

Table of contents

Overview	3
Introduction	4
AI and Data Science in Environmental Research	6
Monitoring the environment	9
AI method in focus: Computer Vision	10
Spotlight on scientific image analysis	12
Scivision, Connecting computer vision model developers to image data providers . .	12
More examples	14
Monitoring systems at scale	14
Forecasting the environmental change	16
AI Method in focus: Deep Learning	17
Spotlight on seasonal ice change	19
IceNet, ice loss forecasts for the season ahead	19
More examples	21
Forecasting to build a sustainable and safe society	21
Simulating the human cost of climate change	23
AI method in focus: Microsimulations	25
Spotlight on producing projections for human population	27
DyME, quantifying the environmental impacts	27
More examples	29
Engagement-led adoption, extension and deployment of AI tools	29

Adapting to climate change	32
Contributing to Sustainable Development Goals	33
Spotlight on the impact of climate change on agriculture	35
Building the UK's Crop Modelling Framework	35
More examples	38
Multi-level interventions for mitigation and adaptation	38
Recommendations	41
Addressing challenges in conducting AI/ML research	42
Interdisciplinary and inclusive collaboration	43
Open, Reproducible and Ethical AI Technology	45
Cross-disciplinary Data Integration and real-world deployment	47
Common Understanding of AI Applications	49
Acknowledgements	51
References	52
Overview	52
Introduction	52
Monitoring	52
Forecasting	53
Simulating	54
Adapting	55
Recommendations	57

Overview

[The Alan Turing Institute](#) brings together researchers, data scientists and expert practitioners to explore, understand and apply Artificial Intelligence (AI) to address global challenges spanning environmental, healthcare, social, economic, political and engineering systems. The Turing, as the national centre for data science and AI in the UK, convenes experts from different domains to collaborate at the national and international levels, raising awareness of the ethical use of AI, influencing policy-making and integrating digital efforts across sectors to build a more sustainable future for our society.

In this report we highlight projects from [AI for Science and Government \(ASG\)](#), a five-year, £38.8 million research programme that has accelerated and scaled AI efforts through multi-disciplinary, collaborative and innovative data science.

This report highlights ASG projects under the Environment and Sustainability theme. All projects have been carried out in partnership with UK-wide stakeholders in academia, industry, and government sectors, highlighting their outcomes and impacts on data science and AI to tackle environmental and climate challenges.

Under the following areas of environmental and climate research, we have highlighted examples from ASG research in this report:

- Monitoring the environment.
- Forecasting environmental change.
- Simulating the human cost of climate change.
- Adapting to climate change.

This report provides successful examples and evidence for AI/Machine Learning implementation in environmental research while addressing the following challenge areas:

- Facilitate interdisciplinary, inclusive and equitable collaboration.
- Create open-source, reproducible and ethical AI technology.
- Pioneer integration of cross-disciplinary data and enable its real-world deployment
- Build a common understanding of AI applications in the environment.

A white paper summarising the learnings and recommendations drawn from this work has been published online as [Tackling climate change with data science and AI](#). Please cite that as *Conner, A., Hosking, S., Lloyd, J., Rao, A., Shaddick, G., & Sharan, M. (2023). Tackling climate change with data science and AI. Zenodo. doi: 10.5281/zenodo.7712969*

Introduction

The effects of climate change and environmental instability constitute some of the most pressing issues facing our society – affecting a large-scale interconnected system spanning nature, society and economic infrastructure. This has wide-ranging consequences for the local population, natural ecosystems and global climate. High-quality, accessible, and reliable information from different sectors can help build a comprehensive view of these complex issues. Modern data science techniques have already advanced our ability to integrate data and methodologies to find sustainable solutions from across all sectors. AI and Machine Learning (ML) can further deepen our understanding of climate change and inform evidence-based decision-making across a suite of policy areas.

AI for Science and Government (ASG) is a 38.8 million programme established at The Alan Turing Institute through UKRI's Strategic Priorities Fund in 2018. ASG research brought together several organisations and experts to build national leadership in data science and AI, developing innovative techniques for fundamental and applied sciences as well as humanities research. Delivered in partnership with the Engineering and Physical Sciences Research Council (EPSRC) and in collaboration with a number of other research councils, ASG's five-year vision was to demonstrate, via a diverse range of activities organised and presented as an integrated whole, how AI can be applied to effectively address urgent societal challenges and transform environmental research, engineering and government.

The Environment and Sustainability theme, one of three overarching areas of research in ASG, researchers undertook the role to understand and address one of the defining crises of our time — the climate emergency. This cross-theme collaborative initiative has enabled the development of AI technologies combining scientific, industrial, public and governmental efforts in environmental research. Under this theme, ASG has funded several initiatives led through multi-stakeholder partnerships, with the most significant strategic investments made towards the Data Science for Science and Humanities (DS4SH), Data-Centric Engineering (DCE) and Tools, Practices and Systems (TPS) programmes. DS4SH and DCE have evolved into large research programmes, instrumental in extending AI research at the institute and fostering an extensive portfolio of work in environment and sustainability. Experts are developing tools, methods and models to draw meaningful insights from data from across disciplines, improving risk management and guiding our response to climate change. TPS was established as a cross-disciplinary research programme offering leadership in open research and reproducibility by enabling inclusive collaboration and deploying interdisciplinary computational frameworks for managing the integration of data, methodology and expertise from different fields and sources. The long-term goals of these programmes will build on the legacy of ASG, shaping national and

international strategies for building a sustainable future. ASG is supported entirely by public funds, through the UK Research and Innovation Strategic Priorities Fund, under EPSRC Grant EP/T001569/1 and EPSRC Grant EP/W006022/1.

In this report, we discuss projects involving data, software and methodologies under the Environment and Sustainability theme of ASG. Underpinning these ambitious efforts to tackle our greatest environmental and climate challenges are the processes for reproducible, ethical and collaborative research in AI and data science. Ensuring meaningful participation from diverse stakeholders has remained fundamental in different stages of the data life cycle, from planning and conducting research to developing methods, tools and technology. The Turing’s community building and convening capacity have contributed greatly towards strengthening interdisciplinary AI research and communities, enabling cross-domain relevance and applications by diverse user groups.

AI and Data Science in Environmental Research

News about natural disasters is dominated by unsettling images of melting glaciers, deforestation, forest fires, extreme floods and other potentially destructive events that are becoming more frequent due to climate change. These news stories raise awareness of and concern for our environment, creating a sense of urgency for public. Research-based evidence and scientific data along with public-led advocacy, have been compelling government and policymakers to take immediate action on sustainability crises that have existed for decades.

In environmental research, images used for social awareness are only a small part of the climate change discourse and a fraction of data gathered daily from different sources. An ocean of scientific data serves as an invaluable source of information that not only shapes our understanding of climate catastrophe but also guides data-based monitoring, forecasting and simulations of different conditions.

Insights from data analysis and predictive models help us understand the earth's processes and systems better and prepare to respond to climate change. Stronger mitigation measures for reducing greenhouse gas emissions, and changes in electricity networks, transportation, buildings, industry and land use are required. Research-based mitigation strategies should guide the qualitative and quantitative increase in energy efficiency, enhancing the capacity of carbon sinks (such as forests and oceans) and sustaining natural systems. The science shows that even the most effective climate change mitigation will not be enough to entirely prevent further climate change impacts, thus making the need for adaptation unavoidable [Nelson, Adger and Brown, 2007]. Climate change adaptation aims to build resilience by reducing harms caused by extreme weather events and other natural disasters by building the capacity to prepare for, recover and adapt to the changing climate. This necessitates cross-disciplinary research that combines knowledge of our environmental systems utilising heterogeneous cross-sector data through the development of climate models and predictions.

In this report, we discuss AI in Environment and Sustainability and provide examples from ASG under the following four categories:

- **Monitoring the environment:** AI tools, data models and workflows have advanced environmental monitoring capability by combining earth observations from satellites with those from the ground sensors, bringing together data of different quality, timescale and resolutions, with novel mechanisms to reduce uncertainty.

- **Forecasting environmental change:** The application of deep learning on accumulated data has helped forecast environmental changes accurately and use of energy systems effectively, creating possibilities to mitigate the extent of climate change.
- **Simulating the human cost of climate change:** Beyond giving deeper insights into environmental changes, multi-domain data have been used to develop effective AI approaches to build climate interventions, simulate their impact on populations and inform effective policies for implementation.
- **Adapting to climate change:** Bringing combined knowledge from other areas of AI, Environment and Sustainability projects have undertaken a broad role in responding to the environmental crisis through technology and policies for mitigation and adaptation.

To demonstrate the successful implementation of AI-based solutions in these areas, among many exciting projects, we highlight Scivision which applies computer vision models and datasets at various scales – from microscopic to satellite; IceNet, a deep learning-based AI predictive tool that produces accurate Arctic sea ice forecasts, outperforming physics-based models in forecasting Arctic sea ice change; and DyME-CHH microsimulation-based toolkits that assist local authorities in creating action plans to address issues ranging from pandemic to climate crisis.

Following the standard of “as open as possible, as closed as necessary”, researchers have involved different stakeholders to build user-friendly open-source computing platforms, provide open access to resources for understanding AI technology as well as open up underlying data and software. Further practices have been systematically applied for sharing outputs and knowledge to enhance accessibility, interpretability and reuse of different components from the projects. Considerable attention has been given to building broadly applicable, reliable and explainable models that take uncertainties and limitations of curated data into account. Development and dissemination of reusable tools, methods and data have ensured that solutions from environmental research can be deployed across a wide variety of applications in different contexts, compounding the long-term impact of the work.

Building a resilient future will require increasing public awareness of climate change, providing effective interventions and accelerating the adoption of modern solutions in different sectors. By offering data-intensive insights and sustainable solutions for different climate change scenarios, the AI/ML work discussed in this paper has informed policies and decision-making at all levels, from individuals and organisations to government and society. Impactful bodies of work include a solar energy generation forecasting project that used citizen science approaches to crowdsource solar panel location information in the UK and forecast the production of solar-based renewable energy in order to minimise the circulation of fossil fuel-based electricity production. Another project called Synthetic Population Catalyst (SPC) integrates population data to create an open dataset of synthetic populations that can be used in complex simulations to help inform policymakers. Tools like EnergyFlex have allowed local authorities to explore inequalities in energy efficiency and target homes in need of retrofits as well as predict which residents may require support for fuel poverty. The Turing’s leadership on the

applications of AI in environmental research have built stronger evidence for innovative interventions, taken alongside nature-based measures for restoring natural ecosystems and making our society more resilient to climate change.

Monitoring the environment

Space-based radars or satellites, unmanned aerial vehicles (drones) and ground-based sensors capture data and measure changes in our environment at various levels. Data from both remote sensing and *in situ*, measurement techniques are combined to observe different ecological systems, such as oceans, forests or volcanoes, and can monitor their conditions over time. AI applications using data from monitoring systems are helping us characterise environmental quality, establish parameters to identify when the quality is compromised and understand the impact those changes have on an environment. AI monitoring technologies are proving to be exceptionally important in translating data into timely responses and improving measures for protected areas and species vulnerable to natural hazards.

AI tools and data-driven models for analysing satellite data allow us to look at the patterns caused by climate change providing a granular level of detail of the Earth’s surface. Ground-based measurements have been critical to understanding specific changes in the atmosphere as well as their impact on local populations (Gallant et al, 2018). However, their limited scope and area of inference make it challenging to reliably validate those observations in isolation. To make new leaps in understanding environmental change and to improve predictions based on environmental monitoring systems, we need intelligent ways to combine satellite data with surface sensors and the output from physics-based climate models. Computer vision approaches provide an AI-based solution for intelligently making sense of visual data such as satellite images. In combination with algorithms to infer outcomes, computer vision demonstrates huge potential for applications for retrieving and interpreting visual information from environmental data.

AI method in focus: Computer Vision

Scientific images, videos and other visual information are routinely generated from microscopy, scanning devices, sensors and satellite technologies capturing measurements of changes in different natural systems – ranging from seeds to geospatial landscapes. Manually analysing, classifying and combining information from such large collections of visual data on a varying scale is not only impractical and error-prone but also highly inefficient. Researchers across all areas of interest increasingly rely on AI/ML technology to sift through big data efficiently and reliably. AI technology has also made it possible to automate the detection, classification, tracking and measurement of objects or regions of interest in images using computer vision techniques. Computer vision, like human vision, learns from previous observations or training data to make predictions and decisions in future observations. To avoid errors and biases, computer vision tools are trained on large amounts of visual data to process images at a pixel level and characterise them correctly through pattern recognition.

The remarkable evolution of AI/ML methods and computing power, alongside the generation of large amounts of public image data, has led to the recent developments in computer vision technology. Nonetheless, the majority of datasets used for training computer vision models and algorithms come from transportation, autonomous or computer-assisted automobiles,

healthcare, agriculture, retail and construction industries. Largely driven by uneven incentive structures for advancing computer vision in industrial settings, applications of computer vision in scientific and exploratory research have remained underutilised. Another challenge of adopting computer vision in scientific settings is that established models are often trained on RGB (red, green, and blue) data, whereas most image sensors use scientific monochrome image data generated from microscopy images, medical scans, thermal imagery, hyperspectral imagery and radar. There are also challenges associated with reusing existing computer vision tools, models and algorithms trained on natural images that are typically oriented according to our own viewpoint and not on objects in all orientations. For instance, pedestrian or traffic flows tend to have a fixed orientation; images of ice flows, forests or landscapes as seen by satellite are not captured from a consistent viewpoint. Trained to recognise patterns from multidimensional visual data, computer vision technology can significantly improve the possibilities to monitor and observe changes in environmental behaviour by identifying factors influencing climate change.

We recommend [General References on Computer Vision](#) from The Computer Vision Handbook, maintained by, and copyright by, Margaret Fleck of the Computer Science Department at Harvey Mudd College for gaining technical details on Computer Vision.

Spotlight on scientific image analysis

Scivision, Connecting computer vision model developers to image data providers

Scivision is an open-source computer vision toolkit. Scivision provides researchers with an interface to search, discover, and integrate datasets and AI algorithms from a wide range of research areas. Scivision grew organically out of the Environment and Sustainability theme to support our objectives for open and reproducible research by creating data pipelines that limit duplication of development work. Scivision is collaboratively co-designed by researchers from within the Turing as well as external partners. Embracing the interdisciplinary model of data science, Scivision brought together researchers from different domains who use computer vision to study image data including plant phenotyping experiments, biomolecular characterisation by electron microscopy and satellite-imaging-based landscape monitoring. Their specific focus in Scivision was to address the common need to identify generalisable digital infrastructure that could make it easy for them to apply AI/ML methods across various data challenges. Their collaboration led to the development of an analytic framework that is agnostic to the choice of the algorithm as well as the intended application of the outputs.

The overarching aim of this project is to develop and deploy technological solutions for computer vision-based image analysis problems commonly found across the sciences and humanities, with a specific focus on the environment. Experts in diverse scientific domains use Scivision's general-purpose Python toolkit to improve their framework for collecting and analysing data. The Scivision platform contains searchable catalogues of pre-trained computer vision algorithms and image datasets, allowing users to test out a range of algorithms in order to choose the one that best fits their data and application. The core features of Scivision are: 1) the Scivision catalogue, containing pre-trained computer vision models and datasets from science and the humanities, and 2) the Scivision model and data source API, a simple, standard interface to models and data that works with the entries in the catalogue.

Scivision members have developed tools and infrastructure that democratises AI/ML and encourages exploratory use of existing toolkits for their applications from one field to another. The Scivision team collaborates with researchers in the UK to build examples showing how the tool can be used to perform tasks such as coastal vegetation detection and the identification of trees from satellite images. Researchers with any level of programming skills can pick up these examples and try them out with minimal intuitive changes to the executable code.

Testing and demonstration of Scivision’s features have been made possible via interactive Jupyter notebooks from research projects at the [Scivision Gallery](#) along with tutorials for new users. These notebooks can be run in the cloud through a service called [Binder](#), which is supported in part by cloud infrastructure provided by the Turing. A standard graphical interface providing images and visual cues further allows environmental researchers, along with other non-coding user communities, to search for datasets and models compatible with their data of interest. Focusing on tackling real-world challenges, Scivision provides reproducible and reusable solutions for computer vision problems commonly found across the sciences and humanities. By enabling exploratory analysis, Scivision will help guide decisions on AI tools and underlying algorithms to be applied in the wider contexts of environmental research.

More examples

Monitoring systems at scale

The observational data available to researchers is growing substantially and is vast and varied, helping us understand physical processes on scales ranging from centimetres to kilometres. AI-powered autonomous robotic platforms provide complementary solutions to monitoring activities requiring human presence. AI tools don't fully replace the need for human-led collection and validation of observational data but can be used to unify datasets from those various monitoring platforms in order to provide a holistic picture.

Since the middle of 2019, researchers at the Turing and [British Antarctic Survey](#) (BAS) have been harnessing powerful AI algorithms to extract information from vast atmospheric, oceanic and sea ice datasets from the polar regions. With tens of millions of data points from our satellite-derived datasets, we are training our AI algorithms to [predict future sea ice](#), with the ability to learn physical relationships between climate variables over both space and time. Additional AI applications in ASG research that have significantly improved environmental monitoring capacity include the [project for monitoring iceberg populations in the Southern Ocean](#) from space. This project applies ML techniques and synthetic aperture radar (SAR) satellite imagery to identify icebergs and track their disintegration in the Amundsen Sea, Antarctica. Approaches from this work will improve the identification of icebergs in both the open ocean and within the sea ice pack.

Closely related is [Seals from Space](#), another project led by BAS for automating Antarctic ecosystem monitoring via high-resolution satellite imagery, tracking seals as potential indicators for the Antarctic ecosystem's health. An AI-enabled system for classifying sea ice and mapping seals can be used to transform satellite images into numbers representing observed ice and seals. Using these environmental features, researchers can address ecological questions concerning seals' preferred habitat and how their surroundings are changing over time.

The DeepSensor toolkit leverages cutting-edge advances in probabilistic AI modelling to intelligently fuse data from gridded satellite imagery and point-based surface sensors. In particular, DeepSensor enables new solutions in two problem areas: (1) increasing the spatial granularity of surface climate variables, generalising across spatially sparse and dense environmental sensor networks, and (2) optimising sensor placement to reduce uncertainties in climate variables and providing robust information to environmental monitoring groups. As the toolkit is being developed, the project team are focusing on three distinct use cases: temperature over Antarctica; soil moisture across the UK; and temperature in London.

Another relevant example for environmental research comes from Ecosystem of Digital Twins — an ASG theme with a broader scope to incorporate digital twins (multiple digital representations of physical infrastructure) at the component, asset, system and process levels. Researchers in the “[Digital Twins of Fleets](#)” project propose population-level analysis for modelling engineering systems by combining abundant data from systems in operation for a long time with sparse data from systems established more recently ([Bull et al, 2022](#)). Their approach encodes domain expertise to constrain the ML model via assumptions (and prior distributions), allowing the methodology to automatically share information between similar assets. This study has been optimised on interpretable fleet models of different in situ data, where ‘fleet’ refers to a population of assets that constitute engineering infrastructure such as civil structures of roads and bridges or trains in a transportation network. Showcasing examples from survival analysis of an operational truck fleet and wind-power predictions for an operational wind farm, they demonstrate the wide applicability in practical infrastructure monitoring.

Forecasting the environmental change

The Arctic is warming four [times as fast as the rest of the planet](#). Even if we can limit our emissions to less than 2 degrees Celsius below pre-industrial levels, Arctic warming could [increase by 4-5 degrees Celsius](#) due to complex positive feedback mechanisms. This could have serious consequences for the region and the planet. Perhaps the biggest change to the region over recent decades has been the [loss of almost half the summer sea ice since the early 1980s](#) from the Arctic Ocean. The most substantial change however is the loss of the thick sea ice that usually circulates within the Arctic Ocean for many years, known as [multiyear ice](#). The sensitivity of sea ice to increasing temperatures has caused the summer Arctic sea ice area to see massive shrinkage over the past forty years, equivalent to the loss of an area around 25 times the size of Great Britain. Climate projections show that by the middle of the century, the Arctic could have [seen its first ice-free summer](#). These accelerating changes have dramatic consequences for our climate, Earth's ecosystems, and the lives of Indigenous and local communities whose livelihoods are tied to the seasonal sea ice cycle.

Reliable sea ice forecasting systems have the potential to help conservation workers, local communities, and research expedition planners anticipate changes in sea ice conditions. Researchers, meteorologists and climatologists depend on rich historical data from different sources to make high-resolution forecasts for a certain region. Modern forecasting techniques use mathematical models of multiple atmospheric conditions that can accurately project longer-term climate predictions (reference: Brenner, Laurie. [Four Types of Forecasting](#)). Even with large datasets, physics-based predictive models and advanced computers to run them on, it is, nonetheless, extremely difficult to forecast conditions in remote and extreme environments due to their complex relationship with their surroundings in a rapidly changing climate. AI has the potential to transform forecasting techniques by enhancing the predictive ability of forecasts for a wide range of environmental phenomena, serving as an early warning system for local populations living in close proximity to sea ice.

Combining large-scale in situ and remote sensing data from environmental monitoring systems at different scales collected over a long period, AI algorithms can be trained to accurately forecast long-term and short-term fluctuations in vulnerable regions. Additional data from physical and digital infrastructure, as well as various open data, citizen science and crowd-sourcing efforts can further enhance the accuracy of forecasting tools. Applications of these AI tools could range from providing accurate daily weather forecasts, forecasts of a city's energy consumption, the wider development of crisis management plans, data-assisted policy-making, timely interventions for supporting communities in high-risk zones, local capacity for renewable energy and protection for native fauna or wildlife.

AI Method in focus: Deep Learning

The average levels of Arctic sea ice in summer months have halved since satellites began monitoring sea ice in 1979. Interactions of the sea ice environments are challenging to capture

using purely physics-based models, which led researchers to explore advanced deep learning-based AI solutions to forecast sea ice changes.

Deep learning is a subfield of ML that uses a complex system of mathematical algorithms called an artificial neural network, inspired by the biological neural network of the human brain. Data sources such as satellites, ground-based sensors and other similar devices routinely generate a large amount of data that are too vast and complicated to draw meaningful conclusions from without sophisticated computational solutions. Deep learning uses a layered approach in its computations, giving AI tools the ability to process unstructured data from multiple data sources to draw conclusions with little to no human intervention. Although deep learning requires large datasets and training models can take days or weeks, once trained, deep learning systems run significantly faster than their physics-based counterparts and reduce the testing time by multiple orders of magnitude.

We recommend [Deep Learning References by Pablo Mesejo](#) from Inria Grenoble Rhone-Alpes, Perception team from 2017, and a more recent and up-to-date references on GitHub repository [Awesome Deep Learning papers](#) by Terry T. Um to gain technical details on Deep Learning.

Spotlight on seasonal ice change

IceNet, ice loss forecasts for the season ahead

[IceNet](#) is an AI tool that enables scientists to more accurately forecast Arctic sea ice conditions months into the future. Developed for understanding Arctic sea ice loss in collaboration with BAS researchers, IceNet has provided the first demonstration of an AI framework that can predict the concentration of sea ice more accurately and more efficiently than traditional physics-based approaches. The improved predictions could underpin new early-warning systems that protect Arctic wildlife and coastal communities from the impacts of sea ice loss.

IceNet provides a solution built on a deep learning model called a convolutional neural network (CNN), which is the technology behind facial recognition systems, medical imaging analysis, and self-driving cars. IceNet has been trained to automatically forecast the next six months of monthly average sea ice maps, based on climate simulations for 1850–2100 and satellite observations from 1979 to 2011. IceNet directly predicts probabilities of sea ice occurring, displaying a confidence level for each prediction. When evaluated on unseen data after training, IceNet was more accurate than similar physics-based prediction systems. IceNet offers a simple framework for probabilistically bounding the ice edge within a region of lower predictive confidence, which has added utility over deterministic ice edge forecasts. Finally, a variable importance method was used to identify the climate variables most important for IceNet’s forecasts. The next generation of IceNet will build on this exciting result, predicting at a more granular daily timescale and running in real-time across the polar regions ([Andersson, Hosking et al, 2021](#)).

IceNet began as a small research project at BAS, but soon expanded at an early stage in collaboration with the Turing researchers, REG and RAM team members from ASG. To ensure rigour across all aspects of this highly interdisciplinary project, new international domain experts were further involved to help address any knowledge gaps in the team. Relying on complementary skills, the IceNet team was able to improve both the forecasting ability of the software and its speed of execution, as well as develop tools for forecast dissemination. Typically the numerical physics-based forecasting systems running on supercomputers require hours of run-time to produce forecasts. IceNet, on the other hand, can run on a laptop 2,000 times faster than equivalent numerical forecast models, taking less than ten seconds on a single graphics processing unit. By training AI models on larger more regionally diverse datasets, IceNet will be able to provide larger coverage and use cases for both the Northern

and Southern Hemisphere. The goal for the next generation of IceNet is to go beyond monthly Arctic forecasts by making daily predictions of sea ice levels in both hemispheres.

Future extensions of IceNet will couple novel deep generative modelling and lifelong learning AI forecasting research with digital infrastructure developments to deliver data products and services, underpinning the development of polar digital twins. In the next stage, IceNet will deliver on three specific aims: *weather-to-seasonal timescale* sea ice forecasts to meet Arctic and Antarctic conservation challenges; research and operational tools providing forecasts to decision-focused software frameworks; and support for users to integrate these technologies into their day-to-day workflows. This ecosystem will draw data generated from deep learning pipelines to demonstrate real-world use cases and make them available through integrations and interfaces for programmatic access. By making it a fully interoperable module to work with other environmental systems and “digital twins”, IceNet will be easy to deploy, energy efficient and a powerful tool for instantaneous and real-time decision-making.

More examples

Forecasting to build a sustainable and safe society

The UK is the world's first major economy to set a legally binding target of being net zero by 2050. With renewable energy already part of our overall energy supply, achieving a balance between the greenhouse gases emitted and removed from the atmosphere will require a transition to a zero-carbon electricity system. According to the [UK National Grid](#), the biggest proportion of electricity entering the national grid is generated by natural gas, which is largely from imported fossil fuel that emits harmful greenhouse gases into the atmosphere when burnt to produce energy. To meet the UK's demand while reducing carbon emissions, considerable efforts are being made by the local and national authorities to increase the amount of energy from renewable and low-carbon technologies. In 1991, renewable energy accounted for just 2% of the UK's electricity generation, which in 2020 increased to 43%. For the first time in UK's history, in 2021 electricity came mainly from renewable energy: a mix of wind, solar, bioenergy and hydroelectric sources, overtaking the use of fossil fuels consumption. The UK endeavour to establish a carbon-neutral infrastructure for energy supply has been greatly assisted by data science and AI technologies.

In 2017, ASG researchers collaborated with the UK National Grid [Electricity System Operator](#) (ESO) to improve their work in 'balancing' the electricity demand with electricity being produced, while accounting for intermittent generation such as wind and solar power. While there is data on the location of UK government-supported solar farms, there is less information on smaller domestic installations of photovoltaic (PV) solar panels. In the absence of accurate forecasting of solar energy input from solar PV installed capacity, electricity produced by natural gases can unnecessarily enter the energy circulation (due fossil fuelled generators running in the background) adding to the carbon emission. Attempting to improve the forecast accuracy for ESO and run the National Grid more economically, researchers and the Turing's doctoral students at a Data Study Group in 2017 contributed to the adoption of [PV solar panel forecasting](#) models for the National Grid. The initial model was further validated, tested and implemented by ESO, alongside a range of other new forecasting approaches, achieving [33% more accurate day-ahead forecasts](#). These more accurate forecasts will help balance energy supply and demand, lowering costs for consumers and ultimately meeting the 'net zero' goals globally ([Stowell et al., 2020])(<https://www.nature.com/articles/s41597-020-00739-0>). Using data from citizen science and crowdsourcing efforts, researchers have created a fully open geographic data source for solar PV within the UK, suitable for reuse in different geographical data and other applications such as machine vision and short-term solar forecasting.

Another example comes from a project dedicated to [quantifying the effects of climate change on extreme weather events](#) using the ‘distributional downscaling’ approach, an approach to infer high-resolution information from low-resolution data. Extreme precipitation events and floods caused by climate change expose over 1.8 billion people to significant flood risk worldwide, causing nearly \$520 billion in economic losses including well-being costs per year ([Rentschler et al. 2022](#)). These events disproportionately affect low-income countries lacking infrastructure investments. Climate models are used for forecasting and predicting the risk of floods and heavy precipitation events, acting as an important source of information for policy-makers. To improve the prediction of high-resolution precipitation, researchers in this project apply a deep learning approach using model fields (weather variables) that are more predictable and generalisable than local precipitation ([Adewoyin et al. 2021](#)). To this end, they present **TRU-NET** (Temporal Recurrent U-Net), an open source Python-based deep learning framework to effectively model multi-scale spatiotemporal weather processes and provide high-resolution predictions of rainfall. Funded by Alan Turing Institute under Climate Action Pilot Projects, TRU-NET uses HPC Facilities at the University of Warwick. The project was benchmarked against existing tools and its suitability was assessed through rigorous experimentation predicting seasonal performance metrics for each model tested on the whole country and predictive errors on 5 specific cities across the range of precipitation profiles. Based on their analysis, the TRU-NET team concluded that the available data is sufficient for applying a deep learning approach to produce robust, high-quality predictions for rainfall across seasons and varying regions. Climate prediction methods such as TRU-NET are important for guiding the decision of policy-makers reducing the financial and societal risk posed by flooding ([Shukla et al. 2019](#)).

Simulating the human cost of climate change

Annual temperatures in the UK average a daily low of 6 degrees and a high of 14 Celsius, which during the summer months ranges from 15 to 25 degrees Celsius. In 2021 and 2022, temperatures during the summer months recorded the highest UK-wide average maximum daily temperatures since 1960. Specifically, in 2022, the extreme heat experienced between June and August was far more intense and widespread than previous comparable heatwaves, with a maximum of 40°C recorded for the first time ever in the UK’s history ([data from currentresults.com](#)). Whilst temperature increases for a few days of the year may not seem significant, when assessed against historical data, it is evident that extreme heat is a result of climate breakdown and has prolonged impacts on wider populations. According to the [Department of Health and Social care](#), severe heat will affect everyone’s health and increase heat-related mortality, which will inordinately affect vulnerable communities with chronic and severe illness, disability and poor housing conditions.

Traditional means of estimating the risks associated with higher temperatures are based on aggregate measures obtained from epidemiological studies, which are both sparse and expensive to replicate. Sensor-based monitoring systems and crowdsourced reporting efforts often help fill gaps in the data, but their qualities remain varied with limited measures for error corrections. Furthermore, measures such as the average temperature within an urban area assume that all members of the population experience the same temperature. However, how heat is experienced by an individual differs depending on their activities; their health conditions and lifestyle, the amount of time they spend outdoors and indoors, their working and living conditions, their daily activities and movements, and their modes of transport. Exposure, therefore, differs with activity, which subsequently interacts with their underlying vulnerabilities, such as existing co-morbidities and lack of resilience to health stressors, to increase or decrease an individual’s relative risk of adverse health outcomes due to heat exposure.

To predict how climate change or extreme heat might impact a given population and support creating coherent response strategies, decision-makers in local and national governments need granular knowledge of the dynamics and demographics of a changing population, the various communities within it and the potential consequences of policy and mitigation plans proposed for them. This requires a comprehensive analysis of information drawn from multiple domains including urban infrastructure, urban planning, transportation networks, land use, air quality, regional energy systems, public health (such as population exposure to heat), existing mitigation plans and other sources. While it would be highly challenging, and potentially invasive or unethical, to gather all of this data for all individuals within an area, researchers can build synthetic populations using aggregate data representative of the population of interest, such as from censuses and traffic movements, and simulate population characteristics using an analytical technique known as microsimulation.

AI method in focus: Microsimulations

Simulations are a computational approach to mathematical modelling, designed to derive conclusions from data by applying a set of algorithms predicting the behaviour of real-world or physical systems. Monitoring and forecasting are data-based modelling techniques that represent real-life systems to help capture data reflecting a specific state, such as temperature in a specific location, or effects of changes such as measuring the impact of climate change on the environment. Microsimulation techniques, unlike other modelling approaches, operate at the level of individual units such as persons within a population, rather than aggregate population data. There are two widely known simulation approaches - agent-based models and microsimulations (DOI: 10.1146/annurev-statistics-010814-020218). Agent-based models are always probabilistic, whereas microsimulations__ are always data-driven. Before implementing a proposed intervention in the real world, microsimulation is often used for evaluating the impact of a possible outcome of a suggested policy. Despite the differences in applications-level, both approaches are identical from a mathematical and computational perspective – they both apply a series of algorithms to data (reflecting a specific state) and address the problem of probabilistically reproducing real-life phenomena.

Microsimulation models use unit-level data to construct a collection such as a group of individuals to represent a target population and their behaviours. Applying a series of algorithms and predetermined probabilistic rules, the model simulates changes in an individual’s characteristics over time giving different versions of the outcomes. Microsimulation models can be classified into ‘static’ and ‘dynamic’. Static microsimulation is a snapshot at a single point in time that can apply a change (such as ageing) to a representative sample, and look at the impact based on the expected characteristics derived from known data. Dynamic microsimulation models apply specific processes (other factors related to ageing) to a representative sample and predict the distributional impact of some intervention of interest, such as the impact of a health risk factor on the future likelihood of disease, or a policy change to improve healthcare access for different populations.

Microsimulation was first applied to [demographic problems in 1957](#) and has remained a popular approach in demographic research ([Schonfield et al, 2018](#)). This technique is increasingly favoured for studies that are generally expensive, unsafe or unethical (such as in clinical trials), impractical or time-consuming for their implementations in the real world. It flips the angle of policymaking by modelling the interacting behaviour of individuals, families, and organisations within a larger system, where algorithms represent behavioural processes changing through time for each individual in the entire population of decision-makers. Taking complex interactions across subpopulations and considering the behaviours of an individual in connection with their environment or society (measured at multiple points in time), microsimulations predict the effects of policy changes on individuals.

Spotlight on producing projections for human population

DyME, quantifying the environmental impacts

The [Dynamic Modelling for Environment - Climate, Heat and Health](#) project (DyME-CHH) has set out to build a microsimulation model (DyME) applied to population-level exposure to extreme heat (CHH). The project aims to predict the health consequences of climate change and rising temperatures on various populations within the UK. The DyME model was first developed in collaboration with academic and industry partners by the Turing's [Urban Analytics programme](#) at the start of the COVID-19 pandemic - where 'E' in DyME initially stood for 'Epidemics'.

In 2020, the Royal Society established the [Rapid Assistance in Modelling the Pandemic](#) initiative (RAMP). It brought together epidemiologists and those with skills in computer modelling in fields such as urban transportation planning, financial markets and individualised marketing on social media to help model non-pharmaceutical interventions to stop the spread of the coronavirus. The Turing researchers participated in the initiative and developed DyME by integrating two previous open source projects affiliated with the Turing: [QUANT \(Quantitative Urban Analytics\)](#), a land-use transportation model; and [SPENSER \(Synthetic Population Estimation and Scenario Projection\)](#), a microsimulation model that produces high-resolution geographical projections of human populations. The work was done in collaboration with universities in the UK and in Denmark, as well as with data-science companies in the UK.

Using DyME, the researchers demonstrated that in 2020, going into lockdown in response to COVID-19 a week earlier would have significantly reduced the rate of infection and transmission in England's Devon county ([Spooner et al., 2021](#)). The success of DyME led to the subsequent adaptation of microsimulation models and tools for use in environmental research. The DyME model was released as open source software and was later expanded to a national level in England under the name Agent-based Simulation of Epidemics at Country Scale (ASPICS) by the ASG programme's Urban Analytics team. Under the name 'Dynamic Microsimulation for Environment - Air Quality' or DyME-AQ, the model was repurposed for use in predicting air quality and associated health hazards. The initial DyME-CHH model will be calibrated for different trial areas, with future work planned with the Turing REG team to upscale for the rest of the UK, developing more efficient R code that the DyME model will be written in initially.

DyME-CHH is building off this body of work and experience created throughout the programme to model the interaction of activity, rising temperatures and vulnerabilities on the health of different populations within the UK, in collaboration with the University of Exeter and the Cornwall Council. Using a synthetic population provided by the [Synthetic Population Catalyst](#) (SPC) project, researchers can already estimate how different individuals might move through the world, and predict how these conditions along with their age, health and socioeconomic factors affect their risks in the face of environmental breakdown. DyME-CHH outputs will allow policymakers to identify the specific risk to an area based on what the population and physical aspects (such as housing or urban design) of the area looks like. An exciting impact of the project is the ability for decision makers to model different scenarios of local adaptation to support developing policy in an accessible, but data-driven way. The goal is to enable local authorities to use this information for identifying and prioritising key mitigation and adaptation strategies, limiting the adverse effects on the health of the most vulnerable people in a population.

More examples

Engagement-led adoption, extension and deployment of AI tools

Academic engagement with the public and governmental sectors will continue to be a central component of data science and AI development. These collaborations accelerate access to large datasets from various domains and of different resolutions required to build AI tools and train models that are widely applicable. A crucial yet missing step that many stakeholders, such as in local councils, public health and grassroots community initiatives, need to take before using data for their purposes is localising data for their areas of applications. Currently, [UKCP18 Met Office climate projections data](#) exists at multiple temporal-spatial resolutions and Representative Concentration Pathways (RCPs). Additionally, climate projections are not just available for the future, but also past and present. Re-calibration at local levels is required to bias-adjust data in these different periods using on-the-ground weather measurements. This recalibration process is currently performed ad-hoc by climate scientists for specific areas and duration, rather than at scale due to computational requirements for manipulating and processing large datasets and the domain expertise needed to apply sophisticated downscaling techniques. As such, when it has been performed, datasets, and knowledge of the implementation of methods, tend to be restricted to small user groups and shared directly within that network rather than from a single source of truth that all can directly access. This leads to issues with the reproducibility of findings, transparency of methods, and lack of meaningful comparison, or more generally the ability to perform comparisons, of the range of bias adjustment methods used for calibrating projections with local climate data. The CLIM-RECAL project addresses these classic cross-domain data challenges. Developed by DyME researchers, CLIM-RECAL integrates data from a variety of different online and offline places, aligns their scales and standards across domains, and creates understandable and accessible documentation for non-domain expert users. Intending to save a large chunk of research time for anyone working with UKCP18 climate projections data, this has been a rigorous, labour-intensive and extremely necessary undertaking by the researchers in ASG. Taking the learnings from the initial RAMP/DyME work to SPC/ASPICS and now specific applications of DyME simulation models to air quality and extreme heat, the DyME team have evidenced the importance of integrating existing datasets, and are enabling future adoption through CLIM-RECAL. Access to high-quality, directly usable, climate data will allow more people to learn about environmental research, and different stakeholders working in climate change spaces to achieve more.

The Turing has continued to facilitate extensive interdisciplinary exchange, including with government and local authorities, in environmental research. Researchers at the University of

Exeter are working with stakeholders within local councils to build an app to help councils plan and mitigate the impact of climate change. The Local Climate Adaptation Tool (LCAT), though initially developed for Cornwall, has now extended to include other local government partners across the UK. Based on a qualitative evidence review, LCAT builds health impact pathways and displays relevant data of interest to stakeholders. The LCAT tool developers work with DyME-CHH to source the appropriate climate data for their analysis. Future iterations of the app will include outputs from DyME-CHH and the related Turing project to provide appropriately adjusted climate data at high granularity, as well as estimates of future population demographics. These integrations from DyME-CHH to the app were identified in collaboration with the Exeter group while working on different projects, which would have been missed if not for collaborative meetings and events. Furthermore, DyME-CHH outputs can now leverage the extensive network built by LCAT to disseminate and ensure findings, which are integrated into management decisions. The projected demographic output DyME-CHH was of particular interest to collaborators and will provide a new application of SPC beyond its initial application in health & environment, amassing software engineering expertise. This is also an opportunity to co-produce documentation that is understandable to stakeholders in social, environmental and health research, ensuring the work is disseminated to its maximum potential.

Another example is [EnergyFlex](#) developed under the ASG project for [simulating energy efficiency opportunities for households](#). Energy flexibility is key to delivering a reliable, sustainable energy system, reducing pressure on energy production systems to continue using non-renewable energy sources ([Energy Efficient Cities initiative](#)). EnergyFlex is an agent-based microsimulation approach that models the energy performance of the housing stock by estimating when people in an urban area are likely to be using energy. Working with local authorities, EnergyFlex also generates synthetic housing stocks and provides a way to explore inequalities in energy efficiency, helping target households in more vulnerable socio-economic circumstances in need of retrofit. To address the challenges in residential decarbonisation in the UK, supported by the RAM team, a workshop with stakeholders from government and industry was convened to identify action plans, in addition to the possible avenues for using microsimulation models such as EnergyFlex. Since the workshop, several steps have been taken that include documentation, dissemination and engagement with stakeholders in the energy and housing domains to explore the use of the tool for strategic planning and policy-making. This project has involved a close collaboration with the [Data & Analytics Facility for National Infrastructure](#) (DAFNI), through which modules of code have been uploaded as standalone executable versions of the model which can be run in the cloud by users without any coding requirement. EnergyFlex models will also be pooled by DyME-CHH, which has already integrated the SPC and QUANT models.

Although it's clear that mathematical modelling and simulation methods are pushing environmental research forward, researchers often lack the high throughput resources required to train and run those computationally intensive models. [RADDISH](#), an ASG-funded project, with collaborators from University College London, provides a fast implementation of state-of-art data assimilation code targeted at High-Performance Computing (HPC). RADDISH,

Real-time Advanced Data assimilation for Digital Simulation of numerical twins on HPC, was initially developed to cope with the timescales, data streams, and complex nature of modelling in different scenarios for weather forecasting. Recognising the lack of tools and capacity to efficiently run simulation models, RADDISH provides an effective solution for the computational implementation of modelling and simulation in HPC. The tool has been rigorously optimised for forecasting tsunamis and modelling the evacuation of people and vessels in coastal areas ([Multilevel Bayesian Quadrature Preprint](#)).

Recent collaborations with the government and public stemming from the COVID-19 pandemic response have provided new opportunities to work across different fields. These modes of working are also helping find applications for existing technology and interdisciplinary approaches to address urgent societal issues related to climate change. Microsimulation techniques, as investigated in the Environment and Sustainability theme, improve our understanding of the impact of environmental policies and other interventions on a population through long-term projection. As already in practice, AI tools and app-based solutions will continue to assist local governments, individuals and communities in identifying ways to eliminate or minimise human-induced heat, air pollution and activities that add to the greenhouse gas emissions and damage our planetary ecosystem.

Adapting to climate change

The United Nations (UN) [Intergovernmental Panel on Climate Change \(IPCC\)](#), [Net Zero Strategy](#) by the UK government and [COP26 reports](#); all point in the same direction: the world has to achieve net-zero emissions by 2050 and immediately scale up climate response strategies. Mitigation and adaptation options will effectively reduce climate risks posed by natural disasters like rising heat, drought and floods while building energy-efficient solutions for biodiversity, food security, health and carbon sequestration. Mitigation strategies focus on managing the causes or minimising the impacts of climate change, whereas adaptation strategies lead to an adjustment in systems and social behaviours that restrict potential damages and maximise opportunities to prepare for a greener future. AI is already playing critical roles in addressing climate risks through applications of advanced methods like environment monitoring, forecasting future events and building better policies assisted by microsimulation techniques. These developments have laid a strong foundation for long-term mitigation and adaptation options, rather than only dealing with near-term risks reactively. Interdisciplinary approaches will remain crucial in engaging domain experts and local communities with sustainable response plans, scaling up the adoption of successful measures from different sectors.

Contributing to Sustainable Development Goals

The UN's [2030 Agenda for Sustainable Development](#) urges all strategies for climate change to go hand-in-hand with ending poverty and other societal deprivations. AI-based responses to climate change, therefore, need to be aligned with sustainable development goals, going beyond building high-tech predictive technologies. Food security, responsible production and consumption, quality education, urgent infrastructure and industry, affordable energy and reduced inequalities are some of the biggest societal concerns that are worsened by climate change, and when not addressed, these further add to the climate-related risks ([Vinueza et al. 2020](#)).

Solutions to climate change that integrate complex knowledge from these social and sustainable development areas can be potentially accelerated through AI technologies. ASG's work is not only demonstrating the success in this area but also extending our considerations for fully embracing AI-based approaches to climate change by drawing insights from the natural, social and digital worlds, and how they interact with each other. Building on the successful implementations in the areas of monitoring and forecasting, researchers have already delivered adaptation solutions in a variety of sectors, such as providing energy flexibility through EnergyFlex and Solar Panel forecasting, and limiting the use of greenhouse gas. Several more AI/ML projects are uncovering the potential of AI applications in enhancing mitigation plans such as by informing agriculture and food production patterns, infrastructure retrofitting, reducing waste and repurposing materials. To ensure the success of all climate response plans, many projects are dedicated to enhancing public awareness of data science, empowering them to understand the advantages and limitations of AI applications through public engagement, education and capacity building.

Generalisable, accessible and explainable AI tools can benefit all stakeholders in research, government, industry, education and society by enabling them to make data-informed decisions for combating the impact of climate change. Amplified by the ongoing efforts at the Turing more broadly, AI will massively improve the effectiveness of climate response plans in the UK, and with their potential to be adopted worldwide will contribute to building a sustainable future for everyone.

Spotlight on the impact of climate change on agriculture

Building the UK's Crop Modelling Framework

Agriculture has important socio-economic implications for creating food security by ensuring the availability, accessibility and affordability of nutritious food for everyone. Since the production patterns of crops rely on specific climate conditions, agriculture is severely vulnerable to climate change. Agriculture is also a direct cause of 8.5% of greenhouse gas emissions ([IPCC report](#)), while current agriculture practices additionally constitute about 15% of emissions from land use change, such as deforestation or land clearing for food production. Further exploitation of agricultural land through industrial mono-planting, over-irrigation and use of harmful chemicals to control evolving crop diseases has resulted in deteriorating soil health. It is hardly a surprise that the [United Kingdom Food Security report](#) (2021) identifies climate change, alongside the loss of biodiversity and misuse of natural capital resources as the biggest threats to food security. In coming years, it is likely that the current food system in the UK will no longer be viable to meet consumer demand.

Under the '[impact of climate change on agriculture](#)' project, the Turing researchers are building a crop modelling framework for the UK in collaboration with Rothamsted Research, John Innes Centre, University of Exeter/Met Office, The National Phenomics Centre at Aberystwyth University and the Centre for Ecology and Hydrology (UKCEH). Turing's researchers are finding optimum ways to combine data and models efficiently, and using these insights to inform the development of the Scivision tool along with knowledge gained from the DeepSensor and Cryo-em projects, as well as building tools to extract more detailed data on plant development. Multi-disciplinary information from plant science, hydrology, soil science, insect population dynamics, economics, consumer behaviour, plant pathology, crop yields and climate models are being integrated into models of plant development and crop yield and scaled up to make nationwide predictions to support the new policy and management practices.

AI applications are embedded to widen our understanding of agricultural aspects including plant breeding conditions, management practices such as fertiliser treatment, as well as crop protection measures against diseases, weeds and pests. To build knowledge about disease and climate-resistant crops, researchers are integrating crop yield data and disease information from Agriculture and Horticulture Development Board (AHDB), and weather variables from the UK's Met Office. By further incorporating satellite data and soil data researchers are

constructing large-scale training data sets in collaboration with the UKCEH. These complex datasets are logically filtered and combined for exploratory visualisations and subsequently analysed to understand the association between meteorological and soil variables and the pathogen-associated yield gaps. This training data has been rigorously aligned on various temporal and geographic scales and further anonymised of commercially sensitive data from UK farmers – making it a significant output from this project.

Integration of plant protection data into crop models has already been achieved for five major AHDB crop varieties (winter wheat, spring wheat, winter barley, spring barley, and spring oats) ([Raza and Bebbler 2022](#)). It predicts future yields based on the levels reported by AHDB and characterises pathogen-associated yield gaps. The gold standard crop simulation model, [APSIM](#), can reproduce historic yield data predicting mean yield accurately. To also account for the variability of observed crop yields, An R codebase of optimization frameworks is also being developed at John Innes Centre, with data integration aided by the Turing to calibrate the most important parameters of APSIM and improve prediction of variability in crop yield. These parameters include the integration of data on temperature, water stress, and phenological development. The wheat yield models provide an exploratory analysis of long-term wheat experiment data such as from genotyping, as well as associated information on disease and weather to forecast wheat yield in the UK. In-depth modelling will be conducted to understand the wheat yield response to nitrogen application and the impact of weather conditions such as unfavourable temperatures and precipitation. The Met Office data will be explored to model fungal pathogen spore dispersal under climate change scenarios. Additional drone images of crops would improve current process-based yield models to gather longitudinal data and classify disease linkage more efficiently.

Open source AI/ML approaches are also enabling the analysis of plant images to extract phenotype data such as characterisation of the size, shape and distribution of seeds. A deep learning pipeline for the detection and segmentation of seeds in 3D X-ray CT images of seed pods was built based on an open source model StarDist and has been integrated into the Scivision tool to analyse cell microscopy images. Further application of open source computer vision methods will enable the exploration of data to better understand the architecture of oilseed plants ([Corcoran et al. 2023](#)). The project team is also finding applications for techniques developed for the plant phenotypes in analysing other image data, such as for the Living with Machine programme in collaboration with DEFRA and BBSRC. All models and data will be integrated into the Scivision and the demonstrator notebooks will be made available via the Environment Data Science Book allowing users to easily load new data and perform inference with these models.

Identifying, reconciling and integrating datasets and modelling methods, as well as exploring how machine learning tools could improve crop yield and plant development models could allow us to both predict the impact of climate change on UK agriculture and identify arable crops that are resilient to climate change. The Turing’s research in this area will contribute to ensuring future food security by helping identify ways for farmers, policymakers and local stakeholders to mitigate future risks and adapt to changing climate conditions shaping the

consumption and production patterns of our society in the near future. Data-driven insights from the crop modelling framework will improve our ability to predict plant development and crop yield indefinitely, strengthening food security over many years to come.

More examples

Multi-level interventions for mitigation and adaptation

Approaches for climate response will involve adaptation and mitigation options for securing not only our food production capacity but also other aspects of the environment to protect our existing carbon stocks and reduce carbon emissions. By taking a data-centric approach, it is possible to integrate long-term time-series data with causal knowledge of physical and socio-economic influences. Multi-disciplinary integration has allowed AI researchers to build better models for predicting future climate conditions for urban farming, material development, peatland preservation and assessing interventions for climate action.

A striking example of adaptation from the ASG repertoire is the [Growing Underground Farm](#) project led in collaboration with researchers in the University of Cambridge, the British Geological Survey and Imperial College London. As wireless sensor technology, large cloud databases, and computer processing power become more available, digital twins are becoming attractive for integrating agriculture into built environments. ‘[Growing Underground Farm](#)’ is utilising digital twin for building a template for space-efficient local farms ensuring the production of greens all year round, while ensuring minimal environmental impact. The world’s first underground farm has been established 33 metres below the pavements of London’s Clapham High Street in London. A former World War II air raid shelter, proposed but never joined the tunnels to the London Underground system post-war, is now stacked with racks of fresh green leaves and microgreens thriving under banks of LED lights. [CROP, the digital twin of this farm](#), represents the reality of the environment through real-time streams of data and provides feedback for optimal management of the ‘twinning’ object. This includes three crucial elements. First, data from an extensive and robust monitoring system that tracks the observable environmental conditions in the underground farm is supported by data curation to ensure the quality and tractability of data. Second, analysing observable data along with information collected by farm operators helps identify key parameters of the farm environment and thereby crop yield. Third, using data modelling techniques to identify critical trends and changes, forecast potential future operational scenarios, and provide feedback on the influence of recent events on the farm environment. CROP’s cloud-based web application provides a 3D model of the farm with links to get live data streams from sensors within the farm. A database stores incoming data from the sensors and the associated models providing access to visualisations of current and historic conditions in the farm and crop growth data. Subsequent versions of CROP included the implementation of a temperature forecasting model, a scenario evaluation

tool and the incorporation of more detailed crop monitoring with yield data, giving insight into improving farm efficiency and crop yield in a built environment (Ward et al., 2020).

In addition to building synthetic environments, knowledge from natural and living systems can be applied in creating adaptation options that improve our ability to engineer new materials that meet advanced functional and sustainability requirements. Molecular simulation investigation inspires biomimicry approaches that transfer knowledge from biological systems in human designs to create bio-inspired energy-efficient materials that don't add to greenhouse gas emissions. AI can speed up the search for alternative or new materials by sifting through millions of potential molecules or biological states and identifying candidates for lab-based testing and development. Researchers within the 'molecular structure from images under physical constraints' project have built a deep learning-based framework, affinity-VAE, where VAE stands for Variational AutoEncoder, a class of CNN used to reduce the dimensionality of 3D volume data while encoding it into a low-dimensional latent representation. Affinity-VAE uses Convolutional Neural Networks (CNN) to determine the probability of a given class member in high throughput imaging data for reconstructing the molecular structure. This approach can automatically cluster and classify objects in multidimensional (2D and 3D) image data, including simulated biological electron cryo-tomography (cryo-ET) data based on their similarity. Affinity-VAE shows potential for the discovery of new candidates in experimental data that can contribute to designing biomimetic materials with environmental-friendly properties and functions. (Mirecka et al., 2022).

Mitigation responses need to focus on high-risk natural resources such as peatlands that store more carbon than all other vegetation types in the world combined. Peatlands are vulnerable to rapid climate change and disturbances such as wildfire and drainage, with the risk that their huge carbon stocks could be released into the atmosphere, further accelerating global warming. To counteract this threat, government, environmental organisations and industry are investing in climate actions to protect undisturbed peatlands and to restore those that have been damaged by human activities. An ASG project on understanding the risk and uncertainty in peatland carbon emissions is drawing information on various qualities of peatlands and their carbon emissions to scope a high-level Bayesian network. In collaboration with UKCEH and the Queens Mary University of London, the research team works to sketch out government policy, specific management interventions and natural mechanisms of resilience that are linked to peatland carbon emissions via causal relationships. This causal knowledge is built on numerical modelling, a scientific understanding of peatland functioning and expert advice on management activities and socio-economic factors. The project is developing a framework to integrate different types of information on peatlands: quantitative data on greenhouse gas fluxes and environmental drivers, such as temperature and precipitation; semi-quantitative understanding of controls and interactions, such as feedback between peat formation and water loss; and qualitative information on socio-economic factors, such as land-use policy and site-level management actions. Part of this research assimilates long-term records of carbon emissions from two Scottish peatlands into a process-based model enabling the identification of key environmental drivers and quantifying uncertainties on peatland carbon emissions. The other part of the work develops a causal framework that links carbon emissions predicted by

the process-based model to the broader ecological and hydrological functioning of peatlands and human interventions. Predictive models developed in this project can accurately predict future peatland carbon emissions and the benefits of specific policies and interventions that drive policy and management of peatlands. <citation needed>

Another nature-based mitigation project comes from the [Data Science for Climate Resilience in East Africa](#) project carried out in collaboration with [the International Small Group and Tree Planting Program](#) (TIST) and the University of Exeter. The project uses the cloud-based Google Earth's catalogue of satellite imagery and geospatial datasets to identify indicators of plant productivity and drought. The specific satellite data for this project is taken from LandSat-7 which gives a low-resolution but long-term survey of the areas of interest, primarily East Africa and Sentinel which gives high-resolution images for directly studying individual farms. By combining these indices with the detailed organisational data from TIST, models are being generated to infer tree cover and study landscape-scale changes occurring due to TIST and Kenyan [Northern Rangelands Trust](#) conservation practices over of their histories of community-led tree planting activities. The satellite data analysis combined with forest cover change datasets shows that community-led activities are observable as increasing greening trends with clear secondary effects on the landscape, such as overspilling or an increase in forest cover near TIST groves. The high temporal resolution of the satellite data is allowing the development of novel tools for assessing the impacts of tree planting and conservation farming, using indicators of plant productivity, vegetation structure and drought indices. These computational methods potentially provide cheap and effective ways to monitor progress in tree growth, resilience, and conservation as well as enhance field surveys and land management. TIST regularly seeds an area by bringing existing members to speak with locals about the benefits of the program that have a significant social and environmental impact. Improved and transparent metrics for assessing the success of their activities have a direct value to tree planting and conservation organisations, supporting their growth and making the benefits of participation clear to members and funders. Addressing some primary concerns regarding drought and soil erosion, the project increases an understanding of the impact of tree planting programmes and engages members within the growing organisation sustainably. An understanding of the social factors at play: for example how information spreads, how the diversity or gender make-up of the engaging groups affects their success, and how the organisation grows, giving invaluable lessons for the future of similar schemes in different countries ([Buxton et al., 2021](#)).

AI-assisted mitigation and adaptation approaches from a variety of solutions can be transferred to other ecosystems, building a better understanding of interactions between environmental, social, biological and economic processes and extending the potential to transform predictions of climate change and intervention options.

Recommendations

Addressing challenges in conducting AI/ML research

Data science and AI have incredible potential for applications in large-scale data analysis to understand the natural environment from various angles, such as for biodiversity monitoring, detecting icebergs in vast polar oceans and tracking wildlife from space. However, not everyone involved in environmental research and data analysis has the technical expertise necessary to use the latest available tools. Furthermore, researchers and practitioners often lack a practical understanding of best practices in data science and lack access to an interdisciplinary network of experts essential for addressing global challenges affecting different parts of our society. These, in addition to broader challenges presented by climate change, add to technical challenges associated with the applications of data science and AI.

Categorised under these four overarching objectives, in this paper, we describe some of the most important challenge areas related to building, testing and implementing AI/ML solutions for monitoring, forecasting, simulating and responding to climate change.

- (Facilitate) Interdisciplinary and Inclusive Collaboration
- (Create) Open, Reproducible and Ethical AI Technology
- (Pioneer) Cross-Disciplinary Data Integration and Real-World Deployment
- (Build a) Common Understanding of AI Applications

Interdisciplinary and inclusive collaboration

Interdisciplinary AI approaches in the context of environmental research and climate change studies are crucial. In environmental research, [interdisciplinarity has been proven to accelerate the adoption of AI solutions](#) from one sector to another — removing the need to start from scratch, avoiding massive reinvestment into challenges that are already addressed, and above all, providing faster solutions for climate change response through collaboration. Collaboration in this context is not a one-time intervention but is embedded throughout the project lifecycle. The success of all processes, from open communication to stakeholder engagement, and from guiding the co-development of project outputs to managing access to all research outputs ultimately depend on inclusive approaches to decision-making. Achieving these, nonetheless, can be extremely challenging, especially among distributed groups of stakeholders representing different objectives from their individual sectors and the problems their communities face.

Efforts across academia, government, private sectors and grassroots often remain disconnected, leading to ad-hoc development of tools for specific applications that cannot be easily adopted by others. Hence, it is important to engage all stakeholders through different modes of collaborative approaches that allow them to explore problems through multiple lenses, ask relevant questions, raise concerns, and collaboratively assess the needs of data-based solutions. The Turing institute is uniquely placed to bring together partners and contributors from academia, industry, government and public sectors, as well as to engage members of the public in building modern, innovative and state-of-the-art technology. The Turing has specifically championed infrastructure roles in the Research Engineering Group (REG) and Research Application Management (RAM, a product manager equivalent at the Turing) team, each leading on specific collaborative aspects of big-team science ([Forscher, Patrick S., et al., 2020](#)). Members in these roles are contributing to building and deploying toolkits and examples from the real world and successfully implementing collaboration models that support inclusive and equitable interdisciplinary research. Research engineers and research applications managers typically work across multiple initiatives, creating new opportunities to connect projects and exchange practices. For example, Scivision is connected to 6 more existing projects at the Turing, all of which have collaborators from academia, industry and government beyond the institute. The Scivision platform enables researchers and algorithm developers from these projects to quickly change between algorithms or datasets, thus decreasing the time needed to adapt methods to datasets or vice versa, leaving greater time for scientific experimentation. In the near future, Scivision will integrate DeepSensor codebase for increasing spatial granularity for real-time multimodal environmental monitoring adding a new challenge to Scivision. REG members assisted with internal practices for code sharing across projects (Scivision, Living with Machine).

One of Scivision’s data models comes from the [MapReader](#) (Hosseini, McDonough et al., 2021) developed under the ASG’s sibling in humanities, *Living with Machines*, a large-scale collaborative effort studying the impact of technology on people during the industrial revolution. MapReader was originally developed as an open source tool to analyse large collections of historical maps. Developed as a generalisable computer vision pipeline, it has found its use in a wide variety of domains, including classifying plant patches in geospatial images that are too large to closely investigate. Scivision’s adoption of MapReader has made it easy for users to run its model with the Scivision tool. [Centre for Environment, Fisheries and Aquaculture Science](#) (CEFAS) adoption of Scivision forms a striking example for both interdisciplinarity and the multi-domain use of generalisable tools. This interaction of CEFAS with ASG research work was led by the RAM team under the Turing’s [Data Study Group](#), a collaborative hackathon event which brings together multi-disciplinary researchers to explore data analysis on practical challenges. In less than 6 months after the Data Study Group event, the RAM team catalysed Scivision’s deployment on CEFAS research vessel – having improved plankton classification accuracy rates to over 90%.

These accelerated routes for solving real-world problems demonstrate the bigger roles of interdisciplinarity, infrastructure roles and cross-domain solutions in speeding up the deployment of research-based technology in different contexts. Future development work of the Turing research will take inspiration from these success stories to better inform the foundational development of AI/ML technology.

Open, Reproducible and Ethical AI Technology

Reproducibility is necessary to ensure the highest quality of research outcomes, ensuring that the same analysis when applied to the same data produces the same outcome. Open Source approaches support reproducibility by enabling access not only to the software or tools but also to their building blocks such as methods, data, code, analysis workflows and documentation. Open and reproducible research practices are not often integrated into projects from the beginning, making it extremely challenging to ensure that all results will be easily accessed, openly examined, reused and built upon by others. Furthermore, the importance of ethics in AI, though widely acknowledged, is not practically taught. Open, reproducible and ethical data practices require a common practical framework for development, computational reproducibility, project management, co-creation and transparent communication processes.

Open source code and open data are part of open science practices to ensure that code and data are available under a permissive licence allowing third-party users to reuse different components of research outcomes for any purpose. Commitment to applying open source practices strengthens the transferability of data models by making them discoverable via online repositories and understandable through thorough documentation.

Recognising the importance of open science in environmental research and data science more widely, ASG boasts an array of projects that integrate open research and reproducibility principles from the start. Scivision's adoption of MapReader developed for humanities researchers by the [Living with Machines](#) programme is already a great example of open source creating new opportunities for technological solutions. The software underlying [IceNet](#), [TRU-NET](#) and [PV solar panel forecasting](#) are also open source tools that have been shared online in public GitHub repositories. Each project accompanies data and additional code to fully reproduce all the results and figures from the published articles, allowing users to evaluate or understand the study and identify their future use. For instance, data generated by IceNet are [published on the Polar Data Centre](#), whereas TRU-NET uses model field data from [IFS-ERA5](#) linked to the [Copernicus Knowledge Base](#). Currently hosted under the [Open Climate Fix](#) ([GitHub repository](#)), The PV solar panel forecasting project, alongside government data, makes extensive use of open data collected by massive online crowdsourcing projects such as [OpenStreetMap](#) and [PVoutput.org](#).

The IceNet codebase and infrastructure closely adhere to open source and reproducibility practices for creating sustainable software used across multiple problem areas. IceNet is already venturing into new directions and has found another practical use in a BAS project to plan the route for the [Royal Research Ship - Sir David Attenborough](#) (RRS-SDA) research vessel

in the Antarctic. IceNet forecasting is being used to identify the most fuel-efficient routes for the RRS-SDA to take through sea ice. The data products from IceNet are being incorporated as part of efforts to build a Digital Twin of the ship and the polar regions. In the case of the RRS-SDA route planner, this receives real-time data from the actual ship and provides a decision-making aid to the navigator on the bridge. Efforts are also underway to use IceNet in conservation with an active collaboration with the [WWF](#). Lead developers of this project have also participated in a Cambridge Venture Project at the Judge Business School to explore how to build future sustainability pathways for the project as a non-profit initiative.

Developed under the Turing’s Climate Action project [Solar nowcasting with machine vision](#), the PV solar panel forecasting data was established in collaboration with Open Climate Fix (an open science initiative to computationally address climate change issues), and OpenStreetMap (a free community-edited geographic database of the world) through a crowdsourcing effort. Volunteers and citizen scientists tagged the locations of solar panels mapping 25% of all the solar panels in the UK on OpenStreetMap. The team is also working on machine learning methods to detect solar panels from satellite images, which would fill gaps in the PV solar panel location data due to unregulated or independent service providers. This project will establish a worldwide open data “clearinghouse” – data collection from various sources to systematise the billing or service allocation. Offering solar PV geodata, and further formatting and transforming them for machine learning algorithms, data will be useful for the prediction of energy production by regional and national operators like National Grid, or commercial market participants. These outputs will help provide short-term solar power forecasting, demand forecasting, and fleet management and enable new demand management and energy-trading innovations, subsequently helping to cut carbon emissions.

Open science workflows extend beyond open data and open source code. The transparent processes for development have allowed researchers to learn about software engineering, while real-world applications of their work result in new collaborations and the co-creation of innovative tools. will extend the use of predictive models in creating new forecasting technologies and enable users to develop need-based solutions – contributing significantly to reducing the impact of the climate crisis on a regional, national and international scale.

Cross-disciplinary Data Integration and real-world deployment

While climate change is evident through observational data, many details about the underlying mechanisms and potential impacts on society have not been studied in detail. Data gathered at different Spatio-temporal resolutions provide additional features and information, each adding a new dimension to help assess how climate change impacts a specific area, such as agriculture. Data also come with different limitations in terms of discipline-specific standards, formats, biases, quality, missingness, contexts and scope of their applications, as well as unknown factors resulting from human and technological interventions at different stages of the data lifecycle. By building a complete digital picture of our natural environment, we can find better ways to track the impacts of climate change on agriculture, biodiversity, oceans, land, water, and the cryosphere. In order to do this we need to computationally integrate data from different scales, modalities, sources and disciplines taking a range of parameters into consideration making it a challenging and extremely expensive endeavour, both in terms of time and resources required. These requirements make it even more important to embed interventions throughout the project lifecycle to enable real-world deployment of existing data-based solutions. It is critical to work with stakeholders across academia, industries, government and the public to respond to climate crises on an unprecedented scale. Unfortunately, it is extremely risky in practice.

The Turing researchers have been able to apply innovative approaches to facilitate the integration of datasets across disciplines, build robust tools and technology and enable the deployment of existing software tools in ways that the original creators may not have originally envisioned. ASG projects have actively convened data scientists, algorithm developers, data owners, software engineers, practitioners and potential end-users from across disciplines. Working on projects like Scivision, DeepSensor, IceNet, Dyme-CHH and ‘Impact of climate change on agriculture’, they have created highly successful frameworks for data and model integration. Open source, reproducible and ethical practices further amplified the iterative improvement and reuse of those computational frameworks by different users. For instance, many models are supplemented by a collection of interactive notebooks with research narratives and code describing real-world applications of data models, exemplifying reproducible processes. [Environmental Data Science Book](#) has published multiple of them as executable python-based notebooks showcasing sensor data and models which could be highlighted through Scivision. [Scivision’s tree crown use case](#) was initially published as an Environmental Data Science notebook. Thanks to its incorporation in Scivision’s public catalogue, the tree crown model and

data are discoverable for the scientific image analysis community. These projects share common challenges in communicating science to experts and non-experts through reproducible and shareable computational notebooks.

Leading the real-world deployment and reuse of research components, RAMs bring new synergies and collaborative opportunities for core developers and external stakeholders. For examples, the RAM team members in DyME-CHH helped identify the value of creating the CLIM-RECAL project as a starting point so that Turing research team could share their extensive data processing and literature review work on climate projections data with other researchers working in the space so they would not need to duplicate efforts.. Through engagement with potential adopters of ASG projects through workshops (AIUK 2022), hackathon-like events (Data Study Group) and partnership meetings, RAM members further amplified the values of the project’s outcomes and impacts as well as created a feedback loop for the potential users to get involved in the projects during the development stage. For several projects described in this paper, the RAM team spearheaded the adoption of an open, regularly maintained Github repository so that progress, code, and outputs are available to external stakeholders.

Common Understanding of AI Applications

Applications of AI technology are essential for addressing the multifold challenges associated with environmental and sustainability issues including climate change, biodiversity, air quality and human activities further adding to the challenges. AI technologies to identify solutions is very important to ensure timely response, but it is equally important to build a better understanding among all stakeholders of the environment itself. There are growing concerns about the lack of transparency in AI, leading to a notion of an AI black box. Often users of AI tools do not know how these tools actually work — the underlying data or method they use, uncertainty and limitations associated with them, social or algorithmic biases they might perpetuate and harmful behaviour or injustice they reinforce. We need to acknowledge the fragmentation of efforts by different stakeholders, each holding a narrow focus on climate efforts. It is essential to build a shared understanding of data science and AI among all stakeholders including the users of AI technologies, especially by demystifying their advantages and limitations.

As the national centre, the Turing is committed to raising awareness of and building public trust in data science by developing AI and data-led technologies that are explainable, ethical and beneficial to our society. Cross-cutting research programmes such as Tools, Practices and Systems (TPS) and Public Policy Programme (PPP) offer projects that are driven by the interest of research communities and public sectors.

In a recent impact report, [Better together: The people-centred approaches driving forward data ethics](#), we discuss how practices and principles developed through ASG's cross theme projects are paving way for other Turing projects. Training efforts, including the Data Study Groups and [data training for biomedical scientists](#) supported by ASG, also work towards building data skills and public engagement in AI-related conversations.

[The Turing Way](#) and [Turing Commons](#) are community-oriented learning resources hosted by TPS and PPP respectively, that teach best practices in data science for use in academia, health-care, industry and government sectors. The Turing Way has also enabled the integration of best practices and supported the replication of its community framework in building the [Environmental Data Science book](#), a project led by researchers from Scivision and IceNet using real-world data, we demonstrate how our actions are impacting our environment, how technological advances can help combat environmental change and how to bring diverse stakeholders together to respond to the climate crisis.

AI in environmental research will help researchers around the globe to access and make more efficient use of the vast volumes of environmental data collected by national research facil-

ities. This work will support broader scientific communities and inform future planning of design, development and implementation of digital infrastructure to monitor our environment efficiently. Furthermore, using [AI explainability](#) methods - which provide us with the ability to interpret the predictions - researchers will 'open up the black box of the AI and conclude what it has learned from the data, potentially providing new scientific insights contributing to sustainable solutions.

Through community-oriented and open ways of working, the ultimate goal of these projects is to empower the beneficiaries of AI to build tools, reuse methods and adopt from successful projects in other sectors to tackle common high-level real-world challenges around climate response.

Acknowledgements

This report has been written by Malvika Sharan, with direct contributions from following ASG researchers:

- Scott Hosking
- Alejandro Coca (Scivision Project)
- Tom Andersson and - James Byrne (IceNet Project)
- Jennifer Ding (DyME Project)
- Evangeline Corcoran (CROP Project)
- Achintya Rao (Case study draft on Scivision and IceNet)
- Aida Mehonic (Overview)
- Alden Conner (Editing)
- Ben McArthur (Advisory and supervision)
- Hannah Sweeney and Harriet McCann (Programme Management)
- James Lloyd (Communications team, Editing)

A white paper summarising the learnings and recommendations drawn from this work has been published online as [Tackling climate change with data science and AI](#). Please cite that as *Conner, A., Hosking, S., Lloyd, J., Rao, A., Shaddick, G., & Sharan, M. (2023). Tackling climate change with data science and AI. Zenodo. doi: 10.5281/zenodo.7712969*. Thanks to James and Alden for editorial work.

Thanks to also those who have spoken in regards with this report or other related communications outputs:

- Gavin Shaddick
- Kirstie Whitaker
- James Lloyd
- Hayley Bennett
- Zaynab Ismail (project manager)
- Zeynep Engin
- Dave Chapman (review of the draft)

References

Overview

- The Alan Turing Institute: www.turing.ac.uk/
- AI for Science and Government (ASG): <https://www.turing.ac.uk/research/asg>
- Conner, A., Hosking, S., Lloyd, J., Rao, A., Shaddick, G., & Sharan, M. (2023). Tackling climate change with data science and AI. Zenodo. doi: 10.5281/zenodo.7712969: <https://zenodo.org/record/7712969>

Introduction

- Nelson, D. R., Adger, W. N., & Brown, K. (2007). Adaptation to Environmental Change: Contributions of a Resilience Framework. *Annu. Rev. Environ. Resour.*, 32(1), 395–419. doi: 10.1146/annurev.energy.32.051807.090348: <https://www.annualreviews.org/doi/abs/10.1146/annurev.energy.32.051807.090348>

Monitoring

- Gallant, A. L., Sadinski, W., Brown, J. F., Senay, G. B., & Roth, M. F. (2018). Challenges in Complementing Data from Ground-Based Sensors with Satellite-Derived Products to Measure Ecological Changes in Relation to Climate—Lessons from Temperate Wetland-Upland Landscapes. *Sensors*, 18(3), 880. doi: 10.3390/s18030880: <https://www.mdpi.com/1424-8220/18/3/880>
- General References on Computer Vision. (1998, January 27). Retrieved from <https://www.cs.hmc.edu/~fleck/computer-vision-handbook/general-vision.html>
- Scivision Example Gallery. Retrieved from <https://github.com/scivision-gallery>
- Supporting organizations — Binder 0.1b documentation. Retrieved from <https://mybinder.readthedocs.io/en/latest/index.html>
- Homepage - British Antarctic Survey: <https://www.bas.ac.uk>
- Artificial intelligence to help predict Arctic sea ice loss: <https://www.turing.ac.uk/news/artificial-intelligence-help-predict-arctic-sea-ice-loss>

- Improving tracking of iceberg populations in the Southern Ocean: <https://www.turing.ac.uk/research/research-projects/improving-tracking-iceberg-populations-southern-ocean>
- Data Study Group Final Report: British Antarctic Survey: (<https://www.turing.ac.uk/news/publications/study-group-final-report-british-antarctic-survey>)
- Dynamic digital twins of interconnected energy and transport networks: <https://www.turing.ac.uk/research/research-projects/dynamic-digital-twins-interconnected-energy-and-transport-networks>
- Bull, L. A., Di Francesco, D., Dhada, M., Steinert, O., Lindgren, T., Parlikad, A. K., ...Girolami, M. (2022). Hierarchical Bayesian Modelling for Knowledge Transfer Across Engineering Fleets via Multitask Learning. arXiv, 2204.12404. Retrieved from <https://arxiv.org/abs/2204.12404v2>

Forecasting

- Rantanen, M., Karpechko, A. Yu., Lipponen, A., Nordling, K., Hyvärinen, O., Ruosteenoja, K., ...Laaksonen, A. (2022). The Arctic has warmed nearly four times faster than the globe since 1979. *Commun. Earth Environ.*, 3(168), 1–10. doi: 10.1038/s43247-022-00498-3. <https://www.nature.com/articles/s43247-022-00498-3>
- Global Linkages: A Graphic Look at the Changing Arctic (rev.1) | GRID-Arendal. <https://www.grida.no/publications/431>
- Kwok, R. (2018). Arctic sea ice thickness, volume, and multiyear ice coverage: losses and coupled variability (1958–2018). *Environ. Res. Lett.*, 13(10), 105005. doi: 10.1088/1748-9326/aae3ec. <https://iopscience.iop.org/article/10.1088/1748-9326/aae3ec/meta>
- Quick facts, basic science, and information about snow, ice, and why the cryosphere matters. <https://nsidc.org/learn>
- Jahn, A., Kay, J. E., Holland, M. M., & Hall, D. M. (2016). How predictable is the timing of a summer ice-free Arctic? *Geophys. Res. Lett.*, 43(17), 9113–9120. doi: 10.1002/2016GL070067. <https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2016GL070067>
- Four Types of Forecasting. <https://sciencing.com/four-types-forecasting-8155139.html>
- Deep Learning References, Pablo Mesejo, Inria Grenoble Rhone-Alpes, Perception team, 2017. <https://project.inria.fr/deeplearning/files/2016/05/deepLearning.pdf>
- Awesome-deep-learning-papers, @terryum. <https://github.com/terryum/awesome-deep-learning-papers>
- Andersson, T. R., Hosking, J. S., Pérez-Ortiz, M., Paige, B., Elliott, A., Russell, C., ...Shuckburgh, E. (2021). Seasonal Arctic sea ice forecasting with probabilistic deep learning. *Nat. Commun.*, 12(5124), 1–12. doi: 10.1038/s41467-021-25257-4. <https://www.nature.com/articles/s41467-021-25257-4>
- DT FAQ: Digital Twin & Digital Engineering FAQ. <https://digitaltwinworks.com/faq>

- How much of the UK's energy is renewable? | National Grid Group. <https://www.nationalgrid.com/stories/energy-explained/how-much-uks-energy-renewable>
- ESO National Grid. <https://www.nationalgrideso.com> and <https://www.nationalgrideso.com/news/eso-and-turing-institute-use-machine-learning-help-balance-gb-electricity-grid>
- Towards a greener grid. <https://www.turing.ac.uk/about-us/impact/towards-greener-grid>
- Stowell, D., Kelly, J., Tanner, D., Taylor, J., Jones, E., Geddes, J., & Chilstrey, E. (2020). A harmonised, high-coverage, open dataset of solar photovoltaic installations in the UK. *Sci. Data*, 7(394), 1–15. doi: 10.1038/s41597-020-00739-0. <https://www.nature.com/articles/s41597-020-00739-0>
- Effects of climate change on extreme weather events. <https://www.turing.ac.uk/research/research-projects/effects-climate-change-extreme-weather-events>
- Rentschler, J., Salhab, M., & Jafino, B. A. (2022). Flood exposure and poverty in 188 countries. *Nat. Commun.*, 13(3527), 1–11. doi: 10.1038/s41467-022-30727-4. <https://www.nature.com/articles/s41467-022-30727-4>
- Adewoyin, R. A., Dueben, P., Watson, P., He, Y., & Dutta, R. (2021). TRU-NET: a deep learning approach to high resolution prediction of rainfall. *Mach. Learn.*, 110(8), 2035–2062. doi: 10.1007/s10994-021-06022-6. <https://link.springer.com/article/10.1007/s10994-021-06022-6>
- TRUNET GitHub @Rilwan-Adewoyin. [<https://github.com/Rilwan-Adewoyin/TRUNET>]](<https://github.com/Rilwan-Adewoyin/TRUNET>)
- Shukla, P. R., Skeg, J., Buendia, E. C., Masson-Delmotte, V., Pörtner, H.-O., Roberts, D. C., ...Malley, J. (2019). Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems. Retrieved from <https://philpapers.org/rec/SHUCCA-2>. <https://philpapers.org/rec/SHUCCA-2>

Simulating

- UK Climate Projections (UKCP). <https://www.metoffice.gov.uk/research/approach/collaboration/ukcp>
- Simulating energy efficiency opportunities for households. <https://www.turing.ac.uk/research/research-projects/simulating-energy-efficiency-opportunities-households>
- Energy Efficient Cities initiative Post - EnergyFlex - Towards A Flexible, Sustainable Urban Energy System. https://eeci.github.io/home/docs/projects/energyplanning/Conversation_Viz
- Average Temperatures in the United Kingdom - Current Results. <https://www.currentresults.com/Weather/United-Kingdom/average-annual-temperatures.php>
- UKHSC Agency - Hot weather and health: guidance and advice. GOV. <https://www.gov.uk/government/collections/hot-weather-and-health-guidance-and-advice>

- Orcutt, G. H. (1957). A New Type of Socio-Economic System on JSTOR. *Rev. Econ. Stat.*, 39(2), 116–123. Retrieved from <https://www.jstor.org/stable/1928528?origin=crossref>
- A brief, global history of microsimulation models in health: Past applications, lessons learned and future directions | International Journal of Microsimulation. <https://microsimulation.pub/articles/00175>
- Impacts of climate change and heat on health. <https://www.turing.ac.uk/research/research-projects/impacts-climate-change-and-heat-health>
- Urban analytics. <https://www.turing.ac.uk/research/research-programmes/urban-analytics>
- Rapid Assistance in Modelling the Pandemic: RAMP | Royal Society. <https://royalsociety.org/topics-policy/Health%20and%20wellbeing/ramp>
- Quantitative Urban ANalyTics (QUANT). <https://www.turing.ac.uk/research/research-projects/quantitative-urban-analytics-quant>
- Synthetic population estimation and scenario projection. <https://www.turing.ac.uk/research/research-projects/synthetic-population-estimation-and-scenario-projection>
- Shukla, P. R., Skeg, J., Buendia, E. C., Masson-Delmotte, V., Pörtner, H.-O., Roberts, D. C., ...Malley, J. (2019). Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems. <https://philpapers.org/rec/SHUCCA-2>
- Spooner, F., Abrams, J. F., Morrissey, K., Shaddick, G., Batty, M., Milton, R., ...Birkin, M. (2021). A dynamic microsimulation model for epidemics. *Soc. Sci. Med.*, 291, 114461. doi: 10.1016/j.socscimed.2021.114461. <https://www.sciencedirect.com/science/article/pii/S0277953621007930>
- Li, K., Giles, D., Karvonen, T., Guillas, S., & Briol, F.-X. (2022). Multilevel Bayesian Quadrature. *arXiv*, 2210.08329. <https://arxiv.org/abs/2210.08329v3>

Adapting

- Climate Change 2022: Impacts, Adaptation and Vulnerability, Climate Change 2022: Impacts, Adaptation and Vulnerability. <https://www.ipcc.ch/report/ar6/wg2>
- Net Zero Strategy 2022: Build Back Greener. GOV. <https://www.gov.uk/government/publications/net-zero-strategy>
- UN Climate Change Conference (COP26) at the SEC – Glasgow 2021. Published in 2022. <https://ukcop26.org>
- Transforming our world: the 2030 Agenda for Sustainable Development | Department of Economic and Social Affairs. <https://sdgs.un.org/2030agenda>
- Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., ...Nerini, F. F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nat. Commun.*, 31932590. <https://pubmed.ncbi.nlm.nih.gov/31932590>

- Carbon: greenhouse gas emissions from agriculture | AHDB. <https://ahdb.org.uk/carbon>
- Department for Environment, F. & R. A. (2021). United Kingdom Food Security Report. GOV. <https://www.gov.uk/government/collections/united-kingdom-food-security-report> Impact of climate change on agriculture. <https://www.turing.ac.uk/research/research-projects/impact-climate-change-agriculture>
- Raza, M. M., & Bebbler, D. P. (2022). Climate change, biotic yield gaps and disease pressure in cereal crops. *bioRxiv*, 2022.08.12.503729. (<https://doi.org/10.1101/2022.08.12.503729>)[<https://doi.org/10.1101/2022.08.12.503729>]
- Home - APSIM. (2019, September 27). <https://www.apsim.info>
- Corcoran, E., Siles, L., Kurup, S., & Ahnert, S. (2023). Automated extraction of pod phenotype data from micro-computed tomography. *Front. Plant Sci.*, 14. doi: 10.3389/fpls.2023.1120182. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9998914/>
- Optimising the world's first underground farm. (<https://www.turing.ac.uk/about-us/impact/optimising-worlds-first-underground-farm>)[<https://www.turing.ac.uk/about-us/impact/optimising-worlds-first-underground-farm>]
- CROP: the Crop Research Observation Platform. <https://www.turing.ac.uk/research/research-projects/crop-crop-research-observation-platform>
- Ward, R., Choudhary, R., Gregory, A., Jans-Singh, M., & Girolami, M. (2020). Continuous calibration of a digital twin: comparison of particle filter and Bayesian calibration approaches. *arXiv*, 2011.09810. <https://arxiv.org/abs/2011.09810v3>
- Molecular structure from images under physical constraints. <https://www.turing.ac.uk/research/research-projects/molecular-structure-images-under-physical-constraints>
- Mirecka, J., Famili, M., Kotańska, A., Juraschko, N., Costa-Gomes, B., Palmer, C. M., ...Lowe, A. R. (2022). Affinity-VAE for disentanglement, clustering and classification of objects in multidimensional image data. *arXiv*, 2209.04517. <https://arxiv.org/abs/2209.04517v1>
- Peatlands and climate change. <https://www.iucn.org/resources/issues-brief/peatlands-and-climate-change>
- Data science for climate resilience in East Africa. <https://www.turing.ac.uk/research/research-projects/data-science-climate-resilience-east-africa>
- TIST Program | 23+ million living trees planted by 130,000+ farmers in India, Kenya, Tanzania & Uganda | Award-winning, time-tested and participant-led afforestation model. (2022, July 01). <http://www.tist.org/welcome>
- Northern Rangelands Trust. <https://www.nrt-kenya.org>
- Synthetic Population Catalyst. <https://alan-turing-institute.github.io/uatk-spc>
- Buxton, J., Powell, T., Ambler, J., Boulton, C., Nicholson, A., Arthur, R., ...Lenton, T. M. (2021). Community-driven tree planting greens the neighbouring landscape. *Sci. Rep.*, 11(18239), 1–9. doi: 10.1038/s41598-021-96973-6. <https://www.nature.com/articles/s41598-021-96973-6>

Recommendations

- Kusters, R., Misevic, D., Berry, H., Cully, A., Le Cunff, Y., Dandoy, L., ...Wehbi, F. (2020). Interdisciplinary Research in Artificial Intelligence: Challenges and Opportunities. *Front. Big Data*, 3. doi: 10.3389/fdata.2020.577974. <https://www.frontiersin.org/articles/10.3389/fdata.2020.577974/full>
- Forscher, P. S., Wagenmakers, E.-J., Coles, N. A., Silan, M. A., Dutra, N. B., Basnight-Brown, D., & IJzerman, H. (2022). The Benefits, Barriers, and Risks of Big Team Science. *PsyArXiv*. doi: 10.31234/osf.io/2mdxh. <https://psyarxiv.com/2mdxh/>
- MapReader. <https://living-with-machines.github.io/MapReader>.
- Hosseini, K., McDonough, K., van Strien, D., Vane, O., & Wilson, D. C. S. (2021). Maps of a Nation? The Digitized Ordnance Survey for New Historical Research. *Journal Vic. Cult.*, 26(2), 284–299. doi: 10.1093/jvcult/vcab009.
- Hosseini, K., Wilson, D. C. S., Beelen, K., & McDonough, K. (2021). MapReader: A Computer Vision Pipeline for the Semantic Exploration of Maps at Scale. *arXiv*, 2111.15592. <https://arxiv.org/abs/2111.15592v1>
- Living with Machines – Harnessing digitised collections and collaborations between historians, data scientists, and curators, to model the effects of mechanisation on society. <https://livingwithmachines.ac.uk>
- Centre for Environment, Fisheries and Aquaculture Science (CEFAS)). <https://www.cefasc.co.uk>
- Data Study Groups. <https://www.turing.ac.uk/collaborate-turing/data-study-groups>
- openclimatefix. GitHub Repo. solar-power-mapping-data. <https://github.com/openclimatefix/solar-power-mapping-data>
- IceNet.ai GitHub repo. <https://github.com/icenet-ai>
- TRUNET GitHub repo, Rilwan-Adewoyin. <https://github.com/Rilwan-Adewoyin/TRUNET>
- Full Record - Forecasts, neural networks, and results from the paper: ‘Seasonal Arctic sea ice forecasting with probabilistic deep learning’ - British Antarctic Survey. <https://data.bas.ac.uk/full-record.php?id=GB/NERC/BAS/PDC/01526>
- ERA5: data documentation - Copernicus Knowledge Base - ECMWF Confluence Wiki. <https://confluence.ecmwf.int/display/CKB/ERA5%3A+data+documentation>
- CAMS and C3S Knowledge Base (CKB) - Copernicus Knowledge Base - ECMWF Confluence Wiki. <https://confluence.ecmwf.int/pages/viewpage.action?pageId=55116796>
- Open Climate Fix. <https://openclimatefix.org>
- OpenStreetMap. <https://www.openstreetmap.org/#map=6/54.910/-3.432>
- PVOutput. <https://pvoutput.org>
- RRS Sir David Attenborough - Wikipedia. https://en.wikipedia.org/w/index.php?title=RRS_Sir_David
- Solar nowcasting with machine vision. <https://www.turing.ac.uk/research/research-projects/solar-nowcasting-machine-vision>
- Raquel Carmo, Jamila Mifdal, and Alejandro Coca-Castro. “Detecting floating objects using deep learning and Sentinel-2 imagery (Jupyter Notebook) published in the Environ-

- mental Data Science book.” ROHub. Jan 28 ,2022. <https://doi.org/10.24424/g1bk-dv49>.
<https://the-environmental-ds-book.netlify.app/welcome.html>
- Better together: The people-centred approaches driving forward data ethics. <https://www.turing.ac.uk/about-us/impact/better-together-people-centred-approaches-driving-forward-data-ethics>
 - Introduction to Data Science for Biomedical Scientists: Turing-Crick Partnership Project, @malvikasharan GitHub repo. <https://github.com/alan-turing-institute/data-training-for-bioscience>
 - The Turing Way Community. (2022). The Turing Way: A handbook for reproducible, ethical and collaborative research (1.0.2). Zenodo. <https://doi.org/10.5281/zenodo.7625728>.
<https://the-turing-way.netlify.app/index.html>
 - The Turing Commons. <https://www.turing.ac.uk/research/research-projects/turing-commons>
 - Schmelzer, R. (2019). Understanding Explainable AI. Forbes. <https://www.forbes.com/sites/cognitiveworld/2019/07/23/understanding-explainable-ai>