Proposed ML Methodology:

ML related tasks:

1. Explore, identify and evaluate techniques to downscale 12km RCM simulations to 5km resolution or higher.
2. Apply techniques to downscale 12km simulations for 6 GCM to create high-resolution precipitation products.
3. Explore, identify and evaluate techniques to downscale 100km GCM simulations to 5km resolution or higher.
4. RCM emulation (if SI is funded).

RA1: Explore, identify and evaluate techniques to downscale 12km RCM simulations to 5km resolution or higher.

This will be aimed at increasing the resolution of the RCM simulations, but also adding value to the 12km data instead of bi-linear interpolation. Our approach will aim to explore various methodologies to enhance this resolution from 12km to 5km next FY, which will follow via a publication. We intend to downscale a wide variety of different RCM variables (e.g. VCSN variables). RA1 is aimed at ensuring that we can achieve our deliverables.

1A: Identify metrics of success for downscaling the simulations to 5km

First, we need to a series of metrics to benchmark the performance of any downscaling model that we create (12km to 5km).

Validation datasets: 500m VCSN (with Council data), 5km VCSN, site-specific observations.

Metrics: Distribution matching, correlation coefficient, root-mean-squared error, bias in the 90th percentile of rainfall, trends in climate projections (if applied to CMIP6 run).

The challenge of using any "learned" method is that it might smooth the dynamical information in the RCM runs, so we need to ensure that we are preserving this with the metrics that we select. These metrics will then be used in conjunction with the models and methods developed in CS1.1.2 and CS1.1.3 to assess what approach is the most successful.

Timeline: See proposed Gannt chart

1B Downscale existing reanalysis driven run using deep-learning (RCM) – (B1,B2)

1. Downscale with deep learning from coarsen resolution (100km) by coarsening the RCM.
2. Downscale from high-resolution (12km), from the direct outputs rom the RCM.
3. Evaluate the performance on the VCSN dataset and focus on examining the performance at high elevation stations (e.g. by extracting rain gauge data).
4. Extend the approach to multiple variables

Other Potential Options:

- We will initially use our preliminary results from DL for downscaling, however later in the financial year, we may explore other alternative techniques also. Other examples of well-known architectures that will likely help are GANs. GAN's maybe able to resolve these high-resolution features that are not possible otherwise.
- Use existing ML downscaling methodology, and train a model to map from 12km to 5km (VCSN observations), and also separately train a model to 500m. Use (1) a CNN to downscale, and also use a new novel method a (2) GAN.
- Clean through the data and identify known issues (e.g. when there is snowfall and the rain gauges record precipitation several days later). Remove these instances from the training data and examine the change in performance.

1C Validate Abha's current downscaling methodology

We will:

1. Use Abha's downscaling methodology applied to a reanalysis run from CMIP5 / other projection
2. Get Abha to apply her bias correction scheme to a **NorESM -MM** model, for a given RCP. This will be used to benchmark our overall model, as it does not use any interpolation scheme.

1D Explore other Statistical Methods for Downscaling

We will (lead by Nico):

1. Leverage Q-Q mappings and other techniques for bias correction (trained against observations).
2. Examine GEV type techniques for a better production of climate extremes.

1E-1F Benchmark and decide the preferred approach

Here we will compare all three approaches (1B,1C, 1D), and determine the most suitable approach for downscaling CMIP6 projections. We will test each of these approaches on a reanalysis forced RCM run and a CMIP6 simulation. We will thoroughly evaluate the performance of these algorithms. We will also examine other methods.

Other research items: Test ML approaches for downscaling a 2.2km run (from 12km) over a short period of time. Explore other ML approaches? Write a paper summarizing the key results.

Timeline: July to December

RA2: Apply techniques to downscale 12km simulations for 6 GCM to create high-resolution precipitation products.

Create a pipeline that automatically downscales Peter's RCM simulations.