

Data-driven solutions

In an age with more climate data than ever before, how can we harness it to drive genuinely evidence-based climate action?

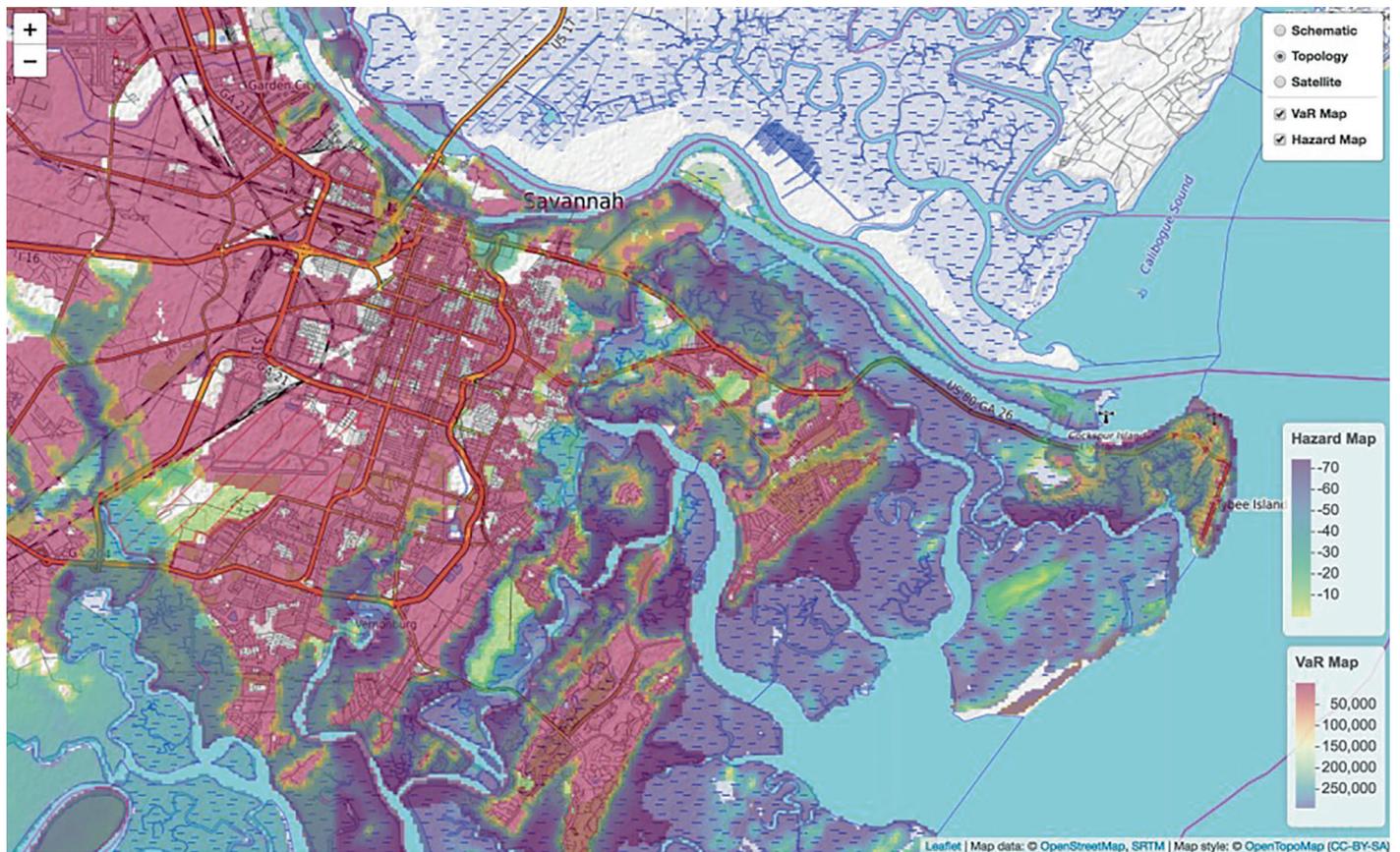
By Auroop R. Ganguly, Evan Kodra, Udit Bhatia, Mary Elizabeth Warner, Kate Duffy, Arindam Banerjee and Sangram Ganguly

The fourth and fifth assessment reports by the Intergovernmental Panel on Climate Change have declared global warming to be “unequivocal” and anthropogenic drivers to be “extremely likely” as the dominant cause. These conclusions, expressed in cautious scientific language governed by strict criteria, need to serve as a clarion call for action – from urban communities to world bodies – to



© ESA/NASA-A. Gerst

Figure 1. Blended hazard and value-at-risk map of Savannah, GA, USA



prepare for what is likely, and to prevent the worst from happening.

Climate adaptation and mitigation have been respectively called “managing the unavoidable” and “avoiding the unmanageable”. While individuals, communities and nations may understand that there are inherent costs to both climate action and inaction, developing a comprehensive, evidence-based, scientifically credible and risk-informed action framework is not straightforward.

National and global climate mitigation policies include investments in renewable energy, carbon capture and storage solutions, divestments in fossil fuels, and environmental and land-use regulations. The will of nations to act may depend on perceptions of climate

impacts and vulnerability on their respective societies and institutions – and on the challenges in adaptation.

Impacted sectors include natural resources (such as food, energy, water, ecosystems), hazards and humanitarian aid (for example, critical infrastructures resilience), as well as population growth and movement (such as environmental refugees).

However, while predictive insights from climate models and data are usually more credible at aggregate scales in space and time, climate action may be better suited at the local scale such as within urban communities. Urbanisation contributes significantly to emissions and land-use change, and hence to climate change, while urban areas are significantly impacted by climate change.

Communities in urban, peri-urban or rural regions need to understand, adapt to and mitigate risk elements. These include global-scale and locally exacerbated hazards (such as global warming and urban heatwaves),

▲ Figure 1. Here we show an example of risk assessment by our startup, risQ, which blends diverse concepts and disparate data sources. Coastal floods are modelled for multiple return periods under sea-level rise. Inundation depths and extents are intersected with geospatial layers, including property value and industrial workforce concentrations, over a municipality. This allows for estimating the municipal tax revenue at risk from climate change. Analysis on comparative tax dollars at risk under climate change could help bond investors and financial rating agencies better quantify the true municipal and state credit risk. This map exemplifies modelled 10-year tidal flood depth and extent (‘hazard map’) under current sea-level conditions in Savannah, GA, overlaid with a workforce density index (‘value-at-risk map’) to help stakeholders estimate industrial output risk

vulnerabilities of infrastructures and lifelines (including natural-built and grey–green infrastructures), as well as exposure of economic assets, ecosystem services and human populations.

Data-driven understanding and predictive insights can improve risk-informed adaptation and mitigation in three ways: (i) through improved understanding of earth

◀ Hurricane Florence approaching the east coast of the USA, as viewed from the International Space Station. Although climate is a ‘big data’ science, only a small amount of the data has been systematically analysed

systems science and engineering, translating to the probabilities and attributes of stresses and shocks; (ii) through risk frameworks, including risk assessments, which consider threat, vulnerability and exposure, emergency management (including preparedness and recovery), as well as time-phased and flexible adaptation strategies; and (iii) mitigation at multiple scales, from global and national to urban and community.

Climate model simulations and remotely sensed observations already exceed petabyte scales (one petabyte equals 1,000 terabytes) and are expected to reach a few hundred petabytes within the next couple of decades. But even though climate is now a ‘big data’ science, only a small fraction of the available data has been systematically analysed. Furthermore, even as ‘big models’ are increasing in space-time resolutions and complexity, this is not necessarily leading to more certainty in stakeholder-relevant insights.

Machine learning

So, a group of climate modellers have resorted to machine learning (ML) – a

subfield of artificial intelligence – to estimate parameters for high-resolution atmospheric processes such as convection. Others have explored ML-based post-processing of model simulations, often guided by physics to obtain finer-scale projections. Beyond atmospheric science, terrestrial ecology has benefited from ML through the creation of a global plant attribute database, which in turn used an advanced data-driven parameter estimation method within numerical models.

Despite these efforts, our lack of understanding of complex climate processes and feedback, as well as sources of irreducible uncertainty, may persist.

First, greenhouse gas emissions and land-use change scenarios that drive the models are not precise predictions with probabilities, but are what-if scenarios.

Second, gaps in our knowledge of the climate system may not be easily plugged. Uncertainties result from variabilities across model simulations as well as their lack of correspondence (when models are hindcast into the past) with observations.

Third, inherent variability exists in the climate system, including extreme sensitivity

to initial conditions, which contributes to the irreducible component of uncertainty.

These three components contribute to the overall uncertainty. In the crucial 0–30 year near term, the projected climate change signal may be within the bounds of this overall uncertainty, which may in turn be dominated by the inherent variability.

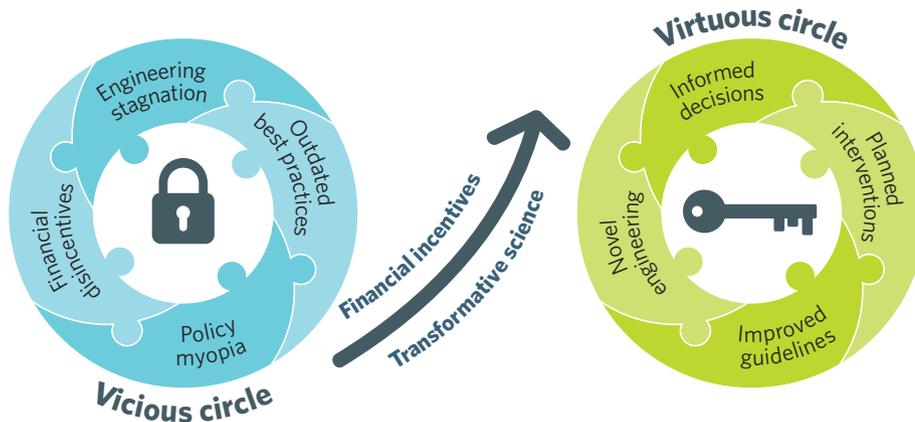
Climate challenges go beyond ‘big’ models and big data. Indeed, climate science is also dominated by what may be viewed as ‘small data’ challenges. Historical records from the data-poor eras of earth science are sparse,

▼ **Figure 2.** In 2012, Hurricane Sandy inundated New York City, causing billions of dollars in damage and costing the Metropolitan Transportation Authority (MTA) an estimated \$124 million in lost revenue. Using a system-recovery modelling approach, the MTA could have hypothetically saved an estimated \$56 million dollars and restored service in less time by strategically prioritising station recovery. Left top: percentage of service restored is shown step-by-step as stations are recovered in sequence. The yellow line is based on the actual sequence of recovery, and the blue is driven by the recovery model. Left bottom: same but for estimated MTA revenue, as inferred using historical MTA data. Middle and right: a snapshot of the modeled MTA recovery mid-process, for the model and the actual historical recovery, respectively

Figure 2. Recovery model for the New York City metro after Hurricane Sandy



Figure 3. Breaking the maladaptation cycle



making historical reconstructions difficult to validate. Climate signals exhibit temporal fluctuations ranging from sub-seasonal to multi-decadal and even longer time scales (low-frequency variability).

The lack of historical data makes it challenging to understand changes or delineate signals from longer-term variations, although reconstructions based on models, instrumented records (even if relatively sparse), and proxy data (such as tree rings, fossils and ice cores) help to partially address aspects of the information gap.

Furthermore, climate data challenges are made worse by complex dependence. Tobler's first law of geography states that "everything is related to everything else, but near things are more related than distant things". While climate data exhibit this property in space and in time, long-range spatial dependence and persistence in time are also common.

Finally, climate change is not just a matter of mean change (for example, global warming) but also about changes in the patterns of extremes such as heatwaves, cold snaps, heavy rain, droughts, floods and hurricanes. Weather extremes turn to catastrophic disasters when hazards (e.g., a hurricane) are aligned with infrastructural (e.g., an inadequately designed dam or levee) and societal (e.g., economic disparity) vulnerability, exposure of people and assets (e.g., businesses and natural resources), as well as lack of emergency management plans. The relative rarity of such extremes adds to

the 'small data' challenge and brings to the fore the need to manage, assimilate, analyse and interpret heterogeneous information.

Nonetheless, extracting predictive insights about the statistics of change and extremes is possible based on specialised data-driven methods such as extreme value theory, network science and signal processing. Thus, our research has examined complex dependence patterns, low-frequency variability in climate (including for extremes) and developed predictive insights. We have studied heatwaves and cold snaps, heavy precipitation, high winds, droughts and urban climate extremes.

Our work on droughts and heavy precipitation has examined long-memory processes and teleconnections. We have discussed deep uncertainty (i.e., where probabilities cannot be easily assigned) and non-stationarity (i.e., significant and fundamental change) in climate and hydrology, as well as in climate adaptation and resilient engineering, and the possibility of blending physics and data sciences to address these challenges. Figure 1 illustrates how simulations and observations from earth system science combined with ancillary information may help generate predictive insights in climate and develop risk assessments.

Attribution studies, where change patterns are related to possible causes, are typically based on observations and model simulations. This is one area where we

believe the climate science community can benefit significantly by interacting with a wider group of interdisciplinary scientists.

Stakeholders such as the US Department of Defense indicate that climate change is a threat multiplier across many sectors, and hence adaptation and mitigation are urgent and necessary. Data challenges in adaptation and mitigation sectors are diverse and disparate – ranging from big data to small data, information gaps and confidentiality issues – and are exacerbated by gaps in understanding processes and the possibility of cascading failures.

Data-driven methods, data-informed, process-based approaches and physics-informed, data-science methods have all been found to be useful. Robust decisions and flexible-planning pathways have been suggested. We have examined coastal processes, water-energy nexus, transportation networks, public health and urban heatwaves, and regulatory principles. Figure 2, for example, shows how recovery strategies designed in anticipation of weather extremes can help save lives and money.

Future work may need to creatively leverage data from the public domains or from well-crafted simulations and testbeds. It may also incorporate confidential data that could still be used either through anonymisation or by following privacy regulations. The state of the art in critical infrastructures resilience offers specific examples.

Finally, the importance of economic incentives to overcome hurdles to best practice or to engineering innovation, as well as to policy myopia, cannot be overemphasised. These in turn may require analysis of financial, demographic and socio-economic data. Figure 3 suggests how a vicious cycle of maladaptation may be transformed to a virtuous cycle through improved incentives and innovations. ●

Note: AR Ganguly, Bhatia, Warner and Duffy are at Sustainability and Data Sciences Laboratory (SDS Lab) of Northeastern University (NU) in Boston, MA, USA; Kodra is at the startup risQ (spinout of the SDS Lab) in Cambridge, MA; Banerjee is at the University of Minnesota (computer science) in Twin Cities, MN; S Ganguly is with the NASA Ames Research Center at Moffett Field, CA. An accompanying list of references is available online at www.climate2020.org.uk