Using Machine Learning to Cut the Cost of Downscaling Global Climate Models

Key messages

Machine learning can be used to emulate Regional Climate models (RCM).

An RCM emulator can effectively downscale global climate models significantly reducing time and cost.

This method presents an affordable way to support resilience and adaptation planning.

Global Climate Models (GCMs) are useful in showing us how our climate may change over the next 50-100 years under different scenarios of anthropogenic emissions of greenhouse gases. Unfortunately, GCMs provide information at a 50 – 250 km scale, whereas decision makers are typically looking at how impacts will be experienced at a local level (Figure 1), ranging from individual properties to postcode level detail. GCMs are therefore commonly 'downscaled' to a fine spatial scale. The science of how this is done is not settled.

Since GCMs have biases, downscaling a large number of GCMs helps understand how uncertain a projected change will be. Downscaling a GCM has traditionally been done by a process called 'dynamical downscaling'.

Figure1: Downscaled temperature over south-east Australia (NARCliM)

Dynamical Downscaling

Dynamical downscaling involves running a Regional Climate Model (RCM). This can be computationally very expensive and time-consuming. This limits the number of GCMs that can be downscaled and poses challenges to the delivery of information needed for decision-making. Crucially, downscaled products still retain biases from the GCM used to provide information, as well as biases inherent in the RCM – that is, all RCM downscaling products are uncertain. Due to the computational costs, this uncertainty cannot be addressed by using many RCMs to downscale many GCMs.

Research (Ref.2) from the ARC Centre of Excellence for Climate Extremes shows how machine learning can be used to cut the cost of dynamical downscaling, by building an 'RCM emulator', i.e. a collection of machine learning models that can emulate a RCM. The RCM emulator can downscale many GCMs at the same cost of downscaling a single one using a traditional RCM. By combining dynamical downscaling with machine learning, it becomes affordable to produce very large ensembles of climate projections, which can help to provide better decision-making support for resilience and adaptation planning.

Emulating Dynamical Downscaling

Machine learning models are built with a training set. (see Explainer Box What is Machine Learning?). Our research uses a training set that can downscale daily Evapotranspiration (ET) data from the Bureau of Meteorology. ET is the amount of water that is transpired by plants and evaporated from soil. This is important to understand the impact of climate change on agriculture, water availability and droughts.

- Our training set included daily ET and other meteorological variables, as well as local features of the land, including, elevation, plant cover, and soil properties. These local features inform the model how ET changes across a large region.

- The labels used for training were local ET information sourced from RCM output. This meant that a simulation with a RCM was needed to build an RCM emulator, but only about 10 years of RCM outputs (Figure 2), compared to the ~150 years of data that might be generated by dynamical downscaling.

- Those 10 years were carefully chosen using a sampling method to ensure that the RCM emulator would not be asked to downscale ET outside the range it was trained for.
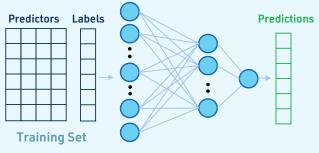
The RCM emulator reduces the computational costs and time of producing detailed climate information by using a short run of the RCM followed by a relatively quick downscaling with machine learning.

What is Machine Learning?

Machine learning is a subfield of artificial intelligence that focusses on building algorithms that can learn and apply what has been learnt autonomously.

The fundamental idea of machine learning is to identify meaningful patterns and structures in data and to use this information to make predictions. This is done by training a model on a set of samples in an iterative way, so that the model adapts itself and improves with experience, until it can perform its designated task accurately.

A common type of learning is called 'supervised learning'. This is when the machine learning model is given a training set with labels, i.e. the answers we want it to predict. The model learns the relationship between the input variables (predictors) and the labels. Other types of learning include unsupervised learning, semi-supervised learning, and reinforcement learning. Each of these approaches has its own specific techniques and applications.

Supervised Learning with a Neural Network

Advantages of RCM emulator

- The RCM emulator can produce similar results to a standard RCM, but is faster and more cost-effective (as shown in Figure 2)

- Many GCMs can be downscaled using the same resources.

- Produces a far larger ensemble of climate projections enabling probabilities of changes to be provided rather than the "best estimate"

- Provides a better assessment of uncertainties in projected changes.

- A better understanding of uncertainty helps support better decision-making for resilience and adaptation planning.

- Can be used to develop RCM emulators to meet specific needs e.g., for farms, agricultural land, or catchments.

- A cost-effective method which enables the assessment of climate change under multiple emission scenarios and the production of climate information for smaller regions.

- Can also be used to fill in gaps in the regional climate models simulations within the Coordinated Regional Downscaling Experiment (CORDEX) framework.
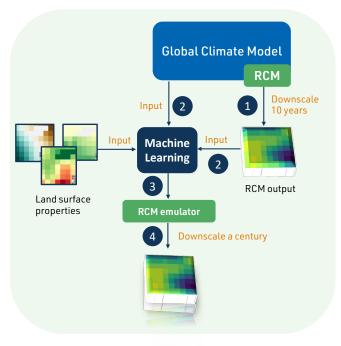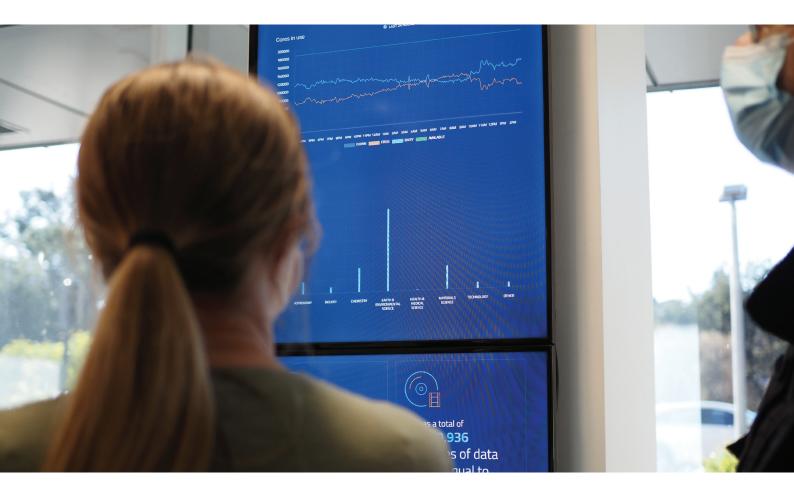

Figure 2: Schematic of how the RCM emulator can be used to downscale Global Climate Models

Future work

This research has the potential to create RCM emulators that can efficiently downscale multiple variables simultaneously, while preserving the physical relationships between them. For example, the water cycle can be downscaled, providing a more comprehensive understanding of how climate change will impact floods, droughts and available water.

We are also exploring how to use machine learning to create ensembles of around 10,000 members with the goal of building a bridge between climate projections and tools used in other communities including catastrophe modelling.

One final but long term goal is that to simulate changes in weather extremes in a climate model requires GCMs to be running at kilometre scales, not 100 km scales. This is simply computationally impossible for the sorts of climate simulations undertaken at present. We wonder if 10 years simulation at kilometre resolution might provide the foundation that, when combined with machine learning, will provide a step-change in what climate modellers can provide to impacts and adaptation research.

1. [1] Olson, Roman, Jason P. Evans, Alejandro Di Luca, and Daniel Argüeso. "The NARCliM project: model agreement and significance of climate projections." Climate Research 69, no. 3 (2016): 209-227. https://doi.org/10.3354/cr01403

2. [2] Sanaa Hobeichi, Nidhi Nishant, Yawen Shan, Gab Abramowitz, Andy Pitman, Steven Sherwood, Craig Bishop, & Samuel Green. (2023). Using machine learning to cut the cost of dynamical downscaling. Earth's Future, 11, e2022EF003291. https://doi.org/10.1029/2022EF003291

Contact

Sanaa Hobeichi
s.hobeichi@unsw.edu.au

Follow Climate Extremes:

in 𝕏 f