VIEWPOINTS ON MODELLING: COMMENTS ON "ACHILLES AND THE TORTOISE: SOME CAVEATS TO MATHEMATICAL MODELLING IN BIOLOGY"

JINZHI LEI

ABSTRACT. Mathematical modelling has been proven to be useful in understanding some problems from biological science, provided that it is used properly. However, it has also attracted some criticisms as partially presented in a recent opinion article [1] from biological community. This note intends to clarify some confusion and misunderstanding in regard to mathematically modelling by commenting on those critiques raised in [1], with a hope of initiating some further discussion so that both applied mathematicians and biologist can better use mathematical modelling and better understand the results from modelling.

Mathematical models have been widely used in the biological science in the recent decades. Meanwhile, there are quite a lot of debates and discussions on the roles that mathematical modelling can play in biology, including the application ranges and limitations of models. In a recent opinion paper published in the journal Progress in Biology and Molecular Biology [1], Dr. S. F. Gilbert raised five caveats for mathematical models in biology, and they are

1) Mathematical models are limited by the science known at the time.
2) Mathematical models can tell what can happen, not what does or did happen.
3) Real-world models can provide a better explanation than the mathematical model.
4) In abstracting reality, the things left behind can be very important.
5) Mathematics models can be Platonic rather than evolutionary.

These caveats are somehow reasonable and suggestive, and we need to be aware of these issues when we try to explain or apply the results obtained from mathematical models. However, the above paper contains many inaccurate descriptions about mathematical modelling, and these inaccuracies may mislead readers and cause confusion, and therefore, they should be clarified. It is unfair and unscientific to expand observations from some models to all mathematical models without solid and convincing arguments. The purpose of this viewpoint note is to clarify the above five caveats, by commenting on each of them. I hope these comments, together with possible following-up discussions, will help us better understand the roles and limitations of mathematical models in biology.

First of all, "mathematical modeling" is not "a set of technologies". In scientific researches, mathematical modeling is a way of thinking and language of describing the hypotheses according to the observed phenomena. The roles of mathematics in scientific research should mainly be reflected in two aspects:

(I) mathematics is a way of logical thinking, which is an ability of helping people to understand the logical relationship behind the experimental facts in biology (the same for other disciplines);
(II) mathematics is a language of science, which can accurately express human's understanding of the laws of nature.

The charm of mathematics for biology (and other sciences) lies in that it is the best way to express people's understanding of natural laws and their underlying mechanisms, and infer possible results based on these understandings through logical deductions. The understanding of a mathematical model and its application in biology should be put into such a frame to see more essential connotation. It is with the above considerations that I will comment on each of the above critiques below.

On Critique 1): Mathematical models are limited by the science known at the time.

This sentence should be modified as "*Mathematical models are limited by our understanding of the science known at the time*", to be more appropriate. Mathematical models are not only descriptions of the world we can see, but also descriptions of our understanding of natural laws behind the observed phenomena. The word "understanding" here should be the core of the model. Sometimes, there are different understandings based on the same phenomenon, and hence different mathematical models can be established based on different hypotheses. Some of them are wrong, and some are more reasonable. Our understandings of natural laws (e.g., the basic laws in life science) are usually gradual and often depend on new discoveries. Therefore, the mathematical models for describing the same phenomenon can also evolve; the limitation of mathematical models does not come from the limitation of mathematics itself, but the limitation of human's understanding of the natural phenomena at the time.

For biology, biologists observe new phenomena, discover the facts, explore the processes of life, and often engage in the work of discovering the known sciences. Scientists from other fields, such as physics, mathematics, chemistries, etc., can play important roles in understanding these known sciences, and discovering the underlying laws behind the phenomena that biologists cannot see clearly. For example, in the discovery of DNA double helix structure, physicists played important roles in several aspects: from discovering the genetic material and the DNA diffraction image, to the discovering of the double helix structure. Another example is the establishment of the mathematical model that describes the mechanism of action potential in neurons, that is, the Hodgkin-Huxley equation [2]. This equation was established based on a series of experiments of dynamic information of the neuronal electrical signals under various stimulations. In order to understand the underlying mechanism behind these experimental phenomena, Hodgkin and Huxley proposed the gating mechanism of ion channels, and proposed the possible rules for the gating mechanism of sodium and potassium ion channels based on the experimental data. The Hodgkin-Huxley equation was then established by combining the gating mechanism with the equivalent circuit model of cell membrane. The Hodgkin-Huxley equation nicely describes the mechanism of action potential and predicts a series of experimental results, and has become a basic equation in computational neuron science. Nevertheless, at the time when this equation was established, people knew nothing about the structure of ion channels, not to mention the occurrence of gating mechanism. Hodgkin and Huxley understood the mechanism of ion transportation before they see the ion channels, and the proposed mathematical model is a tool to describe their hypothesis. If Hodgkin and Huxley did not speculate about the gating mechanism behind the changes of the electrical signals they have seen, but tried to explore the structure of the ion channels and the detailed processes of ion transportation, the understanding of the mechanism of action potential would be postponed to at least decades later. The Hodgkin-Huxley equation serves as an excellent example demonstrating that we can establish a reasonable and useful mathematical model through a better understanding of science known at the time. More interestingly, people found that similar mechanisms can also be applied to describe the electrical signal behaviour of cardiomyocytes.

On Critique 2): Mathematical models can tell what can happen, not what does or did happen.

It is one-sided or biased to criticize that "mathematical models can tell what can happen, but not what did happen". It may be true for a statistical models that usually tell the probabilities of the occurrence of some events/phenomena. Many mathematical models are given in a form of deterministic formulations, such as the Hodgkin-Huxley equation mentioned above, and the reaction-diffusion equation model for Turing patterning. The deterministic models can somehow tell us what can happen and what cannot happen, and help us to understand why something did happen. Hence, for most deterministic models established based on our understanding of the underlying mechanisms, the above criticism should be rephrased as *"Mathematical models can tell what can happen, and explain why some phenomena may or may not happen"*.

As for the relationships between mathematical models and experiments, if one can predict what will happen through simple experiments, it may not be significant to introduce mathematical models, such as the example of of X-linked disease mentioned in Gilbert's article. However, for many experiments with more complex dynamics and multiple scale interactions, it is not trivial, if not impossible, to get insights behind the experimental observations. In this case, a proper mathematical model can help us to understand the experimental results based on different hypotheses, and to select a more reasonable mechanism. For example, in the field of developmental biology, morphogens are important signaling molecules that govern the tissue pattern development, however there were debates on how the morphogen gradients are formed. Biologists proposed that morphogen gradients can be established by a process of repeated cycles exocytosis and endocytosis, that is, so-called planar transcytosis. Later, a mathematical model with reaction-diffusion equations shows that diffusive mechanisms of gradient formation may be more convincing [3, 4]. In this example, the mathematical model can not only tell us what did happen, but also help us understand why other mechanisms are unreasonable.

A good mathematical model should provide reasonable explanations in accordance with experimental facts, and the rationality of the mathematical model needs to be able to withstand the verification of the experimental facts. When different models can explain the same experimental results, we may need to design new experiments to distinguish these models. Therefore, mathematical models should be constantly revised with the discovery of new experimental facts, which is the only way for us to understand the laws of nature. However, we should not deny the role of mathematical modelling merely because they are inconsistent with some experimental facts; instead, the deviation between model predictions and experimental facts can sometime stimulate us to explore the ignored factors or wrong hypothesis in the model, corrections to the model hypotheses can eventually lead us to a better understanding of the experimental findings.

On Critique 3): Real-world models can provide a better explanation than the mathematical model.

In [1], the author introduced two examples to illustrate the limitations of the mathematical models: Kepler's law in planetary motion and gravitational interpretation, and binary code model in Drosophila embryo development. Through these two examples, the author tried to criticize that "mathematics can give a model that explains phenomenon, but which does not work as nature actually does". I think this point is basically correct, and is also a reason why many biologists do not trust mathematical models. However, any model is an approximate description of the real world, especially for those mathematical models in the field of biology. It is impossible to consider all the details. The question is, do we need to consider all details? Are different models helpful for the problems under consideration? Even in physical sciences, many models are perfect and can stand the test of experiments, but they are also only real-world approximations. When we study the motion of objects with conventional scales (such as the flight of rockets, the motion of planets, etc.), it is sufficient for us to employ the Newtonian mechanics models.

When we study problems in the cosmic scale, such as the movement near the black holes or super stars, we need mathematical models based on general relativity. When we study quantum behaviour, we need to consider mathematical models derived from quantum mechanics. These models are very good descriptions of the problems at a certain scale. However, there is no unified theory that can describe all the basic interactions from quantum to gravitation force. Even if there is such a theory, it is impossible (and unnecessary) to apply it to study most practical problems. For a better understanding of our world, we need a balance between a workable mathematical model and the real-world models.

In fact, it is not appropriate to use Kepler's law and gravitation theory to discuss the point here. Kepler's law is the law of planetary motion induced by astronomical observation data. The law of universal gravitation can be strictly deduced according to Newton's second law and Kepler's three laws, and it is a theoretical improvement of the observed laws. On the basis of recognition of Newton's second law, the Law of Universal Gravitation is equivalent to Kepler's laws. However, Kepler's laws are only a summary of the observed data of planetary orbits. These laws cannot be judged in a wider range, and the predictability of these laws are very limited for the planets that are not easy to be observed. When we obtain the law of gravitation, we can not only explain Kepler's three laws, but also predict the orbits of planets, such as the occurrence of comets, the discovery of Pluto, the orbit of the moon landing rocket, and so on. In [1], the author said that the law of universal gravitation cannot explain why planets choose the orbit we can see. This involves another problem of the formation process of planets that may lead to an extended mathematical model that includes the change of material distribution during the birth of the star systems.

In the example of the binary code model in Drosophila embryo development in [1], the simple model of binary code model may be over simplified for the real world problem. However, in latter studies, more related genes and regulatory factors were found, different types of mathematical models were constructed based on these regulatory relationships which provided good explanations to the relevant experimental phenomena [5].

To better understand the gap between mathematical models and the real world, we need to look back to the first point about the limitation of mathematical model: mathematical models are limited by our understanding of the science known at the time. The gap between the mathematical model and the real world mainly due to our understanding of the real world. Peoples may propose different hypotheses and hence different models when they try to formulate the experimental facts. An important ability of experienced applied mathematicians is that they can put forward appropriate explanations and great guesses about the underlying mechanism through their insightful observations into the problems being studied. A successful mathematical model often includes a nice guess on the mechanism. There are many examples in this respect, such as Hodgkin and Huxley's guess on the mechanism of gated channel [2], Turing's hypothesis on the mechanism of pattern formation caused by morphogenesis[6], Heisenberg's conjecture on the mechanism of energy cascade in turbulence spectrum, Lin and Shu's density-wave structure of disk galaxies [7], etc. For mathematical models in biology, we may need to limit our research to a certain range under specific conditions due to the complexity of biological systems. To work out with a good mathematical model, close cooperations and discussions between mathematicians and biologists are required to make appropriate guesses/hypotheses, and a good model can only be established through iterative model predictions, experimental verifications, and corrections. In this process, mutual trust and cooperation between mathematicians and biologists are very important.

Each mathematical model has its scope of application. It is not appropriate to deny a mathematical model with facts beyond the scope of application of the model. The applicability of a mathematical model should be considered from several aspects: (1) Is the hypothesis proposed in the mathematical

model reasonable? (2) What is the scope of application of the mathematical model? (3) How can we improve/remove the unreasonable assumptions in the model or extend the scope of the model?

On Critique 4): In abstracting reality, the things left behind can be very important.

This point of view is partially correct, especially for complex biological systems, we have to make simplification and abstraction when we establish a mathematical model. Hence, there are some factors that cannot be included in the model. The issue is that whether or not these excluded factors are important for the problem being studied? In developing mathematical models, mathematicians often focus on some main factors that are most important for the problem, and ignore the unimportant or less important effects. This process is somehow subject to the judgement on the importance of the factors, and hence different factors may be included in different mathematical models developed by different people. In this case, the things left behind can be very important. When we established mathematical models, we often point to a main problem to be answered, and hence make assumptions according to the problem. In this way, we need to focus on the main interactions that are important for the issue under consideration, and unfortunately neglect other less important factors. For example, in modelling cell behaviour, we can establish mathematical models for a motif of signalling pathways if our goal is to understand the role of a specific gene in the pathway; or we can model the dynamical response of a more complex signalling pathway if we want to know the cellular response to external stimuli; we may also need to include cell-to-cell interactions if we want to study the interactions between cells. There are models of different scales, which may depend on the questions we try to answer. Moreover, some important factors may be missing from a mathematical model because of the limitation of current technologies. These issues are common to all researchers in all areas in natural sciences. On the other hand, when important factors are found missing in a mathematical model (or theories), it may motivate us to modify the model and/or even develop new theories.

On Critique 5): Mathematical models can be Platonic rather than evolutionary.

For this point, the author stated that mathematical models can only be used to describe the ideal state, but cannot describe the complex and changeable environment and long-term evolution. As the author stated: "But one must remember, that mathematics can only model regulations, and that evolution has a large component of containment of contingencies". From the perspective of mathematical methods, mathematics can describe more complex situations and regularities, including the changing of environment, random process, and the evolution of dynamics systems, etc. There are various mathematical methods that can be used to describe these complex processes. Furthermore, in terms of our understanding of the biological world, mathematical models cannot be created out of nothing, and cannot provide good models for facts or phenomena that people do not understand. For example, if we do not know the mechanism behind complex behaviours in biological phenomena and cannot make reasonable assumptions either, we will not be able to establish appropriate mathematical models. Mathematical models are not magic, it cannot be used to study the process that we cannot provide reasonable understanding.

Above are comments on Gilbert's five caveats on mathematical models in biology. From the perspective of applied mathematics, there is no essential difference between the mathematical models in biology and the mathematical models in physics. Despite the various aspects of the problems we may face, there is no essential difference in research methods, ideas and processes in mathematical modelling. I would conclude this note by quoting Prof. C. C. Lin's general themes of establishing mathematical models in researches of applied mathematics, which are essential to understand the role of mathematical models in different disciplines. As these themes are abstracted from a private communication, I provide them in the form of a scanned figure (Fig. 1).

GNERAL THEMES

General themes for a plan of research in Applied Mathematics
After the subject or topic of research is selected

1. Empirical knowledge collected in a systematic manner
2. Analysis and synthesis of empirical knowledge and data
3. Formation of a systematic "physical picture"

4. Formulation of a corresponding mathematical model
5. Development of a mathematical theory with predictive power
6. Adoption of appropriate mathematical methods for the study of selected topics and cases in detail

7. Comparison of theoretical conclusions with empirical knowledge and data.

应用数学研究的主要思路

FIGURE 1. General themes of of applied mathematics proposed by Prof. C. C. Lin in [8].

REFERENCES

[1] Gilbert, S. F., *Achilles and the tortoise: Some caveats to mathematical modeling in biology*, Progress in Biophysics and Molecular Biology **137**(2018), 37-45.

[2] Hodgkin, A. L. and Huxley, A. F. , *A quantitative description of membrane current and its application to conduction and excitation in nerve*, Journal of Physiology **117**(1952), 500-544.

[3] Lander, A. D., Nie, Q. and Wan, F. Y. M., *Do morphogen gradients arise by diffusion?* Dev. Cell **2**(2002), 785?796.

[4] Matthew, F., *Morphogen gradients in theory*, Dev Cell **2** (2002), 689-690.

[5] Morelli, L. G. L., Uriu, K. K., Ares, S. S. and Oates, A. C. A., *Computational approaches to developmental patterning*, Science **336** (2012), 187-191.

[6] Turing, A.M., *The chemical basis of morphogenesis*, Phil Tran Roy Soc London B **237**(1952), 37-72.

[7] Lin, C. C. and Shu, F. H., *On the spiral structure of disk galaxies*, Astrophysical Journal **140**(1964), 646.

[8] Lin, C.C., *Private communications*.

SCHOOL OF MATHEMATICAL SCIENCES, TIANGONG UNIVERSITY, TIANJIN 300387, CHINA
E-mail address: jzlei@tiangong.edu.cn