

# Wheat from the chaff

Mathematical modelling is almost as old as science. But the nature of applied mathematics is changing — in physics as well as in biology and medicine, engineering and even economics. The rise of powerful computation has made it possible to create and study models with unprecedented complexity; and to gather equally massive data to use in testing those models.

But a model with 30 or 100 free parameters has enormous flexibility. Can it really ever be tested? Or can a sufficiently clever scientist always manage to adapt it to the data? Science has always worked by putting creative hypotheses and ideas to the experimental test. In the face of high-dimensional modelling, could this practice run into trouble?

I've wondered about this question in the context of recent efforts to build models of whole economies. Physicists have been influential in doing so, making significant advances over traditional economic models. But a lingering worry remains — how can you trust a model with a dozen or more variables, and avoid fooling yourself about its explanatory capacity?

Well, physicist Mark Transtrum and colleagues may have found an answer — or, at least, they make an encouraging suggestion (M. K. Transtrum *et al.*, preprint at <http://arxiv.org/abs/1501.07668>; 2015). The problem may not be as vexing as it seems, they argue, because many, if not most, high-dimensional models, as well as real processes, are 'sloppy' — that is, their behaviour depends on very few parameters or details, and the rest are mostly irrelevant. And, they suggest, this isn't just a happy accident; it happens for deep reasons.

Physicists understand well enough how complexity often reduces to simplicity; lots of physics, after all, relies on low-dimensional effective models and theories that work despite ignoring masses of microscopic detail. It's true in statistical physics, thermodynamics or fluid dynamics, and this seeming miracle rests on well understood theory — the continuum limit, or renormalization group arguments. Cartoon theories, Transtrum and colleagues note, can be useful even if they ignore a lot — as long as they get a few crucial things right.

For example, liquids like water or petroleum differ completely at the molecular level, yet on macroscopic



The world in all its detail is overwhelmingly complex, yet also much simpler than it looks.

scales flow in patterns determined only by viscosity and density. So too with models of collective behaviour such as magnetism: many microscopic models yield very similar, and very simple, macroscopic behaviour.

In other words, sloppy theories work well surprisingly often. As Transtrum and colleagues argue, this pattern reaches right across science. They analysed a set of models running from radioactive decay to systems biology, using information theory to characterize the sensitivity of these models to the variation of particular parameters. In every case, they found that the distribution of the magnitudes of eigenvalues — reflecting the relevance of different parameters — fell off roughly log-linearly, with each parameter being less important than the previous by a constant factor. Quite generally, a few parameters (or combinations thereof) tend to be of much greater importance than all others.

That's an empirical result. But the authors use an elegant geometric interpretation of statistics to argue that we should actually expect this. High-dimensional theories, in a certain sense, tend to have a low effective dimension, which is closely associated with the few important or 'stiff' parameter combinations. Their analysis also leads to a new technique for generating low-dimensional, reduced models from initial models of much higher dimension — a recipe for identifying the emergent variables of greatest interest.

How useful this particular technique will be in practice remains unclear. But the general perspective has potentially huge implications. It may, for a start, explain a number of otherwise puzzling examples of modelling good luck. Why, for example, does the simple technique of principal component analysis work so well in applications ranging from molecular biology to demographics? Given a high-dimensional data set — the spatiotemporal dynamics of the human heart, for example — such analysis often finds that a description of

low dimension, retaining only a few basic spatiotemporal modes, captures a large fraction of the variation in the data. The generic sloppiness of natural processes may be a natural explanation.

Sloppiness, the authors argue, may also explain why many biological systems are so robust to environmental variations. The circadian rhythm in cyanobacteria, for example, maintains a 24-hour cycle over a wide range of temperatures, even as chemical reaction rates of the key proteins involved double over this range. This may tell us something about how evolution has exerted control, focusing on a few stiff parameters, yet also benefiting from the low sensitivity of the circuit to other parameters. Engineering such control may not be as difficult as it seems, given the existence and availability of many neutral dimensions for adjustment.

Transtrum and colleagues also argue that sloppiness probably explains why humans — and many other animals — are so good at visual pattern recognition, and reliant on it for interacting with the external world. Pattern recognition means locking on to low-dimensional representations, and ignoring huge volumes of other data. We don't need any further detail, perhaps, because the objects we need to identify — faces, for example — admit low-dimensional representations capturing most of the important information.

Finally, sloppiness might even explain why science itself is possible. The world in all its detail — in anything from ecology to macroeconomics to astrophysics — is overwhelmingly complex, yet also much simpler than it looks. Vastly simplified models can always capture important features, although they of course leave others out (in many systems, for example, influences on single elements truly can lead to macro changes). We do science with low-dimensional models because more complex models are typically less efficient, dwelling on details of marginal importance.

Transtrum and colleagues, it seems, started out thinking about how to analyse data and build models useful for systems biology. Rather than just applying to biology, however, their findings may be much more general. Complexity is a barrier to understanding, but it's not nearly as impenetrable as it seems. □

MARK BUCHANAN