Data, Science, and Global Disasters

John M. Chambers

Abstract.   The spread and impact of COVID-19 have disrupted human activities and energized a response of scientific activity on a remarkable, nearly unprecedented scale. This has somewhat distracted attention from a broad range of less immediate but fundamentally more serious global threats resulting from human actions. These can be collectively labelled the *anthropocene disasters*.

Science cannot itself prevent or mitigate them. To do so requires a global policy resolve not currently existing. When and if that resolve emerges, science will be essential for guiding action. This science will be radically data-intensive, global and inclusive. Teams will be required that include the best and most motivated individuals from all relevant scientific disciplines, plus members knowledgable about implementing likely policy recommendations. Such participants must be attracted to join and then properly supported and rewarded–not likely with current academic structures. Some insights can be gained from the recent experience with COVID-19 and the much less recent example of research at Bell Labs.

Key words and phrases: Data science, climate change, biodiversity, pandemic, Bell Labs.

1. INTRODUCTION

The COVID-19 pandemic is widely and justifiably considered a disaster on a global scale. Its severity, scope and duration have impacted people's lives to a degree not experienced over the preceding few decades at least. The response has included a wide range of scientific studies. The papers in this issue are examples of the challenges encountered and insights obtained from the resulting data-intensive science.

Without at all minimizing this impact (millions of deaths, social and psychological damage and economic disruption), other threats are more fundamental to the future of humanity and of life on the planet. The pandemic has to some extent lessened attention to these threats but the very involvement of science with it has lessons for future efforts to deal with them.

2. GLOBAL DISASTERS

Much on our planet is changing for the worse due to the impact of human activity. Recently, these changes have grabbed the attention of the general public to a lesser extent than a new disease, in part because their damage often

John M. Chambers is Adjunct Professor in Statistics and Senior Advisor for Data Science, Stanford University, Stanford, California 94305-4020, USA. He was in Bell Labs Research from 1966 to 2005 (e-mail: jmc4@stanford.edu).

occurs less rapidly, on a scale of years or even generations, and so looks less like a "disaster". Disasters there will be, however: our children will suffer, our grandchildren even more. The challenge is to act and to choose the action wisely.

The phenomenon most frequently noted, justifiably, is human-caused *climate change* and its consequences [5]. But our activity is having similarly dire affects on other aspects of the natural environment and the life that inhabits it.

The term *anthropocene* (e.g., [2]) has come to denote the current age, characterized by the dominance of human influence on climate and the environment. I propose to label the various resulting threats as the *anthropocene disasters*.

Alarms have sounded in the scientific literature and more widely in response to examples including:

- the loss of farmland to urbanization [12] threatening food supply and environmental quality, at the same time that the loss (partly *to* farmland [14]) of wild spaces and other human intrusions threaten the species using them [9];
- destructive changes in the world's oceans, vast as they seem, due to a variety of pollution, overfishing and direct destruction (e.g., bottom dredging) [17];
- rapid population loss and even imminent extinction for fundamental elements of the environmental food chain, notably insects [15];

- destructive effects of waste products from modern human society, notably plastics [10] but many others as well.

The anthropocene disasters are deeply interconnected in the "fabric of life", to use the powerful metaphor [3]. They will require direct, coordinated responses from us, the source of the problem.

So far, the overall global actions lag catastrophically short of being adequate. For all of the examples above, the direction of change remains harmful (in many cases at an accelerating rate) as noted in many of the references cited. Science is deeply involved in understanding the source and evaluating any proposed response to each of these.

However, one fundamental caveat needs to be stated (and emphasized).

> For none of these problems can science alone provide a solution. Only coordinated global policy, wisely chosen and strongly enforced, will be an adequate response to these global disasters.

The best response will be difficult to define, but almost certainly will require a variety of strong actions. These actions are likely to result in significant pain for large sections of the human population. Voluntary efforts will not be adequate, admirable as they are.

Given the will to adopt an effective policy, data-intensive science will be essential to choose the detailed actions to implement the policy. Here are some opinions on how such science should be organized and supported.

3. SCIENCE FOR ANTHROPOCENE DISASTERS

All the anthropocene disasters are global and fundamental, arising not from some local eruption or small-group activity, but from aspects of the vast human ecosystem. None will be easily mitigated.

What can be done, how best to do it, and the resulting consequences: these are questions that will need the best and most specific quantitative estimates. Mistakes will be made; policy must continually evolve to do better. The best, perhaps the only suitable paradigm for such evolution is the self-correcting pattern of science.

This will be *data-intensive* science needing more and better data than provided, for example, for the current pandemic. To make useful policy predictions will require varied and global data on humanity and its activities. And not merely nominally global (collected in different areas without coordination), but consistent, complete and accurate to the highest possible degree. As data collection for COVID has illustrated (e.g., [7]), data related to society and human activity tends to be far from this standard. Data can be improved by analysis (e.g., [6]) but with difficulty and only to an extent. Support for the science will need

to include policy action where crucial data is absent or inconsistent in some regions.

For much *non*-human data, the continuing progress in automating and instrumenting observations of the planet and its various life forms refines our ability to obtain the raw information needed (e.g., [8]). Challenges for using these sources come from the volume and varied nature of the input. Science for anthropocene disasters will need to extend the ability to handle the information and then turn it into scientific data.

The analysis applied to the data must guide policy actions in response to the threats. Experience with the pandemic has shown the benefits of coordination between science and policy (e.g., [13]). The policy actions needed for the anthropocene disasters will be more fundamental and more likely to imply economic or social disruption. Coordination of science and policy must be integral to the investigation.

The anthropocene disasters are all complex in their causes but critical for our future. The science needed to guide policy will be challenging to perform and will be required to produce important predictions integrated over the wide range of relevant disciplines. It follows that the scientific effort must attract the best and brightest minds and be given the essential resources to succeed.

From these observations, two essential requirements follow, in my opinion.

1. The science to deal with such challenges must be done by *teams* whose skills cover all specialties relevant to analyzing the problem and to implementing recommended responses. This includes science to understand the problem, science to assess the effects of resulting policy, and policy expertise itself.
2. The efforts must be *strongly supported* in all respects, including providing the data needed, implementing recommended actions, and rewarding the individuals who commit their careers to the cause. This last implies an approach to career evaluation and reward significantly different from typical current practice, particularly in academia.

Teams

The science described above can only be carried out by combining a wide range of skills. Its efforts must be integrated both longitudinally and laterally.

Longitudinal means in the *question–data collection–analysis–policy recommendation* temporal sequence. Too many inefficiencies and misunderstandings result when the science is done as "pure" science and then used to communicate with policy makers. Both scientists and decision makers must be active members of the teams that address these crucial problems. From a purely technical perspective, this allows both groups to see the detailed

picture of how policy affects likely results. More fundamental, however, is the increased understanding both sides will have from learning something of the other's approach and concepts. This is a lesson learned, for example, from joint projects involving research scientists and developers at Bell Labs.

The challenges posed also impose lateral integration across fields of scientific expertise. Here too, inefficiencies and inaccuracies rule out the typical one-discipline-at-a-time approach. The impacts of actions, or equally of the absence of action, will rarely fall conveniently in the expertise of any limited range of academic disciplines.

A scenario of possible actions to limit some human activity that currently contributes to an anthropocene disaster (for example, destruction of natural habitat) will have wide implications, both for the natural world (species survival directly plus possible climate and environmental quality impacts) and for human activities (social, economic and political). Future actions will need to be sizable, not just tweaking convenient targets but intervening on a scale that could reverse the direction of current disastrous changes. For that reason, all the impacts need to be modeled and predicted using the best data-intensive science for each aspect.

Real integration is needed. Just taking the final result from, say, an ecological analysis as input for studying the economic implications of a recommended action comes up short in both science and inter-personal relations. The discipline responding to the proposal is likely to see their task as measuring the bad news to decide whether to protest.

The data-intensive science needed will not be well served either. A full evaluation requires an analysis that explores and models possible actions simultaneously over as full a space of variables as possible. This requires integrating the analyses themselves; for example, to visualize outcomes of a range of actions as they jointly affect things to fix (e.g., species decline from wildspace becoming farmland) and resulting pain (food supply and price).

Integration at the analysis stage will be challenging if the disciplines differ fundamentally in their computational approach. Reaching a shared view of data structures, models and workflow will be worth the effort, however, since the analysis will then be able to respond better to its integration with policy questions.

Beyond this technical integration, experience with cross-discipline studies (certainly at Bell Labs, but also elsewhere I believe) has shown the essential value from participants understanding the key concepts of other participants' methodology. The teams must nurture this communication.

The collegial environment in such teams can be richly rewarding. All the same, goal-directed teams need leaders more broadly than does purely discovery-driven science,

both as a focus for communication and a buffer against undue interference. Management in Bell Labs research during its most productive period had somewhat similar requirements, in my experience, most importantly to understand and support the team members' activities, with careful nudges towards the overall goals when some adjustment seemed needed. The same mix seems likely to arise for the future teams but with explicit strategic questions more frequently relevant.

Support

To repeat once more: Science can only save our future once effective commitments are in place worldwide at the highest level. Then it must have the resources to succeed.

Resources basically require money. The scale will likely dwarf current research budgets but the needs and arguments for them are reasonably familiar from grant requests: mainly, infrastructure including computation, salaries and facilities.

The more serious challenge is in attracting the best people and rewarding them adequately. Particularly critical is the category of early-career researchers across the wide range of disciplines needed for these studies. They are likely to provide key ideas and energy toward the success of the effort. At the same time, particularly in the academic environment, they are currently caught in stressful career pressures. These and the typical evaluation criteria for advancement in academia would work against fully committing to such studies or in fact joining the team at all.

The response to the COVID-19 pandemic has generated a number of remarkable joint efforts, such as those reflected in the papers for this special issue. These teams have generally formed spontaneously from the initiatives of the participants, who have remained in their original academic or other research environment. Very substantial efforts have been needed to obtain and analyze valid data, and then to report the results effectively to those making decisions, as well as to the public. Early-career contributors risk suffering for less time spent on academically rewarded research and publication, as noted in the perspective by Mukherjee [11].

The scientific efforts needed for the anthropocene disasters will require longer term commitments to a distinctly different enterprise from ordinary scholarship. No quick fixes are likely. The teams need very strong support and prestige, justified by the importance of their contribution to our planet's future. Serious involvement will be a career choice. It must not be a damaging one to a promising researcher. But the present academic merit and promotion system, with its discipline-specific evaluation and tendency to measure quantity of publications, is not positioned to adequately reward such commitments.

Undoubtedly there will be significant scientific insights gained along the way, but the focus of the team must be

on finding a path for action, likely the result of collective contributions from multiple disciplines. This is not the way to maximize contributors' publications or personal prominence. In academic merit assessment, faculty from a single discipline will be challenged to evaluate an individual's contribution to such a multi-discipline effort. Evaluation by a broader faculty body will be hampered by lacking general expertise on the technical content, tending to result in a fallback to publication records or other documentable summaries.

Reform of the academic merit and promotion system is an active topic, but often focusing (justifiably) on lack of fairness or insufficient regard for non-research activities (e.g., [16]). General support for inter-disciplinary research is often expressed, but how this will work here is unclear.

A relevant model comes from Bell Labs research during a period when that organization was the source of a number of significant advances. Gertner [4] gives a non-technical but reasonably accurate overview; a perspective for data science is briefly outlined in my paper [1], Section 2. At this time, Bell Labs was able to offer competitive salaries and equipment. The common recruitment strategy was to attract talented new PhDs and support them in largely independent research within the general areas of interest to Bell Labs. There were enough notable successes to provide the option of taking up a tenured academic position on the basis of the Bell Labs research record.

As a result, Bell Labs was an attractive target for young researchers anxious to pursue novel directions. My own decision to go there was an example: in the mid-1960s statistical computing as an area of research was not generally recognized in academia, but the research leadership at Bell Labs were excited to be involved. Then and over the next several decades a substantial number of early-career researchers, including several of my colleagues, followed a route from Bell Labs to valuable academic careers.

The scientific teams tackling the anthropocene disasters can have a similar role. Two aspects are needed: prestige for the science being done and an internal mechanism for proper recognition and reward.

Merit rewards must be part of the teams' support. Outstanding contributions must be sufficiently recognized to make them valid coin for a participant's professional advancement. Such recognition must be reflected both in compensation and in explicit, publicized awards–all while maintaining a team spirit: the overriding goal remains scientific contributions to dealing with the disaster. Making clear and well-justified merit judgements is challenging in teams with the wide range of disciplines required here.

Some characteristics of merit review in Bell Labs research are relevant. Individuals' merit ranking was jointly resolved by the managers of a center containing related departments (for example, data analysis research in the mathematics center). Ideally, the reviewers balanced technical knowledge of the research with appreciation for breadth. The review process was nested hierarchically in a couple of layers up to the whole of research. Funds to be used for raises in salary were divided up following the same hierarchy from the top down. When the system worked well (which was fairly often) good research was rewarded appropriately, at an encouraging but not spectacular financial level. The research management at all levels came from Bell Labs research themselves, helping to build a sense of community.

There was in fact quite a strong feeling for what constituted "good research", varying somewhat among areas but broadly consistent. It was generally easier to argue for an "important" idea than for several minor ones; that is, for an idea that was either conceptually path-breaking or had substantial practical prospects (ideally both, of course). Joint research was encouraged, indicating that ideas were being shared and evaluated; cross-discipline joint research was an additional plus.

Behind all the evaluations was the ultimate notion of contributing to a common goal. In this case, technically, the interests of the parent corporation. Fortunately, in this period AT&T in its semi-monopoly position was anxious to show that it was advancing science and contributing to the good of the community at large.

Similar characteristics for managing research are needed for science dealing with the anthropocene disasters: rewards linked to the value of a contribution to the common goal; thorough inclusion of diverse disciplines in the evaluation; and encouragement for ideas that make a difference.

4. CONCLUSION

Much that is best on our planet faces grave threats—the anthropocene disasters—for which we humans are responsible. Only strong global action can mitigate these disasters. Any chosen approach will have significant consequences, some of which are likely to be both unexpected and painful. Policy must be a learning process. Science is the essential guide to this through its ability to improve on imperfect actions via further data collection and analysis.

Responding to the anthropocene disasters will require a scientific effort having a scale and breadth not achieved before. The teams carrying it out must combine strong research capability with a focus on shared goals, while providing an attractive career path for the talented individuals who will be essential. The response to the current pandemic and the earlier Bell Labs research environment provide some guidelines.

Right now, hopes for the future require finding the will to act (towards which I have no useful advice) while nurturing the data-intensive science needed to guide us, given the will.

REFERENCES

[1] CHAMBERS, J. M. (2020). S, R, and data science. In *Proc. ACM Program. Lang.* **4** Article 84. Association for Computing Machinery, New York, NY, United States. https://doi.org/10.1145/3386334

[2] DAVIES, J. (2016). *The Birth of the Anthropocene*. Univ. California Press, Oakland, CA, USA.

[3] DÍAZ, S. (2022). A fabric of life view of the world. *Science* **375** 1204. https://doi.org/10.1126/science.abp8336

[4] GERTNER, J. (2013). *The Idea Factory*: *Bell Labs and the Great Age of American Innovation*. Penguin, Baltimore.

[5] IPCC (2022). Sixth assessment report. International Panel on Climate Change.

[6] JAHJA, M., CHIN, A. and TIBSHIRANI, R. J. (2022). Real-time estimation of Covid-19 infections: Deconvolution and sensor fusion. *Statist. Sci.* **37** 207–228.

[7] JHA, P., DESHMUKH, Y., TUMBE, C., SURAWEERA, W., BHOWMICK, A., SHARMA, S., NOVOSAD, P., FU, S. H., NEWCOMBE, L. et al. (2022). COVID mortality in India: National survey data and health facility deaths. *Science* **375** 667–671. https://doi.org/10.1126/science.abm5154

[8] KEITT, T. H. and ABELSON, E. S. (2021). Ecology in the age of automation. *Science* **373** 858–859. https://doi.org/10.1126/science.abi4692

[9] LUGHADHA, E. N., BACHMAN, S. P., LEÃO, T. C. C., FOREST, F., HALLEY, J. M., MOAT, J. et al. (2020). Extinction risk and threats to plants and fungi. *Plants*, *People*, *Planet* **2** 389–408. https://doi.org/10.1002/ppp3.10146

[10] MACLEOD, M., ARP, H. P. H., TEKMAN, M. B. and JAHNKE, A. (2021). The global threat from plastic pollution. *Science* **373** 61–65. https://doi.org/10.1126/science.abg5433

[11] MUKHERJEE, B. (2022). Being a public health statistician during a global pandemic. *Statist. Sci.* **37** 270–277.

[12] NARDUCCI, J., QUINTAS-SORIANO, C., CASTRO, A., SOM-CASTELLANO, R. and BRANDT, J. S. (2019). Implications of urban growth and farmland loss for ecosystem services in the western United States. *Land Use Policy* **86** 1–11.

[13] NICHOLSON, G., BLANGIARDO, M., BRIERS, M., DIGGLE, P. J., FJELDE, T. E., GE, H., GOUDIE, R. J. B., JERSAKOVA, R., KING, R. E. et al. (2022). Interoperability of statistical models in pandemic preparedness: Principles and reality. *Statist. Sci.* **37** 183–206.

[14] ROBINSON, S. K., THOMPSON III, F. R., DONOVAN, T. M., WHITEHEAD, D. R. and FAABORG, J. (1995). Regional forest fragmentation and the nesting success of migratory birds. *Science* **267** 1987–1990.

[15] SÁNCHEZ-BAYO, F. and WYCKHUYS, K. A. (2019). Worldwide decline of the entomofauna: A review of its drivers. *Biol. Conserv.* **232** 8–27.

[16] SCHIMANSKI, L. A. and ALPERIN, J. P. (2018). The evaluation of scholarship in academic promotion and tenure processes: Past, present, and future. F1000Research 7.

[17] STAFFORD, R. and JONES, P. J. (2019). Viewpoint–ocean plastic pollution: A convenient but distracting truth? *Mar. Policy* **103** 187–191.