

Preface

Biology is concerned with living systems and their constituents at all scales, namely, molecule, cell, tissue, organ, individual, organism, and ecosystem. Studies which aim at describing, understanding, and monitoring these systems are grounded in the data provided by experiments and observations. While such activities had a prominent qualitative component, as illustrated by biological classifications and taxonomies, the advent of novel experimental techniques has triggered a shift to the quantitative side. Two examples will illustrate these dramatic changes. At one end of the scale, where crystallizing a protein used to be a tour de force, for which Kendrew and Perutz were awarded the Nobel Prize for Chemistry in 1962, high-throughput structural genomics projects nowadays allow crystallizing hundreds of new proteins and complexes per week. At the other end of the scale, sequencing a genome used to be another tour de force, for which Sanger was awarded a Nobel Prize (his second) in Chemistry in 1980. Nowadays, the genomes of whole populations are uncovered by meta-genomics projects, shedding light on a given ecosystem as a whole. Similar examples can be found at the intermediate scales as well.

This spectacular progress is due to a synergy between technological advances, allowing data acquisition on biological objects at complementary scales, and conceptual syntheses of ideas from biology, physics, chemistry, mathematics, statistics, and computer science. Such advances are motivated not only by outstanding scientific challenges but also by the potential value added by new protocols and techniques in biotechnologies and (bio-)medicine. A principal goal of this book is to illustrate that in modeling biological systems, deeper insights can be gained using more advanced mathematical and algorithmic developments that implicate a wide spectrum of techniques from applied mathematics and computer science. But if modeling in biological sciences is expanding rapidly, the specificities of the data dealt with and the different natures of the mother disciplines must be understood in order to develop a full synergy. Before discussing the contents of the book, we therefore briefly address selected properties of biological systems, as well as characteristics of mathematical and algorithmic modeling.

Specificities of Biological Systems

Biological Systems Integrate Multiple Scales

The fact that biological systems span multiple scales is obvious from the enumeration given above, which encompasses molecules to ecosystems. But in complex organisms, apprehending a complex function may also involve multi-scale and global considerations. One such example is the regulation of blood pressure in vertebrates. This regulation indeed involves molecules (e.g., the thyroid hormones affecting the calcium homeostasis), tissues (e.g., the cardiac cells accounting for electric phenomena), as well as whole organs (the heart of course, but also the kidneys which regulate electrolytes and the volume of the blood). Modeling such a complex phenomenon requires integrating across these scales, as well as coupling together models of the individual building blocks.

Biological Systems Are Governed by a Mix of Deterministic and Probabilistic Behaviors

In physics, the study of a perfect gas at the atomic level starts with the computation of its mean free path, while a macroscopic description is given by the law of perfect gases. The same holds for a number of biological systems: while their description at the finest scale requires probabilistic models, integrating across scales typically yields deterministic behaviors. This integration explains the (seemingly) deterministic phenotypes of living systems. But the probabilistic nature of computations at a fine scale may also affect coarser ones. An example is perceptual multi-stability which is observed in different sensory modalities such as vision. If the underlying neural mechanisms are still unclear (i.e., physiological correlates of perceptual switches, origins of transitions, or populations dynamics), it is clear that the influence of noise underpins multistability. Thus, deterministic and/or probabilistic models may be developed independently or jointly, depending on the phenomenon of interest.

The Complexity of Biological Systems Warrants Phenomenological Models

Biological systems have evolved under the dual mechanism of mutation and selection. They often involve redundant features, which may be seen as backups. In metabolic phenomena, for example, alternative, competing pathways related to a particular function often exist, and one may replace the other in case of failure, such as a mutation impairing a particular protein. This complexity touches on the very nature of biology itself, and it is unclear whether simple laws governing the behavior of complex systems will ever be discovered. It is in sharp contrast with physics, where Newton's or Coulomb's laws, just to mention two, have been instrumental

in modeling gravitation and electrostatic interactions, respectively, and in triggering technological developments. This observation explains why modeling for biology comes in two guises. On the one hand, selected models are derived from first principles. On the other hand, phenomenological models are developed based on the investigation of correlations between parameters describing the system of interest, calling for methods in the realm of machine learning and inferential modeling.

The Variability of Biological Systems Calls for Statistical Assessments

A cell in a given tissue shares the phenotype of its mates concurring in the function contributed by the organ, but may have specific features as well, for example, if one of its genes has undergone a mutation. The hearts of two humans share the same design pattern, but each may also have specific features. These two examples illustrate two key features of biological models. First, generic models accommodating individual-specific variations are called for. Second, the parameters used to single out specific properties should be scrutinized under the lens of statistical assessments.

Modeling for Biology and Biomedicine

The fact that biological systems provide a mine of complex modeling problems should not come as a surprise. Yet any mathematical or algorithmic development for biological sciences requires reconciling somewhat different perspectives, and a number of specificities need to be accepted in order to appreciate the contributions discussed in this book.

System-Centric Versus Generic Developments

Biology is often a system-centric activity, with a focus on a cell, an organ, or a pathology, while mathematics and computer science aim at exhibiting general properties and algorithms which may be instantiated in a number of settings.

Ill-Posed Problems, Models, and Their Validation

While mathematics and computer science have traditionally been concerned with well-posed problems, biology essentially raises ill-posed problems. In fact, modeling in computational biology and biomedicine is equally, if not more, about designing models, than it is about solving well-posed problems. Also, since models are simplified representations of complex phenomena, validations are called for. In fact, any model should be confronted to experiments at some point, in order to be confirmed or falsified. Numerous such examples will be discussed in this book.

Multidisciplinary Models

Models for complex biological systems are often multidisciplinary as they involve mathematics, physics, biology, and computer science.

Biology is of course the starting point, as any modeling study requires circumscribing the properties of interest and the data involved, but it is also the yardstick to be used for assessment (knowledge from the literature, additional data, etc). In making these specifications, one needs to compromise between biological accuracy on the one hand and conceptual simplicity on the other, as the latter warrants uncluttered models retaining the essential features (leaving aside the amenability of such models to calculation and simulation). Mathematics allows specifying abstract properties of the system studied (e.g., geometry, topology, hierarchical organization, and dynamics) and allows solving systems of equations, performing statistics, etc. Physics and chemistry are used to endow this abstract model with selected properties, for example, mechanical or electrical. Biology provides the semantics of the model by embedding it in a biological context. Finally, computer science allows automating certain tasks, running simulations and performing analyses. The role of simulations in investigating complex systems cannot be overstated; such simulations can be seen as numerical experiments, from which properties can be inferred.

We note in passing that the systems simulated may be continuous or discrete, and the question of discretizing continuous biological processes while retaining essential properties is always a critical one.

Mathematical and Algorithmic Contributions

From a transverse, rather than applied, perspective, mathematics and computer science can contribute to biology in at least two ways. First, an existing methodological development may be improved in terms of robustness and efficiency. These aspects include mastering the numerics involved in floating-point calculations, improving the convergence properties of algorithms, for optimization in particular, and designing algorithms with enhanced asymptotic properties – which *scale better* as biologists might put it.

Second, and perhaps more importantly, concepts and algorithms from mathematics and computer science may lay the groundwork for more advanced and more accurate models, and several topics in mathematics and computer science are currently undergoing major developments. One of them is stochastic modeling, which is especially prominent in problems from neurosciences, such as interpreting spike trains. Another is inverse problem solving, for inverse problems are faced not only in physiology and neurosciences but also in image processing. In the former, one wishes to infer the inner structure of an organ, for example, the heart or the brain, from peripheral measurements. In the latter, one is concerned with the enhancement of a blurred image acquired by a microscope or by tomography. Yet another contribution is machine learning and statistical inference, as the

investigation of correlations between the various parameters describing a system is a ubiquitous challenge.

Software and Programs as Telescopes

An algorithm is an effective method which needs to be implemented in software in order to become operational. This coding process is in general nontrivial, as it may be influenced by a number of factors such as the need to certify selected critical tasks, requirements on the numerical accuracy of the calculations carried out, constraints inherent to the size of the data processed, portability and parallelization issues, compatibility with programs upstream and downstream, etc. This complexity explains the emergence of highly specialized libraries which are often community-wide efforts, and which can be seen as research instruments equivalent to telescopes for astronomers.

Motivation to Write This Book

Computational biology and biomedicine is a vast field where intensive research is currently being carried out, with outstanding perspectives both in terms of the complexity of the scientific problems to be addressed and technological developments to be made. Taking up these challenges requires developing an enhanced synergy between biology and biomedicine on the one hand and applied mathematics and computer science on the other hand.

In line with this observation, the motivation to write this book has been to show that researchers trained in more quantitative and exact sciences can make major contributions in this emerging discipline, and those with roots in biology and biomedicine can benefit from a true leveraging power tailored to their specific needs.

The need to train actors with multidisciplinary skills explains the ever-increasing number of advanced-level classes at the master's level which have begun over the last few years. This book itself emerged from a series of lectures given within the scope of the *Computational Biology and Biomedicine (CBB)*¹ program at the Université Nice Sophia Antipolis, France. The goal of this Master is to provide advanced training at the crossroads of biology, biomedicine, applied mathematics and computer science. This book adopts the organization of these class units, namely, bioinformatics, biomedical signal and image analysis, and modeling in neuroscience.

This book deals with the problem of modeling certain biological phenomena, and we believe that is well suited for two communities. The first is the vast community

¹MSc in Computational Biology and Biomedicine, Université Nice Sophia Antipolis, France, <http://cbb.unice.fr>

of applied mathematicians and computer scientists, whose interests should be captured by the added value generated by the application of advanced concepts and algorithms to challenging biological or medical problems. The book also highlights some unsolved and outstanding theoretical questions, with potentially high impact on these disciplines. The second is the equally vast community of biologists, and more generally life scientists, concerned by the three fields covered in this book: bioinformatics, biomedicine, and neuroscience. Whether scientists or engineers, they will find in this book a clear and self-contained account of concepts and techniques from mathematics and computer science, together with success stories on their favorite systems.

Book Organization

While the topics discussed span a wide spectrum, the presentations of the chapters adopt a unified plan. In short, each chapter, which has been written in tandem by computer scientists/applied mathematicians and biologists/biophysicists/medical doctors, presents the biological or medical context, follows up with mathematical or algorithmic developments triggered by a specific problem, and concludes with one or two *success stories*, namely, new insights gained thanks to these methodological developments.

At the end of each chapter, the reader will find a list of online resources grouped into three categories: tutorials, databases, and software. In assembling these resources, the goal has been to provide pointers to more specific topics, but also to expand the horizon and to list important programs related to the material presented.

To facilitate the reading experience, a list of acronyms is also supplied at the end of each chapter. Finally, the book ends with two thematic indexes corresponding to biology, medicine, physics, and biophysics on the one hand and to mathematics and computer science on the other hand. These indexes are also accompanied by a flat index amenable to alphabetical search. Again, this index shows the plurality of the contributions and the complementarity of the disciplines involved.

Let us now briefly characterize the content of each chapter.

Part I: Bioinformatics

Chapter 1: Modeling Macromolecular Complexes: A Journey Across Scales.

Motivated by the question of unraveling the key features of the macromolecular complexes which account for biological functions, this chapter discusses phenomenological models for protein complexes, namely, atomic resolution models for binary complexes, and coarse-grain models for large assemblies. In both cases, accurate geometric and topological models allow uncovering subtle biological and biophysical signals. These approaches are validated through enhanced correlations

between biophysical and structural properties of protein complexes in the Protein Data Bank.

Chapter 2: Modeling and Analysis of Gene Regulatory Networks. This chapter addresses the problem of modeling networks of interacting genes and macromolecules, a central topic in systems biology. Having recalled classical models based on systems of ordinary differential equations, whose nonlinear features are not easily dealt with, and which also pose parameter identification problems, the authors proceed with linearization schemes of such systems and with formal analysis based on finite-state transition graphs. An application to the elucidation of the reason why the bacteria *Pseudomonas aeruginosa* over-secretes mucus in lungs affected by cystic fibrosis is discussed.

Part II: Biomedical Signal and Image Analysis

Chapter 3: Noninvasive Cardiac Signal Analysis Using Data Decomposition Techniques. This chapter discusses techniques to analyze electrocardiogram signals, with applications to the prevention and the cure of cardiac arrhythmia. More precisely, the authors develop statistical techniques, based on principal component analysis and independent component analysis, to analyze signals incurring subtle variations and burdened with recording noise. Some of these techniques are currently used by cardiologists to make clinical decisions.

Chapter 4: Deconvolution and Denoising for Confocal Microscopy. Fluorescence light microscopes such as confocal microscopes have become powerful tools in life sciences for observing biological samples, in order to measure the spatial distribution of proteins or other molecules of interest. However, there are some inherent imaging limitations in confocal images such as blurring due to the diffraction limit of the optics and low signal levels. To overcome these limitations, the method developed in this chapter consists of improving the resolution by deconvolution. This illustrates the concept of regularized solutions which allow solving ill-posed problems (in the variational framework or in the Bayesian probabilistic frameworks as proposed here). This chapter develops general concepts that can be applied to any similar imaging problem, allowing one to go beyond current limitations in imaging systems.

Chapter 5: Statistical Shape Analysis of Surfaces in Medical Images Applied to the Tetralogy of Fallot Heart. This chapter develops tools to model the morphology and the physiology of the heart, when affected by the so-called tetralogy of Fallot, a severe congenital defect requiring open-heart surgery in infancy followed by subsequent operations. The methodology developed consists of modeling the heart using the mathematical formalism of differential forms and currents, from which correlations between the morphology of the heart and physiology properties can be inferred, thus guiding clinical decisions. This chapter emphasizes the need to develop generic models amenable to patient-specific tailoring.

Chapter 6: From Diffusion MRI to Brain Connectomics. Diffusion MRI (dMRI) is a unique modality of MRI which allows one to indirectly examine the microstructure and integrity of the cerebral white matter *in vivo* and *noninvasively*. In this chapter, the authors present an overview of the mathematical framework for dMRI. Two methodologies can be distinguished: The first is based on physics and aims at modeling the imaging process from a physical point of view. The second is based on computer science and aims at extending approaches from computer vision to dMRI images seen as images in higher dimensions. Thanks to these recent mathematical and modeling advances dMRI is becoming a key element in the study and diagnosis of important pathologies of the cerebral white matter, such as Alzheimer's and Parkinson's diseases, as well as in studying its physical structure *in vivo*.

Part III: Modeling in Neuroscience

Chapter 7: Single-Trial Analysis of Bioelectromagnetic Signals: The Quest for Hidden Information. This chapter deals with the analysis of multitrial electrophysiology datasets coming from neuroelectromagnetic recordings by electroencephalography and magnetoencephalography (EEG and MEG). In order to correctly capture the inter-trial variability, the authors develop techniques such as nonlinear dimensionality reduction and extended Matching Pursuit methods. These techniques to interpret multitrial signals are already largely applied in clinical or cognitive science research. The rapidly growing field of Brain Computer Interfaces is also driving research on the online interpretation of EEG signals, so that much progress on single-trial biosignal is expected in the future.

Chapter 8: Spike Train Statistics from Empirical Facts to Theory: The Case of the Retina. This chapter focuses on methods from statistical physics and probability theory allowing the analysis of spike trains in neural networks. Taking as an example the retina, the authors present recent contributions aiming at understanding how retina ganglion cells encode the information transmitted to the visual cortex via the optical nerve, by analyzing their spike train statistics. Understanding the neural code remains an open challenge and this chapter makes a stride in this direction, based on the joint advances of MEA recording devices, spike sorting algorithms, statistical methods, and efficient algorithms accommodating the huge amount of data inherent to neural activity.

Outlook

Each topic covered in this book has of course been the subject of different publications presenting in-depth treatments of the state of the art. But these highly specialized texts often exhibit a bias inherent to the discipline the authors originate from, and tend to target experts and researchers already in the field. The same

holds on the mathematical and computer science side; a large number of books are available, each focusing on one specific set of techniques. But none strikes a balance between biological and methodological developments.

As should be clear from this Preface, this book instead touches upon a limited number of topics in the vast realm of computational biology and biomedicine, and for each of them seeks to balance the applied and the formal considerations. Along the way, it clearly shows that a panoply of complementary techniques are needed to apprehend the complex issues encountered in life sciences.

We believe that our initiative will be a successful complement to existing material, both from the biological and modeling sides, and we hope that this book will serve as a useful source of reference, inspiration, and motivation for both students and fellow researchers in applied mathematics and biology.

Sophia Antipolis, France
May 2012

Frédéric Cazals
Pierre Kornprobst



<http://www.springer.com/978-3-642-31207-6>

Modeling in Computational Biology and Biomedicine

A Multidisciplinary Endeavor

Cazals, F.; Kornprobst, P. (Eds.)

2013, XXVI, 318 p., Hardcover

ISBN: 978-3-642-31207-6