Applied mathematics

Finding the state of a system – using only data

Brendan Keith

Pinpointing the state of a complex system is tricky, especially when the underlying mathematical equations aren't known. But a data-driven technique makes light work of it – and could even change the way that models are formulated. **See p.261**

Imagine you're on a hill, flying a kite, and someone asks you to describe exactly what the kite is doing. "It's over to the right, just above the treeline," you might say, "and the wind is making it spin anticlockwise and move to the left." This is an example of state estimation: the process of determining the true state of a complex system, typically from noisy, incomplete and often indirect observations. In a world replete with data, progress in data-driven modelling is rapid, but compatible state-estimation techniques are lagging behind. On page 261, Course and Nair[1] propose a state-estimation technique that enables efficient forecasts using data, without the details of any underlying model. In doing so, they offer a fresh take on the general approach to model discovery.

In the case of the kite, its state is encapsulated by your description of its flight at that moment – where it is, the direction in which it's moving and its angular momentum. States can similarly be defined to describe how planets are moving, for example, or what processes are occurring inside a cell. The state captures all the essential details about a system to help us to understand, predict and control what happens next.

Unfortunately, few physical systems are as easily observable as a kite. Most can be observed only indirectly. To understand this fundamental limitation, imagine describing what the kite is doing without ever looking up; indeed, imagine basing the entire description on the tension of the line in your hand. Would you be as confident in your description? Clearly, not being able to see the kite makes describing its state much harder.

This is precisely why it is crucial to develop state-estimation techniques when measurements are scarce or observations are indirect. State estimation is used to reconstruct the position and velocity of an aircraft using radar measurements, and to monitor the safety and efficiency of industrial processes. It helps to determine the current state of Earth's climate, and estimate fluctuations in the pricing of assets on the stock market. It can even be used to ascertain the internal states of a biological cell. State-estimation techniques provide a mathematical framework to merge information from various sources and previous knowledge, resulting in an improved understanding of the world around us.

Classical state-estimation techniques are typically used to study systems for which the mathematical equations describing their behaviour are known. These methods were applied as long ago as the 1960s, when one such technique, known as the Kalman filter[2], was used to help NASA engineers guide the Apollo spacecraft to the Moon[3]. However, that feat was accomplished with ample knowledge of the structural form of the equations of motion, which were derived from physical laws. By contrast, general models can now be derived automatically and algorithmically from observational data alone.

Indeed, techniques for data-driven modelling have been developed to describe complex systems ranging from biological processes[4] to astrophysical phenomena[5]. Course and Nair's state-estimation technique is designed precisely for this new age of data-driven model discovery. The approach is based on a technique called Bayesian inference, which is used widely, but which can be computationally challenging for complex systems. The authors' main innovation was to calculate the set of parameters required for such inference by using stochastic approximations that can be computed in parallel, thereby maximizing efficiency. This contribution is exciting – not only because it is more efficient than classical state-estimation techniques, but also because it does not require the complicated equation solving that these methods rely on. It therefore

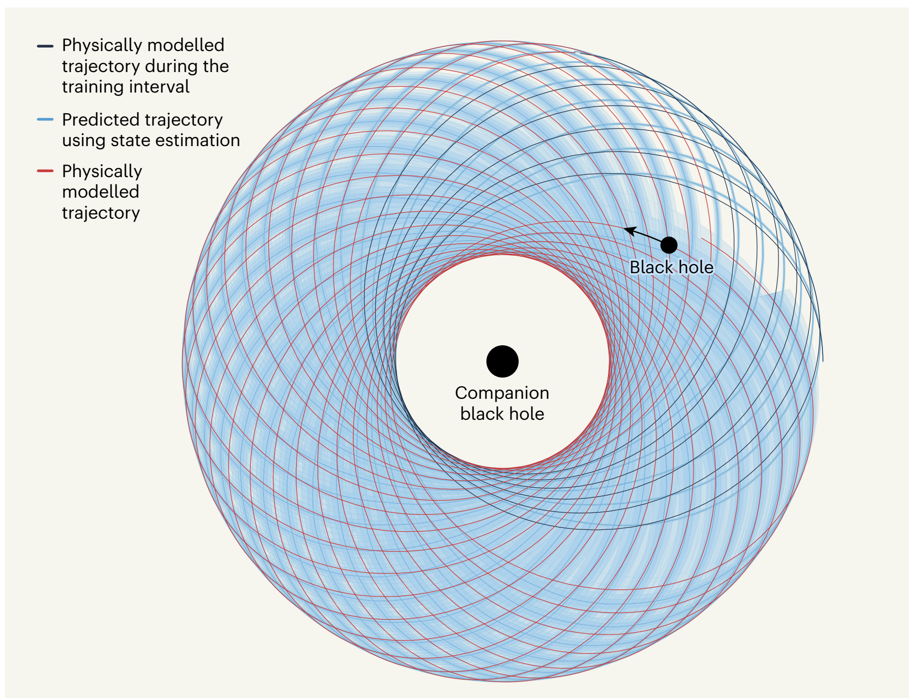

Figure 1 | Estimating the state of two orbiting black holes. Course and Nair[1] developed a method that makes it possible to estimate the state of a system from data, without having an underlying mathematical model. As an example, they used the technique to predict the trajectories of two black holes orbiting each other, as well as the governing equations, on the basis of observations of the gravitational waves that the black holes generate as they spiral towards each other. The gravitational-wave observations over a short time period are used as training data to infer governing equations for the black holes' motion, which are then used to predict the complete orbital trajectory. The authors' estimates are well matched to a set of test data drawn from a physical model, and can be predicted far beyond the duration of the observations. The trajectories depicted lie in the orbital plane of a small black hole in motion around its more massive companion. (Adapted from Extended Data Fig. 4 of ref. 1.)

presents a fresh approach to model discovery in mathematics.

Owing to its many uses, state estimation influences science, engineering and even public policy. For this reason, quantifying uncertainty is crucial for making informed decisions and risk assessments on the basis of any state prediction. Course and Nair showed that their method is widely applicable by considering examples ranging from modelling fluid flow to predicting the motion of black holes (Fig. 1).

The latter application built on previous work[5] to show that the equations describing the motion of two black holes in orbit around each other can be reverse-engineered from measurements of the gravitational waves that the black holes generate before merging. The gravitational-wave observations over a short interval are used as training data to infer the orbital equations, and the uncertainty associated with this inference can then be used to predict and quantify the uncertainty in a complete orbital trajectory over a much longer interval.

State estimation already serves as a foundation for predicting and controlling states in myriad applications, ensuring that devices, machinery and complex systems operate safely and efficiently. But the nature of scientific enquiry is evolving, and interest in developing methods to handle the flexibility and complexity of data-driven models is growing. Thus, whereas research in the past focused on creating a synergy between state-estimation techniques and known physical laws, current research demands that both model formulation and state estimation adapt to the intricacies of our increasingly complex, multifaceted and data-centric world. Like an airborne kite above a hill, Course and Nair's work will help to test the winds of change as they begin to blow scientific progress in a new direction.

Brendan Keith is in the Division of Applied Mathematics, Brown University, Providence, Rhode Island 02912, USA.
e-mail: brendan_keith@brown.edu

1. Course, K. & Nair, P. B. *Nature* **622**, 261–267 (2023).
2. Kalman, R. E. *J. Basic Eng*. **82**, 35–45 (1960).
3. Grewal, M. S. & Andrews, A. P. *IEEE Control Syst. Mag*. **30**, 69–78 (2010).
4. Yazdani, A., Lu, L., Raissi, M. & Karniadakis, G. E. *PLoS Comput. Biol*. **16**, e1007575 (2020).
5. Keith, B., Khadse, A. & Field, S. E. *Phys. Rev. Res*. **3**, 043101 (2021).

The author declares no competing interests.

Evolution

A foundational view of the physics of evolution

George F. R. Ellis

How can physics underlie the emergence of biology's complex functionality? A powerful interface between physics and biology that describes the processes of evolution by natural selection provides a compelling answer. **See p.321**

Everything we see around us, including ourselves, emerges out of physical interactions between fundamental particles. But because physics does not have any concept of function, it cannot distinguish the emergent functional features that are central to biology[1] from random fluctuations. The complex structures of proteins, all of which have emerged to perform specific biological functions, are a case in point[2,3]. In addition, the laws of physics are timeless and eternal, unaffected by historical events, so cannot be used to describe how the past evolution of species affects their present and future. On page 321, Sharma *et al.*[4] present what they call assembly theory as a way to fill this gap, providing a framework to unify descriptions of evolutionary selection across physics and biology.

"Attempts to describe evolution quantitatively do not relate to the underlying physics."

The existence of living beings that are well adapted to their environment is explained by Charles Darwin's theory of natural selection. At a macro level, natural selection states that species evolve by initially random variants being selected for survival over many generations through their relative reproductive success[5]. But attempts to describe this process quantitatively, for example through Hamilton's Rule and the Price equation[6], just describe outcomes and do not relate to the underlying physics. The same is true of Fisher's fundamental theorem of natural selection[7] and of mathematical formulations of population genetics.

Assembly theory fills this gap in an innovative way by quantifying the degree of evolution and selection in an ensemble of objects. Conventionally, an object is defined by the material particles from which it is made. Assembly theory instead defines an object through its possible formation histories in an 'assembly space' in which objects are made by joining elementary building blocks together recursively to form new structures.

The assembly universe is the space that contains all of the conceivable pathways for assembling any object from the same building blocks. But the parts of this space that are actually accessible are limited, first by the laws of physics, and second by historical contingency: new things can be built only on the basis of what is already there, further constraining what is possible.

The authors build a quantity they call 'assembly' from two variables: copy number, meaning the number of copies of an object in an ensemble; and assembly index, the minimum number of steps needed to produce an object. These combine to give an equation that determines the amount of selection that was necessary to produce an ensemble of objects. The authors' key contention is that a transition from no selection to selection – such as happened when inanimate matter became animate – changes the pathways taken in assembly space in a mathematically definable way embodied in this equation. In essence, an object with a high assembly index that has a high copy number is evidence of selection. Two timescales determine the dynamics of the assembly process: the rate at which new, unique objects are formed, and the rate at which those objects are copied after they exist. If the relationship between these two timescales is such that resources are available for making more copies of existing objects, then selection can occur.

The assembly index of a molecule could possibly be determined experimentally, which would allow a check on theoretical calculations. Sharma *et al.*[4] give examples of assembly pathways for molecular processes, including the joint assembly space for polymeric chains and processes catalysed by enzymes, as well as spaces in which selection has generated ensembles of high complexity.

The authors state that assembly theory