelling procedures assume predetermined behaviour of systems.

(iv). **System non-linearities:** Most biosystems are moderately or highly non-linear in their dynamics. For example, figure 1 shows the biological activity in compost versus temperature **(**Hall, 1998). At low temperature, there is low activity. As the temperature rises, the activity increases non-linearly. At higher temperatures (above about 70° C in this case), activity drops dramatically as proteins become denatured and cellular activity shuts down. This figure depicts typical behavior for microbial systems. However, most fields of system dynamics and control have been developed assuming linear or pseudo-linear models to describe and predict system behavior.

(v). **Temporal and Spatial Heterogeneity:** Many biosystems change dramatically in time and space. Most biosystems exhibit rhythmic behavior characterized by rising and falling populations, rising and falling rates of internal temperature, hormone level and many other activities. At organism level, sleep and reproductive cycles are good examples biorhythms. On the other hand birth, growth, senescence, and death are cyclical at the population level. Biosystems also vary in their properties and activity from one location to another. For instance, within a cell there are organelles that have very different functions and properties. Larger organisms have organs that are very different from each other, while communities or ecosystems include different organisms, each of which plays different roles.

(vi). **Emergent Properties:** Biological phenomena and interactions result in new or "emergent" system properties and dynamics as smaller/simpler components interact to form more complex systems. For instance, at the sub-cellular level, proteins are described by their amino acids sequence. However, as these proteins take on three-dimensional characteristics that further define their properties as they bend and twist. In the cell, organelles interact to perform various functions that collectively define properties at the cellular level. At the organism level, interactions among cells may give rise to other emergent properties, resulting in phenomena commonly described at the organ or systemic levels. Thus, to predict (via a scientific approach) or control (via an engineering approach) the behaviour of biosystems, some knowledge of their system properties and characteristics is necessary.

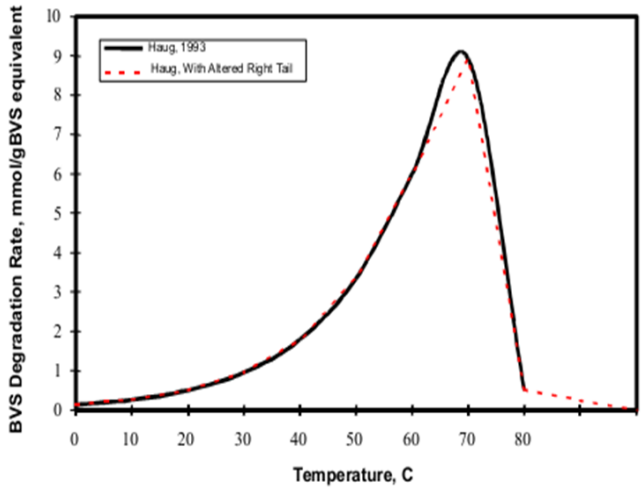


Figure 1: Biological activity in a composting system (source: Hall, 1998)

1.3 Importance of Modelling Biosystems

Of recent times, there has been an increased interest in model-based understanding of biosystems. This has been largely driven by the need for a system level understanding of biological systems. A system level understanding of biosystems is motivated by:

(i). Increasingly large amounts of biological data: Due to detailed studies and recent high throughput technologies, biology has been able to accumulate large amounts of data for important biosystems like humans and other organisms.

(ii). Little understanding of biosystem function and design: Though there are vast amounts of data on biosystems, there is no systematic way of bringing together all the known knowledge in order to comprehensively understand the functions it serves. Furthermore, there is no systematic way to predict what the system does or modify the system to do something else (engineer the system). This is brought about by the fact that: 1) biological systems are very complex 2) they have different organisational levels (i.e. molecular-cellular-tisuelar-organ-organism-population-ecosystem biospehere). But there is no good way of bringing this hierarchy together.

(iii). Tremendous potential impact for systems approach: Being able to understand and manipulate biosystems will bring tremendous positive impacts. It will enable:

i. Gaining insights into fundamental biological processes such cell cycles, biorhythms, disease and aging process
ii. Breakthroughs in human health care through improved diagnosis and drug design; new therapies (e.g. gene therapy); personalized Medicare and development of biologically based consumer products and pharmaceuticals
iii. Engineering of biosystems such as advanced biosensors, bio-generated energy (like bioenergy) and bio-factories (producing things like drugs, edible vaccines, bio-materials etc)
iv. Advancement of the science of biomimetics (bio-inspired computing, engineering and technology) and biomaterials (e.g. bio-steel)
v. Development of control systems and structures for biological systems (Li et al., 2010; Tyson et al., 2003)

1.4 Types of Models

Depending on the knowledge of physical laws employed and the amount of data available models for describing biosystems can be white-box, grey-box or black-box models.

(i). White box models: The White Box models represent cases when it is possible to construct a model from prior knowledge and physical insight, in other words, the models are perfectly known. First principles information is used to derive the full models whose complexity can be reduced using system identification or other techniques.

(ii). Grey box models: The Grey Box models represent cases when some physical insight is available but several parameters remain to be determined from observed data. Certain structured information about the model (system) is already available however there are a number of unknown free parameters which need to be estimated.

(iii). Black box models: The black box models represent cases in which no physical knowledge is available or used but the chosen model structure belongs to families that are known to have good flexibility and have been "successful in the past". Since no prior model is available all parameters must be determined using different techniques like system identification, inverse modeling and universal approximation among others.

1.5 Modeling Techniques

There are different techniques in which modeling can be done. These include but not are limited to system identification, artificial neural networks as well as morphology based techniques. These techniques present traditional as well as novel tools which can be used for modeling biosystems.

(a) System Identification Techniques

System identification is comprised of mathematical tools and algorithms that build dynamical models from measured data. This is mostly done in time domain though it can also is done in frequency domain. Depending on the knowledge of physical laws employed in their development, system identification models can be termed as white-box, grey-box or black-box models. For white box models, system identification is used to reduce the order or complexity of full models derived by first principles information. For grey box models, certain structured information about the model (system) is already available. However, such models still have a number of unknown free parameters which can be estimated by system identification. Since no prior model is available in black box models, all parameters need to be determined. The general procedure involved in system identification model building involves: generating data set D; selecting a (set of) model structure and selecting the criteria (least squares for instance) used to specify the optimal parameter estimates and validate the model. The main elements of system identification include:

(i). Statistical view of induction and inference because all work is relative to the finite samples and noisy data set.
(ii). Optimization- continuous parameter estimation and discrete model selection
(iii). Dynamics (control/state space):- representation of ODE/differential equations

Most system identification algorithms are used to develop black box models. According to Salat (2003), mathematical models resulting from system identification are used for the purpose of controlling, simulation

tests and forecasting. Sjoberg et al., (1995) pointed out that a key problem in system identification is to find a suitable model structure, within which a good model is to be found. System identification models can be termed as white-box, grey-box or black-box models, depending on the knowledge of physical laws employed.

(b) Shape/Morphology/Geometry Techniques

This technique is born of the need to consider the spatial aspects of biological systems such as position, interaction and physical shape. This is because physical shape aspects of biosystems cannot be ignored in the study and understanding of these complex entities. However, biosystems can exhibit features hard to capture with classical quantitative observables (Merelli and Rasseti, 2012). In addition, interactions and collective functions can be driven by not only quantitative features (e.g. space and geometry) but also by higher-level, more abstract qualitative relations. By definition, a shape based system is a tuple (S,B) where B is the behavior of the system, a set of process terms, given by the syntax:

$$P ::= \underset{Choice}{P+P} \quad \underset{Composition}{PIIP} \quad \underset{action\ prefex}{\underset{a\epsilon A}{a.P}} \quad \underset{inaction}{0} \tag{1}$$

And S is the structure of the system, i.e the shape of the interaction space defined by a topological space (A,T) on the set of actions A. A number of models can be developed by using this technique including deferential equations based models, compartmental and lattice based models (Cacciagrano et.al., 2010) aa well as shape calculus based models (Bartocci et. al., 2010; Paoletti et. Al., 2012).

(c) Inverse Modelling Techniques

In principle if a real - valued vector X describes the condition of a phenomenon and because of this condition X, a result Y is produced, a model, m by which Y can be predicted from an appropriate X can be defined such that:

$$Y = (m)X \tag{2}$$

A representative set of inputs of X's can be collected and the model applied for each. However, if for some reason determination of X is not feasible or difficult, while Y can be easily determined, a representative set of Y's can be collected and the model applied inversely such that for every Y (input) a corresponding X is obtained as output (Timothy, 1993). Though explicit inversion of complex models is not possible, neural networks (NN) can solve this problem. Each Y in Eqn (2) is applied to a NN's inputs and the network is trained to produce the corresponding X at its output. Alternatively, if (m) incorporates some coefficients, they can be optimized by any optimization techniques. Inverse modeling is used to develop black box models. These technique represents a novel tool for modeling most biosystems.

(d) Mathematical Techniques

Mathematical modeling is used to develop white box models. This is when there is adequate information and evidence that there exists an identifiable mathematical relationship that describes the behavior of a biosystem. The technique is used to develop simple Linear and Nonlinear models as well as stochastic models. It is also employed in developing representation of fundamental processes and carrying out numerical solution procedures, analysis as well as validation.

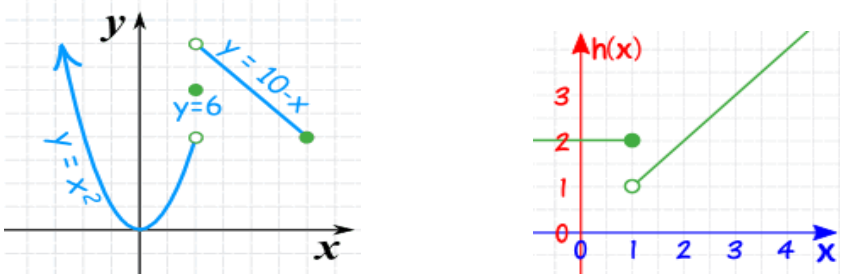
(e) Piecewise Functions Techniques

Coupling of multiple models becomes necessary when modelling complex systems (Duff, 2012). A piecewise function is a function which is defined by multiple sub functions. Each sub function applies to a certain interval of the main function's domain (a sub-domain). Piecewise is a way of expressing the function, rather than a characteristic of the function itself. However, when additional qualification is given, piecewise can describe the nature of the function. For example, a piecewise polynomial function is a function that is a polynomial on each of its sub-domains, but possibly a different one on each. Piecewise functions are defined using the common functional notation, where the body of the function is an array of functions and associated sub domains. For example equqtions 3 and 5 denotes the notation for piecewise functions graphically shown in figure 2.

$$f(x) = \begin{cases} x^2 & \text{if } x < 2 \\ 6 & \text{if } x = 2 \\ 10 - x & \text{if } x > 2 \text{ and } x \leq 6 \end{cases} \tag{3}$$

$$h(x) = \begin{cases} 2, & \text{if } x \leq 1 \\ x, & \text{if } x > 1 \end{cases} \quad (4)$$

A piecewise function is said to be continuous on a given interval if: i) it is defined throughout that interval; ii) its constituent functions are continuous on that interval, and iii) there is no discontinuity at each endpoint of the sub domains within that interval. Figure 2b define a piecewise function that is piecewise continuous throughout its sub domains, but is not continuous on the entire domain as shown by a jump at x_o



(a) (b)

Figure 2: Graphical representation of piecewise functions (a solid dot means "including", an open dot means "not including"): a) quadratic linear function b) different linear functions on either side of xo.

(f) Universal Approximation Technique

This technique is based on the universal approximation theorem: Let ϕ( ) be a nonconstant, bounded, and monotonically-increasing continuous function. Let X_m denote the m-dimensional unit hypercube $[0,1]^m$. The space of continuous functions on X_m is denoted by C(Xm). Then, given any function $f \in C(X_m)$ and ε>0, there exist an integer N and real constants ai, bi ε **R** and wi ε R^m where **i** = 1, *N* such that we may define:

$$\boldsymbol{F(x)} = \sum_{i=1}^{N} \boldsymbol{\alpha_i \varphi\,(w_i^T + b_i)} \quad (5)$$

as an approximate realization of the function f where f is independent of ϕ; that is,

$$|\boldsymbol{F(x)} - \boldsymbol{f(x)}| < \in \quad \text{For all x ε } X_m. \quad (6)$$

In other words, functions of the form *F(x)* are dense in ε(X_m)

Universal approximation models are usually NN models. Fussy logic models are also possible. Neural networks can represent a wide variety of interesting functions when given appropriate parameters; it does not touch upon the algorithmic learnability of those parameters.

Vii Building block techniques

This modelling technique considers cells as building blocks of Biosystems (cells-tissues-organs-systems-organisms- populations-ecosystems). For instance the bio-physiological functions for cells can be modelled as equations 5 and 6.

$$\frac{d}{dt}C = \frac{d}{dt}[M + ST + PS + CR + CG + CM + MT] \quad (7)$$

$$p(t) = \frac{d}{dt}\,p = \sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{k=1}^{P}\sum_{l=1}^{Q}\frac{d}{dt}C, \;\; \forall \frac{d}{dt}C > 0 \quad (8)$$

Where: M = metabolic; ST = Signal transduction; PS = Protein synthesis; CR = cell reproduction; CG = cell growth; CM = cell movement; MT = molecular transport and N = number of cells.

(g) Mass and Energy Balance Techniques

Mass balance techniques are based on the law of conservation of mass (matter cannot disappear or be created spontaneously). The law of conservation of energy (energy cannot be created nor destroyed) forms the basis for energy balance techniques. Techniques based on mass and energy balance also utilize chemical balance as well as equilibrium and metabolic kinematics. These techniques can be used to create white and grey box biosystem models.

1.6 Approaches to Biosystems Modelling

Different approaches can be used to conceptualize biological systems (cells, tissues, systems, organs, organisms, populations and ecosystems) so as to make it possible to model them. Reductionist approaches break down biological systems into simple manageable units. These are the traditional engineering approaches and often involve "brute force"; are application-based and focus on the objective at hand. They often assume time-invariant, linear, and predictable system dynamics. Holistic (system) approaches consider the biological system as a whole. The holistic approach to biosystems modelling is very interdisciplinary and involves biology, biophysics, biological engineering, Synthetic biology, genomics, natural selection, traditional engineering (devices, control, learning, communication) and mathematics among other disciplines.

However, biological systems are complex, dynamic and the data involved is very noisy and limited that modelling them as a whole is very tricky. Hybrid approaches combine both, Reductionist and holistic approaches in various proportions (depending on available data) to model biosystems. Thus they can be considered to be novel. The following is a brief description of these three approaches to biosystems modelling.

1.6.1 Reductionist Approaches

This is the traditional engineering approach which concentrates on the problem at hand i.e. “let the method decide”. It does not adequately address all aspects of a biosystem and how it is related to other systems. This approach heavily depends on analogies that can be grasped or assumed to exist between a given physical system and the biosystem under consideration. Where no clear analogy with a known existing physical system methodology can be established, the biosystem is forced to fit into one via assumptions. Though, computationally convenient, the reductionist approach is not suitable for biosystems. Mechanistic (physical) modelling is the commonest form the reductionist approach.

1.6.2 Mechanistic (physical) Modelling Approach

Mechanistic (physical systems) modelling makes use of certain fundamental laws to build a description of a process. These are the balance equations that describe the conservation of mass and the conservation of energy. Mechanistic modelling is very expensive in terms of human effort and expertise. It requires large amount of representative data and in many cases can only be acquired by perturbing the process via planned experiments. Mechanistic models are used to provide more realistic predictions and more can be done in terms of analysis (Davidovits, 2001). Mechanistic modelling is mostly used to develop white box models. Modelling techniques which can be utilized in mechanistic modelling approach include Shape/morphology/geometry, Mathematical, Piecewise functions, and mass and energy balance as well as chemical balance /metabolic kinematics techniques.

1.6.3 Holistic Approaches

In the holistic approach the biology of system decides the methodology for modelling it: - “let the system decide”. The unique properties of the biosystem including non-linearity of system dynamics, time-variant behaviour and emergent properties are factored in model development. The interaction of the system with other systems is also considered. This requires complete paradigm shift and involves development of completely new modelling methodologies that may incorporate human or other intelligence (Franklin et al., 1992; Hall, 1998). Though most ideal, holistic approaches are difficult to implement in that they depend on prior understanding of the bio-physiological dynamics of the biosystem.
Holistic approaches include phenomenological, statistical and biological equilibrium based modelling methods.

1.6.4 Phenomenological Approach

This holistic modelling approach considers biosystems as whole phenomena. The general behaviour is then modelled as fusion of the different characteristics of the system. It is used to develop black box models. For instance the health status of a human-being can be modelled as shown in figure 3.

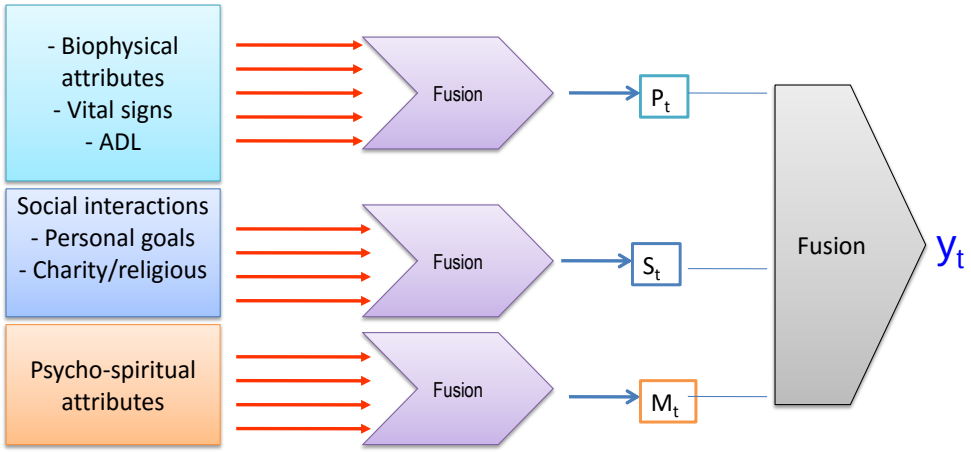


Figure 3: Schematic model of health status of a human

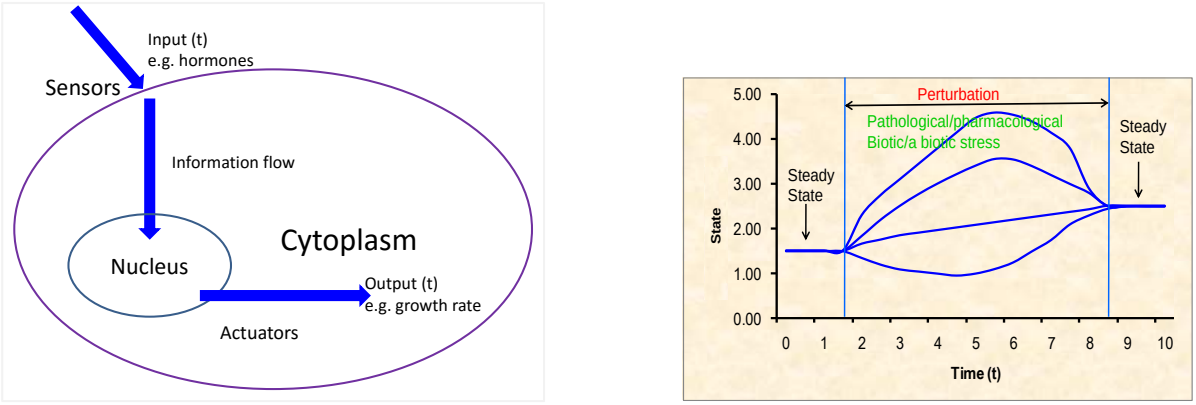
Modelling techniques which can be utilized in phenomenological modelling include inverse modelling, universal approximation and to some extent mass and energy balance techniques.

1.6.5 Statistical Approach

In this approach, the behaviour of a biosystem is modelled as the sum total of the behaviour of its building blocks as shown in equation 8. This approach majorly utilises system identification and building block techniques. The statistical approach can be used to develop white box, black box or grey box models.

1.6.6 Biological Equilibrium Approaches

This approach is based on the fact that biological systems always maintain a steady state and try to attain equilibrium whenever a perturbation from this steady state occurs. For instance a cell is a dynamic system that is continuously sensing its environment, processing information and outputting results or responding to the environment. It is self-correcting in that it tries to attain steady (equilibrium) state after any perturbation as illustrated in figure 4. Building block hypothesis and mass and energy balance modelling techniques can be utilised in biological equilibrium modelling approach.



(a) (b)

Figure 4: Illustration of cell as dynamic system (a) Illustration of a cell system that receives inputs (via sensors) from its environment processes the information received in the nucleus and outputs the results. (b) Illustration of self- correcting nature of a cell system

1.6.7 Hybrid Approaches

In the hybrid approach, the holistic and reductionist approaches are intelligently combined to achieve modelling of biosystems i.e. "let the situation decide". Where an existing engineering methodology can be used after some modification, then such methodology is adopted. If no existing engineering methodology can be used without unrealistically compromising the biology of the biosystem, then a completely new methodology is developed. Modelling techniques which can be utilised to hybrid approaches include building block technique, inverse modelling, piece wise functions, universal approximation and Shape/Morphology/Geometry. Models developed in hybrid approaches includen black box and grey box models.

2.0 Conclusions

In this paper a number of modelling techniques and approaches have been presented. The modelling techniques highlighted include: system identification; shape/morphology/geometry modelling; inverse modelling; mathematical modelling; piece wise functions, universal approximation; building block methods and mass as well energy balance methods. The modelling approaches presented are categorised into reductionist (mechanistic (physical) modelling); holistic (phenomenological modelling, statistical modelling, biological equilibrium modelling) and hybrid (combinations of reductionist and holistic methods). The phenomenological, hierarchical and biological equilibrium approaches can be considered to be novel in modelling biological systems. While inverse modelling, piecewise function and building block hypothesis methods can be considered to be novel modelling techniques. Utilising these novel approaches and techniques will make it possible to develop fairly accurate models for biosystems. This will enable model based systems level understanding of biosystems. A sytems level understanding of biological systems will enable development of engineered biological systems; development of new biologically based consumer bioproducts and biopharmaceuticals; advancement of the science of biomimetics; development of advanced control systems and structures for biosystems and development of personalized Medicare in addition to model based health care systems among other benefits.

References

Bartocci E, Lio P, Merelli E and Paoletti N (2012). "Multiple Verification in Complex Biological systems: the Bone Remodelling Case Study". Transactions of Computational systems Biology XIV LNBI 7625 53076

Bartocci E, Corradini F, Di Berardini, MR, Mereli E, Tesei L (2010). "Shape Calculus, A spatial Mobile Calculus for 2D Shapes". Scientific Annals of Computer Science 20 (1) 1-31

Biomodels Database www.ebi.ac.uk/biomodels-main (assessed, may 2014)

Cacciagrano D R, Coradini, Merelli E, Tesei (2010). Multiscale Bone Remodelling with Spatial P Systems. Procceding of MECBIC 2010, EPTCS 40 70-84

Dividoits P. (2001). Physics in Biology and Medicine (2nd Ed.). Elsevier, UK

Duffy G V (ed.) (2012). Advances in Applied Human Modelling & Simulation (2012). CRC Press

Hall S G; Lima M (2001). Problem Solving Approaches and Philosophies in Biological Engineering: Challenges from Technical, Social and Ethical Arenas. Trans. ASAE 44 (4): 1037 – 1041

Master T. (1993. Practical Neural Networks Recipes in C++. Academic Press, USA

Merelli E, Rasetti M (2012). "The Immune System as a Metaphor for Topology Driven Patterns Formation in Complex Systems". Artificial Immune Systems, LNCS7597/2012 289 – 291.

Li C, Donizelli M, Rodriguez N, Dharuri H, Endler L, Chelliah V, Li L, He E, Henry A, Stefan MI, Snoep JL, Hucka M, Le Novère N, Laibe C (2010) BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models. BMC Syst Biol., 4:92.

Paoletti N, Lio P, Merelli E and Viceconti M (2012). Multilevel Computational Modelling and Quantitative Analysis of bone Remodelling. IEEE/ACM Transactions of Computational Biology and Bioinformatics 9(5) 1366 – 1378

Tyson J J, Chen K C, Novak B (2003). Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. Curr Opin. Cell Biol. 15(2):221-31