High resolution regional climate model simulations available through the ACECRC Climate Futures team: What we have and how they can be used

T.A. Remenyi^a, <u>P.T. Love</u>^a, R.M.B. Harris^a, M. Thatcher^b, T. Rafter^b, N.L. Bindoff^{abcd}

^a Antarctic Climate & Ecosystems Cooperative Research Centre, Hobart, Australia ^b Commonwealth Scientific and Industrial Research Organisation, Oceans & Atmosphere, Melbourne, Australia ^c Institute of Marine & Antarctic Studies, University of Tasmania, Hobart, Australia ^d Tasmanian Partnership for Advanced Computing, University of Tasmania, Hobart Australia Email: <u>tom.remenyi@utas.edu.au</u>

Abstract: The Climate Futures for Tasmania Project began in 2007 and has developed into the Climate Futures team within the Antarctic Climate and Ecosystems Cooperative Research Centre. The original project produced a data archive of dynamically downscaled simulations from 6 CMIP3 global climate models, for both the A2 and B1 emissions scenarios using CSIRO's Conformal Cubic Atmospheric Model (CCAM) over Tasmania. The highest resolution outputs were of 14 km spatial resolution with 6 hrly temporal resolution. This has recently grown to include a further 6 CMIP5 global climate models dynamically downscaled simulations over South-East Australia and Tasmania at 1-hourly temporal resolution. This new set of simulations has a range of spatial resolution over South East Australia (133.95E to 154.05E and -44.05N to - 27.95N, which is approximately Baird Bay in the west to the East Coast, and from the Gold Coast in the North to the southern tip of Tasmania); and 5 km targeting the Australian Alpine Region, but including almost all of Victoria.

This data archive has been used by numerous projects in the fields of ecological research, natural disaster management, local government land-use planning, water management, power generation and to provide agriculture with estimates of the future. We will provide a description of the gridded data products available for use now through the ACECRC Climate Futures team and a general description of how to access them. We will provide examples of how these data have been used in the past and describe the archiving paradigm we have found to be the most effective when planning to revisit or query an archive numerous times for various different purposes.

This presentation aims to introduce the research community to this data archive; describe the archiving method selected and demonstrate the advantages of this approach; and provide example use cases.

Keywords: Regional Climate Model outputs, archive management, Conformal Cubic Atmospheric Model (CCAM), climate futures

1. INTRODUCTION

Climate model projections of the future provide insight into how the global climate system will change under different conditions, providing useful information that can guide long-term, strategic decisions of organisations in both the public and private realms. Climate modelling efforts are coordinated by the Coupled Model Intercomparison Project (CMIP) established by the World Climate Research Programme (WCRP) Working Group on Coupled Modelling (WGCM). Under this leadership, Global Climate Models (GCMs) and their outputs have become easier to access, share, compare, and use.

The outputs from GCMs are useful for estimating broad scale, globally relevant changes, including atmospheric CO_2 levels, global heat budget and large scale climate drivers like the position of atmospheric features such as the subtropical ridge. Global changes like these, reported by the Intergovernmental Panel on Climate Change (IPCC) are important for national governments to understand so they can design robust, protective and constructive policies. However, at regional scales (e.g. state and local levels of government) this information is not sufficient to influence decision making. This is where regionally downscaled climate model simulations add value.

The Climate Futures for Tasmania Project was one of the first collaborations between research, government and industry in Australia to produce regionally downscaled climate simulations. The strong engagement with government and industry ensured the project goals and outputs relevant to government and industry. This project began in 2007 and has since developed into the Climate Futures team within the Antarctic Climate and Ecosystems Cooperative Research Centre. This strong focus on government and industry engagement has continued, with subsequent subprojects developed with significant input from end users and stakeholders. This helps to ensure the outputs produced are relevant and therefore used.

The original experiment – dubbed Climate Futures for Tasmania (CFT), described in section 3.1 – produced 140 model years from two emission scenarios (SRES A2 and B1) of CMIP3 GCMs dynamically downscaled using the Commonwealth Scientific and Industrial Research Organisation (CSIRO) Conformal Cubic Atmospheric Model (CCAM). This archive has been used by more than 100 research projects since it was created and has proven to be a significant asset for the Tasmanian community.

2. METHODS

The experimental design, tools and mathematical methods used have been described in detail by Corney et al. (2010) and McGregor et al. (2005). A brief summary of the methods is provided here.

2.1. Dynamical Downscaling

Dynamical downscaling is a process that uses coarsely gridded numerical model outputs - such as those provided by GCMs - as inputs into a higher resolution numerical model that can provide greater detail and accuracy than the coarser scale parent, due to the more realistic representation of orographic features and a much improved capability to resolve smaller scale physical processes (e.g. atmospheric convection).

CCAM has been actively developed for this purpose since 1994 and was configured for these experiments to use sea surface temperature outputs from parent GCMs to force an atmospheric model. CCAM can be run at varying resolutions: for example, it has been used to create global 50 km spatial resolution outputs; 2 km limited region outputs for domains over Sydney and Brisbane, Australia; and is actively being developed to improve its capacity at finer resolutions ($\sim 100 \text{ m}^2$).

CCAM was released to the community as an open source, free to use product in 2016. If you wish to contribute to the project please contact the project team.

More information is available here: <u>https://confluence.csiro.au/display/CCAM</u>.

2.2. Re-gridding from cubic conformal grid to regular longitude-latitude grid

CCAM operates using a conformal cubic grid which is computationally efficient and useful, the underlying grid resolution of the cubic-conformal grid varies significantly throughout the domain which is inconvenient for applied studies. As such, data is re-gridded onto a regular longitude-latitude grid using a system developed by CSIRO that considers the need to conserve mass wherever necessary, as well as performing the typical area-weighted calculations during the transformation process. The re-gridding is also useful for generating atmospheric layers on constant pressure levels rather than the hybrid sigma-pressure levels used within the modelling framework. All available data through the ACECRC Climate Futures group the post-processed re-gridded product thus, the underlying resolution of the model and must be considered when using these products.

2.3. Data archive management

Regionally downscaled climate model outputs are typically large data archives, usually consisting of 10s to 100s of Terabytes, and often Petabytes of data. Managing archives at these scales requires specialist resources and expertise.

The CFT archive is ~ 200 Tb, stored as monthly files, each containing about 150 climate variables. However, this has recently been found to be an inefficient archiving method for repeated analysis workflows (discussed in section 4), so there are plans to re-configure the CFT archive so that each file contains a single variable (which may be multi-dimensional) with as many time steps as practical.

The CFSEA archive is ~ 300 Tb stored as decadal files of a single variable. A single file of a 4-dimensional variable (i.e. longitude, latitude, pressure and time) can be ~ 100 Gb in size. This makes it more consistent with CORDEX and CMIP archiving conventions, reduces the total number of files in the archive and allows for rapid data extraction for subsequent repeated analysis (discussed in section 4).

3. REGIONAL CLIMATE MODEL OUTPUTS AVAILABLE

3.1. Climate Futures for Tasmania archive (CFT)

CFT CFSEA

The CFT archive contains "medium" and "high" resolution simulation outputs, described in further detail below. These were dynamically downscaled by CCAM using sea surface temperature from the SRES A2 and B1 scenarios of 6 CMIP3 GCMs (ECHAM5, GFDL-CM2.0, GFDL-CM2.1, MIROC3.2 (medres), CSIRO-Mk3.5, UKMO-HadCM3 (Randall et al. 2007)). Reasons for the selection of these models was not documented by the original team, however, these are all at the wetter end of the plausibility spectrum.

CFT medium resolution simulations extend over a limited domain, 124.74E to 170.25E and -50.25N to -24.75N, which is approximately from Balladonia in Western Australia to Mt Cook in New Zealand and from Bundaberg in Queensland down to the sub-Antarctic Southern Ocean. This had a spatial resolution of about



60 km and is archived at a temporal resolution of 6 hours, continuous from 1961 through to 2100.

CFT high resolution simulations extend over a limited domain from 143.45E to 149.05E and -44.05N to - 38.95N, which is approximately the extent of Tasmania and its major islands (King Island and the Furneaux group). This had a spatial resolution of about 14 km and is archived at the temporal resolution of 6 hours, continuous from 1961 through to 2100. Summary data products are available at daily, monthly, decadal and thirty year periods for selected variables. For a select number of cells within the grid (those that aligned with the Bureau of Meteorology observational stations), 6 min data was archived for temperature and rainfall.

3.2. Climate Futures for South East Australia archive (CFSEA)

The CFSEA archive contains output from "medium", "high" and "higher" resolution simulations dynamically downscaled by CCAM from the RCP 8.5 scenario – commonly referred to as a "business-as-usual" scenario – of 6 CMIP5 GCMs (ACCESS1-0, CNRM-CM5, HadGEM2-CC, NorEMS1-M, MIROC5, GFDL-ESM2M (Flato et al. 2013)). These models were selected based on advice from authors of the CSIRO Climate Futures in Australia project, which has a tool that indicates which models best represent each region of Australia.

The CFSEA medium resolution simulations extend over a limited domain 34.75E to 175.25E, -60.25N to 0.25N, which is approximately from Mozambique in the west to New Zealand in the East, and from the Equator in the north, to the Southern Ocean in the south. There are 10 atmospheric levels and ~120 variables available. This is a subset of the global domain, which is archived by the CSIRO (contact Marcus Thatcher at Marcus.Thatcher@csiro.au for details). The archived data have a spatial resolution of 50 km, 3 hourly temporal resolution, with summary variables also available at daily and monthly time steps.

The CFSEA high resolution simulations extend over a limited domain 133.95E to 154.05E and -44.05N to -27.95N, which is approximately from Baird Bay in the west to the eastern edge of Australia, and from the Gold Coast in the north, to the southern tip of Tasmania. There are 20 atmospheric levels and ~120 variables available. The archived data have a spatial resolution of 10 km with 1 hourly temporal resolution, with summary variables also available at daily and monthly time steps. These simulations were post-processed to produce two separate grids, a South East Australia grid and a Tasmania grid to reduce the area of ocean inside the model and outputs, thus reducing total data volumes. In the instance of Tasmania, it is recommended the CFSEA archived is used with preference over the older CFT archive, although both are still valid and relevant.

The CFSEA higher resolution simulations extend over a limited domain 140.475E to 151.025E and - 39.225N to -32.975N, which is approximately from Mt Gambier in the west to Wollongong in the east and Molong in the north to Wilson's Promontory in the south. There are 6 atmospheric levels and ~120 variables available. The archived data have a spatial resolution of 5 km, 1 hourly temporal resolution, with summary variables also available at daily and monthly time steps.

4. **DISCUSSION**

4.1. Data archiving

For applied climate studies, typical workflows require an entire time series for a small number of variables. For example the Growing Degree Days index is calculated from daily minimum temperature and daily maximum temperature, accumulated throughout the growing season. Therefore, it makes sense to archive files that contain single variables with many time steps rather than many variables with few time steps. This is in direct contrast to the way the climate models run and how the outputs are typically generated.

Climate models typically output all variables for each monthly time step as it is generated. This ensures each request for computation is a small discrete piece that can be managed by a shared compute environment and can be easily repeated if it fails. From a modelling perspective, it makes sense to ensure all the available variables are stored together, so that quality control checks can occur efficiently before moving forward to the next time step.

This creates a tension between what is efficient for the modelling environment that produces the simulations and what is efficient for the users of these archived simulations.

For the last 10 years, the CFT archive has been stored as monthly files, directly reflecting the manner in which it was generated. Operationally this has been adequate and workable, although sometimes slow. Recently, we have placed significant effort into identifying the performance bottlenecks and have quantified the two major sources: 1) many small files, and 2) distributed variable storage (one variable across many files).

Too many small files, where few directories containing large numbers of files, has recently been identified by the archive data managers as placing an unnecessary load on the file system. This reduces the performance of simple terminal operations like *ls* and *find*, which visit and interrogate each file when called. Reduced performance in such simple commands impacts on the efficiency of the entire user community of the shared resource.

When analysing a long time series of a single variable, an activity climate researchers repeat often, distributed variable storage (where a single variable is stored in many files), requires a significant amount of I/O. This is a major bottleneck for analysis processes. To demonstrate, using an R programming environment we measured the elapsed time required by the system to create connections to the same data stored in two different formats.

- Many Files Test: Using the R programming language with the *raster* package as a tool kit, create connections to 10 years of gridded data derived from the Bureau of Meteorology observations, stored as monthly files (61 Mb each, 7.32Gb total). **Time elapsed:** ~100 seconds.
- Single File Test: Using the R programming language with the *raster* package as a tool kit, create a connection to 55 years of gridded data derived from the Bureau of Meteorology observations, stored as a single large file (39 Gb total). **Time elapsed:** ~0.9 seconds.

Further tests revealed the I/O for each file takes about 0.9 seconds (in this instance retrieval was from disk storage, retrieval from tape storage would obviously be longer), where 10, 100 and 1000 files take about 9, 90 and 900 seconds to complete the same task, regardless of the file size. Storing variables in fewer, larger files, containing as many time steps as possible (regarding single file size limits) offers an enormous potential improvement in system performance during analysis steps.

4.2. Data access

The Climate Futures data archive is accessible through the Nectar Cloud for Australian Academic Federation (AAF) members. The Nectar cloud allows users to launch an analysis server on a virtual machine using compute facilities adjacent to the data store. The data archive can be mounted directly to the analysis server, which can provide a fully interactive integrated developer environment (IDE) for users. This process removes the requirement for users to download and maintain a local copy of the data archive on their personal computer, as data can be accessed and analysed directly. Through Nectar, analysis server templates are available for (but not limited too) RStudio, Jupyter and Matlab programming IDEs. Contact the Climate Futures team for further details (enquiries@acecrc.org.au).

The Climate Futures team have also collaborated with the Biodiversity Climate Change Virtual Laboratory (BCCVL) to provide data products for direct use within ecological species distribution models. Further information is available here: <u>http://www.bccvl.org.au/</u>

A more primitive access method (i.e. not recommended) is downloading (and maintaining) a local copy of target files through a THREDDS data server. The CFSEA data is not yet ready for public release. The CFT archive is available here: <u>http://portal.sf.utas.edu.au/thredds/catalog.html</u>,

4.3. How to use regional climate model outputs

Climate model outputs can be used to assess the general trend, and sometimes the rate and magnitude of change expected into the future. For some applications bias adjustment of the model output (not discussed) is necessary for it to be useful. These can be estimated using a range of methods, a few are described here.

Calculating climate indices is a common and typical use case for these kinds of simulations. A climate index is a useful way to summarise the concurrent conditions of many variables. For example, Forest Fire Danger Index (FFDI) is used by fire agencies across Australia to estimate the combined influence of antecedent rainfall with current temperature and wind speed and direction on the risk of dangerous fire conditions for a given location. As it happens, FFDI is a complicated index that uses input from two other indices, the Mount Soil Dryness Index and the Drought Factor that are both used to estimate antecedent rainfall influence on soil moisture levels (Griffiths, 1999).

Estimating future extreme conditions is another common use of climate simulations. Extreme conditions can be characterised by a single variable (e.g. maximum daily, monthly or annual temperature) or can be an extreme value for an index that is already calculated (examples in Sillmann *et al.* 2013).

Estimating the distribution of a variable is a very useful tool to illustrate how typical conditions have changed, or will change into the future.

4.4. Example use cases

Example a: Figure 2a is a comparison of historical and future monthly temperatures at Mt Hotham in Victoria, Australia. The projections suggest a fairly uniform warming across the year of about 2-3 ^oC. The monthly representation allows the comparison of adjacent months and illustrates that in the future, conditions will be 1-2 months warmer, where future-July will feel more like historical-May and future-October will feel more like historical-December. The summer months become so hot they exceed historical conditions. Presenting the changes in this way can be useful when engaging with stakeholders who are operationally reliant on annual cycles (e.g. farmers, snow-makers, fire agencies)

Example b: Figure 2b presents how chill-hours are related to elevation, and how this relationship changes into the future. Chilling hours are only accumulated when within lower and upper limits of a species-specific range. Chill hours are important for many agricultural crops, such as black currents, stone fruits and apples. The increase in chill-hours at future high elevations is due to the decreased frequency of temperatures below the lower limit. The decrease in chill-hours at low elevations is due to an increased frequency of temperatures above the upper limit (Holz *et al.* 2010). Together this creates an unintuitive, nonlinear response to a warming climate, demonstrating the value of climate models and higher resolution.

Example c: Figure 2c compares the projections of historical (1961-1990) and future (2070-2100) cumulative Growing Degree Days across the austral growing season (July-Jun). Growing Degree Days is an index used in agriculture to represent the amount of heat available to a plant for growth on any one day. This can be accumulated across a period, such as the growing season. It is a useful metric for comparing different seasons and locations. Figure 2c shows how future GDD values are significantly higher than those during the historical period, with historically maximum seasonal values in the Coal River Valley being reached 2-3 months earlier

by the end of century (Holz *et al.* 2010). These changes impact the types of crops that could be grown in the region. For example, in the wine industry, low seasonal GDD values are ideal for pinot noir varieties. As the climate warms, growers may wish to transition to other more appropriate varieties such as merlot or cabernet.

Example d: Figure 2d compares the historical and future climatic risk of Queensland Fruit Fly (*Bactrocera tryoni*) establishing in Tasmania. Tasmania is currently unsuitable, and as such, mitigation measures are minimal and effective. As the climate warms into the future, the north of Tasmania will become suitable, increasing the risk of invasion (Holz *et al.* 2010). This information is important for quarantine, agriculture and transport sectors to be aware of and develop plans to address.

Example e: Figure 2e shows the modelled maximum extent of snow cover over the Victorian Alpine Region in the decades 1980s, 2000s, 2070s and 2090s (Harris *et al.* 2016). These projections of the number of days of snow cover per year reflect the declines in maximum snow depth that have already been recorded at Australian alpine resorts since the 1950s. By the 2090s only the very highest peaks will experience any days of snow cover, with severe impact likely on natural ecosystems and winter recreational use.

Example f: Figure 2f shows the change in the cumulative number of hours suitable for snow-making at Mt Stirling in Victoria, Australia for each decade from the 1960s to 2090s. There is a decrease through time, although it suggests a rapid decrease will occur in the near future, indicated by the white space in the figure (Harris *et al.* 2016). Hence the challenges presented to Alpine resorts by reduced snow fall will be compounded by significantly reduced opportunities for artificial snow-making.



Figure 2. Examples of how regional climate model output can be used (historical = 1961-1990; future = 2070-2100): **a)** Comparison of historical and future monthly temperatures at Mt Hotham in Victoria, Australia (Harris *et al.* 2016); **b)** Comparison of how historical, mid-future (2015-2045) and future chill hours change at different elevations across Tasmania; **c)** Comparison of historical and future cumulative growing degree days in the Coal River Valley(Holz *et al.* 2010); **d)** Comparison of the historical and future climatic risk of Queensland Fruit Fly (*Bactrocera tryoni*) establishing in Tasmania (Holz *et al.* 2010); **e)** Modelled maximum extent of snow cover over the Victorian Alpine Region in the decades 1980s, 2000s, 2070s and 2090s (Harris *et al.* 2016); **f)** Comparison of projected cumulative snow-making hours at Mt Stirling in Victoria, Australia for each decade from the 1960s to 2090s (Harris *et al.* 2016).

5. CONCLUSIONS

The Climate Futures team at the ACECRC has a significant and valuable archive of regionally downscaled climate model simulations freely available for use. It is accessible through a choice of IDEs through the Nectar Cloud, or directly through THREDDS services. Over the last decade we have trialed a range of data archiving configurations, finding that for applied climate studies, large files that contain many time steps of a single variable are the most useful and most efficient.

Analysis of the CFT climate projections has produced an extensive list of reports and publications motivated by engagement with a diverse range of stakeholders in both the public and private sectors. Recent development of the CFSEA downscaled climate projections provides new opportunities to improve on previous analyses in terms of both higher resolution and greater sophistication of the CMIP5 global driving data.

Acknowledgements: The authors would like to thank for their continued and excellent support the technicians and help-desk at the Tasmanian Partnership for Advanced Computing (TPAC); the National eResearch Collaboration Tools and Resources project (NeCTAR); the National Computational Infrastructure (NCI) facility; and CSIRO High Performance Computing. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

REFERENCES

- Bennet, J. C., Ling, F. N., Graham, B., Grose, M. R., Corney, S. P., White, C. J., Holz, G. K., Post, D. A., Gaynor, S. M. and Bindoff, N. L. (2010). Climate Futures for Tasmania: water and catchments technical report, *Antarctic Climate & Ecosystems Cooperative Research Centre, Hobart, Tasmania*.
- Corney S. P., Katzfey, J. J., McGregor, J. L., Grose, M. R., Bennet, J. C., White, C. J., Holz, G. K., Gaynor, S. M. and Bindoff, N. L. (2010). Climate Futures for Tasmania: climate modelling technical report, *Antarctic Climate & Ecosystems Cooperative Research Centre, Hobart, Tasmania.*
- Flato, G., J. Marotzke, B. Abiodun, P. Braconnot, S.C. Chou, W. Collins, P. Cox, F. Driouech, S. Emori, V. Eyring, C. Forest, P. Gleckler, E. Guilyardi, C. Jakob, V. Kattsov, C. Reason and M. Rummukainen, (2013): Evaluation of Climate Models. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Griffiths, Deryn. (1999) Improved Formula for the Drought Factor in McArthur's Forest Fire Danger Meter. *Australian Forestry* 62 (2): 202–6.
- Harris, R. M. B., Remenyi, T. A. and Bindoff, N. L. (2010). The potential impacts of climate change on Victorian alpine resorts. A report to the Alpine Resorts Coordinating Council. *Antarctic Climate & Ecosystems Cooperative Research Centre, Hobart, Tasmania.*
- Holz, G. K., Grose, M. R., Bennet, J. C., Corney S. P., White, C. J., Phelan, D., Potter, K., Kriticos, D., Rawnsley, R., Parsons, D., Lisson, S., Gaynor, S. M. and Bindoff, N. L. (2010). Climate Futures for Tasmania: impacts on agriculture technical report, *Antarctic Climate & Ecosystems Cooperative Research Centre, Hobart, Tasmania.*
- McGregor, J. L. (2005). CCAM: geometric aspects and dynamical formulation. Technical Paper 70, CSIRO Atmospheric Research, Melbourne.
- Randall, D.A., R.A. Wood, S. Bony, R. Colman, T. Fichefet, J. Fyfe, V. Kattsov, A. Pitman, J. Shukla, J. Srinivasan, R.J. Stouffer, A. Sumi and K.E. Taylor, (2007): Cilmate Models and Their Evaluation. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M.Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Sillmann, J., V. V. Kharin, X. Zhang, F. W. Zwiers, and D. Bronaugh (2013), Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate, J. Geophys. Res. Atmos., 118, 1716–1733, doi:10.1002/jgrd.50203