
An Invitation to Stochastic Mathematical Biology



George Constable, Yvonne Krumbeck, and Tim Rogers

Mathematical biology is unusual as a discipline in being defined by its application area, rather than by the type of mathematics involved. We are now in the era of mass data gathering and cheap computation, revealing the messy details of biological systems with unprecedented clarity. Mathematical biologists are increasingly required to confront the essential randomness and uncertainty of the natural world; a challenge that brings with it the opportunities to ask and answer new questions, sometimes providing surprisingly simple explanations for complex natural phenomena. In this article we survey the growing field of stochastic mathematical biology, revealing the surprising explanatory power of randomness in biology, and highlighting developments and challenges across a spectrum of mathematical areas.

George Constable is a lecturer of mathematics at the University of York. His email address is george.constable@york.ac.uk.

Yvonne Krumbeck is a doctoral student at the University of Bath. Her email address is y.krumbeck@bath.ac.uk.

Tim Rogers is a professor of mathematics at the University of Bath. His email address is t.c.rogers@bath.ac.uk.

Communicated by Notices Associate Editor Emilie Purvine.

*For permission to reprint this article, please contact:
reprint-permission@ams.org.*

DOI: <https://doi.org/10.1090/noti2381>

Three Good Reasons

Modelling biological systems can be tricky. Feedback loops operating across vastly disparate spatial and temporal scales give rise to the huge richness and complexity of the living world. The process of abstracting from reality to a model involves removing much of this intricate detail. The more detail that has been discarded, the more the model is said to be “simple.” The more simple a model is, the better chance we have of analysing it with mathematics to make predictions, gain deep understanding, and derive general principles. Mathematical biologists face the challenge of constructing models that are simple enough to be amenable to analysis, but that still retain the essential features of the biological system under study.

There is one particularly powerful abstraction that is performed routinely and often subconsciously: we replace quantities with our expectation of those quantities, often without regard for the error. This has entered the public consciousness recently with attention on the basic reproductive number of COVID-19. Any single transmission of the coronavirus is governed by a sequence of events across vastly different scales (subcellular to intercontinental) which is so complicated as to be completely unpredictable. Yet these personal stories are aggregated at a population level to produce a single number quantifying the

rate of new infections and thus the exponential growth of the disease. When mathematical models are used to inform decision-making affecting lives and livelihoods, we face a challenging task to publicly justify the chosen models and to explain the uncertainties in our predictions.

By relaxing the assumption of perfect predictability and embracing stochastic thinking, mathematical biology can better model biological systems, can expand the breadth and sophistication of predictions made, and can expose the counterintuitive ways that randomness brings emergent order to the natural world. These are three good reasons to study stochastic mathematical biology which we explore here.

First, all biological systems are inherently stochastic.¹ Plant and animal cells are complicated and crowded environments, with myriad competing and complementary processes and events taking place, where precise measurements are difficult to make. Births and deaths of organisms, which ultimately drive the dynamics of whole populations and the evolution of species, are unpredictable in their timing, causes, and consequences. Interactions between species determine the fate of entire ecosystems, but they are both notoriously difficult to quantify and contingent on unpredictable environmental factors. Empirical scientists deal with these uncertainties using statistics. For mathematical modellers, the correct proxy is the use of random variables to replace unknowns. In this way uncertainty can be carried forward through the model in a quantifiable way.

Second, stochastic models offer the possibility of interrogating a rich set of questions including: (i) *what is the distribution of possible outcomes or future states?*; (ii) *what is the probability of transitioning between two distant states?*; and (iii) *how long must we wait for a certain event to occur?* In a biological context, the answers can have particular significance, such as: (i) *how many species will be supported in an ecosystem?*; (ii) *what is the probability that cancerous epithelial cells become migratory?*; or (iii) *over what time frame do we expect a mutation to occur in a particular gene?* A great deal of the power of stochastic mathematical biology comes from the ability to formulate and answer such questions.

Third, we note that the inclusion of stochastic effects in simple models can often enhance their explanatory power. In what follows, we showcase several examples from a range of different biological applications. We will take care to explore the different forms of stochasticity that may appear in a model. One important example is demographic noise: randomness in the timing of individual events that can play out at the population level to dramatically affect

system behaviour. Another is randomness in model parameters, in which incomplete information about the environment or interactions within a system give rise to an ensemble of potential realisations whose statistics must be calculated.

Many different areas of mathematics have important contributions to make in this growing field; the remainder of this paper is organised by mathematical discipline to emphasise this point, and we pose open questions in each. A few key recent references are highlighted throughout; a more comprehensive reading list for interested researchers is available online [KCR21]. We conclude by discussing further exciting directions for study, including developments that may come with the synthesis of different forms of stochasticity within the same model.

Probability

The construction and analysis of probabilistic models of populations was pioneered around the turn of the 20th century by early proponents of theoretical population genetics, who were keen to understand the factors driving changes in gene-frequencies.² It is perhaps no surprise that many of these researchers were in fact statisticians, including Ronald Fisher and Sewall Wright, although other mathematicians such as the self-affirmed number theorist G. H. Hardy were also influential [KCR21]. As a result, the importance of demographic stochasticity (though not named in these terms at the time) was appreciated in theoretical population genetics from its inception. It took decidedly longer for these ideas to filter out into the realms of ecology and biochemistry.

The most simple probabilistic population models are *branching processes*, which keep track of the family tree of a growing population. In the famous Galton-Watson process, the number of individuals in generation $n + 1$ is given by a sum of independent and identically distributed random variables, one for each member of generation n . Nearly 150 years from its inception, probabilists are still finding interesting things to say about this model, which is a cornerstone of applied probability theory. The analysis of more complex branching processes (for example involving multiple types, or changing environments) gives important insights into the behaviour of biological growth of various kinds, including the early development of disease outbreaks [BD95], or metastatic cancer [YJB⁺10].

The main strategy in the analysis of branching processes is to exploit the independence of different branches of the tree. This is also the main limitation, as the

¹From the Greek *stokhastikos* meaning “guessable,” we use the word *stochastic* to encompass randomness in parameter values, unknown initial data, the timing of and outcome of events, the predictions made by models, and any other element of uncertainty.

²Sadly, some of this early work is linked to the racist and ableist eugenics movement. Institutions (arguably including mathematics as a whole) whose reputations have long benefitted by association with the groundbreaking theoretical contributions made at the time now face an awkward reckoning with the ugly historical context.

theories developed diverge from reality at the point that interaction between branches becomes important. Possible examples of such interactions include competition for limited resources in the growth of populations, or self-reinforcement leading to choice-based swarming behaviour (e.g., foraging in ant colonies).

Challenge: Develop tools to compute the long-time behaviour of locally and non-locally interacting branching processes.

Statistical Mechanics

When populations are naturally limited to a known size, a different probabilistic construction is preferred. Suppose one is interested in a population of discrete “individuals” of one or more types, represented by a “state vector” of integers \mathbf{n} , encoding the number of individuals of each type at a given moment. These “types” could be species abundances [ZGG⁺19] (e.g., the number of foxes and rabbits), genotypes [SST20] (i.e., the number of individuals carrying a specific genetic variant), protein molecules in a cell [JG20] (which can exist at very low numbers), or any number of other biological populations of interest. We could then consider the time-evolution of the probability $P_{\mathbf{n}}(t)$ of being in some state \mathbf{n} at time t :

$$\frac{dP_{\mathbf{n}}(t)}{dt} = \sum_{\mathbf{n}' \neq \mathbf{n}} [T(\mathbf{n}|\mathbf{n}')P_{\mathbf{n}'}(t) - T(\mathbf{n}'|\mathbf{n})P_{\mathbf{n}}(t)], \quad (1)$$

where $T(\mathbf{n}'|\mathbf{n})$ are transition rates that govern the probability per unit time of transitioning from a state \mathbf{n}' to a state \mathbf{n} . Although equation (1) (known as Kolmogorov’s forward equation, or the master equation) describes a continuous time process on a discrete state space, discrete time and continuous state space descriptions can be set up analogously. Models of this type are often referred to as Individual Based Models (IBMs) as they take account of the interactions between discrete individuals. They are also called stochastic compartmental models when considering populations divided into compartments (e.g., representing the stages of infection in an epidemic) with random movements between them.

Solutions to the above problems often come down to solving a large set of simultaneous equations across the different values of \mathbf{n} . For instance, the stationary probability distribution can be obtained from taking the limit $t \rightarrow \infty$ in (1), and tells us about the equilibrium distribution that the system would reach at long times (e.g., the probability of observing a given number of species on an island once migrations to and from a mainland have been allowed to equilibrate). Analogous (although more involved) expressions exist for hitting probabilities and hitting times, which respectively tell us about the probability of reaching some threshold state (e.g., passing a critical

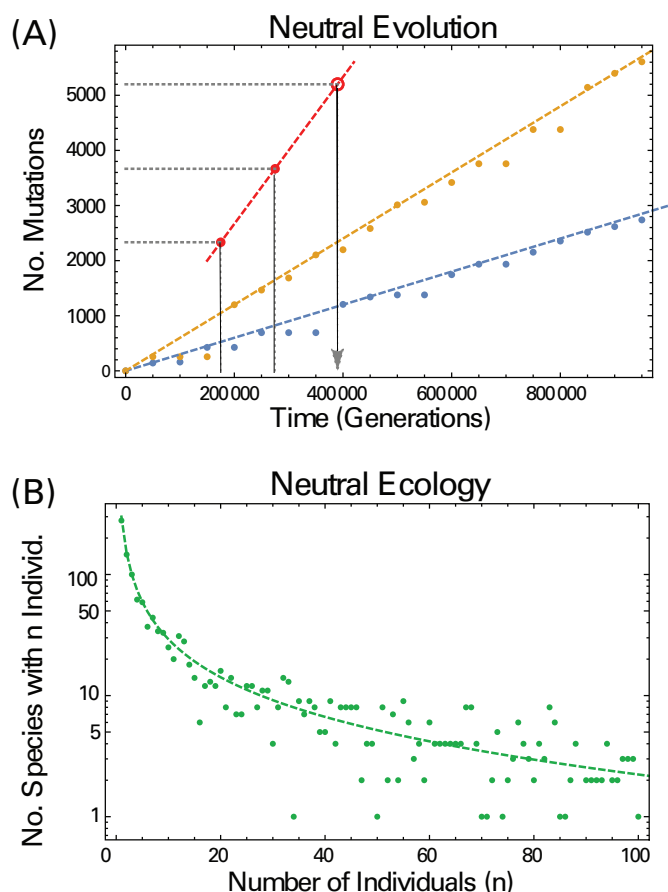


Figure 1. Top: Kimura’s neutral theory of evolution predicts that neutral mutations (see equation (2)) should accumulate at a fixed rate (see theoretical blue and orange dashed lines and corresponding blue and orange markers from stochastic simulations). This linear pattern has been observed across scales from animal haemoglobin to influenza viruses. In red, we show how this predictability can be used to construct a “molecular clock”: if the number of substitutions for two species (red disks) is known, along with the time since their divergence (e.g., from the fossil record), then an inferred substitution rate (red dashed line) can be used to estimate the divergence time for a third related species with known genetics (red circle). Such methods have recently been used to estimate the divergence time of human coronavirus SARS-CoV-2 from animal viruses. **Bottom: Hubbell’s neutral theory of biodiversity** predicts the distribution of species abundances (see theoretical green lines and corresponding markers from stochastic simulations). Multiple studies have used such models to understand the mechanisms underlying ecological diversity, from marine phytoplankton to Amazonian tree communities, to the human gut microbiome.

number of infections in an epidemic model) or the time taken to get there (e.g., for a resistant strain of bacteria to displace its ancestors). As a consequence, for general systems the best that may be hoped for is a set of recurrence relations that can be evaluated numerically. However in one-dimensional systems (in which the vector \mathbf{n} is replaced by a single abundance n) closed-form solutions

are often more forthcoming. This has led to a particular focus on models consisting of two classes of individuals (e.g., cooperators and defectors in a model of social conflict) within a population of fixed size, N (i.e., where the number of individuals of the two classes is given by n and $N - n$).

While closed-form solutions for these quantities of interest are difficult to obtain for general multivariate systems, a particular class of these systems, neutral models, are often more amenable to analysis. These neutral models assume that all the classes in the population behave in an equivalent manner under exchange. For instance, in a neutral model of population genetics, the probability transition rate from a state $\mathbf{n}' = (\dots, n_i, \dots, n_j, \dots)$ to a state $\mathbf{n}'' = (\dots, n_i + 1, \dots, n_j - 1, \dots)$ (i.e., the probability of a birth-death event that increases the number of a gene variant i and decreases that of variant j) is given simply by the product of their frequencies:

$$T(\mathbf{n}|\mathbf{n}') = \frac{n_i n_j}{N N}, \quad (2)$$

where $N = \sum_k n_k$ is a constant. This enforced symmetry increases the ease with which such multivariate problems can be handled.

Far from being simply a concession to analytic tractability, such neutral models have been instrumental in advancing our understanding of emergent biological patterns in the natural world. In the late 1960s Motoo Kimura, building on the early work of Fisher and Wright, demonstrated that just such a model as that illustrated in equation (2) was capable of explaining the “molecular clock” in evolution, whereby mutational differences between distinct lineages accumulate linearly with time (see Figure 1). Three decades later, Stephen Hubbell took a very similar approach to explain the log-normal distribution of species abundances recurrent in ecological data (see Figure 1). While debate still rages about the appropriate messages to take home from neutral models, whether they are indicative of a true underlying absence of structure or simply serve as null hypotheses for more complex theories, they have continued to provide insights in a broad range of fields, including the evolution of language, ancestral “sexes,” and even internet memes [KCR21].

Challenge: Stepping up significantly in complexity from IBMs are so-called Agent-Based Models, composed of interacting agents with complicated internal logic. Work on these models is almost entirely simulation-driven, leaving open important questions about underlying *reasons* for observed behaviour. To what extent can statistical mechanics methods be applied here?

Stochastic Dynamical Systems

While probabilistic modelling flourished in the early population genetics community, contemporary ecological, epidemiological, and physiological theory is rooted in the analysis of differential equations. Amongst some of the most famous successes of this work involve the emergence of patterns in time, as well as in space (see the section, PDEs and Related Fields, below). Such results can loosely be thought of as modern stochastic analogues of results such as deterministic Lotka-Volterra predator-prey cycling (which was thought to explain empirically observed temporal oscillations in predator-prey abundances³). Recently, researchers in a field broadly described as *applied stochastic analysis* have been revisiting the foundational assumptions of these models, and finding that a proper treatment of demographic noise reveals surprising new behaviours.

From a Kolmogorov’s forward equation of the form of equation (1), various techniques exist to extract a stochastic differential equation (SDE) representative of the behaviour for large but finite population size. The typical form is

$$\frac{dx_i}{d\tau} = f_i(\mathbf{x}) + \frac{1}{\sqrt{N}}\eta_i(\tau), \quad i = 1, 2, \dots, S, \quad (3)$$

where the vector field $f(\mathbf{x})$ can be crudely understood as controlling the “deterministic component” of the dynamics, and $\eta(\tau)$ is a series of multiplicative Gaussian white noise terms with zero mean and with a correlator $\langle \eta_i(\tau)\eta_j(\tau') \rangle = B_{ij}(\mathbf{x})\delta(\tau - \tau')$. The structure of both f and B is derived from the set of possible events in the underlying (so-called “microscopic”) probabilistic model.

Analysis of the deterministic system $\dot{\mathbf{x}} = f(\mathbf{x})$ provides insight into the dynamics in the limit of infinite population size. However, “infinite” is often not a good proxy for “large,” and the behaviour (even on average) of an ensemble of populations may be very different to the infinite system limit due to demographic noise.

So how can we extract usable predictions from equation (3)? In the one-dimensional case, essentially all questions of interest can be answered by integration. Alternatively, if f is linear and B constant, then equation (3) describes an Ornstein-Uhlenbeck process, for which a complete characterisation is known. In all other cases, the main line of attack available is to somehow reduce the multivariate non-linear stochastic dynamical system to an effective system that is either univariate or linear.

We first address linearisation, with the aid of an illustrative example: Lotka-Volterra predator-prey cycling. Simple

³These periodic fluctuations became apparent from data collected by fur trading companies in North America during the 1800s. Despite the ecological devastation this industry caused, these detailed data sets have continued to find use in modern ecological studies.

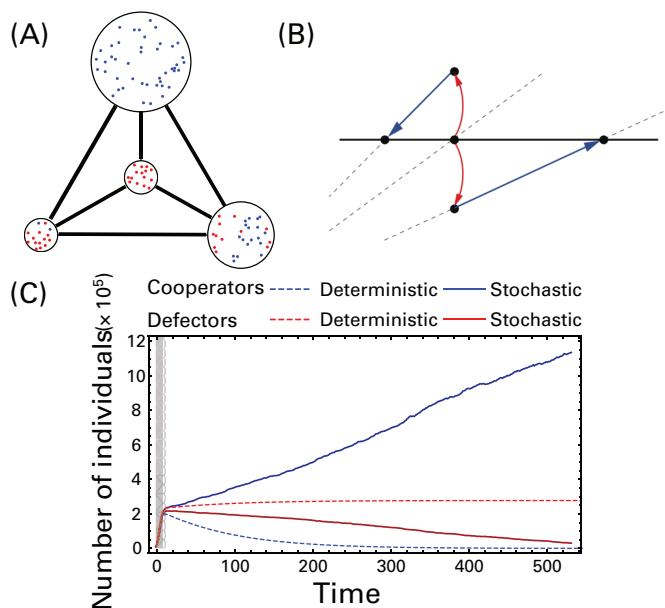


Figure 2. Noise-induced selection for cooperation. **Panel A:** A model in which cooperators (blue) and defectors (red) compete for resources on a series of patches. Cooperators reproduce more slowly, but can reach larger numbers on each patch by efficiently sharing resources [CRMT16]. **Panel B:** A discrepancy between deterministic and stochastic formulations of the model arises; stochastic fluctuations (red arrows) are equally probable in either direction from a deterministic slow manifold (black solid line), yet “kick” the stochastic system along fast manifolds (dashed lines) in a different direction (blue arrows, here bias towards the right) to that predicted deterministically [PR17]. **Panel C:** The cumulative result of these “kicks” is that while a deterministic model robustly predicts that cooperators should be driven to extinction, an equivalent stochastic model shows that cooperators can in fact fixate.

deterministic ODE models of the situation feature either neutral cycling (which is unstable to perturbations and thus leads to species extinction) or damped oscillations that die out as the system approaches a stable fixed point of coexistence between predators and prey (and thus does not match empirically observed oscillations). More complex models feature limit cycles that predict unreasonably low species abundances at certain points in the cycle. This difficulty in recapitulating periodic fluctuations in abundances with simple ecological models persisted until 2005, with the seminal work of McKane and Newman [MN05] analysing stochastic fluctuations in an individual-based Lotka-Volterra model.

For large but finite population sizes, demographic noise repeatedly “kicks” the system away from equilibrium, re-exciting the damped oscillations present in the deterministic limit. The essential features of these so-called quasi-cycles are captured in the power spectrum of fluctuations (the Fourier transform of the time-series autocorrelation function). McKane and Newman showed how the power

spectrum can be computed from an analysis of the linearised form of the SDE system. A short form of their result is as follows: if \mathbf{x}^* solves $f(\mathbf{x}^*) = \mathbf{0}$, then introducing fluctuations $\xi = \sqrt{N}(\mathbf{x} - \mathbf{x}^*)$, one may expand in large N to find $\dot{\xi} = A\xi + \eta(t)$. Here A is the Jacobian matrix of f at \mathbf{x}^* , and the noise has covariance matrix $B = B(\mathbf{x}^*)$. Standard computations then yield the power spectrum

$$P(\omega) = (A - i\omega)^{-1}B(A^T + i\omega)^{-1}. \quad (4)$$

A peak in the power spectrum implies the presence of quasi-cycles in the population dynamics with that frequency. This approach has since seen application to a wide range of situations in stochastic mathematical biology, from the seasonality of measles to synchronised oscillations in neural networks [KCR21].

Challenge: In models with many interacting types (e.g., species in a complex ecosystem), the matrix function (4) is impossible to evaluate by hand. What other theoretical approaches are possible in this regime?

A second approach to analysing stochastic dynamical systems is to reduce their dimension. This is usually achieved through a separation of time scales argument using the theory of slow manifolds. If one can identify quantities that are approximately conserved on short time scales, it is possible to integrate out faster fluctuations. This technique is well understood in deterministic dynamical systems, but an important difference occurs when applied to stochastic systems: the second-order features of the approach to the slow manifold contribute a biasing term to the reduced dimension model that must be carefully computed. A systematic procedure for this is laid out in [PR17], leveraging powerful results of Katzenburger.

A large class of important biological models exhibit natural timescale separation between the fast response of ecology and slow adaptation via evolution. Surprising and sometimes counterintuitive results have been found in models of the evolution of cooperation (see Figure 2), viral strain evolution, sex chromosome evolution, and others [KCR21].

Challenge: Away from ecological/evolutionary dynamics, most biological models do not have a clean separation of time-scales that is known in advance. Can we develop more widely applicable techniques to reduce the dimension of complex models?

PDEs and Related Fields

In the 50s, Turing introduced the concept of spatial patterns that are induced by a diffusion process of an activating and inhibiting component (e.g., prey and predator species). At the time, this mechanism was hypothesized

to explain the origin of animal coat patterns. While modern experimentation has revealed that animal coat patterns in fact have a genetic basis, emergent Turing patterns have helped explain many other phenomena, from vegetation patterning to mussel distribution on the sea bed [KCR21].

Although it is a simple mathematical model that suggests how certain natural patterns could emerge, the Turing model has been highly criticised for its restrictive assumptions (e.g., the diffusive rate of the inhibitor is required to be much larger than of the activator, and often parameter choices are more sensitive than robust patterns in nature would suggest). However, in the context of individual-based models, *stochastic* Turing patterns have been reported to emerge in a far greater range of parameters. With a proper treatment of noise, it appears the simple mechanism proposed by Turing is more robust than previously thought.

Butler and Goldenfeld [BG09] introduced a spatial predator-prey model, and investigated the stochastic effects of demographic noise. Besides the temporal oscillations of species abundances, it was shown that persistent perturbations, coming from intrinsic noise, can sustain transient patterns that would otherwise decay in a deterministic model. This phenomenon has become known as a *stochastic Turing pattern*. Beyond being merely a mathematical curiosity, recent studies have provided evidence that stochastic Turing patterns could potentially explain observed phenomena across many scales, from the coexistence of trees and grass in the mesic savannah [MGCL13], to spatial correlations in gene expression in synthetic bacterial populations [KML⁺18].

Interestingly, the rigorous mathematical treatment of noise in spatial systems lags some way behind that of non-spatial SDEs. The reason for this is the well-known difficulty in the formal interpretation of stochastic PDEs. This question has been the subject of intense recent work since the breakthrough of regularity structures and paracontrolled calculus, although interest in the area is presently almost completely confined to applications in physics.

Challenge: What do the new tools for rigorous analysis of SPDEs have to say about applications in stochastic mathematical biology?

Stochastic spatial models also occur naturally in the study of biological motion, in particular when considering the collective behaviour of groups of animals. Here the stochasticity and randomness can take many different forms and often depend on the model itself, with noise often turning out to be a necessary ingredient to reproduce complex observed behaviours. These models are mainly investigated through computational simulations, a paradigm dating back to Reynold's "boids" program written in 1968. In many models, noise is added to choice of

direction of the moving of individuals that make up the swarm. Alternatively, something as simple as randomising the order of update events when individuals adjust their trajectories can introduce enough of a stochastic element to achieve "realistic" behaviours.

Data from animal experiments support the modelling of update events as themselves being random. Yates et al. [YEE⁺09] used a coarse-grained model to estimate the drift and diffusion coefficients in a 1D model of marching locusts; their findings suggest that randomness in locust decision-making is increased when the alignment is low. This model included intrinsic noise in the form of random diffusion, though Bode et al. [BFJW10] reproduced the same behaviour using asynchronous updating rules instead, from which the stochastic behaviour emerges. Whatever the source, it is believed that noise is a crucial mechanism that allows marching locusts to spontaneously change direction.

Recently Biancalani et al. [BDM14] have shown that noise not only drives a system to switch between stable states, but can create these states in the first place. Their model investigated the foraging behaviour of an ant colony that switched between food sources; similar effects have recently been observed in experiments on the noise-induced schooling of fish [JMAK⁺20].

Challenge: Swarming animals are often engaged in decision-making processes (foraging, predators avoidance, etc.) at both collective and individual levels. How does the interplay of spatial motion and demographic noise affect the flow of information and emergent behaviours?

Random Matrix Theory

The above examples all concern random effects that play out over observable time scales. Some random and unpredictable effects develop so slowly (for example through evolution) that they should be treated as essentially static. The most important example in this class is the heterogeneous characteristics of, and interactions between, species in an ecosystem. The goal of research in this direction is to make predictions about ecosystem stability with only incomplete and uncertain data on its makeup.

In 1972, based on the work of Gardner and Ashby, Robert May introduced the framework of Random Matrix Theory to the world of theoretical ecology. The model proposes that random coefficients in a *community matrix* (which captures the interaction strength between different species in an ecosystem) could replace parameters that would otherwise be too difficult to estimate in practice. Building on the insights developed by Wigner, May derived a stability criterion for such random ecosystems:

$$\sigma\sqrt{NC} < d, \quad (5)$$

where N is the number of species within the ecosystem, C is the probability that a pair of species interact with each other, and σ^2 is the variance of interaction strength. The parameter $d > 0$ on the right-hand side is the magnitude of the negative diagonal elements of the community matrix ($d = 1$, in the model proposed by May). According to this *circular law*, all eigenvalues of the community matrix lie with high probability within a disk of radius $\sigma\sqrt{NC}$, centered at $-d$ on the real axis of the complex plane. In other words, if the inequality in (5) is met, all eigenvalues have negative real part, and thus the community matrix is considered locally stable.

The criterion derived by May implies that complexity (i.e., a large number of species, high connectance, or diverse species interactions) decreases the stability of an ecosystem. However, this result contradicted the common intuition that large and diverse ecosystems are considered more stable, where the removal of a single link can easily be compensated. This discrepancy ignited a complexity-stability debate that has raged ever since. However the model introduced by May clearly serves best as a null model, rather than an accurate representation of real ecosystems.

The simplistic assumptions that make up the random community matrix are insufficient for characterising real structured ecosystems. Instead, one seeks the missing mechanisms that could be key for explaining the stability of complex ecosystems we observe in nature. Such improvements to the random matrix model of ecosystems include accounting for topological structure and the effect of intraspecies competition. Amongst the most influential recent work has been that of Allesina and Tang [AT12], who investigated how the structure of pairwise ecological interactions within ecosystems could affect stability.

Instead of assigning fully random interaction coefficients, they allowed to vary the proportion of predator-prey pairs, relative to mutualistic and competitive interactions. With this new random matrix set up, the stability criterion derived from the circular law in (5) is modified to become an elliptic law for local stability. Given a larger proportion of predator-prey interactions, the eigenvalue distribution of the random matrix is stretched along the imaginary axis, forming an ellipse. Therefore, the stability criterion that requires the real part of all eigenvalues to be negative relaxes. The opposite holds true for systems with large proportions of symmetric interactions such as mutualism and competition. This simple yet powerful random matrix model explains mathematically how predator-prey interactions promote stability of ecosystems, an insight that has been supported by other theoretical studies [Gal18].

While the random matrix model of species interactions is an abstraction of real ecosystems, interrogation of this

model has furthered our understanding of the basic principles behind ecosystem dynamics. Interestingly, it has been shown that a large set of different ecosystem models can be reduced to a single model based on just a few statistical parameters [BABL18]. Progress is also beginning to be made in the challenging task of connecting to empirical studies to shed light on the key features that set real ecosystems apart from their model counterparts [JPR⁺15], and potentially infer structural features of interaction matrices from timeseries data [KYCR21].

Challenge: Can we robustly interrogate real data with RMT models, and vice versa?

An Invitation

Stochasticity has been shown to operate across a huge range of scales in mathematical biology, from intracellular physiological processes and the transmission of genes governing evolution, to the species interactions that govern ecosystem stability and the diversity of life on earth. In this article we discussed examples of mathematical models that have been studied previously from a deterministic point of view, and we showed how more recent studies that incorporate stochasticity in these models open up a new range of questions to interrogate about the biological system at hand. Encouraged by this perspective on mathematical biology, we have illustrated how interrogating the effect of this stochasticity requires drawing on a range of mathematical subdisciplines, including probability, statistical mechanics, stochastic and partial differential equations, and random matrix theory.

Despite a rich history of the application of these techniques in mathematical biology, the field of stochastic mathematical biology remains remarkably fertile. Perhaps due to the perceived technical difficulty of addressing these probabilistic problems in biological modelling, there remain a number of open problems, some of which we have sought to highlight in the course of this article. Interestingly, many of these do not require an overly complicated approach that covers all details of the underlying biological processes. Instead, mechanistically minimal models that are informed by key empirical insights from biologists have been shown to hold surprising explanatory power. Progress does however require knowledge of the comparatively sophisticated mechanics of stochastic processes. It is in this spirit that we offer this invitation to mathematicians across disciplines to join us in working in this exciting field.

References

- [AT12] Stefano Allesina and Si Tang, *Stability criteria for complex ecosystems*, *Nature* **483** (March 2012), no. 7388, 205–208.

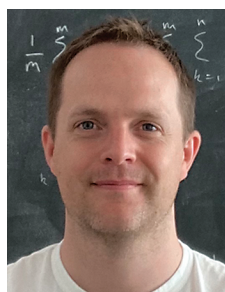
- [BD95] Frank Ball and Peter Donnelly, *Strong approximations for epidemic models*, Stochastic Process. Appl. 55 (1995), no. 1, 1–21, DOI 10.1016/0304-4149(94)00034-Q. MR1312145
- [BABL18] Matthieu Barbier, Jean-François Arnoldi, Guy Bunin, and Michel Loreau, *Generic assembly patterns in complex ecological communities*, Proc. Natl. Acad. Sci. USA 115 (2018), no. 9, 2156–2161, available at <https://www.pnas.org/content/115/9/2156.full.pdf>.
- [BDM14] Tommaso Biancalani, Louise Dyson, and Alan J. McKane, *Noise-induced bistable states and their mean switching time in foraging colonies*, Phys. Rev. Lett. 112 (2014), no. 3, 038101.
- [BFJW10] Nikolai W. F. Bode, Daniel W. Franks, and A. Jamie Wood, *Making noise: emergent stochasticity in collective motion*, J. Theoret. Biol. 267 (2010), no. 3, 292–299, DOI 10.1016/j.jtbi.2010.08.034. MR2974407
- [BG09] Thomas Butler and Nigel Goldenfeld, *Robust ecological pattern formation induced by demographic noise*, Phys. Rev. 80 (2009), no. 3, 030902.
- [CRMT16] George W. A. Constable, Tim Rogers, Alan J. McKane, and Corina E. Tarnita, *Demographic noise can reverse the direction of deterministic selection*, Proc. Natl. Acad. Sci. USA 113 (2016), E4745–E4754.
- [Gal18] Tobias Galla, *Dynamically evolved community size and stability of random lotka-volterra ecosystems (a)*, EPL (Europhysics Letters) 123 (2018), no. 4, 48004.
- [JPR⁺15] Alex James, Michael J. Plank, Axel G. Rossberg, Jonathan Beecham, Mark Emmerson, and Jonathan W. Pitchford, *Constructing random matrices to represent real ecosystems*, The American Naturalist 185 (May 2015), no. 5, 680–692.
- [JMAK⁺20] Jitesh Jhavar, Richard G. Morris, U. R. Amith-Kumar, M. Danny Raj, Tim Rogers, Harikrishnan Rajendran, and Vishweshha Guttal, *Noise-induced schooling of fish*, Nature Physics 16 (2020), no. 4, 488–493.
- [JG20] Chen Jia and Ramon Grima, *Small protein number effects in stochastic models of autoregulated bursty gene expression*, J. Chem. Phys. 152 (2020), no. 8, 084115, available at <https://doi.org/10.1063/1.5144578>.
- [KML⁺18] David Karig, K. Michael Martini, Ting Lu, Nicholas A. DeLateur, Nigel Goldenfeld, and Ron Weiss, *Stochastic Turing patterns in a synthetic bacterial population*, Proc. Natl. Acad. Sci. USA 115 (2018), no. 26, 6572–6577, available at <https://www.pnas.org/content/115/26/6572.full.pdf>.
- [KCR21] Yvonne Krumbeck, George W. A. Constable, and Tim Rogers, *Stochastic mathematical biology reading list (2021)*, available at people.bath.ac.uk/ma3tcr/StochMathBio_reading_list.pdf.
- [KYCR21] Yvonne Krumbeck, Qian Yang, George W. A. Constable, and Tim Rogers, *Fluctuation spectra of large random dynamical systems reveal hidden structure in ecological networks*, Nature Communications 12 (2021), no. 1, 1–14.
- [MGCL13] Ricardo Martínez-García, Justin M. Calabrese, and Cristóbal López, *Spatial patterns in mesic savannas: the local facilitation limit and the role of demographic stochasticity*, J. Theoret. Biol. 333 (2013), 156–165, DOI 10.1016/j.jtbi.2013.05.024. MR3092495
- [MN05] A. J. McKane and T. J. Newman, *Predator-prey cycles from resonant amplification of demographic stochasticity*, Phys. Rev. Lett. 94 (2005Jun), 218102.
- [PR17] Todd L. Parsons and Tim Rogers, *Dimension reduction for stochastic dynamical systems forced onto a manifold by large drift: a constructive approach with examples from theoretical biology*, J. Phys. A 50 (2017), no. 41, 415601, 33, DOI 10.1088/1751-8121/aa86c7. MR3708113
- [SST20] H. Schenk, H. Schulenburg, and A. Traulsen, *How long do red queen dynamics survive under genetic drift? a comparative analysis of evolutionary and eco-evolutionary models*, BMC Evol. Biol. 20 (2020), no. 8.
- [YJB⁺10] Shinichi Yachida, Sian Jones, Ivana Bozic, Tibor Antal, Rebecca Leary, Baojin Fu, Mihoko Kamiyama, Ralph H. Hruban, James R. Eshleman, and Martin A. Nowak, *Distant metastasis occurs late during the genetic evolution of pancreatic cancer*, Nature 467 (2010), no. 7319, 1114–1117.
- [YEE⁺09] Christian A. Yates, Radek Erban, Carlos Escudero, Iain D. Couzin, Jerome Buhl, Ioannis G. Kevrekidis, Philip K. Maini, and David J. T. Sumpter, *Inherent noise can facilitate coherence in collective swarm motion*, Proc. Natl. Acad. Sci. USA 106 (2009), no. 14, 5464–5469.
- [ZGG⁺19] Silvia Zaoli, Andrea Giometto, Jonathan Giezen-danner, Amos Maritan, and Andrea Rinaldo, *On the probabilistic nature of the species-area relation*, J. Theoret. Biol. 462 (2019), 391–407, DOI 10.1016/j.jtbi.2018.11.032. MR3882333



George Constable



Yvonne Krumbeck



Tim Rogers

Credits

Opening image is illustrated by Miranda Bane. Figures 1 and 2 are courtesy of the authors. Photo of George Constable is courtesy of George Constable. Photo of Yvonne Krumbeck is courtesy of Yvonne Krumbeck. Photo of Tim Rogers is courtesy of Tim Rogers.