J Biosci (2022)47:19
DOI: 10.1007/s12038-022-00260-z

Perspectives

Nine things to keep in mind about mathematical modelling in ecology and evolution

Amitabh Joshi

Evolutionary Biology Laboratory, Evolutionary and Organismal Biology Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Jakkur, Bengaluru 560 064, India

(Email, ajoshi@jncasr.ac.in)

MS received 15 December 2021; accepted 26 January 2022

Modelling in ecology and evolution, especially in India, is often done by researchers trained in physics or engineering without much experience of studying living systems. This is partly driven by a fallacious conviction that modelling is largely about mathematical skills and that, consequently, modellers can equally effectively apply their skills to problems in fields as diverse as physics/engineering and ecology/evolution. I discuss why this fallacy arises, and the many ways in which modelling in ecology or evolution is actually a very different endeavour from that in much of physics, even though the form of the equations deployed across disciplines is typically quite similar. Since modelling is not primarily about the mathematics but about the system being studied, I believe that a reasonable degree of comfort with models and modelling is important for those researchers in ecology and evolution who primarily undertake empirical studies, whether in the laboratory or the field. Equally, I suggest that researchers doing modelling in ecology and evolution, who were trained in the mathematical or physical sciences, need to understand the systems they attempt to model and also appreciate how modelling ecological and evolutionary processes differs from much of the modelling done in classical physics and allied fields. I also discuss what models are, whether modelling is subjective or objective, and what modelling entails if it is to meaningfully add to scientific understanding. This article is aimed primarily at young researchers interested in ecological and evolutionary questions, whether coming from a background in the biological or physical/mathematical sciences.

Keywords. Density-dependent selection; generality; mechanistic versus phenomenological models; philosophy of modelling; population dynamics; precision; realism

1. Introduction

> *Modelling is too important to be left to the theoreticians*
>
> *– adapted from Georges Clemenceau*

"War", as Georges Clemenceau, French prime minister during the later part of World War I, famously observed, "is too important to be left to the generals." As an ecologist and evolutionary biologist, I have often similarly felt that modelling is too important to be left to the theoreticians. Much of the modelling in ecology and evolution, especially in India, is often done by people with training in mathematics, physics, or engineering, and not too much understanding or appreciation of the subtle complexities of systems of interacting organisms, or, indeed, of the fact that modelling in ecology and evolution is, in some ways, actually quite different from modelling in physics. I should hasten to add that there are many exceptions: many Indian physicists who model ecological and evolutionary systems have actually taken the effort to come to grips with the biology of those systems but they are still a minority.

The attitude of many of our Indian ecologists and evolutionary biologists is also a bit strange in that they

---

This article is part of the Topical Collection: Emergent dynamics of biological networks.

often exhibit a tendency to avoid, if not actually abhor, any serious engagement with theory in their fields. For example, it has been argued that biologists should be agnostic with regard to theoretical controversies (Gadagkar 2015). Yet, it is often the case that long-standing theoretical controversies resist resolution precisely because the problem is not in the mathematical analysis of the models but in the mapping from the biology to the mathematics. In such cases, the resolution typically comes about by a novel deployment of aspects of our empirical understanding of the system, rather than by mathematical refinements not rooted in such an empirical understanding. Such critical insights leading to a resolution are, therefore, far more likely to come from biologists bringing their empirical experience to bear on the theory, rather than from theoreticians operating largely in the mathematical and not the empirical realm. A good example of this can be seen in the resolution, in the 1990s (Joshi and Moody 1995, 1998), of a then almost twenty-five-year-old debate, involving leading theoretical evolutionary biologists of that time, about whether the so-called two-fold cost of sex is due to the cost of male function (e.g., Maynard Smith 1968) or of genome dilution (e.g., Williams 1975). While both J Maynard Smith and GC Williams were biologists, in this case they were relying on simplistic models and ignoring some relevant biological details. Eventually, the issue got resolved by noting that both classes of models were treating male gamete production by individuals reproducing asexually in their female capacity as being fixed either at zero, or at par with the male gamete production of sexually reproducing individuals, even though most undergraduate botany courses indicate that male gamete production and fertility of such individuals varies widely between these two extremes. Incorporating these two traits as parameters in a slightly more realistic model, rather than fixing their value at 0 or 1, revealed *inter alia* that both investing in male function and genome dilution contribute to the cost of sex, except at the two extremes, where one or the other becomes the sole component (Joshi and Moody 1995, 1998). The point I wish to stress here is that debates about theory in ecology and evolution that hinge upon purely mathematical considerations usually get resolved fairly quickly by theoreticians themselves. The long-standing debates about theory resist resolution by theoreticians not bringing all relevant empirical knowledge to bear precisely because their resolution lies outside the domain of the theoretical, residing in the details of how biological reality has been mapped onto the mathematics in the modelling process. Such problems with the mapping are more likely to be picked up by researchers applying their understanding of the system being modelled, who also have some familiarity with the process of modelling. It is in this context that I feel that it is our job as biologists to understand, engage with, and contribute to modelling in our fields.

Another point worth making is that many researchers, both in physics and biology, seem to think that mathematical modelling is essentially similar across these widely different scientific disciplines. Certainly, mathematical modelling of the kind developed in physics has also been used in many areas of biology, especially ecology and evolution, for over a century (Otto and Day 2007). Yet, while it is true that the structure of the functions or equations used in mathematized models and, consequently, their analysis, is often very similar across biology and physics, the manner in which the models are deployed and tested, and what our expectations from the models are, actually differs quite substantially across these disciplines. My focus here is on modelling in biology, but the general point, that good and effective modelling requires a thorough empirical understanding of how the system being modelled actually behaves in different contexts, is likely to be true for a physical, biological, or any other kind of system. I should make it clear at this point that when I refer to ‘modelling in biology’, I mean ‘modelling in ecology and evolution’. Not being very familiar with modelling in sub-organismal biology, I am not certain of the degree to which my arguments will be applicable to that domain. For a perceptive, albeit slightly dated, review of modelling in sub-organismal biology, the reader is referred to Gunawardena (2014).

I should also make my background, biases, and expectations explicit: I was trained as a geneticist, but for the last thirty-four years I have been working on questions in population ecology and evolutionary biology, especially small-population and metapopulation dynamics, the evolution of population dynamic behaviours, life-history evolution, and the evolution of adaptations to crowding. I do mostly multi-generation experiments with fruit-fly populations in the laboratory, but also do theoretical research, both analytical and simulation-based. I was not formally trained in theoretical research, but enjoyed studying the theoretical analyses of ecological and evolutionary dynamics as a student. When I was a graduate student working on the coevolution of competing *Drosophila* species, I began to do some theoretical work on my own, on the side, and found it greatly satisfying. I have continued to also do theoretical work ever since, though I am primarily an experimental evolutionary biologist. Moreover, the experience of teaching essentially theoretical courses in hypothesis testing, population genetics, and population dynamics to students from a biology background for twenty-six years has also led me to many insights that have been helpful in getting students to appreciate what modelling is and is not. This piece is aimed primarily at young researchers, with a background either in physics/mathematics/engineering or biology, who are interested in theory in ecology and evolution and wish to contribute meaningfully to it. If it perhaps also encourages some young experimental ecologists and evolutionists to step up and ‘own’ the theory in the field, I will be more than recompensed. Secondarily, I hope the piece might also be of some interest to physicists and other non-biologists who undertake the modelling of problems in ecology and evolution.

2. Modelling is subjective

صوُرت ہے تیری، یا عکس میرا
کیا شے ہے، بتا، تصویر، مُجھے
نبراس اکبرآبادی

[*Is it your form, or a reflection of me?*

Tell me, what is the essential nature of this picture

Nabraas Akbarabadi]

A question that implicitly or explicitly underlies many discussions on the philosophy of modelling is whether a model is a human construct, reflecting our thinking, or is a reasonably accurate reflection of reality. In this section, I will argue that modelling is essentially a subjective activity, even though the analysis of the behaviour of a model under differing conditions can be fairly objective. Unfortunately, many of us have a tendency to conflate the modelling with the analysis and, consequently, labour under the misapprehension that modelling is a very objective endeavour.

Modelling has been described by many people in varied ways. Essentially, modelling, in its broadest sense, is an integral part of both mundane decision-making in our daily life and of constructing knowledge in the domain of science. One of the best delineations of what modelling is, and its role in how we construct scientific understanding, was provided by Rosen (1991), and I am largely reiterating his take on the issue. A model, in this view, is essentially a mapping we create between a subset of what we observe in the real world and its representation in our minds. In particular, we attempt to mirror, in patterns of entailment among the entities that make up our model, what we believe to be the cause–effect relationships among the relevant constituents of interest that are involved in the phenomenon under observation. If we conceive of the creation of scientific understanding about some phenomenon as a process in which we go back and forth between a reality space and a concept space, we can appreciate the role of modelling in the scientific process (figure 1). For example, in the reality space on the left side of figure 1, we observe the entities *a*, *b*, *c*, *d*, *e*, *f*, *g*, *h*, and *i*, at least some of which seem to be related to one another by cause–effect relationships. Of these nine entities, following Pablo Picasso’s dictum “Art is the elimination of the unnecessary”, we choose to ignore *b*, *e*, and *h*, based on our question of interest

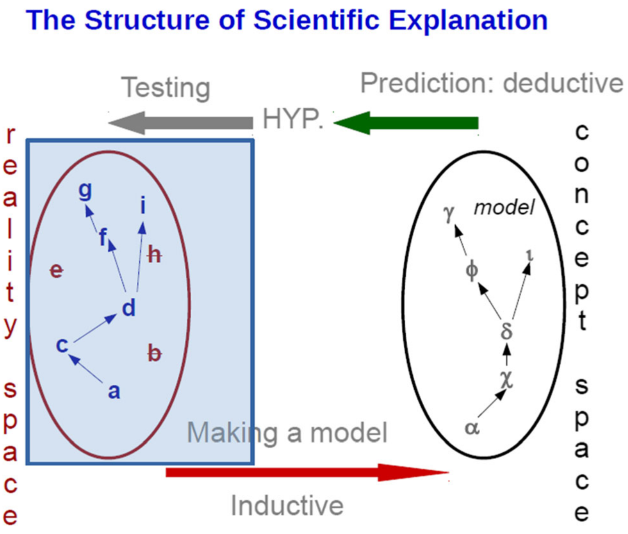

Figure 1. A schematic depiction of the process by which scientific understanding is created. On the left of the figure, enclosed in the dark red oval, is a subset of some phenomenon in the real world (reality space) that we wish to understand. Understanding, in this context, typically means attaining the ability to make reasonably reliable predictions about the phenomenon. In this phenomenon, entities *a*, *b*, *c*, *d*, *e*, *f*, *g*, *h*, and *i*, which seem to be related to at least some of one another by cause–effect relationships, appear to play a role. Of these nine entities, we choose to ignore *b*, *e*, and *h* in our modelling, and map, by induction, the remaining cause–effect relationships onto a pattern of entailment among entities $\propto$, χ, δ, φ, γ, and ι in concept space (on the right of the figure). This is the act of modelling (*sensu* Rosen 1991). From the model, which may be mathematized, we deductively derive predictions that then become hypotheses to be subjected to empirical testing. If our predictions are reasonably close to observations, preferably repeatedly and under various conditions, we conclude that the model is useful.

and our understanding of what may or may not be relevant to it. Consequently, we focus on the remaining six entities and the cause–effect relationships we can discern among them (represented by arrows in the reality space). We note that the choice of question and the elimination of what is thought unnecessary are both subjective actions. Another researcher, trying to understand the same phenomenon, may eliminate a slightly different subset of entities from consideration, leading to a subjectively different model of what is essentially the same phenomenon.

The next step is the modelling per se, which, in the case of mechanistic models (e.g. many population genetics models), involves the creation, in our minds, or concept space, of a set of entities (α, χ, δ, φ, γ, and ι on the right side of figure 1) with some relationships of entailment among them, which we hope accurately embody the critical features of the cause-and-effect relationships among the entities of our interest in the reality space. In the case of phenomenological models (e.g. many simple models of population growth), we often set up higher-order entities (e.g. per capita population growth rates, carrying capacities) in the concept space that do not necessarily correspond to directly measurable entities in the reality space. In such cases, we try to find patterns of entailment relationships among these higher-order entities that yield outcomes that are concordant with those we observe in the real system. For more on mechanistic versus phenomenological models, see section 5.

This step of modelling is essentially an exercise in inductive reasoning, and is also fairly subjective in that different researchers may opt for different mappings from the reality space to the concept space as representations of the same phenomenon. Sometimes, different mappings can be more or less useful, depending on the context. This hopeful mapping is the model and, in its essence, is always verbal. In many cases, we can choose to formally and explicitly depict the nature of the entailment relationships by mathematical statements. Consequently, the same model can have alternate mathematizations, and just like the models themselves, alternate mathematizations of a model can be relatively more or less useful in differing contexts. I return to this issue in more detail in section 3.

The next step is to deduce, using the model/mathematization, the consequences of altering some aspect of the phenomenon or its context. This is the activity of developing predictions from the model, and these predictions take the form of 'if X, then Y' statements, where both X and Y can be anything from point values to qualitatively described states. The predictions, in turn, must then be tested against observations or experiments. The empirical testing of predictions from a model requires engaging with the central question of statistical hypothesis testing: how close is close enough? Depending on the nature of the prediction, and of its testing, this often becomes an issue of assessing how well the model fits the observed data. In experiments in many areas of physics, it is not uncommon to get extremely good fits ($R^2 > 0.9$), whereas in ecology and evolution even an R^2 value of 0.6–0.7 is considered a pretty decent fit. Indeed, in most ecological systems, not just the magnitude of the errors but also the nature of the error distribution is quite different from that in physical systems. The error distribution can also interact in seemingly counter-intuitive ways with data transformations and back-transformations. For example, mathematically, if y is an exponential function of x, then taking a log-transform of y renders it linear in x. One can then use linear regression to fit the data, estimate the intercept and slope, and then back-transform to get the parameters of the original exponential function. Prior to the advent of high-powered computers and algorithms for nonlinear curve fitting, this was the standard method for estimating the parameters of exponential models. However, as shown by Mueller *et al.* (1995), in the context of comparing patterns of ageing across populations using the Gompertz (1825) equation, log-transformation can sometimes yield extremely biased results. The Gompertz equation, $\mu(t) = Ae^{\alpha t}$, links age-specific mortality rate ($\mu(t)$) to age-independent mortality (A) and a rate of ageing (α). When parameters of this equation are estimated by linear regression, following log-transformation, the results can be very biased, because the error distributions can differ substantially across populations. For example, if populations with different evolutionary histories vary in α, the constant rate of change of transformed y, assumed in linear fitting, can vary greatly within and among populations. Similarly, for populations with varying α, the relative magnitude of errors at different ages will be very different, and these will get differentially compressed during transformation of y values. These problems introduce significant bias in the parameters estimated by log-transformation followed by linear fitting, consequently rendering the linear-fitting approach useless for among-population comparisons (Mueller *et al.* 1995).

Model fitting is also beset by the complications that (i) the best fitting model need not necessarily be the 'true' model, and (ii) often models that are clearly 'wrong', in that their assumptions go against what we know of nature, can nevertheless fit data pretty well. A good example of the latter is Ptolemaic astronomy, which despite assuming circular planetary orbits around the earth, nevertheless managed to successfully predict celestial phenomena by the creative addition of epicycles. This problem of ‘over-parameterization’ is discussed in more detail in section 4.

Overall, then, we can see that the model is basically a mapping from a reality space to a concept space, and it is tested by deducing predictions from it and comparing them to observations. If the predictions match observations reasonably well, and what is reasonable can be very different between ecology/evolutionary biology and physics, the model is deemed to work well. Subjectivity enters into models at many stages of the process of model making and testing. The choice of phenomenon to study, how to model it, which mathematization of the model to use, which predictions to test, and how to test them, are all decisions with large subjective components. It is important, therefore, to bear in mind that the knowledge gained from a modelling enterprise is not strictly objective in any meaningful way.

3. A model and its mathematization are different

هر لهذه به شکلی بُتان عیّار بر آمد، دِل بُرد و راون شُد
هردم بہ لباسی دیگران یار بر آمد، گہ پیِر و جوان شُد
مولانا جلال الدین رُومی

[*Each instant, my artful Beloved appears in varied idol forms; steals my heart, and departs*

Every moment, the Beloved appears before me in a different garb; young one time, another old

Maulana Jalaluddin Rumi]

I mentioned in the previous section that it is important to differentiate between a model and its various mathematizations. For example, standard textbooks (e.g. Hartl and Clark 1989) present simple one-locus population genetics models with allele or genotype frequencies at that locus as the state variables, and mathematically deduce their change over generations, under the effect of factors like mutation, migration, inheritance or selection, through the notion of a gamete pool from which pairs of gametes are picked at random to form zygotes. This is, of course, not what actually happens during reproduction in most species of animals and plants. One can, however, also mathematize this phenomenon in terms of a mating table, explicitly tracking each kind of mating, its frequency, and the offspring distribution resulting from it. It can then be seen that the assumption of random mating is formally equivalent to the assumption of random union of gametes in the gamete pool (Hartl and Clark 1989). Change in the composition of a population can also be mathematized in terms of phenotypic frequencies, with their change being modelled in terms of the transmission fidelity of each phenotype under a given mating system (Vidya *et al*. 2022). Which mathematization is most useful varies according to context. Under random-mating contexts, the gamete pool mathematization works well and gives fairly accurate predictions of how the genetic composition of the population at that locus changes over a few generations, with the added benefit of permitting the use of one or a few allele frequencies, rather than a larger number of genotypic frequencies, as state variables. If mating is non-random, the mating table mathematization can accommodate the more complex scenario and allow prediction of genotypic frequencies in the next-generation zygotes. If there is a non-genic inheritance mechanism at play, then a more general phenotype-based mathematization invoking transmission fidelities can be used to predict the phenotypic frequencies in the next generation, provided we have knowledge of the transmission fidelities obtaining under the given inheritance system.

An example of different mathematizations yielding very different kinds of insights can be seen when comparing population genetic models of selection to the Price (1970) equation (Joshi 2020), or even comparing alternate mathematizations of the Price equation to one another (Rice 2004). The process of selection acting on a set of genotypes at a locus can be mathematized via examining allele frequency changes over time, as in basic population genetics models (e.g. Hartl and Clark 1989), or via the covariance approach pioneered by Price (1970). The two kinds of mathematization have different purposes and strengths. The allele frequency mathematization of selection takes the well-established classical physics approach of trying to derive equations of state variable (allele frequency) change over time, under some simplifying assumptions. The key here is to relate recursively, under the simplifying assumptions of an ideal large population, the allele frequency in one generation to what it was in the previous one. The Price equation, on the other hand, takes the very different approach of asking what mathematical structure arises from a consideration of the process of selection qua selection, without any simplifying assumptions (Rice 2004). The allele frequency mathematization is useful for short-term predictions of the change in the genetic composition of a population at one or a few loci, in well-controlled systems. The Price equation, in contrast, offers an insight into the fundamental attributes and consequences of the selective process, independent of the system of inheritance or the trait that is the focus of investigation (Rice 2004; Joshi 2020).

4. Models are couplets, not photographs

فاصلہ ء خُد و بیخُد ناپ لے اِک شعر سے
ایسا کیا کوئی سُخنور زیِست کی محفِل میں ہے
نبراس اکبرابادی

[*One who can measure the distance between self and non-self with a couplet*

Is there any such accomplished poet in this soirée of life?

Nabraas Akbarabadi]

In a Monty Python sketch (*https://www.youtube.com/watch?v=l9Aj7W3g1qo*), the Pope, who has commissioned a painting of the Last Supper from Michelangelo, keeps insisting that he wants a faithful depiction of the gathering, much to the chagrin of the painter, who expects and argues for artistic license. Finally, a frustrated Michelangelo exclaims, in a strangely befitting Australian accent, "You want a bloody photographer, that's what you want!" There is often a tendency to conflate models in biology with photographs, especially in the case of relatively mechanistic models. What I am terming photographs and models are sometimes contrasted as analogical versus conceptual models, respectively (Mazarati 2007). The famous paper of Levins (1966) on the strategy of modelling in population biology, which first articulated in detail the notion that it is impossible to combine realism, generality and precision in a model of an ecological process (see also section 5), arose in part as a reaction to the elaborately detailed, photograph-like models being promoted by systems ecologists (discussed in detail by Odenbaugh 2006).

Apart from the problem that a very detailed and realistic model typically sacrifices generality, there is also a deeper philosophical issue here: the difference between description and explanation. If one seeks a description, detailed realism is, no doubt, desirable. However, an explanation implies that a large subset of the observed aspects of a natural phenomenon is being shown to be predictable from our knowledge of a small set of effectors. In statistical terms, explanation requires that a substantial fraction of variation in the dependent variable(s) be accounted for by variation in one or a few independent variables, with a small number of parameters. If we include a sufficiently large number of parameters in our model, we can explain ever-increasing proportions of the variation in the dependent variable but with severely diminishing returns in terms of our understanding of the phenomenon. This is the spirit in which John von Neumann was quoted by Enrico Fermi as saying, "With four parameters I can fit an elephant, and with five I can make him wiggle his trunk." Eventually, in the limit that the number of parameters equals the number of data points, we can explain the entire variation in the system under study with our model, thereby explaining everything and, therefore, paradoxically, nothing. Model fitting, thus, requires a judicious balance between the opposing requirements of goodness of fit and a reasonably small number of parameters.

It is also in this context that I believe models are more like couplets in Urdu or Persian poetry, or Japanese haiku, than photographs. Models do not, and indeed should not, aim to describe. They are succinct metaphorical representations of some aspect of reality, with the metaphors typically being mathematical and distilled from our experience, i.e., from our intuition about the system being modelled. Except for the bit about the metaphors being mathematical, that is also a fairly accurate description of a couplet or a haiku. Moreover, good models – like couplets – are usually beautiful and evocative, delineating fruitful avenues of thought rather than providing definitive answers. As Bertrand Russell said,"A good notation has a subtlety and suggestiveness which at times make it almost seem like a live teacher"; the same is true for models, more generally.

Just as couplets convey fundamental ideas clothed in metaphors – think of the burning candle and the tiny moth in the Persian-Urdu poetic tradition – models, too, sing to us (in mathematical lyrics) of some insight into the natural world. The Breeders' equation ($R=h^2S$) in quantitative genetics, for example, does not just embody a relationship between selection response, heritability and the selection differential in the context of plant or animal breeding. It also conveys to us a fundamental insight that evolutionary change in the mean trait value (R) arises from an interaction between the ecology of the system (encapsulated by S) on the one hand, and the mechanism of inheritance of the trait and the mating system (both distilled into h^2) on the other. The underlying message of this little equation is, thus, that evolutionary change is generated by the interaction of ecology with heredity.

5. Realism, generality and precision

کبھی کِسی کو مُکمّل جہاں نہیں مِلتا
کہیں زمیں، کہیں آسماں نہیں مِلتا
نِدا فاضلی

[*Nobody ever gets a perfect world*

Either the earth or the sky is missing

Nida Fazli]

In a very influential paper on strategies of modelling in population biology, Levins (1966) posited the existence of trade-offs among the model attributes of generality, realism and precision, writing, "It is of course desirable to work with manageable models which maximize generality, realism, and precision toward the overlapping but not identical goals of understanding, predicting, and modifying nature. But this cannot be done." He then went on to classify modelling approaches in ecology and evolution into three types, each sacrificing one of generality, realism and precision for the sake of the other two (Levins 1966). He also introduced a somewhat idiosyncratic notion of a 'robust' theorem (with some similarity to the notion of structural stability of a model), a similar result obtained from a cluster of different models of the same problem, sharing a major biological assumption but differing in other simplifying assumptions made, an approach he summed up eloquently, but somewhat vaguely, with the assertion, "Hence our truth is the intersection of independent lies" (Levins 1966). The contrast was with a 'fragile' or 'non-robust' theorem, wherein the result from a model depends crucially on the specific simplifying assumptions made and, therefore, lacks generality.

In Levins' (1966) three-way classification of modelling strategies, models sacrificing generality for realism and precision are akin to the 'photographs' discussed in section 4, whereas those sacrificing realism for generality and precision are characteristic of the physicists' approach, incorporating many simplifying assumptions and leading to insights about how different factors interact to affect outcomes of ecological processes. Against these two kinds of model, Levins (1966) juxtaposed the third approach, favoured by him and some others (e.g. MacArthur 1962; MacArthur and Wilson 1967). In this approach, precision is sacrificed for generality and realism, with the models often not specifying precise mathematical functional forms for the relationships among component entities, thereby yielding 'rules of thumb' that Levins (1966) termed 'qualitative predictions', as opposed to the quantitative predictions arising from models following the previous two strategies. Incidentally, a very nuanced perspective on the relative merits of quantitative versus qualitative predictions from ecological models permeates GF Gause's (1934) insightful but now, unfortunately, rarely read book. I discuss one practical advantage of this aspect of modelling in ecology in section 7.

While both major claims made by Levins (1966) – the generality–realism–precision trade-offs, and robust theorems – have been severely critiqued from a philosophy of science perspective (Orzack and Sober 1993; Orzack 2012), they do continue to have considerable resonance among ecologists. In particular, I believe that the alternative strategies of sacrificing precision versus realism (in the sense of Levins 1966) reflect an important difference between classical physicists and ecologists/evolutionary biologists with regard to the perspectives and priorities of modelling. In my opinion, this difference, which permeates their characteristic disciplinary mindsets, is rooted in the far greater inherent complexity, variability and environmental sensitivity of living systems as compared to physical ones, as discussed in section 6.

Another pertinent issue that Levins (1966) discussed is that of 'sufficient parameters', which is especially relevant in the context of mechanistic versus phenomenological models, both of which are widely used in classical physics and statistical physics, respectively. In ecology and evolution, however, with the limited exception of population genetics, mechanistic models are rare. Mechanistic models, though also abstract simplifications, attempt to map causal relationships among natural entities involved in the phenomenon of interest onto patterns of entailment among component entities of the model in the manner depicted in figure 1. Parameters in such models are often attributes or properties of system components that can be measured. This approach can also be termed as 'modelling by cause'. In phenomenological models, which use a black-box type of approach, the mapping produced in concept space only attempts to mimic outcomes of inputs, and not the causal relationships among the natural entities involved in mediating between inputs and outputs. Consequently, one can term this approach as 'modelling by consequence'. In these kinds of models, parameters are often complex emergent properties at a relatively high hierarchical level, with little one-to-one correspondence to directly measurable attributes of the components of the system at lower levels of the structural hierarchy. These sorts of parameters are what Levins (1966) referred to as 'sufficient parameters', with some widely used examples in ecology and evolution being heritability, fitness, per capita population growth rate, and per capita rate of immigration/emigration. Of course, purely mechanistic or phenomenological models correspond to two ends of a spectrum, rather than embodying a strict dichotomy.

An important attribute of such ‘sufficient parameters’ emphasized by Levins (1966) is that they contribute to imprecision in the models in which they appear, as a direct consequence of their being the result of a many-to-one mapping from simpler parameters at lower levels of the structural hierarchy. In this many-to-one mapping, information is lost. Therefore, such models may need to be supplemented by slightly more mechanistic models at lower hierarchical levels. As Levins (1966) put it, “Thus a satisfactory theory is usually a cluster of models. These models are related to each other in several ways: as coordinate alternative models for the same set of phenomena, they jointly produce robust theorems; as complementary models they can cope with different aspects of the same problem and give complementary as well as overlapping results; as hierarchically arranged ‘nested’ models, each provides an interpretation of the sufficient parameters of the next higher level where they are taken as given.” Another reason for which phenomenological models contribute to overall imprecision in the modelling process is that their ‘sufficient parameters’ often do not correspond to well-defined and directly measurable biological traits and, consequently, can only be estimated by model fitting, a process that has its own attendant uncertainties. This inherent limitation when dealing with ‘complex phenotypes’ permeates most modelling in the broad areas of population and community ecology, and evolutionary biology.

6. Context-specificity of model performance

گُزر جا بن کے سیلِ تُند رو کوہ و بیاباں سے
گُلِستاں راہ میں اۓ تو جُوۓ نغمہ خواں ہو جا
علامہ اِقبال

[*Rush like a torrent through the mountains and the deserts*

Should you encounter a garden, become a softly singing stream

Allama Iqbal]

At one level, given that models are basically tools for trying to address questions about natural phenomena, it is a trivial truism that model performance will be context-dependent. However, there is a large difference in the way this context-specificity of model performance plays out in models of physical versus ecological/evolutionary systems. The inherently greater complexity, variability, and environmental sensitivity of living systems, as compared to most physical ones, ensures that the mapping between reality and concept space in practically all ecological or evolutionary models (except population genetics) is much fuzzier, and far more of a simplifying abstraction, than it is in many physical models. One practical implication of this difference between models in ecology/evolution and physics is that models in ecology often have very changed behaviour, or altered ability to explain observed data, under slightly altered environmental conditions. This particular complication is, to my knowledge, relatively rare in models of physical systems.

A very good example of this kind of context-specificity of model behaviour in ecology was explored in a paper published in this very journal by Somdatta Sinha and one of her colleagues (Sinha and Parthasarathy 1994), and subsequently explained further by Ruxton (1995). Sinha and Parthasarathy (1994) showed that two commonly used simple discrete-time models of ecological population growth – the Ricker and the logistic – that show very similar dynamic behaviour, and a similar period-doubling route leading to chaos, nevertheless behave in extremely different ways if one incorporates a constant number of immigrants or emigrants per generation. Ruxton (1995), in an extension to their work, traced this difference in response to migration to the different shapes of the return maps of the two models at high population size, using an isocline analysis. On the more practical side, it has also been shown that, despite the close similarity in their dynamic behaviour in the absence of immigration or emigration, the logistic model performs quite poorly compared to the Ricker model when fitted to population size data from closed single-vial fruit-fly populations (Sheeba and Joshi 1998).

The Ricker (1954) model ($N_{t+1} = N_t \exp(r(1 - N_t/K))$), though developed in a fisheries context, has been shown to be a good descriptor of insect population dynamics, especially in species with predominantly scramble competition (Brännström and Sumpter 2005). The model assumes that realised per capita population growth rate falls off exponentially with increasing population size, and condenses all biological details into two parameters, r and K, where e^r is the maximal per capita growth rate and K is the equilibrium population size. It has also been shown to capture at least the gross features of the dynamics of many single-species fruit-fly cultures fairly well (Sheeba and Joshi 1998; Dey 2007, 2012; Dey and Joshi 2018). Yet, once examined in detail, the situation is found to be quite nuanced, and the performance of the Ricker model exhibits considerable context-specificity, even though its performance when fitted to fruit-fly population size data is robust to sex-ratio variation (Mueller and Joshi 2000). In studies wherein fruit-fly populations were maintained on different nutritional regimes factorially combining low and high food levels for larvae and adults, respectively, the Ricker model provided good fits to data only from populations maintained on an LH food regime (low food for larvae, high food for adults), whereas it did not fit data from LL, HL or HH food regimes that well (Sheeba and Joshi 1998; Dey 2007). The cause of this discrepancy appears to be the fact that the realized per capita growth rate as a function of population size varies in shape across the food regimes, with the LH food regime providing the closest approximation to an exponentially declining function as assumed by the Ricker model (Dey 2007). In fact, even in the LH food regime, the data suggest a steeper decline than exponential in the realized per capita growth rate as a function of population size. This kind of environmentally induced variation in the functional form of the model exemplifies the limitations of modelling ecological/evolutionary processes with simple models in which many traits with potentially differing environmental sensitivities are collapsed into a small number of parameters that can, consequently, only be estimated by model fitting but not directly measured (for a general discussion of this issue, see Onstad 1988). It was only with the development of substantially more detailed agent-based simulations, incorporating the major density-dependent feedbacks on life-history traits in fruit-fly cultures (Tung *et al*. 2019), that good fits were obtained to multiple data sets from fruit-fly populations with varying selection histories and different nutritional regimes.

7. The usefulness of rules of thumb and stochastic predictions

جُستجُو کُل کی لئے پھِرتی ہے اجزا میں مُجھے
حُسن بے پایاں ہے، دردِ لا دوا رکھتا ہوں میں
علامہ اِقبال

[*The quest for the whole leads me to wander, searching, among the parts*

But the depths of beauty/reality are fathomless, hence I have an incurable pain

Allama Iqbal]

Living systems are characterized by almost ubiquitous variations, in both attributes and environmental sensitivity, at various hierarchical levels of structural organization. Moreover, functionality in living systems is typically integrated across multiple levels of the structural hierarchy. And, as if that were not enough, variation in biological structure and function can be plastic or heritable (and not just via genes), and has both ecological and evolutionary causes and consequences, leading to the situation encapsulated in the pithy adage that what is noise to the physicist is music to the ecologist! This characteristic of living systems has profound effects on the nature of data sets in ecology and evolution and, consequently, on how we can best test predictions from models.

One consequence of this gloriously rampant variability in living systems is the seemingly counter-intuitive conclusion that qualitative 'rules of thumb' kinds of predictions are often more useful in ecology and evolution than more precise point predictions (Gause 1934; Levins 1966; but see also Orzack and Sober 1993 for a discussion of the limitations of qualitative predictions in distinguishing between the importance and sufficiency of selection in optimization models in evolutionary ecology). A good account of some practical consequences of this tension between qualitative and quantitative predictions for modelling in ecology is given by Wood (2001). One important reason why rules of thumb are more easily testable than point predictions in experimental ecology and evolution is that we typically do not have the ability to set up contrasting experimental treatments on a fine enough scale, whereas we can do so on a coarser scale. This problem is exacerbated by the limitation of having to estimate, rather than measure, 'sufficient parameters', as discussed in section 5. Therefore, if our prediction pertains to different outcomes caused by various values of a 'sufficient parameter' or, conversely, to different 'sufficient parameter' values resulting from varying experimental conditions, our ability to test point predictions is severely compromised. For example, in experimental studies in population ecology, even with controlled laboratory environments and well-characterized model systems like fruit-flies, it is not reliably possible to either produce or detect a difference of magnitude 0.2 in intrinsic per capita growth rate (r), even though simple population growth models might predict very different consequences of such a difference in r, or predict such a difference as the outcome of varying experimental treatments. Similarly, in evolution, a point prediction of heritability is essentially useless, given the large uncertainty of any experimental estimation of heritability. Indeed, one reason why a combination of modelling and controlled multi-generation laboratory experiments led to such an impressively detailed and nuanced understanding of the dynamics of spatially structured and unstructured single species populations of fruit-flies (reviewed in Dey and Joshi 2018) was the deployment of rules of thumb linking nutritional regimes to gross aspects of population dynamics (Dey 2007, 2012; Tung *et al*. 2019), itself made possible as a result of the detailed species-specific *Drosophila* model of Mueller (1988).

Another seemingly counter-intuitive consequence, at least to many students, of the complexity of ecological and evolutionary systems, is the relatively greater usefulness of stochastic rather than deterministic models. This is most clearly exemplified by the situation in population genetics. Most researchers get their first, and often only, introduction to natural selection as a major factor driving adaptive evolution through relatively mechanistic one- or two-locus models of allele and genotypic frequency change under selection. Yet, paradoxically, most evolutionary biologists working on adaptive evolution, whether in the laboratory or in the field, hardly ever have occasion to use such population genetic models of selection in their research. The reason is that practically most traits relevant to adaptation in metazoans are complex phenotypes, affected by both large numbers of loci and multiple environmental factors. In the face of such complexity, the deterministic one- or two-locus models simply do not apply, as their simplifying assumptions are violated *en masse*. The population genetic models that actually work reasonably well, and are widely used in actual research in evolution, are stochastic models of genetic change in the absence of selection, as exemplified by the models used in molecular phylogenetics.

In ecology, too, stochastic models are widely used in diverse contexts like biodiversity and community ecology (e.g. Fisher *et al.* 1943; Hubbell 2001), animal movement (e.g. Gupta *et al.* 2017), and population ecology (e.g. Chapman 1967; Tung *et al.* 2019). Stochasticity can be included into simple analytic models as well as individual-based models, and can be operationalized through state variables, parameters, or both. Good general overviews of stochastic formulations of ecological models and their applications can be found in Allen (2003), and Black and McKane (2012).

8. Models may not just facilitate but also constrain our thinking

تُو نے دیکھا سِطوَتِ رفتارِ دریا کا عُروج
موجِ مُضطر کِس طرح بنتی ہے اب زنجیر دیکھ
علامہ اِقبال

[*You have seen the glorious rise of the speeding current of the ocean waves*

Now, see how the turbulent wave becomes the chain around its own feet

Allama Iqbal]

While mathematical models are undoubtedly often very helpful in clarifying our thinking about ecological and evolutionary phenomena (Otto and Day 2007), they can also sometimes end up constraining our thinking because certain biologically possible outcomes may not be accommodated within the framework of the model. This problem tends to be more serious in the case of phenomenological rather than more mechanistic models. A classic example is provided by the field of density-dependent selection. After the first verbal enunciation of the notion that chronic high versus low density would often select for very different traits (Dobzhansky 1950), some formalization of the theory began with the work of MacArthur (1962), and MacArthur and Wilson (1967). Subsequent one-locus two-allele population genetic models of evolutionary change were all based on the discrete-time version of the logistic model of population growth (discussed in detail by Joshi *et al.* 2001). The logistic equation, in its most popularly used form in ecology, models population size in a generation as a quadratic function of population size in the previous generation, with two parameters: r (the maximal per capita population growth rate) and K (the equilibrium population size, also often misleadingly termed the carrying capacity) (Joshi *et al.* 2001). It is also a phenomenological model, in that the parameters r and K are not measurable biological traits; they are higher level (population) attributes that can only be estimated by model fitting. In the logistic model, the only way of substantially increasing realized population growth rate (fitness) at high density, especially in the face of r–K trade-offs, is via an evolutionary increase in K (Joshi *et al.* 2001). The two major ways in which K can increase in a population are through the evolution of either smaller body size or a greater efficiency of food to biomass conversion. It is easier to imagine the evolution of greater efficiency rather than small size, because size also trades off with fecundity, a major fitness component.

It is, perhaps, not surprising that from the late 1960s to the mid-1990s, the dominant view among those studying adaptations to crowding in *Drosophila* was that such adaptations would involve the evolution of greater efficiency of food-to-biomass conversion. So widespread was this view that it took the demonstration of the evolution of competitive ability via increased larval feeding rate, at the cost of efficiency, in two separate sets of crowding-adapted populations of *D. melanogaster* (Mueller 1990; Joshi and Mueller 1996) for it to begin to change. Moreover, this was despite the fact that a detailed and mechanistic species-specific model of *Drosophila* population dynamics and evolution under crowding had already shown that a greater *K*, achieved via increased efficiency, was not the only way in which populations, wherein larvae competed for limiting food resources, could evolve to become more competitive (Mueller 1988). In the *Drosophila* model, competitive ability could also increase through the evolution of greater rates of food acquisition by larvae under conditions of overall food limitation, without any change in *K* (Mueller 1988), contrary to the limitation inherent in the phenomenological logistic model in which increased *K* was necessary for greater competitive ability.

This kind of situation, where a model (verbal or mathematical) becomes a constraint on our thinking by limiting us to considering what is possible in the model, rather than in nature, is quite common in ecology and evolution. For decades, the widespread, almost axiomatic, acceptance of the competitive exclusion principle meant that ecologists did not consider the possibility of competitor coexistence seriously, as the principle implied that competitors would either have to diverge so as to cease being in competition, or else, one would drive the other extinct (reviewed by Arthur 1982). Yet, the competitive exclusion principle arises from a limitation of the Lotka–Volterra model of competition, based on the continuous time logistic equation, rather than a biological constraint upon competing species (Joshi 2001). Once again, it took quite some time after models different from the Lotka–Volterra formulations indicated that reasonably long-term competitor coexistence was likely in nature, for the notion of possible competitor coevolution to become broadly accepted (Arthur 1982; Joshi and Thompson 1995, 1996).

There are many such examples in ecology and evolutionary biology in which wide acceptance of a particular model ends up blinding researchers to the possibility of alternative mechanisms, not possible under the constraints of the dominant model, that might be enabling organisms to adapt to certain ecological challenges. It is in this context that Michael Rose, a pioneer of experimental evolution with fruit-flies, has often remarked during talks that "the moral of the story is that *Drosophila* do not read papers in the *Journal of Theoretical Biology*!" Consequently, while it is important to be familiar with the models commonly used in one's area of work, it is also sometimes very helpful to go back to one's understanding of the biology of the system and reason things out *de novo*, as though the previous models did not exist.

9. Agent-based simulations, though useful, are not really models

کاری پاکان را قیاس از خود مگیر
گر چه مانَد در نبِشتن شیر و شیر
مولانا جلال الدین رومی

[*Do not presume to hold the actions of the pious as being similar to yours*

For even the lion and milk, though so different, stay the same in writing

Maulana Jalaluddin Rumi

Note: in Persian, both lion (sher) and milk (shīr) are written identically]

In recent decades, simulations of agent-based 'models' to investigate questions in ecology and evolution, and also in the social sciences, have become increasingly popular (e.g. Railsback and Grimm 2011; Salamon 2011; Wilensky and Rand 2015). In such 'models', typically called individual-based models (IBM) in ecology, large-scale computer simulations are used to explicitly model the actions and interactions of many lower-level constituents of a complex, hierarchical system, in an attempt to address how emergent properties and behaviours at higher hierarchical levels arise as a consequence of what is happening at lower levels in the structural hierarchy. Such an approach, at least in principle, offers the prospect of linking lower-level mechanistic models to higher-level phenomenological ones, thus adding detail and nuance to our insight into emergent processes in complex systems. In my own research area, an agent-based 'model' of *Drosophila* population dynamics has been very successfully used to understand how the dynamics and stability of laboratory cultures of fruit-flies are shaped by complex interactions among resource availability, demography, and life-history traits (Tung *et al.* 2019). This model is also the first one to provide very good fits to data from fruit-fly populations with differing evolutionary histories, as well as from those subjected to different nutritional regimes.

Rather than discussing more examples of the usefulness of the agent-based approach in ecology and evolution, I want to draw attention to the fact that these ‘models’ are fundamentally very different from the sense in which I have been using the term here. I have been using the term ‘model’ to refer almost exclusively to less detailed, often equation-based, models, even if the analysis of the behaviour of some of those models requires computer simulations. Agent-based simulations are, in many ways, closer to experiments, under even more controlled conditions than possible in the laboratory, rather than models, leading them to often be referred to as ‘*in silico* experiments’. However, the point I wish to stress is that, especially in the fields of ecology and evolution, with their preponderance of relatively phenomenological models, the agent-based simulations approach has great potential to allow us to link interactions between measurable phenotypes to their reflection in the behaviours of the ‘sufficient parameters’ in response to different conditions, thereby shedding some light on the hitherto unseen innards of the black boxes of our phenomenological models. Such illuminating insight into what drives the behaviour of ‘sufficient parameters’ can, however, also be attained by more classical modelling, as exemplified by recent attempts to model the evolutionary process via births and deaths rather than relying upon composite measures like fitness (Doebeli *et al.* 2017). Given that very large agent-based simulations can sometimes approximate photographs rather than models, incorporating and yielding so much fine-grained detail that it can obscure a simpler mechanistic understanding of the emergent properties, the best progress might be made by a judicious mix of both relatively simple and complex agent-based simulations, together with relatively more classical modelling ‘bridges’ between lower- and higher-level phenomena in ecological and evolutionary systems.

In one practical respect, agent-based simulations are also very helpful for experimental ecologists and evolutionists, such as myself, in that they permit the screening of large numbers of experimental scenarios of interest in a reasonably short time. Multigeneration experiments in ecology and evolution are extremely labour-intensive, and last for months or years, making it very difficult to do many different such experiments. Given the investment of time and labour in such studies, it is very helpful to be able to determine in advance, using agent-based simulations, which specific investigations among a set of potential experiments are most likely to yield useful insights, rather than ruefully realizing, after a few years of hard work, that not much insight was gained from the experiment after all.

10. Modelling should not be thought of as a domain expertise

تُمہاری تہذیب اپنے خنجر سے آپ ہی خودکُشی کریگی
جو شاخِ نازُک پے آشِیانہ بنیگا نا پایدار ہوگا
علامہ اِقبال

[*Your culture and progress will end up killing itself with its own dagger*

For a nest built on a shaky branch cannot persist for a long time

Allama Iqbal]

One of the most pernicious fallacies that has gripped the imagination of many in science today is that modelling is a primary domain expertise. In other words, that there is a category of accomplished people called modellers, who are capable of competently and usefully modelling phenomena ranging from physics or engineering to biology. This notion is both incorrect and harmful, and arises from a dangerous conflation of the distinct activities of model formulation and model analysis. To put it another way, the crux of this fallacy is that it assumes that the essence of modelling is the analysis of the model and the consequent deduction of predictions from it. This is both ironic and sad, as it privileges the essentially technical over the truly scientific. A similarly fallacious and unfortunate elevation of an approach to analysis to the perceived status of a domain expertise, rather than a context-specific method to be applied to problems in particular domains, is presently also taking place with regard to data science. It is likely to lead to the same sorts of problems that we are already experiencing as a result of modelling being thought of as a domain expertise in itself.

As we saw in section 2, modelling is (a) a mapping from reality space to concept space, and (b) subjective, to a considerable degree. That is why experience of empirically studying the system one is attempting to model becomes so important: the informed intuition about which details are important and which can be ignored, acquired through a detailed study of a specific system, can only be complemented, and not substituted, by a facility in manipulating and solving differential equations or other such mathematical entities. This problem is of particular and practical significance in India, where ecological/evolutionary modelling, including in epidemiology, is largely done by physicists or engineers, many of whom do not have even a passing acquaintance with the systems they model with aplomb. When applied to problems in academic ecology/evolution, such an approach is not harmful per se: theoreticians can enjoy themselves exploring the properties of mathematical objects that they think represent ecological phenomena, a certain genre of theory journals get papers to publish, and ecologists and evolutionists, with a happy equanimity, just ignore such studies. However, when this kind of modelling approach spills over from academic ecology/evolution to applied areas like epidemiology (e.g. Agrawal *et al.* 2021), it is often accompanied by a distinct possibility of undesirable practical consequences.

Why exactly we persist with the fallacy of modelling being a primary domain expertise remains a bit of a mystery to me, although I suspect it is part of the insidious seepage of notions from the discipline of management to other endeavours like research, health care, education, and journalism, leading in each case, to very harmful effects. It is a tenet of contemporary management thinking that specific knowledge of, and expertise in, manufacturing a product or providing a service is not very necessary for running a successful business enterprise, as compared to fancy marketing and glib sales pitches. Incidentally, this is fallacious, even in the context of business management, as most experienced business persons agree. However, it is easy to imagine this logic entering into research and resulting in the elevation of modelling or data science to the level of primary domain expertise, while downplaying the importance of actually understanding the systems or data sets in question.

Even a cursory look at the history of modelling in ecology and evolution suggests very strongly that the most useful and insightful models have come from researchers who either had a solid grounding in biology and acquired the necessary mathematical skills (e.g. Sewall Wright, Motoo Kimura, and Alexander Nicholson) or had a basic training in the mathematical sciences and acquired the necessary biological knowledge and insight (e.g. Ronald Fisher, John Maynard Smith, and Robert May). Modelling in ecology and evolution, carried out with a degree of detached disdain for the messy complexities of living systems, does not usually add much to our understanding in those fields, though it may be pleasurable for those indulging in it, and may even suffice to support a 'successful' academic career.

11. Conclusions

بس اب سو جاؤ، نیند آنکھوں میں ہے، کل پھِر سُنایں گے
ذرا سی رہ گیی ہے رات، افسانے بہت سے ہیں
قمر جلالوی

[*It is time to sleep, the eyes are drowsy, we will continue tomorrow*

The night is almost gone, and there is still many a tale to tell

Qamar Jalalvi]

In this paper, I have discussed certain aspects of models and the modelling process, and touched upon some of the ways in which modelling in ecology and evolutionary biology substantially differs from modelling in much of physics. These differences are primarily a consequence of the far greater complexity, variability and environmental sensitivity of biological systems, as compared to physical ones. My motivation for writing this piece stems from two complementary frustrations that have grown over the years. On the one hand, I keep encountering fellow researchers in ecology and evolution who are a bit overwhelmed by models in our field and tend to avoid close encounters with them once they have passed their required courses, or parts thereof, in ecological and evolutionary theory. This saddens me for two reasons. The first is that empirical research not substantially informed by theory often remains descriptive and, while providing information about some specific system, never really goes beyond that to contribute to the growth of knowledge in the field. The second, in some ways more important to me, is that if ecologists and evolutionary biologists abdicate all responsibility for modelling to those with the requisite mathematical, but not biological, training, we will miss out on all kinds of interesting and useful theory. This is because the choice of questions for theoretical investigation is often very different depending upon one's background and training. Many theoretical questions of interest to ecologists and evolutionary biologists may be mathematically relatively unchallenging but, nevertheless, of considerable relevance to practitioners in the field. Such questions often will not get taken up at all by theoreticians who do not have a close engagement with empirical research in that area. I recall that GH Hardy, embarrassed that his model was mathematically trivial, had to be persuaded very hard by Reginald Punnett to publish his insight that the mechanism of Mendelian genetics, by itself, would not tend to alter the genetic composition of a large, random-mating population.

On the other hand, mirroring my frustration with fellow ecologists and evolutionary biologists who shy away from engaging with theory, is my frustration with colleagues trained in the physical sciences who engage in modelling ecological and evolutionary processes without seriously coming to grips with the empirical reality of the systems themselves. This often leads to theoretical work that is either addressing questions that are not of much interest to empirical researchers in the field, or hammering problems in ecology or evolution into a form that is amenable to analysis by whichever class of models the researcher is comfortable analyzing. This state of affairs is especially frustrating given the fundamental and profound contributions made to our understanding of ecological and evolutionary systems by those whose primary training was in mathematics or physics but who took the effort to also engage with the biological systems they modelled, such as my personal intellectual heroes like Ronald Fisher, Robert MacArthur, and Robert May.

In addition to the issues raised in the nine sections echoing the title, I would like to briefly touch upon one more aspect of models in general, especially for the benefit of students reading this piece. It is that models are not judged by whether they are 'true' (i.e., accurate mimics of reality) but by whether they are 'useful', reflecting the fact that science, although preferably done with great idealism, is essentially a pragmatic activity, even when it is of academic interest and not explicitly applied to solving real life problems. As Levins (1966) has beautifully said, "The validation of a model is not that it is 'true' but that it generates good testable hypotheses relevant to important problems. A model may be discarded in favor of a more powerful one, but it usually is simply outgrown when the live issues are not any longer those for which it was designed." The same sentiment was often expressed by the statistician George Box (1976, and subsequently) as "all models are wrong, but some are useful." Models are invariably 'wrong' in the sense of not being accurate representations of reality, but can be 'right' in the sense of allowing us to empirically test our assumptions about real world processes or phenomena by testing the consequences we draw from our assumptions through modelling (Enderling and Wolkenhauer 2021). Another relevant corollary to this 'true' versus 'useful' distinction is that it behoves us to be very clear that a model is useful for understanding some aspect of reality but, ultimately, should not be conflated with that aspect of reality: "the best material model for a cat is another, or preferably the same cat" (Rosenblueth and Wiener 1945). In general, it is important as a scientist to not just focus on doing science but also be cognizant about issues in the philosophy of science. The same is true for modelling. There is a fairly extensive literature on the philosophy of modelling (e.g. Orzack 2012 and references therein), which many of us who make and work with models should acquaint ourselves with, but often do not, to our own loss.

Finally, I would like to share with readers a little gem of a story from the canons of Zen Buddhism that I first read when I was a beginning graduate student, fully (over)-confident – with what Agatha Christie memorably termed 'the arrogance of youth' – that with mathematical models of ecological and evolutionary processes, we could, in fact, perfectly explain most everything of interest in organismal biology. The story describes the reply of a Zen master when asked by novice monks what his life-long and deep study of Zen has revealed to him. The master says, "When I was just beginning to study Zen, the trees were just trees, the river was just a river, and the mountains were just mountains." "Then," he continues, "as I began to delve deeper into Zen, the trees became more than just trees, the river became more than just a river, and the mountains became more than just mountains." Finally, he says, "Then, once I immersed myself deeply in Zen, the trees went back to just being trees, the river went back to just being a river, and the mountains went back to just being mountains." Over three decades have gone by since my early graduate student days, and I have now come to realize that this story also encapsulates my changing appreciation of the role of mathematical models in ecology and evolution rather well. After a brief period of glorious omnipotence, in my youthful thinking, the models have gone back to being just what they were: mappings from reality space to concept space. Nothing more but, equally importantly, nothing less either!

Acknowledgements

My views on mathematical modelling have partly been shaped over the years via interesting discussions with a number of people. For these, I thank Prabhakar G Vaidya, Michael E Moody, Laurence D Mueller, Somdatta Sinha, Sutirth Dey, TNC Vidya, NG Prasad, and my father, Devi Datt Joshi. I also thank TNC Vidya and an anonymous reviewer for very helpful comments on the manuscript. The preparation of this manuscript was supported in part by funds from a JC Bose National Fellowship from SERB, India.

References

Agrawal M, Kanitkar M and Vidyasagar M 2021 SUTRA: a novel approach to modelling pandemics with applications to Covid-19. *ArXiv arXiv:2101.09158*

Allen LJS 2003 *An introduction to stochastic processes with applications to biology* (Upper Saddle River: Pearson/Prentice Hall)

Arthur W 1982 The evolutionary consequences of interspecific competition. *Adv. Ecol. Res.* **12** 127–187

Black AJ and McKane AJ 2012 Stochastic formulation of ecological models and their applications. *Trends Ecol. Evol.* **27** 337–345

Box GEP 1976 Science and stastistics. *J. Am. Stat. Assoc.* **71** 791–799

Brännström Å and Sumpter DJT 2005 The role of competition and clustering in population dynamics. *Proc. R. Soc. Lond. B* **272** 2065–2072

Chapman DG 1967 Stochastic models in animal population ecology; in *Proceedings of the fifth Berkeley Symposium on mathematical statistics and probability*, *Vol. 4: Biology and problems of health* (eds) LM Le Cam and J Neyman (Berkeley: University of California Press) pp. 147–162

Doebeli M, Ispolatov Y and Simon B 2017 Point of view: towards a mechanistic foundation of evolutionary theory. *eLife* **6** e23804

Dey S 2007 *Experimental and theoretical investigation of the dynamics and stability of single populations and metapopulations of* Drosophila melanogaster *in the laboratory*. PhD thesis, Jawaharlal Nehru Centre for Advanced Scientific Research, Bengaluru

Dey S 2012 *Coherence, dynamics and stability in spatially structured and unstructured populations: effects of dispersal and crowding*. PhD thesis. Jawaharlal Nehru Centre for Advanced Scientific Research, Bengaluru

Dey S and Joshi A 2018 Two decades of *Drosophila* population dynamics: modelling, experiments, and implications; in *Hand book of statistics, Vol. 39: Integrated population biology and modelling, Part A* (eds) CR Rao and ASR Srinivasa Rao) (Amsterdam and Oxford: Elsevier) pp. 275–312

Dobzhansky T 1950 Evolution in the tropics. *Am. Sci.* **38** 209–221

Enderling H and Wolkenhauer O 2021 Are all models wrong? *Comput. Syst. Oncol.* **1** e1008

Fisher RA, Corbet AS and Williams CB 1943 The relation between the number of species and the number of individuals in a random sample from an animal population. *J. Anim. Ecol.* **12** 42–58

Gadagkar R 2015 How should biologists engage with controversial mathematical theory? *Curr. Sci.* **108** 1869–1873

Gause GF 1934 *The struggle for existence* (Baltimore: Williams and Wilking)

Gompertz B 1825 On the nature of the function expressive of the law of human mortality and on a new mode of determining life contingencies. *Phil. Trans. R. Soc. Lond. A* **115** 513–585

Gunawardena J 2014 Models in biology: 'accurate descriptions of our pathetic thinking.' *BMC Biol.* **12** 29

Gupta M, Joshi A and Vidya TNC 2017 Effects of social organization, trap arrangement and density, sampling scale, and population density on bias in population size estimation using some common mark-recapture estimators. *PLoS One* **12** e0173609

Hartl DL Clark AG 1989 *Principles of population genetics*, 2nd Edn (Sunderland: Sinauer)

Hubbell SP 2001 *The unified neutral theory of biodiversity and biogeography* (Princeton: Princeton University Press)

Joshi A 2001 Interspecific competition; in *Encyclopaedia of life sciences* (London: John Wiley & Sons) doi: https://doi.org/10.1038/npg.els.0003286

Joshi A 2020 The Price equation and the mathematics of selection. *Resonance* **25** 495–512

Joshi A and Moody ME 1995 Male gamete output of asexuals and the dynamics of populations polymorphic for reproductive mode. *J. Theor. Biol.* **174** 189–197

Joshi A and Moody ME 1998 The cost of sex revisited: effects of male gamete output of hermaphrodites that are asexual in their female capacity. *J. Theor. Biol.* **195** 533–542

Joshi A and Thompson JN 1995 Alternative routes to the evolution of competitive ability in two competing species of *Drosophila*. *Evolution* **49** 616–625

Joshi A and Thompson JN 1996 Evolution of broad and specific competitive ability in novel versus familiar environments in *Drosophila* species. *Evolution* **50** 188–194

Joshi A and Mueller LD 1996 Density-dependent natural selection in *Drosophila*, trade-offs between larval food acquisition and utilization. *Evol. Ecol.* **10** 463–474

Joshi A, Prasad NG and Shakarad M 2001 *K*-selection, α-selection, effectiveness, and tolerance in competition: density-dependent selection revisited. *J. Genet.* **80** 63–75

Levins R 1966 The strategy of model building in population biology. *Am. Sci.* **54** 421–431

MacArthur RH 1962 Some generalized theorems of natural selection. *Proc. Natl. Acad. Sci. USA* **48** 1893–1897

MacArthur RH and Wilson EO 1967 *The theory of island biogeography* (Princeton: Princeton University Press)

Maynard Smith J 1968 Evolution in sexual and asexual populations. *Am. Nat.* **102** 469–473

Mazarati A 2007 The best model for a cat is the same cat...or is it? *Epilep. Curr.* **7** 112–114

Mueller LD 1988 Density-dependent population growth and natural selection in food-limited environments: the *Drosophila* model. *Am. Nat.* **132** 786–809

Mueller LD 1990 Density-dependent natural selection does not increase efficiency. *Evol. Ecol.* **4** 290–297

Mueller LD and Joshi A 2000 *Stability in model populations* (Princeton: Princeton University Press)

Mueller LD, Nusbaum TJ and Rose MR 1995 The Gompertz model as a predictive tool in demography. *Exp. Gerontol.* **30** 553–569

Odenbaugh J 2006 The strategy of 'The strategy of model building in population biology.' *Biol. Philos.* **21** 607–621

Onstad DW 1988 Population-dynamics theory: the roles of analytical, simulation, and supercomputer models. *Ecol. Modell.* **43** 111–124

Orzack SH 2012 The philosophy of modelling or does the philosophy of biology have any use? *Philos. Trans. R. Soc. Lond. B* **367** 170–180

Orzack SH and Sober E 1993 A critical-assessment of Levins' 'The strategy of model-building in population biology' (1966). *Q. Rev. Biol.* **68** 533–546

Otto SP and Day T 2007 *A biologist's guide to mathematical modelling in ecology and evolution* (Princeton: Princeton University Press)

Price GR 1970 Selection and covariance. *Nature* **227** 520–521

Railsback SF and Grimm V 2011 *Agent-based and individual-based modelling: a practical introduction* (Princeton: Princeton University Press)

Rice SH 2004 *Evolutionary theory: mathematical and conceptual foundations* (Sunderland: Sinauer)

Ricker WE 1954 Stock and recruitment. *J. Fish. Res. Board Can.* **11** 559–623

Rosen R 1991 *Life itself: a comprehensive inquiry into the nature, origin and fabrication of life* (New York: Columbia University Press)

Rosenblueth A and Wiener N 1945 The role of models in science. *Philos. Sci.* **12** 316–321

Ruxton GD 1995 The effect of emigration and immigration on the dynamics of a discrete-generation population. *J. Biosci.* **20** 397–407

Šalamon T 2011 *Design of agent-based models* (Repin: Bruckner Tomas)

Sheeba V and Joshi A 1998 A test of simple models of population growth using data from very small populations of *Drosophila melanogaster. Curr. Sci.* **75** 1406–1410

Sinha S and Parthasarathy S 1994 Behaviour of simple population models under ecological processes. *J. Biosci.* **19** 247–254

Tung S, Rajamani M, Joshi A and Dey S 2019 Complex interaction of resource availability, life-history and demography determines the dynamics and stability of stage-structured populations. *J. Theor. Biol.* **460** 1–12

Vidya TNC, Dey S, Prasad NG and Joshi A 2022 The Darwinian core of evolutionary theory and the Extended Evolutionary Synthesis: similarities and differences; in *Evolutionary biology: contemporary and historical reflections upon core theory* (eds) TE Dickins and BJA Dickins (Berlin: Springer) (in Press)

Wilensky U R and W 2015 *An introduction to agent-based modeling: modeling natural, social, and engineered complex systems with NetLogo* (Cambridge: MIT Press)

Williams GC 1975 *Sex and evolution* (Princeton: Princeton University Press)

Wood SN 2001 Partially specified ecological models. *Ecol. Monogr.* **71** 1–25

Corresponding editor: SUSMITA ROY