



EDITORIAL

‘A novel model blah blah blah’

Anahid Basiri

School of Geographical and Earth Sciences, The University of Glasgow, Glasgow, UK.

E-mail: ana.basiri@glasgow.ac.uk, journal@rin.org.uk

Some time ago, in response to a review request for a manuscript, I received a video link. A video link instead of usual textual review seemed odd. I opened the link and watched the video, actually quite a few times until I finally got the message/review. It was a video of someone repeatedly breaking things and trying to fix them in the weirdest possible way; at first the guy cut a plastic ruler with scissors and then melted the two edges to be able to glue them back together and ‘fix’ the ruler. Then he drilled a milk packet and when the milk started pouring, he used waterproof dressing strips to stop it. I watched it several times until I understood what it actually meant; it seemed to me that the reviewer wanted to imply that the authors of that paper solved a problem that did not exist in the first place. I contacted the reviewer to ensure that he did not send the link by mistake or I that I had misunderstood his ‘honeyed poison.’ We had a much longer conversation about the role of scientific method in the time of open data, big data, industry 4.0. He told me that many researchers seem to come up with problems that do not exist, have been solved already, or do not matter at all. Even if the problem exists or matters, researchers often provide a solution that may solve that specific and narrow problem, but does not make sense overall or is not extendable or applicable to slightly different scenarios. Having answered too specific research questions, one may wonder if we are changing our research interest based on availability of data, capability of computers and development of new models and methods. Should this not be this the other way around?

That conversation reminded me of another email from an internationally recognised researcher in an engineering discipline a few years ago. He had decided not to review any research papers because of the same sort of fundamental issue. In his email, he wrote and I quote:

‘The vast majority of submissions in our field seem to be either a problem that was essentially solved 20 years ago (or more) and compare one algorithm (“a novel blah blah blah”) with another algorithm that is generally not the state of the art. Or whatever is fashionable (often machine learning these days) and apply it without thinking about what the best way is actually of solving the relevant problem.’

What this colleague was referring to does not seem to be specific to any specific research area. These are common concerns in several disciplines, including social science, engineering and medical science, whose researchers use the statistical and machine learning models for domain-specific and applied research questions. It is not even new concern; so we should not blame the ‘fancy’ machine learning models; they have been our concerns since we started using ‘models’ and ‘simulations’. We can hardly find a lecture nowadays without any reference to the famous quote from George Canning, ‘I can prove anything by statistics except the truth’, or from George Box, ‘All models are wrong, but some are useful’ (1976). However, with the advent of big data, open data, and industry 4.0, we hear greater concerns around blind use of data and models. But are they valid concerns? Do we solve the problems that are already solved or do not exist/matter? Do we use fancy machine learning and artificial

intelligence blindly just ‘because we can’? Do we look for presentable and ‘statistically significant’ results, regardless of validity of the methods? Do we appreciate what model can say and what can’t do?

1. Streetlight effect

We understand and get to know phenomena, processes and our world through sensing, making measurements using our five basic senses or sensors we make. Our knowledge about phenomena depends heavily on how much data we can collect about them. As Lord Kelvin famously said (Popular Lectures and Addresses, 1889):

‘When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind: it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science, whatever the matter may be’.

Many qualitative researchers may find that the quote, ‘If you cannot measure it, then it is not science’, to undermine their research. However, many argue that qualitative research also generates nonnumerical data, such as text. So, the view may still stand and be the basis of many disciplines and research areas. However, the potential challenge with this mindset can be the illusion of knowing something or generating knowledge by just accruing some measurements. This can become an issue at the time of big data, when we are ‘drowning in data but starving for insights’ (Ferraioli and Burke, 2018). This can potentially introduce several issues, but here I would like to focus on two:

- Solving problems that either do not exist or do not matter, what we call the ‘streetlight effect’.
- Changing the course of scientific method and framing our research questions based on what we can do and not necessarily what we should.

The streetlight effect refers to the situation that we forget our main job, pushing the boundaries of science and engineering, and try to look for some results, patterns or a model that is simply publishable or presentable or simply because we have data about it.

The streetlight effect refers to the story of policeman who sees a man searching for his keys under a streetlight and tries to help by looking for the keys under the streetlight with him. After a few minutes with no success, the policeman asks if he is sure he lost them there, and the man replies, no, that he lost them in the park. The policeman asks then why he is searching there, and the man replies, ‘*This is where the light is!*’

My colleague and that reviewer were concerned that we look for our ‘keys’ wherever ‘light’ is, looking for a pattern in data in order to predict a future of the phenomena, or to model a relationship between objects, solely because we have access to data and not out of curiosity or for the sake of solving a problem in the real world. We find solutions for problems that do not exist.

The availability of data at an unprecedented rate and scale has given us measurements and data about a plethora of objects, processes and phenomena. This allows us to express them with numbers even though we do not actually define them because we did not design data collection for the sake of getting to know these objects, processes or phenomena. Let me clarify: some decades ago, in order to do our jobs, we had to design surveys, build sensors, and do field work to collect data and measure things. This cost time, money, energy and labour. The only way to get to know things based on Lord Kelvin’s view was to design a research framework and optimise what we need to collect using our data collections methods, optimally designed instruments and make good enough measurements, so that we could then examine our hypothesis or create models that described the observed phenomena. In the era of big data, open data, Internet of Things (IoT), ubiquity of wearable sensors, and in general in the time of explosion of data, this strategic planning phase might have lost its position. Why do we need to plan for data collection and optimally use instruments and sensors if we already have them for free-ish? We may not need to focus on a specific points and time to make measurements as we can sense ‘everything’.

We are drowning in data; IBM in 2020 estimated that 90% of all the data in the world have been created in the last two years! This ubiquity of data results in our being able to convolute ‘expressing with numbers’ with ‘making measurements’. They are not the same. Having access to large polls of data allows us to model, simulate and predict using potentially ‘fancy’ models, simulation capabilities and prediction methods. However, this can bring the risk of changing the course of scientific method or studying only where data exist and not necessarily where we should try to research. It is not surprising to know that almost half of the start-ups that fail do so because they are solving a problem that does not really exist and creating solutions that nobody wants.

The role of framing a theory, defining and developing research questions, having hypothesis and choosing methodology regardless of availability of data seem to be increasingly important. If we have data, that is great. Otherwise, we must collect. But one should not look for results and patterns wherever data can be found simply because data already exist. In other words, we should ‘either find a way or make one’ but must not blindly follow any passage that exists. This is different from curiosity research; that is very valid and must remain to push boundaries of science. But if we find a solution just because it is publishable or just because we can, we may end up with *streetlight science*.

3. Can the ends justify the means?

The second part of the argument that my colleague put forward is about applying any fashionable models (often machine learning and artificial intelligence (AI) these days) without thinking about what and if they are actually the best way of solving those problems. Model selection has been an important part of designing research and scientific method. However, the models can also be selected based on their performance and the quality of the output rather than the characteristics and nature of the phenomena.

Selecting models based on their outputs is not wrong and is a common method for experimental work. In fact, many projects in statistics and computer science work on fitting the best model. We have all tried to fit a line through a scatter plot of data points that *best* expresses the relationship between those points models. So, the outputs comparison is a common and useful way for selecting and comparing models in terms of their performance. Also, we know that the uncertainty of the output results can propagate back through the models and tell us more about the input data, i.e. the data that tell how much we know the phenomena.

While output results can be potentially a good way to compare the model, with the explosion of data and having easy and free access to software packages and libraries that allow us to play with parameters, it has been possible to compare the results or outputs of the models potentially *blindly*. By ‘blindly’ I mean ‘playing with parameters and variables’, so one can run several models and compare their output against each other or against some quality measures, such as accuracy, response time, power consumption or scalability. Blindly comparing results and fitting models based on data – and not the theory, context, or characteristics of the problems – may result in overfitting and potentially lead to p-hacking. We may end up with some models that work quite well for that specific dataset but are almost useless for slightly different datasets. Sometimes we see high accuracy for the model, but with slight change in the study area (moving to another region or temporal epoch) we get a very unsatisfactory result. This may question the transferability and interpretability of the model. Please note that these are the potential issues of even simple model, e.g. linear regression. The problem can become bigger when we look at some of the ‘black box’ models, such as deep neural network.

We raise these issues not because they are widespread in our discipline. No, exactly the opposite; many of the models that we use are mathematical and physical and are built upon the nature of the phenomena we study. For example, we use signal pathloss models for positioning and navigation that incorporate the signal components, such as reflections, scattering, and diffraction into Maxwell’s partial differential equations. So, the models are not blindly fitted to data. Even in statistical models, we widely see the integration of context or discussion about the justification or interpretations of results. For example, in this single issue of the *Journal of Navigation* we have the paper from Rawson (2021) on Developing Contextually Aware Ship Domains Using Machine Learning that takes context into account.

So, the contextual information is used to tailor the models. We also see the others who use or develop the models to improve the navigation services through patterns that they recognised or future they try to predict. For example, Özlem et al. (2021) developed a simulation model which is capable of mimicking actual vessel arrival patterns and vessel entrance decisions based on the expert opinions, in response to the whole new AI challenge of ‘keeping human in the loop’. The model is customised based on different policies for day and night traffic and is run for a period of seven years with 20 replications for each year to ensure re-applicability of the model. They also validate the model’s results with the actual results of performance measures to be able to suggest that the proposed algorithm can be a guide for operators regarding scheduling decisions in congested, narrow waterways. Amir Ali Nikkhah (2021) looks at coarse alignment of marine SINS using the location of fitted parametric circle of gravity movement. Fitri Trapsilawati et al. (2021) study the integration of conflict resolution automation and vertical situation display for on-ground air traffic control operations. Carral et al. (2021) modelled both the operative and routine learning curves in manoeuvres in locks and in transit in the expanded Panama Canal, and for this they monitored the whole of the canal continuously for 42 months of operation, to make sure the seasonality and different context are all factored in. Djani Mohovic (2021) reduces the risk of collision between ships in a close-quarters situation by simulating collision-avoidance actions. Qi et al. (2021) developed a cellular automaton-based model for ship traffic flow in busy waterways. For this they first justified the research questions, i.e. spatial–temporal discretisation, safe distance and collision avoidance timing, as the three core components of ship traffic flow are difficult to determine using current simulations. When they established this is not a ‘streetlight science’, then they proposed a novel traffic flow model that can integrate the rules for ships’ motions by considering safe distance and collision avoidance timing. Finally they evaluated the performance by comparing the results with actual observed ship traffic data. We are actively pushing the boundaries of science and engineering using the availability of data, better computation capabilities, and of course domain knowledge that result in development of new devices, better theories, more navigable world

References

- A Dictionary of Thoughts.** (1908). edited by Tryon Edwards, p. 587.
- Box, G. E. P.** (1976). Science and statistics. *Journal of the American Statistical Association*, **71**(356), 791–799. doi:10.1080/01621459.1976.10480949
- Carral, L., Tarrío-Saavedra, J., Sáenz, A., Bogle, J., Alemán, G. and Naya, S.** (2021). Modelling operative and routine learning curves in manoeuvres in locks and in transit in the expanded Panama Canal. *Journal of Navigation*, 1–23. doi:10.1017/S0373463320000727
- Ferraioli, J. and Burke, R.** (2018). *Drowning in Data, but Starving for Insights*. Deloitte Insights.
- Mohovic, D., Mohovic, R., Suljic, M. and Njegovan, M.** (2021). Reducing the risk of collision between ships in a close-quarters situation by simulating collision avoidance actions. *Journal of Navigation*, 1–16. doi:10.1017/S0373463321000114 <https://blog.politics.ox.ac.uk/when-does-traditional-statistics-become-machine-learning/>
- Özlem, Ş., Or, İ., and Altan, Y.** (2021). Scheduling and simulation of maritime traffic in congested waterways: An application to the Strait of Istanbul. *Journal of Navigation*, 1–17. doi:10.1017/S0373463320000715
- Popular Lectures and Addresses vol. 1.** (1889). ‘Electrical Units of Measurement’, delivered 3 May 1883.
- Qi, L., Ji, Y., Balling, R. and Xu, W.** (2021). A cellular automaton-based model of ship traffic flow in busy waterways. *Journal of Navigation*, 1–14. doi:10.1017/S0373463320000636