

# Bias Correction of Downscaled CMIP6 Output

*Prepared for Ministry for the Environment*

*June 2024*

**Prepared by:**

Isaac Campbell, Peter B. Gibson, and Neelesh Rampal

**For any information regarding this report please contact:**




Peter B. Gibson  
Climate Scientist  
Climate Variability  
peter.gibson@niwa.co.nz

National Institute of Water & Atmospheric Research Ltd  
Private Bag 14901  
Kilbirnie  
Wellington 6241

Phone +64 4 386 0300

NIWA CLIENT REPORT No: 2024154WN  
Report date: June 2024  
NIWA Project: BSP22301

Revision	Description	Date
Version 1.0	Final version sent to client	26 June 2024

Quality Assurance Statement		
	Reviewed by:	Andrew Tait
	Formatting checked by:	Alex Quigley
	Approved for release by:	Alison MacDiarmid

© All rights reserved. This publication may not be reproduced or copied in any form without the permission of the copyright owner(s). Such permission is only to be given in accordance with the terms of the client's contract with NIWA. This copyright extends to all forms of copying and any storage of material in any kind of information retrieval system.

Whilst NIWA has used all reasonable endeavours to ensure that the information contained in this document is accurate, NIWA does not give any express or implied warranty as to the completeness of the information contained herein, or that it will be suitable for any purpose(s) other than those specifically contemplated during the project or agreed by NIWA and the client.

## Contents

<b>Executive summary .....</b>	<b>5</b>
<b>Guidance on usage .....</b>	<b>6</b>
<b>1     Theoretical Background .....</b>	<b>8</b>
1.1   Preserving physical characteristics .....	9
<b>2     BCSD Methodology .....</b>	<b>10</b>
2.1   Consistent Designation of Land and Ocean Grid Cells.....	12
2.2   Temperature variables.....	12
2.3   Precipitation.....	14
2.4   Potential Evapotranspiration .....	16
<b>3     BCSD Performance Evaluation .....</b>	<b>17</b>
3.1   Annual Climatologies .....	17
3.2   Seasonal cycle .....	18
3.3   ETCCDI.....	20
3.4   Climate Change Signal.....	23
3.5   Temporal correlation .....	26
<b>4     Summary and conclusions .....</b>	<b>26</b>
<b>5     Acknowledgements .....</b>	<b>28</b>
<b>6     Glossary of abbreviations and terms .....</b>	<b>29</b>
<b>7     References.....</b>	<b>30</b>
<b>Appendix A             Details of Quantiles Mapping Configuration.....</b>	<b>33</b>

### Tables

Table 1:	Summary of bias corrected variables with the methods used.	9
Table 2:	List of available bias corrected CCAM output for precipitation, tasmax, tasmin & PET for a given driving GCM and shared socio-economic pathway.	10
Table 3:	ETCCDI indices for climate variability and extremes used in this study.	11
Table 4:	Performance indicators used to select the best performing method for each variable considered.	12

## Figures

Figure 1-1:	An example of a distributional based bias correction method.	8
Figure 2-1:	Summer and winter CCS for bias corrected and raw ACCESS-CM2—CCAM data using multiplicative detrended quantile mapping.	14
Figure 2-2:	Heatmap comparing performance of several methodological variants for 9 indices.	16
Figure 3-1:	Annual climatologies for tasmin for VCSN, bias corrected CCAM output, LOYO CV and raw CCAM output, along with bias from VCSN.	17
Figure 3-2:	Annual climatologies for accumulated precipitation for VCSN, bias corrected ACCESS-CM2—CCAM output, LOYO CV and raw ACCESS-CM2—CCAM output, along with bias from VCSN.	18
Figure 3-3:	Winter climatologies for tasmax for VCSN, bias corrected EC-Earth3—CCAM output, LOYO CV and raw EC-Earth3—CCAM output, along with bias from VCSN.	19
Figure 3-4:	Long-term monthly mean accumulated precipitation for 12 locations across NZ for bias corrected GFDL-ESM4—CCAM output.	20
Figure 3-5:	Annual climatologies for TXx for VCSN, bias corrected EC-Earth3—CCAM output, LOYO CV and raw EC-Earth3—CCAM output, along with bias from VCSN.	21
Figure 3-6:	Annual climatologies for highest intensity rainfall for a single day for VCSN, bias corrected EC-Earth3—CCAM output, LOYO CV and raw EC-Earth3—CCAM output, along with bias from VCSN.	22
Figure 3-7:	Perkins skill score comparing the histograms of wet spell lengths against VCSN for the bias corrected EC-Earth3—CCAM output, the corresponding cross-validated corrected output, and the raw output.	23
Figure 3-8:	Climate change signal between the historical and SSP3-7.0 experiments for summer and winter for accumulated precipitation over those seasons.	24
Figure 3-9:	Annual climatologies number of frost days for the historical and SSP3-7.0 experiments and the CCS.	25
Figure 3-10:	Temporal correlation between bias corrected ACCESS-CM2—CCAM output and the corresponding raw model output for daily accumulated precipitation in the historical and SSP3-7.0 experiments.	26
Figure A-1:	Quantile spacing for linearly spaced nodes, log-spaced nodes and sigmoid spaced nodes.	33
Figure A-2:	Dummy data, reference and simulated data drawn from distribution with the same mean and high variance.	35
Figure A-3:	Dummy data, reference and simulated data drawn from distribution with the low mean and high variance.	36
Figure A-4:	Monthly means for multiplicative precipitation dummy data with groupers.	37
Figure A-5:	Summer and winter climate change signals of ACCESS-CM2-CCAM under the ssp370 scenario.	38
Figure A-6:	Climate change signal over the Mahanga station under the ssp370 scenario, highlighting the inflation effect of EQM on trends, without explicit trend preservation.	39

## Executive summary

The ‘National climate projections for Aotearoa New Zealand’ project has produced dynamically downscaled model projections from six global climate models (GCMs) from the Coupled Model Intercomparison Project phase six (CMIP6) across a range of future scenarios. The dynamical downscaling primarily employs the Conformal Cubic Atmospheric Model (CCAM) at ~12km resolution over New Zealand. Due to the enhanced resolution, dynamical downscaling generally improves the representation of climate at the regional scale relative to GCMs. However, biases can remain in the output after downscaling. Bias correction and statistical downscaling (BCSD) provides a relatively simple and effective method for reducing CCAM biases whilst also further increasing the spatial resolution of the output (to ~5km).

This report summarises the bias correction methods applied to four key climate variables from CCAM output at daily temporal resolution, namely:

- Accumulated precipitation
- Maximum near-surface air temperature
- Minimum near-surface air temperature
- Potential evapotranspiration

We use a common underlying bias correction technique known as quantile mapping for all four climate variables. Here, quantile mapping operates by mapping the distribution of CCAM modelled output onto a reference (or observed) distribution. NIWA’s Virtual Climate Station Network (VCSN) data, arguably the most comprehensive high-resolution daily national product, is used here as the reference dataset. Additional processing steps are taken so that the bias correction preserves key physical characteristics (e.g., the climate change signal) of each target variable.

The bias correction vastly improves the distributional properties of the CMIP6 downscaled data relative to VCSN. Most notably:

- Biases in the mean and variance in the historical period are reduced to near-zero.
- Bias correction markedly improves the representation of climate extremes in the historical period, though certain biases remain in some regions.
- Biases in seasonal cycle are significantly reduced, especially for precipitation.
- Climate change signals are well-captured for climatologies of temperature variables.
- Climate change signals are generally well-preserved but can show slight modification for extreme and threshold-based indices.

## Guidance on usage

Bias correction was performed for the following variables independently: daily precipitation, daily maximum (tasmax) and minimum (tasmin) near-surface air temperature, and potential evapotranspiration (PET). The bias correction was produced for each downscaled global climate model (GCM) independently, since certain biases are known to be model-dependent. Correlations between variables are not explicitly considered in bias correction, meaning that, in some instances, it is possible that this introduces minor physical inconsistencies (e.g. altered relationships between temperature and humidity). The only multivariate bias correction explicitly considered here was between tasmax and tasmin, through the daily temperature range (DTR). This was enforced so that daily tasmax is always greater than tasmin.

We recommend that the bias corrected data produced here should primarily be used in applications that require climate variables in isolation. If multiple variables are required to be used in combination (e.g. in downstream hydrological models), the user should carefully check the relationships between variables to ensure that no unexpected irregularities are introduced through bias correction. Alternatively, the user could consider using a bespoke multivariate bias correction specific to their application.

The bias correction methodology has been designed to preserve the climate change signal (CCS) from the Conformal Cubic Atmospheric Model (CCAM) output as closely as possible (described in detail in Sections 1 and 2). Here, the CCS is defined as the amount of change (percentage or absolute change) in a climate variable in a future period (e.g. 2080-2099) relative to a historical base period (e.g. 1986-2005). The rationale for preserving the CCS is the assumption that the physics-based CCS from the CCAM output is more reliable than any artificial modification induced by statistical bias correction over the historical period (Maraun, 2016; Maurer & Pierce, 2014). In subsequent sections of this report, we provide a comprehensive comparison of the CCS before and after bias correction. While the CCS is generally well preserved, for some regions and variables there are minor discrepancies. If these discrepancies are deemed too large for a certain application, the user should consider using the non-bias corrected projections instead. Note that in some cases, the preservation of the CCS will not be as accurate at the extremes of the distribution, further detail is provided in Section 3.

The reference observational dataset for all four variables used to bias correct CCAM output was NIWA's Virtual Climate Station Network (VCSN) data. Bias correction assumes that the reference dataset provides an accurate estimation of the observed climatology and statistical distributions. While VCSN is arguably the most accurate highest-resolution nationwide daily product, the VCSN is known to contain certain biases which remain difficult to precisely quantify, such as warm and dry biases over some parts of the Southern Alps (Tait, Sturman, and Clark 2012; Tait and Macara 2014). Unavoidably, aspects of these biases will remain in the bias corrected CCAM output.

It is important to note that certain variables were not targeted in this bias correction: near-surface wind, relative humidity, and shortwave radiation. This is because on a nationwide scale at daily temporal resolution the observational uncertainty in these variables is relatively large. As such, bias correcting these variables could introduce additional biases in the CCAM output. Future work could consider bias correction of these variables on a regional basis where the observations are deemed reliable.

## Guidance on usage

When using the bias corrected or raw CCAM climate projections, we generally advise the user to look at the spread of projections across multiple GCMs. The spread across GCMs can provide an estimate of uncertainty in the projections. In certain cases, the multi-model mean can dampen plausible changes for certain variables or seasons where there are large differences in the sign of the CCS between GCMs. For example, recent research has highlighted that there is relatively large disagreement in the sign of the CCS for end-of-century projections of summer precipitation over New Zealand (Gibson et al., 2024a).

It is important to note that the bias correction for CMIP6 described here differs from the previous bias correction performed on CMIP5 national climate projections. In summary, the bias correction from CMIP5 adopted a semi-empirical approach to bias correct the dynamically downscaled outputs from a resolution of 30km to 5km grid based on learnt relationships over the historical reanalysis-driven (i.e., ERA-40) run. This semi-empirical approach made local corrections ('enhancement factors') based on elevation, lapse rate, and wind direction, among others, determined from expert knowledge and from existing literature in other regions. Another main point of difference is that here our approach performs bias correction for each GCM independently, instead of using the single reanalysis-driven run. This helps ensure that the bias correction targets both biases stemming from the GCM (i.e. large-scale circulation) and biases stemming from CCAM (i.e. small-scale physics). The reader is referred to Mullan et al. (2018) for further details of the CMIP5 bias correction methodology.

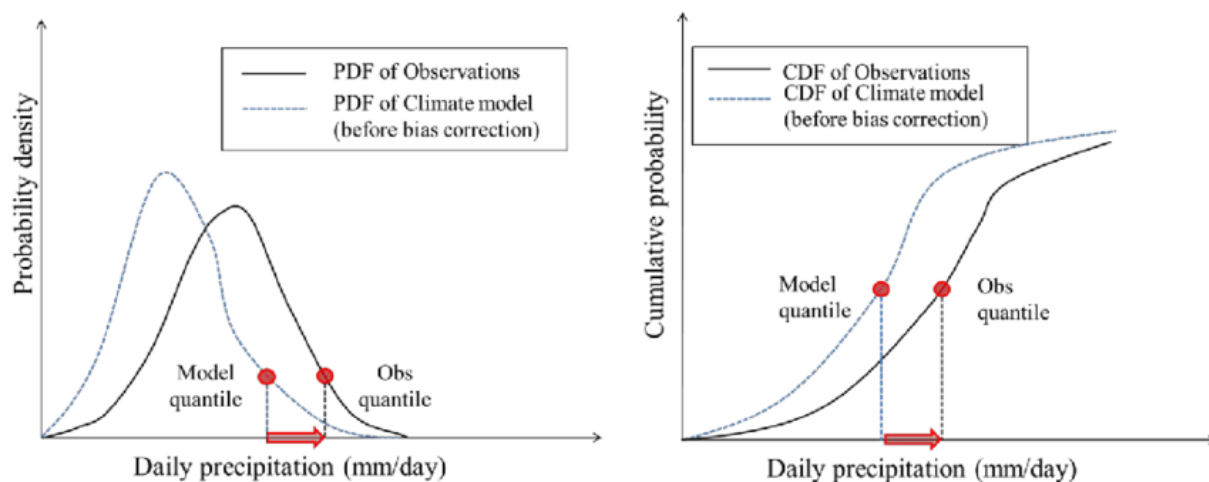
Additional general guidance is available on best practices around using bias corrected output, as applied to other regions (Met Office 2018; Maraun 2016; Alavoine and Grenier 2023; Casanueva et al. 2020; Gudmundsson et al. 2012).

# 1 Theoretical Background

Dynamical global climate models (GCMs) represent the scientific community's best effort to simulate the Earth's climate system. They are founded on physical laws which describe how the Earth's climate responds to external forcings such as increases in greenhouse gas concentrations. However, the resolution (i.e. horizontal grid spacing) of GCMs is a well-known and important limitation, constrained by computational capacity even on large modern supercomputers. As a result, parameterisations are required which reduce physical laws, operating over unresolved scales, to statistical properties and functions. Uncertainty associated with these parameterisations can lead to systematic biases in the modelled climate.

Dynamical downscaling, using regional climate models (RCMs), can resolve processes that are not directly captured at the scale of GCMs. RCMs generally reduce GCM biases; however, important biases can remain in RCM outputs, which can affect their utility in certain climate change applications.

As a result, numerous statistical methods have been developed in an attempt to correct these systematic biases in post-processing. These range from simple methods, such as the delta change method, to complex multi-step, multi-variate methods, such as the N-dimensional probability density function transform [MBCn] (Cannon, 2018). Most approaches are distributional based, correcting the distribution of the model output to match the corresponding observations (see Figure 1-1).



**Figure 1-1: An example of a distributional based bias correction method.** A schematic representation of the quantile mapping method, taken from Kim, Kwon, and Han (2016).

For the bias correction of dynamically downscaled CMIP6 climate projections over Aotearoa New Zealand, we use a commonly used method known as Empirical Quantile Mapping (EQM). This method is popular due to its flexibility and overall accuracy in correcting distributional biases, such as biases in the mean, variance, and other higher order moments. EQM is an empirical, non-parametric method. Previous studies have shown that this method tends to perform better in various settings than those fitting parameterised distributions, such as fitting a gamma distribution for precipitation (Gudmundsson et al., 2012; Gutjahr & Heinemann, 2013; Ivanov & Kotlarski, 2017). EQM operates by deriving empirical cumulative distribution functions (eCDFs) of modelled and observed data and derives a transfer function that 'maps' the modelled eCDF onto the observed eCDF (Figure 1-1). The transfer function is composed of adjustment factors (AFs) that are applied to the corresponding quantile.



Bias correction methods alter the statistical distribution of the modelled output but do not explicitly consider physical processes when doing so. Two important assumptions are made when bias correcting. Firstly, it is assumed that the underlying RCM is capable of accurately capturing the relevant physical processes. Secondly, the RCM biases are time-invariant, i.e., the RCM biases in the historical period are the same as those in the future period. If these assumptions are not fulfilled, statistical artefacts can be introduced through bias correction (see Maraun et al. (2017) for notable examples).

Quantile mapping (QM) methods generally perform well over the calibration period (the period it is trained on). However, the empirical nature of the method can be a source of overfitting, leading to a reduction in generalisability of the transfer function. In other words, the method may perform exceptionally well on the period over which it is trained, but then incur larger biases when applied out-of-sample in a future climate. The transfer function is heavily dependent on the representativeness of the calibration period; if the sample period does not adequately capture variability across timescales, performance is more likely to be diminished for out-of-sample data. To assess whether this criterion is fulfilled, we apply a leave-one-year-out cross-validation (LOYO CV) using the same bias correction method. For details, see Section 2.

A detailed description of the methodology for each target variable is given in Section 2, with particular attention to preservation of the climate change signal and other physical characteristics. We present results for each corrected variable in Section 3, with attention given to various performance indicators. These include skill metrics in correcting climatologies, seasonality, indices developed by the Expert Team on Climate Change Detection and Indices (ETCCDI) [Zhang et al. 2011], and preservation of the climate change signal (CCS). Additional technical details describing the final QM configuration is given in Appendix A, as well as implementation details.

## 1.1 Preserving physical characteristics

It is important that certain physical properties and characteristics of the downscaled data are maintained after bias correction. Here we target two important characteristics: the trend and seasonality. Further variable-specific processing is also discussed in Section 2.

**Table 1: Summary of bias corrected variables with the methods used.**

Variable	Method	Reason
Precipitation	Quantile Delta Mapping	Preserve climate change signal; non-Normal distribution
Daily Minimum Near Surface Air Temperature (Tasmin)	Detrended Quantile Mapping	Preserve climate change signal; Normal distribution
Daily Temperature Range (DTR)	Quantile Delta Mapping	Preserve climate change signal; non-Normal distribution
Daily Maximum Near-Surface Air Temperature (Tasmax)	Tasmin + DTR	Preserve Tasmax > Tasmin
Daily Mean Near-Surface Air Temperature (Tas)	$\frac{\text{Tasmax} + \text{Tasmin}}{2}$	Preserve Tasmax > Tas > Tasmin
Potential Evapotranspiration (PET)	Quantile Delta Mapping	Preserve climate change signal; non-Normal distribution

EQM assumes a stationary distribution and can modify trends of non-stationary variables (Maraun, 2016; Maurer & Pierce, 2014; Pierce et al., 2015). Specifically, where CCAM variance is too low the trend will be inflated. This is a particular issue for non-stationary variables such as near-surface air temperature, which is well known to have a clear and robust positive trend (IPCC, 2021). Therefore, explicit trend preservation is recommended for all variables with clear and robust trends (Alavoine & Grenier, 2023; Maraun, 2016).

In the current report, two methods for trend preservation are explored: Detrended Quantile Mapping (DQM) [Cannon, Sobie, and Murdock 2015]; and a variation on Quantile Delta Mapping (QDM) [Cannon, Sobie, and Murdock 2015]. DQM preserves the mean trend by first removing the trend and normalising the data, applying EQM, subsequently adding the trend back on in the post-processing. QDM preserves the trend in each quantile by calculating a transfer function based on the historical quantiles and applies the transfer function to quantiles recalculated for the future period. Both methods reduce to EQM if applying bias correction to data used in the calibration.

Variables such as near-surface air temperature exhibit a distinct seasonal cycle that should be accounted for. Seasonality can be retained by applying the bias correction to each month-of-year or day-of-year. Applying QM in this way creates two further considerations: decreased sample size; and discontinuities in AFs between points in the seasonal cycle. Both considerations can be accounted for by using a window function. This uses data around a calendar period in question, e.g., a 3-month window around December would include data from November and January, in the calculation of quantiles and AFs. This increases the sample size and smoothens discontinuities between calculated eCDFs.

## 2 BCSD Methodology

We bias correct output from CCAM for six driving GCMs for the historical period and projections based on three shared socio-economic pathways (SSPs). For the dynamical downscaling, GCM selection was based on a balance of data availability, historical evaluation, future warming rate, and GCM independence (see Gibson et al. (2024b) for more details). Table 2 presents a list of available bias corrected data for each GCM.

The VCSN dataset was used as a ground truth, to which all variables were bias corrected. We use the augmented VCSN product for rainfall, and the Norton-corrected daily minimum and maximum near-surface air temperature (Tmin & Tmax) for tasmax and tasmin. The augmented VCSN slightly differs from NIWA's operational VCSN rainfall dataset, as it also incorporates rain gauge data from regional councils. The training period for the QM algorithm was taken to be the maximum available intersecting period between the CMIP6 historical experiment and the VCSN (1972-2014), corresponding to 43 years of data.

**Table 2: List of available bias corrected CCAM output for precipitation, tasmax, tasmin & PET for a given driving GCM and shared socio-economic pathway.** An asterisk indicates the GCM was an SST-driven run, else CCAM was nudged by the driving GCM.

GCM	Historical	SSP1-2.6	SSP2-4.5	SSP3-7.0
ERA5-c192	✓	x	x	x
ACCESS-CM2	✓	✓	✓	✓
AWI-CM-1-1-MR*	✓	✓	✓	✓
CNRM-CM6-1*	✓	✓	✓	✓

GCM	Historical	SSP1-2.6	SSP2-4.5	SSP3-7.0
EC-Earth3	✓	✓	✓	✓
GFDL-ESM4*	✓	✓	✓	✓
NorESM2-MM	✓	✓	✓	✓

Several universal parameters were selected when training the QM algorithm, which are detailed in Appendix A. The number of quantiles used to compute an eCDF was selected to be 50. An eCDF was computed for each day of the year, with a 31-day window to incorporate data from 15 days either side. For a given value in the time series to be corrected, the quantile to which it belonged was interpolated to its nearest neighbour in the 50 quantiles, and the corresponding AF is applied. This interpolation method has greater granularity but has the advantage over other methods of not introducing missing values at the upper and lower limits. For any value lying outside the modelled eCDF of the training period, the transfer function is extrapolated as a constant factor (Boé et al., 2007).

A summary of the QM algorithm for each variable and GCM can be described as followed:

1. The eCDFs are computed for the reference dataset (VCSN).
2. The eCDFs are computed for the CCAM output.
3. The eCDFs from the VCSN and the CCAM output are used to develop a transfer function to map CCAM output to the VCSN.
4. The transfer function is applied across all SSPs to the CCAM output.

**Table 3: ETCCDI indices for climate variability and extremes used in this study.**

ETCCDI	Description	Input Variable
CDD	Consecutive dry days (< 1mm)	pr
CWD	Consecutive wet days (≥ 1mm)	pr
SDII	Simple precipitation intensity index	pr
Rx1day	Annual maximum of daily precipitation	pr
Rx3day	Annual maximum of 3-day precipitation	pr
TXx	Annual maximum daily near-surface air temperature	tasmax
TNn	Annual minimum daily near-surface air temperature	tasmin
DTR	Diurnal temperature range	tasmax, tasmin
FD	Annual number of frost days (< 0°C)	tasmin
TX25	Annual number of days above 25°C	tasmax

The final methodology for each variable was modified to maximize bias correction performance whilst retaining their physical characteristics. To ensure the bias correction method improves performance across the entire distribution, we evaluate performance in representing ETCCDI indices, see Table 3 for details. The finalised variable-specific methodologies presented below were selected based on the performance for several skill metrics (Table 4), as well as the applicability given the variable-specific distributional properties.

Each method was cross-validated (CV) using a leave one year out (LOYO) approach, whereby EQM (plus variable-specific pre- and post-processing techniques) was applied to all years in the training period with the omission of a single year. EQM was deemed sufficient as each method – i.e., DQM and QDM – reduces to EQM inside the training period. The year left out was corrected using the transfer function derived from the other 42 years. This process was repeated for all 43 years, and the CV dataset was concatenated from those years.

**Table 4:** Performance indicators used to select the best performing method for each variable considered.

Target	Measure	Metrics
Annual climatology	Mean	RMSE, MAE, PCORR
Summer & winter climatologies	Mean & seasonal cycle	RMSE, MAE, PCORR
Calendar month climatologies	Mean & seasonal cycle	RMSE, MAE, PCORR
ETCCDI	Extremes	RMSE, MAE, PCORR
Annual CCS	Trend preservation	RMSE, MAE, PCORR
Seasonal CCS	Trend preservation & seasonal cycle	RMSE, MAE, PCORR
ETCCDI CCS	Trend preservation & extremes	RMSE, MAE, PCORR
Temporal correlation	Preservation o/f physical signals	PCORR
Spell lengths histogram	Break-up of wet/dry spells	Perkins skill score

## 2.1 Consistent Designation of Land and Ocean Grid Cells

An issue can arise around the coastline when interpolating from the raw CCAM output onto the VCSN grid. Grid cells in CCAM are designated as either ocean or land cells. On the other hand, the VCSN contains only land cells. The distinction must be consistent when downscaling CCAM output onto the VCSN grid as this can fundamentally alter important physical characteristics of a grid cell. For example, daily temperature range is significantly different over the ocean than over land and we want to preserve the daily temperature cycle over land. To achieve this, we allow only land cells to be interpolated to the VCSN grid with the following steps:

1. Mask out all native CCAM ocean cells and set to missing values.
2. Use nearest-neighbour interpolation to assign all ocean cells to the value of the nearest land cell.
3. Interpolate CCAM output onto VCSN grid with bilinear or conservative re-gridding (variable-dependent).
4. Mask out all ocean cells using the VCSN mask.

## 2.2 Temperature variables

Daily minimum and maximum near-surface air temperature (tasmin and tasmax) provide important information about the daily temperature cycle. As such, we want to preserve the relationships between temperature variables. Daily minimum temperature must, by definition, be less than or equal to daily maximum temperature. To maintain this relationship, we bias correct tasmin and the daily temperature range (DTR; defined as tasmax minus tasmin) independently. Then, we post-calculate tasmax from the two corrected variables (tasmin plus DTR).

Tasmin and tasmax are typically normally distributed with a dynamical range on the order of 300K (~27°C). This range is sufficiently far from any physical limits (e.g., zero Kelvin) to avoid introducing unphysical values. The transfer function can therefore be derived and applied additively (see Appendix A for details).

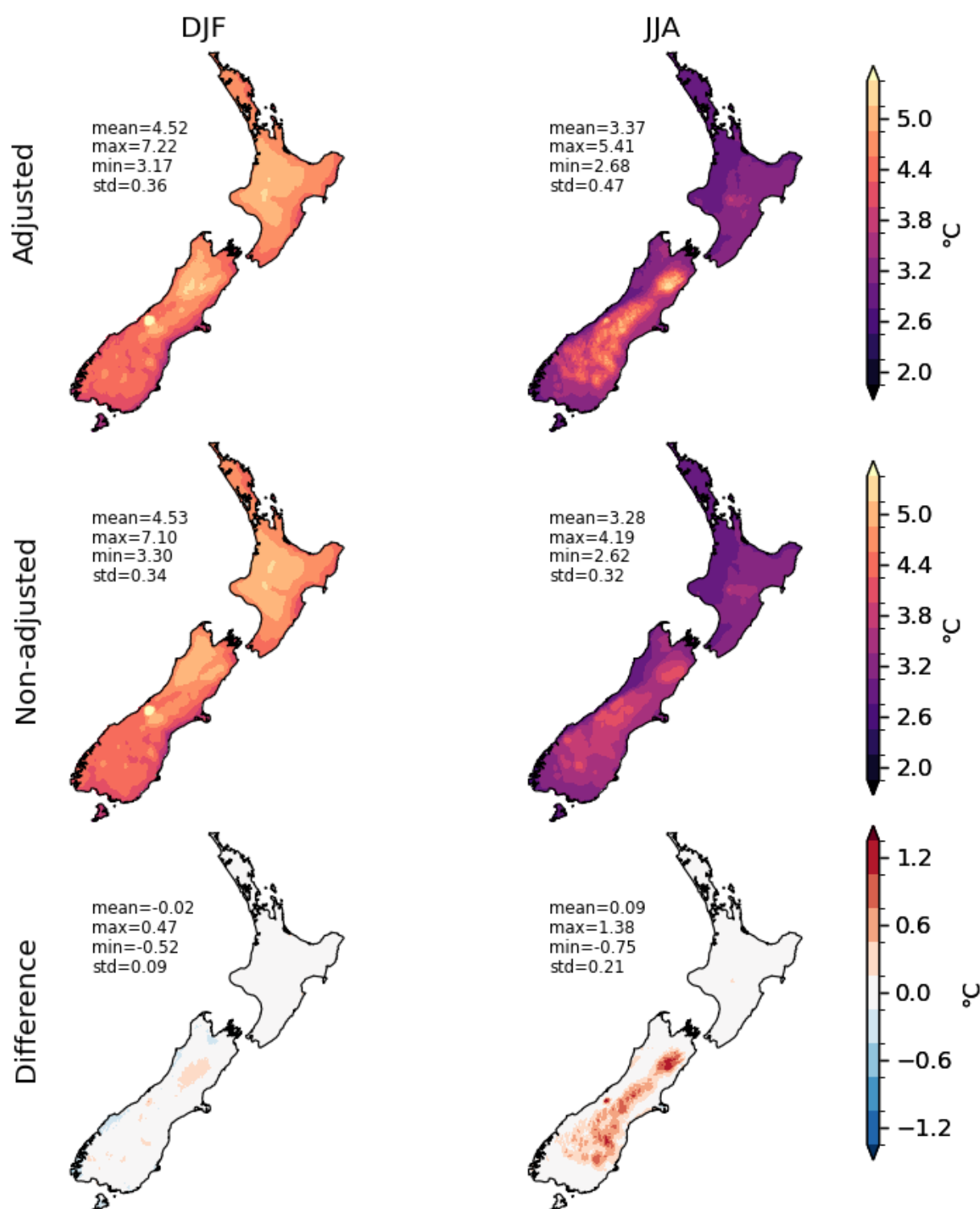
In contrast, the DTR is not normally distributed, and cannot be less than 0°C. To ensure that the DTR in VCSN is always non-negative, we masked out values that fall below a physical threshold (defined as the long-term monthly minima of CCAM DTR for each GCM). By default, QM methods do not have constraints to prevent the introduction of non-physical values. Therefore, several options were explored which apply this constraint, including multiplicative QM, transforming the data into log-space, and using a QM method that does not assume a normally distributed target variable. However, all above methods can result in a strong modification of the CCS in some locations (an example of strong modification of the CCS can be seen in the winter in Figure 2-1).

Consequently, the transfer function was applied to DTR additively, as this best retains the CCS. This did exclude the introduction of negative DTR values, however; the percentage of negative DTR values was acceptably low (<0.05%). Any negative DTR values were subsequently set to zero.

An additional step was taken to preserve the CCS (i.e. trend) throughout the future period, whereby the transfer function was applied using a rolling window. The window was set to contain 20 years of data and rolled over the future period in 5-year steps, i.e., the 85 years of projections (2015-2099) was split into 17 5-year periods, and quantiles were calculated from a 20-year window centred around each 5-year period. This removes the assumption of a linear temperature trend which may not be met at all locations.

In summary, based on the considerations described above, the final methodology for bias correcting tasmin and tasmax was as follows:

1. Bias correct tasmin with additive DQM with a rolling window.
2. Bias correct DTR using additive QDM with a rolling window.
3. Post-calculate tasmax as tasmin plus DTR.



**Figure 2-1: Summer (DJF) and winter (JJA) CCS for bias corrected and raw ACCESS-CM2—CCAM data using multiplicative detrended quantile mapping.** Modification of the CCS can be seen in winter (bottom right).

## 2.3 Precipitation

The distribution of precipitation differs from that of temperature-related variables in several important ways. On the daily timescales considered, precipitation follows a Gamma distribution and possesses a large number of zero values (i.e. dry days). This creates several issues in the context of QM. A variety of methods have been suggested to account for these issues. In the following, we

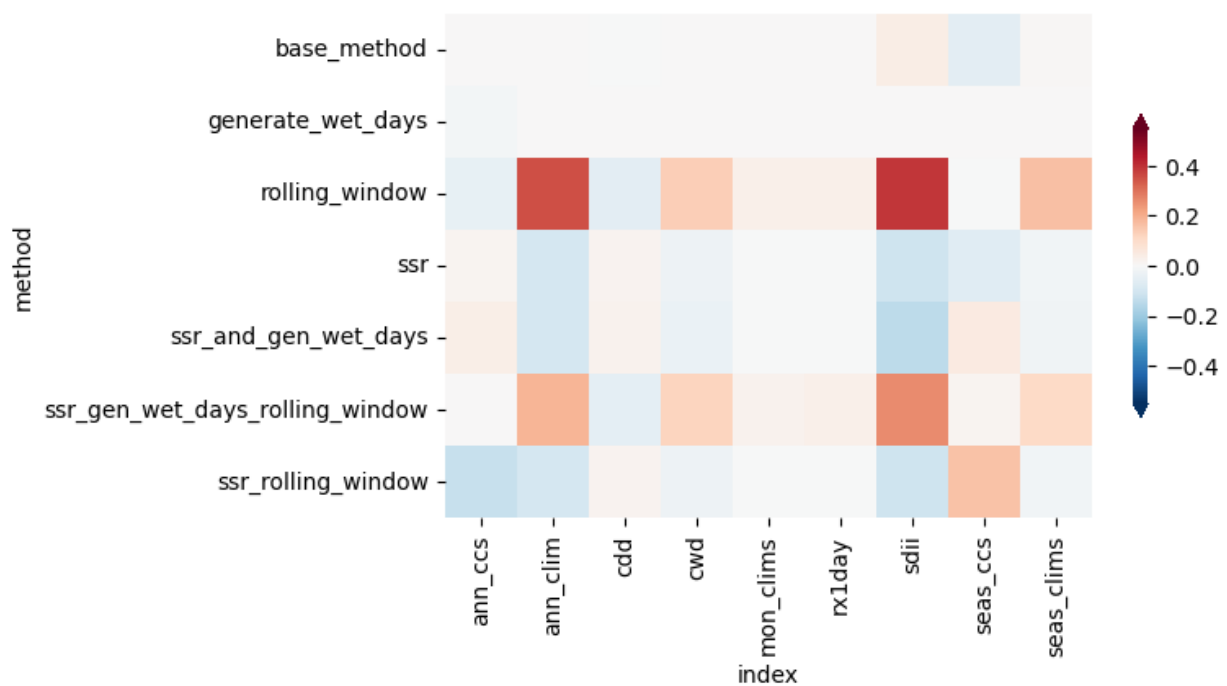
outline some of the most promising methods, and provide an evaluation of them in isolation and in combination with other methods for the NZ region.

Firstly, a naïve application of QM may result in negative values of precipitation. Negative precipitation is not possible in the real world, but the statistical bias correction process does not respect this boundary without imposing additional constraints. To account for this, here the transfer function is applied *multiplicatively* (see Appendix A for details). As a consequence of applying QM multiplicatively, the bias correction can introduce non-finite values. With multiplicative QM, the transfer function is derived as a ratio between the reference and CCAM eCDFs. This can introduce some cases where the denominator (a quantile value from the reference eCDF) is a zero value. Here, we apply the method of Cannon, Sobie, and Murdock (2015), termed Singularity Stochastic Removal (SSR) by Vrac, Noël, and Vautard (2016), to avoid this. SSR removes all zero values by adding normally distributed noise (up to 0.05mm per day) to all zero values, before calculating the transfer function, thereby removing the possibility of any quantile taking the value of zero. In the post-processing, all values below the upper noise limit are set to zero.

We use QDM to bias correct precipitation nation-wide as many regions across NZ can have statistically significant trends in the future period that vary by season (Gibson et al. 2024a). The QDM implementation used here does not assume a linear combination of mean trend and normally distributed residuals, rather the quantiles in the future period are recalculated and the transfer function applied accordingly. This preserves the empirical nature of the QM method – no assumption about the precipitation eCDF is made in the trend preservation. Another advantage of this implementation is the reduction to EQM for instances where there is no trend, i.e., drawn from the same stationary distribution. Therefore, the trend preservation is applied on a grid cell by grid cell basis.

QM has a known artefact which results in an overestimation of the wet-day frequency (the number of wet days in a given year,  $r_0$ ) relative to the observed frequency (Casanueva et al., 2020; Themeßl et al., 2012). In this instance, a dry day refers to a day in which less than 1mm of precipitation occurs. The overestimation occurs as CCAM modelled dry days are mapped to wet days. This can be avoided by explicitly introducing wet days in the modelled time series to match the observed  $r_0$  (Generate Wet Days) Themeßl, Gobiet, and Heinrich (2012) suggest linearly interpolating between zero precipitation and the modelled value at which  $r_0^{\text{model}}$  would be equal to  $r_0^{\text{obs}}$ . Casanueva et al. (2020) build on this method by replacing the linear interpolation by sampling a gamma distribution to better recreate the expected eCDF.

A final method was considered, whereby we apply the transfer function with a rolling window function during the future period to allow for non-linear trends, as described above in Section 2.2. To identify the best performing bias correction technique for precipitation over NZ, we applied each method (or combination of) to three models (ACCESS-CM2, EC-Earth3 & GFDL-ESM4) and calculated the skill metrics for a sub-sample of performance indicators from Table 4. The multi-model ensemble (MME) mean was taken for each metric, presented in Figure 2-2.



**Figure 2-2: Heatmap comparing performance of several methodological variants for 9 indices.** The methods labels indicate the processing method (or combination of) that has been applied, each of which have been described above. The base method in this case is QDM. The colour represents relative performance compared to the median method, with red indicating a weaker performance and blue a stronger performance. The performance score is derived from the three-model ensemble mean, aggregated across four skill metrics (PCORR, RMSE, MAE, and MAPE) which were weighted by their mean value to take a value around unity. Seasonal and monthly indices were averaged across their component calendar periods.

Using QDM as the basic method, the addition of SSR achieved the best performance on balance. Using SSR showed improved performance for several indices, notably the annual accumulated precipitation climatology and the simple daily intensity index (SDII), without seriously compromising performance across other indices.

## 2.4 Potential Evapotranspiration

To maintain a consistent definition of PET we produce offline calculations of PET for both CCAM and VCSN using the Penman method (Penman & Keen, 1997). We calculate PET using the Pyet python package (Vremec et al., 2023) with the following data from the CCAM and VCSN: tasmax, tasmin, tas, relative humidity, 10m wind speed logarithmically scaled to 2m windspeed, downwelling shortwave radiative flux at the surface, elevation, and latitude. The VCSN wind speed data is only available from 1997 onwards, therefore, the maximum overlapping period between CCAM and VCSN from which we can calculate the transfer function is reduced to 1997-2014 for PET.

To address this reduced sample size, a few key parameters in the bias correction of PET were modified. The number of quantiles was here reduced to 25, and the data was grouped by month with a 3-month window. These changes were found to help reduce over-fitting of the bias correction to the training data. As with precipitation, PET cannot be assigned a negative value and tends to follow a non-normal distribution. Therefore, a similar method to the one applied to precipitation was followed for PET. Namely, we use QDM as the base bias correction method, and SSR to prevent the introduction of non-finite values. PET also exhibits a significant trend due to its relationship with temperature in its derivation. The trend was re-applied to the future period using a rolling window, with a 20-year window and a 5-year step size, as described in Section 2.2.



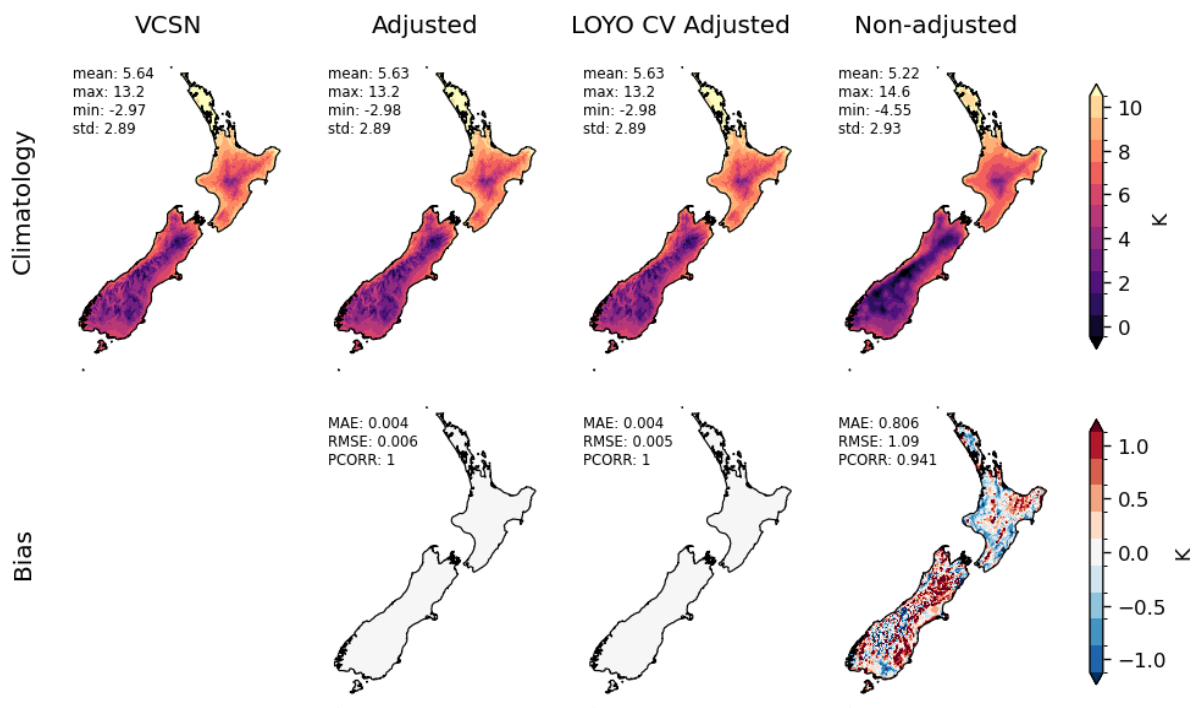
### 3 BCSD Performance Evaluation

This section describes the performance of the final chosen bias correction methods for the indicators outlined in Section 2. The figures presented are selected from three downscaled models: ACCESS-CM2; EC-Earth3; and GFDL-ESM4. These downscaled GCMs were chosen to represent a reasonably large spread of Equilibrium Climate Sensitivity (ECS) and differences in the CCAM experiment design (i.e. SST-driven versus nudged model runs). While there are some differences between downscaled GCMs in overall performance, the results presented here are overall a good representation of the performance of the bias correction on other GCMs in the ensemble.

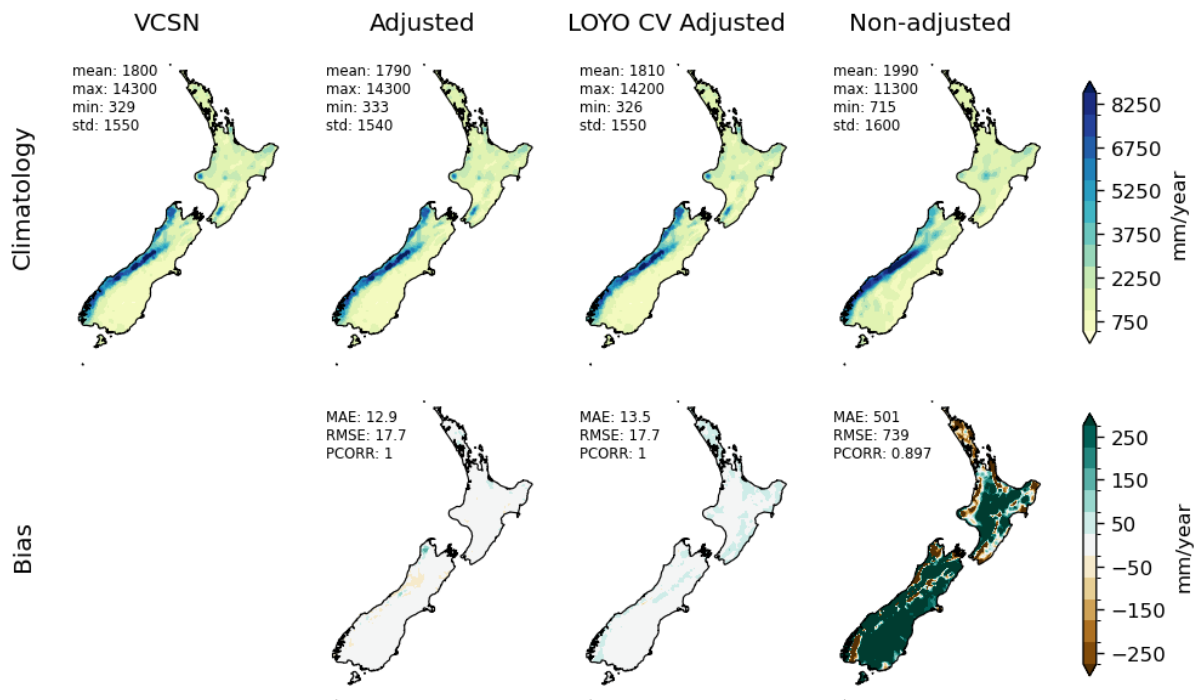
#### 3.1 Annual Climatologies

As we discussed earlier, the bias correction computes transfer functions on a day-of-year basis to ensure the seasonal cycle is retained. However, these day-of-year transfer functions also perform well in correcting annual climatologies (e.g. annual temperature and rainfall). For example, bias correction considerably improves the tasmin annual climatology, with the mean absolute error (MAE) from VCSN decreasing from 0.80°C from the raw CCAM output (non-adjusted) to 0.00°C (to three significant figures) with bias correction (Figure 3-1). Spatial correlation is also greatly improved, subsequently taking a value of 1.00 (to 3 s.f.). The tasmax annual climatology is similarly improved (not shown), with MAE decreasing from 1.32°C to 0.06°C, and spatial correlation of 1. These very small remaining climatological biases after bias correction are considered inconsequential given the observational uncertainty in VCSN itself. The bias correction for precipitation annual climatology performs similarly well, with MAE reduced from 501mm/year in the raw CCAM output to 12.9mm/year after bias correction (Figure 3-2).

The LOYO CV for all three variables shows little deviation from the bias corrected data using the full calibration period, suggesting that the model successfully generalises for those distributions.



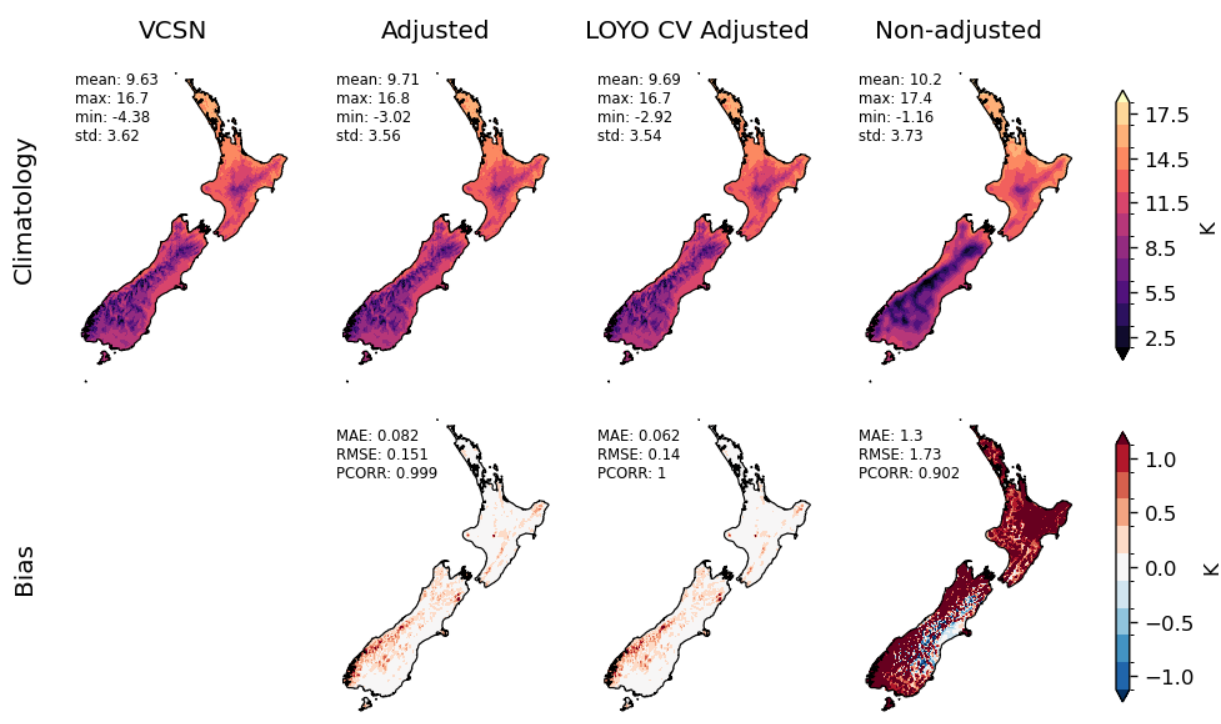
**Figure 3-1:** Annual climatologies for tasmin (1972-2014) for VCSN, bias corrected CCAM output, LOYO CV and raw CCAM output (top row), along with bias from VCSN (bottom row).



**Figure 3-2: Annual climatologies for accumulated precipitation (1972-2014) for VCSN, bias corrected ACCESS-CM2—CCAM output, LOYO CV and raw ACCESS-CM2—CCAM output (top row), along with bias from VCSN (bottom row) .**

### 3.2 Seasonal cycle

Overall, we find that bias-correction significantly improves the representation of the seasonal cycle across all climate variables. For each of the bias corrected variables, MAE and RMSE remain of the same order of magnitude found for annual climatologies, with spatial correlations around 1.00. An example is given for winter tasmax in Figure 3-3.



**Figure 3-3: Winter (JJA) climatologies for tasmax (1972-2014) for VCSN, bias corrected EC-Earth3—CCAM output, LOYO CV and raw EC-Earth3—CCAM output (top row), along with bias from VCSN (bottom row).**



**Figure 3-4: Long-term monthly mean accumulated precipitation for 12 locations across NZ for bias corrected GFDL-ESM4—CCAM output.**

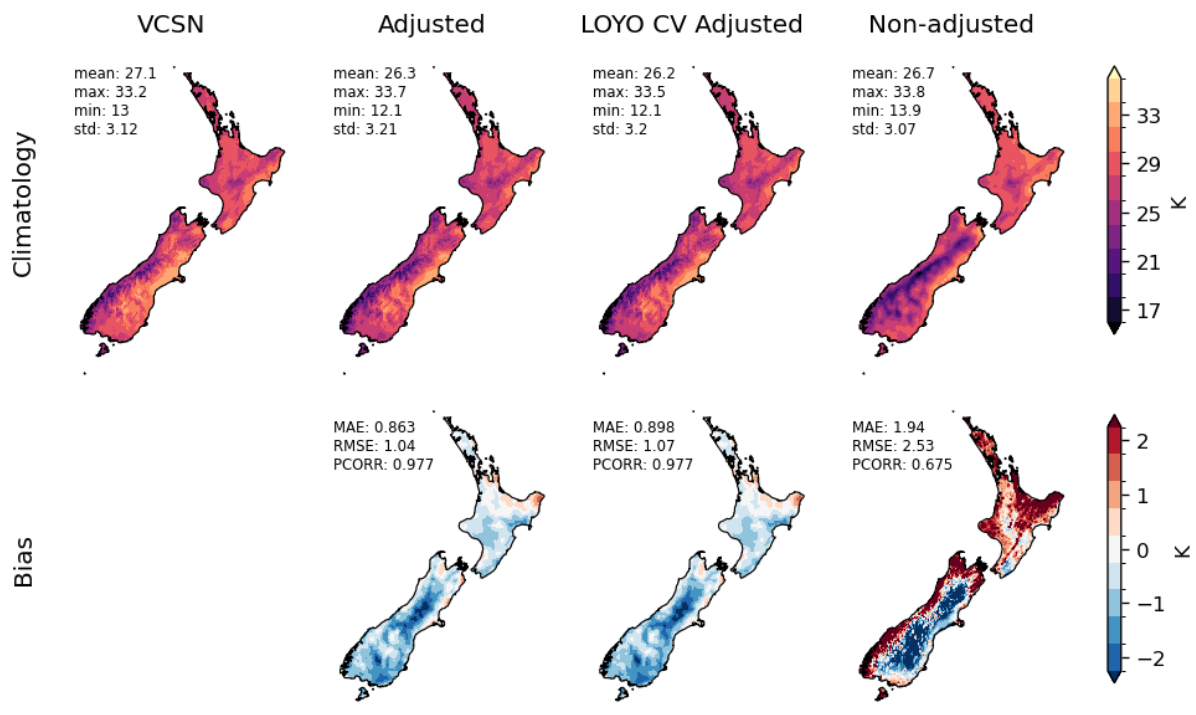
To further illustrate how bias correction improves the seasonal cycle, we selected 12 locations across NZ to more closely inspect the seasonal cycle after bias correction (e.g., Figure 3-4). The chosen sites were either of population and/or economic significance and provided a representative sample of the various climatic conditions nationwide. The long-term monthly means are improved across all variables. The need for bias-correction of the seasonal cycle is particularly evident for precipitation in certain regions and GCMs. For example, for GFDL-ESM4, the raw CCAM output struggles to capture the seasonal cycle over Fiordland with excessive winter precipitation (Figure 3-4), which is subsequently corrected for in the bias corrected output. Note that this seasonal bias is often smaller when CCAM is applied to downscaling other GCMs, but GFDL-ESM4 has been selected shown here to highlight the added value of the bias correction.

### 3.3 ETCCDI

QM methods are highly effective at capturing the climatological means; however, the higher-order moments (i.e. extreme events) of the distribution are more difficult to bias correct, given their infrequency. For example, the 99<sup>th</sup> percentile may be calculated from a relatively small number of data points (relative to the average), and thus the bias corrected output may have a higher error

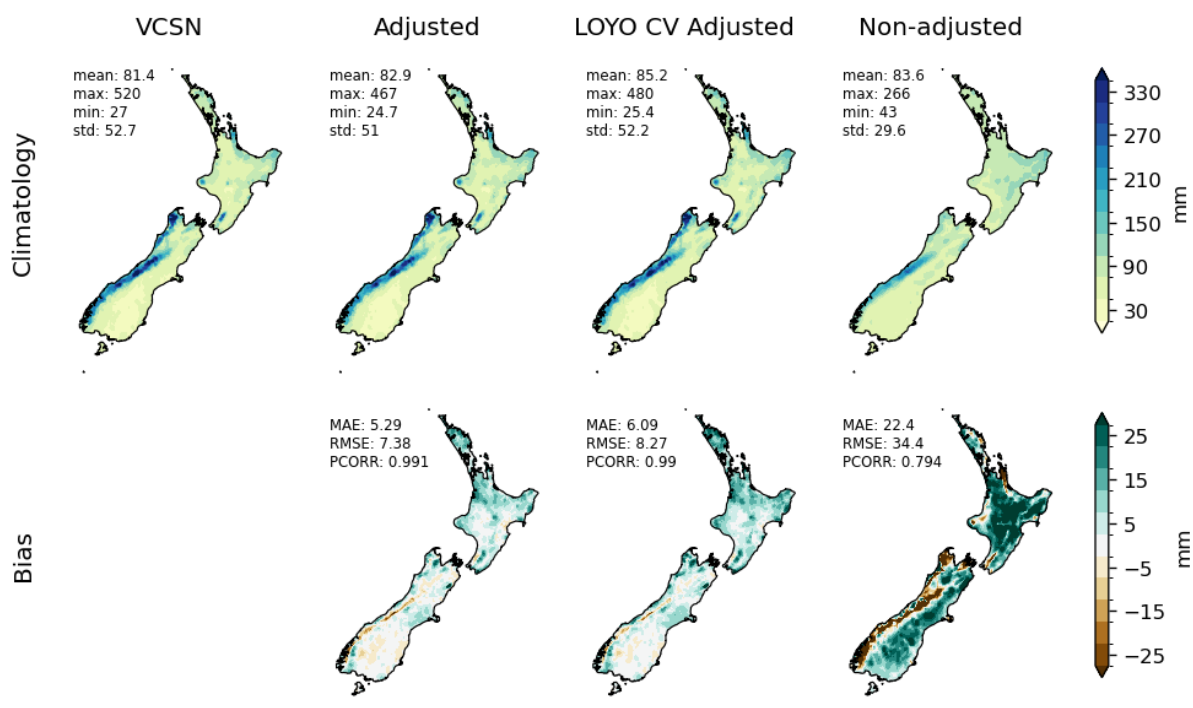
than for the climatological mean. To assess how the bias correction performs for higher order moments we evaluate several ETCCDI indices. The discussion provided here represents a subset of the full ETCCDI indices that were evaluated. Overall, we find that bias correction performs well in capturing higher order moments of the distribution.

The bias correction for hot temperature extremes shows reduced performance relative to the seasonal and annual climatological tasmax, as expected. Figure 3-5 shows the annual daily maximum temperature (TXx) climatology for the EC-Earth3 nudged run. The warm biases found across the North Island (NI) and around the coastal regions of the South Island (SI) are largely eliminated. However, the cold biases across the interior persist after bias correction, although to a lesser degree. Nevertheless, it is important to highlight that there is a marked improvement across skill metrics after bias correction is applied, with RMSE decreasing from 2.53°C for the raw CCAM output to 1.04°C for the bias corrected output.



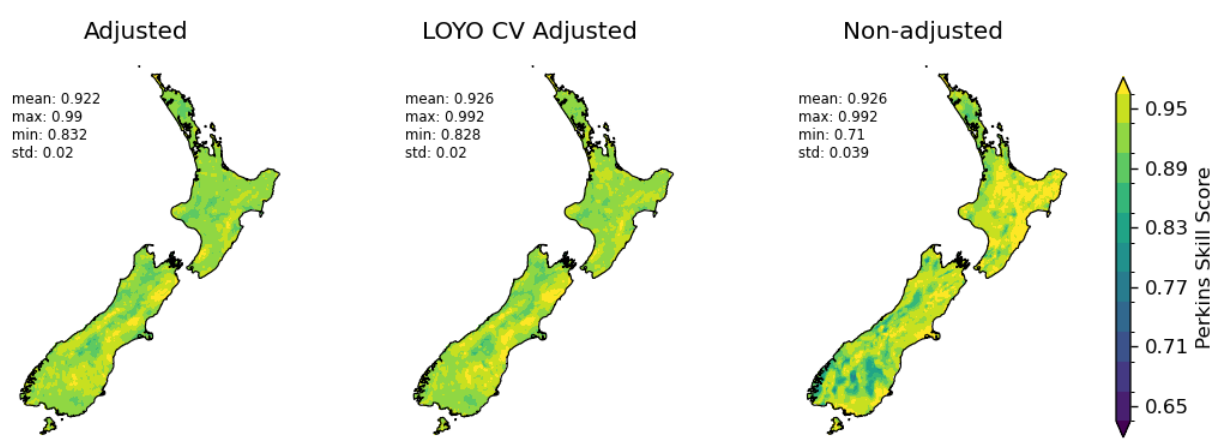
**Figure 3-5: Annual climatologies for TXx (1972-2014) for VCSN, bias corrected EC-Earth3—CCAM output, LOYO CV and raw EC-Earth3—CCAM output (top row), along with bias from VCSN (bottom row).**

High intensity rainfall (Rx1day) across the calibration period (1972-2014) also shows significant improvements from bias correction compared to the raw CCAM output (Figure 3-6). Spatial correlation increases from 0.79 to 0.99, and RMSE significantly decreases from 34.4 mm/day to 7.4 mm/day. While dry biases across the Southern Alps remain, they are much smaller in the bias corrected output. In contrast, the wet biases across parts of Northland have not been improved from bias correction for Rx1day.



**Figure 3-6: Annual climatologies for highest intensity rainfall for a single day [Rx1day] (1972-2014) for VCSN, bias corrected EC-Earth3—CCAM output, LOYO CV and raw EC-Earth3—CCAM output (top row), along with bias from VCSN (bottom row).**

Despite the considerable reductions in biases documented above, one known area of weakness of QM methods is the tendency to break up the length of wet spells, leading to a systematic shortening of wet spells. This overall effect can be seen by comparing histograms of the wet spell length using the Perkins skill score, which measures the overlap of two distributions (Perkins et al., 2007). A high Perkins skill score (a metric which ranges from 0 to 1) in a certain location would here indicate that the length of all wet spell events is comparable between observations and CCAM. Areas of slightly diminished performance (i.e. slightly lower skill scores) are evident across the North Island in Figure 3-7, reflecting the tendency to break-up longer wet spells. On the other hand, in other regions (e.g. Central Otago) bias correction improves the representation of wet spell length. This can be traced back to the raw CCAM output overestimating the frequency and length of wet spells in this region (Campbell et al. 2024, in review).

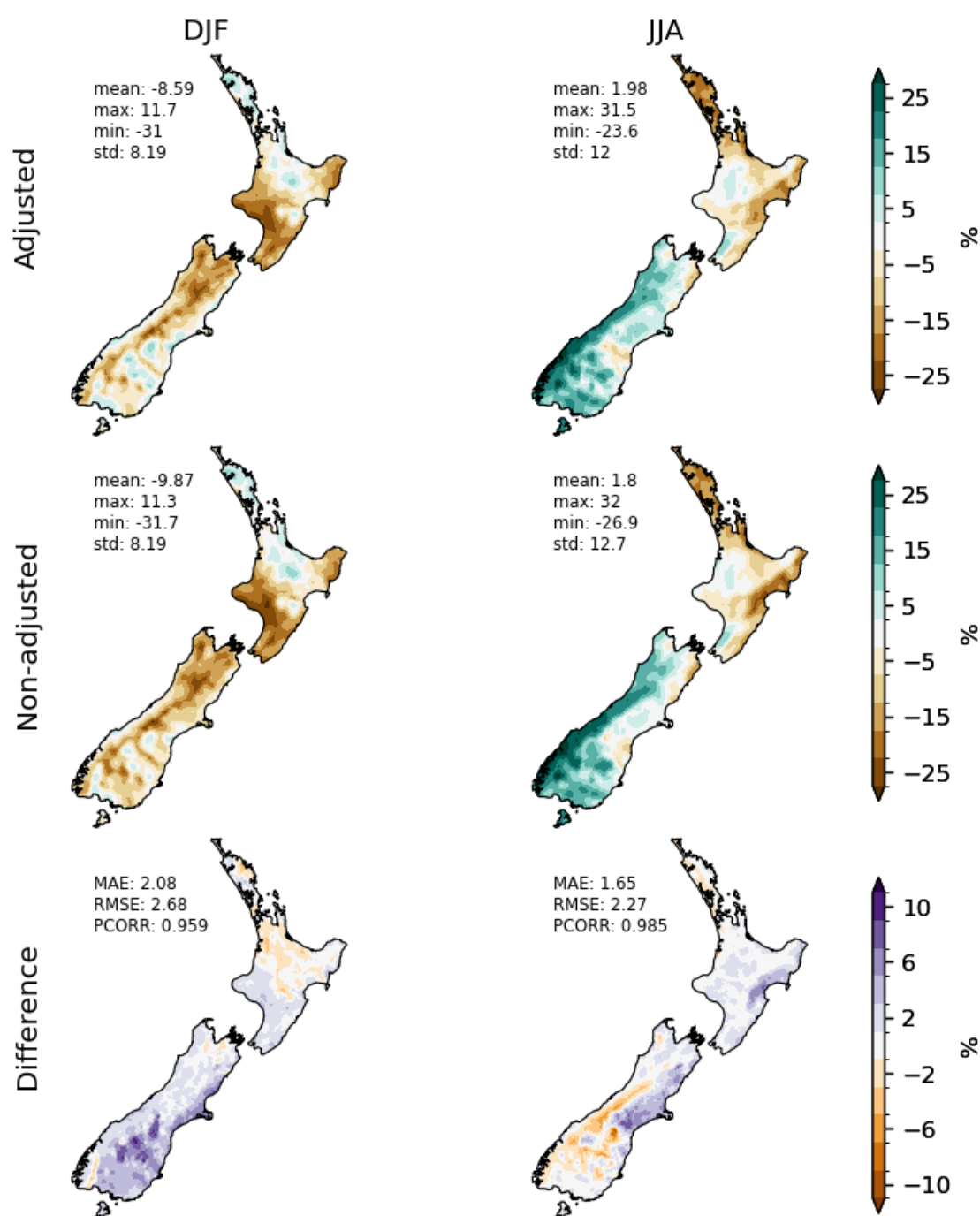


**Figure 3-7: Perkins skill score comparing the histograms of wet spell lengths (consecutive days with greater than 1mm of rain) against VCSN for the bias corrected EC-Earth3—CCAM output, the corresponding cross-validated corrected output, and the raw output (interpolated to 5km). Twenty bins were used to construct the histograms up to 20-day spells.**

### 3.4 Climate Change Signal

As discussed earlier, the climate change signal (CCS) from the raw CCAM output for each variable was a priori assumed to be more reliable than allowing the bias correction to substantially alter the CCS. In some cases, EQM can artificially inflate (or deflate) the CCS when there is a mismatch in variance between the calibration and reference data (Maurer and Pierce 2014; Maraun 2016; see Appendix A for further details). Explicit trend preservation was therefore implemented for all variables (see Section 2 for details). Here, we evaluate the success of that trend preservation relative to the trend in CCAM.



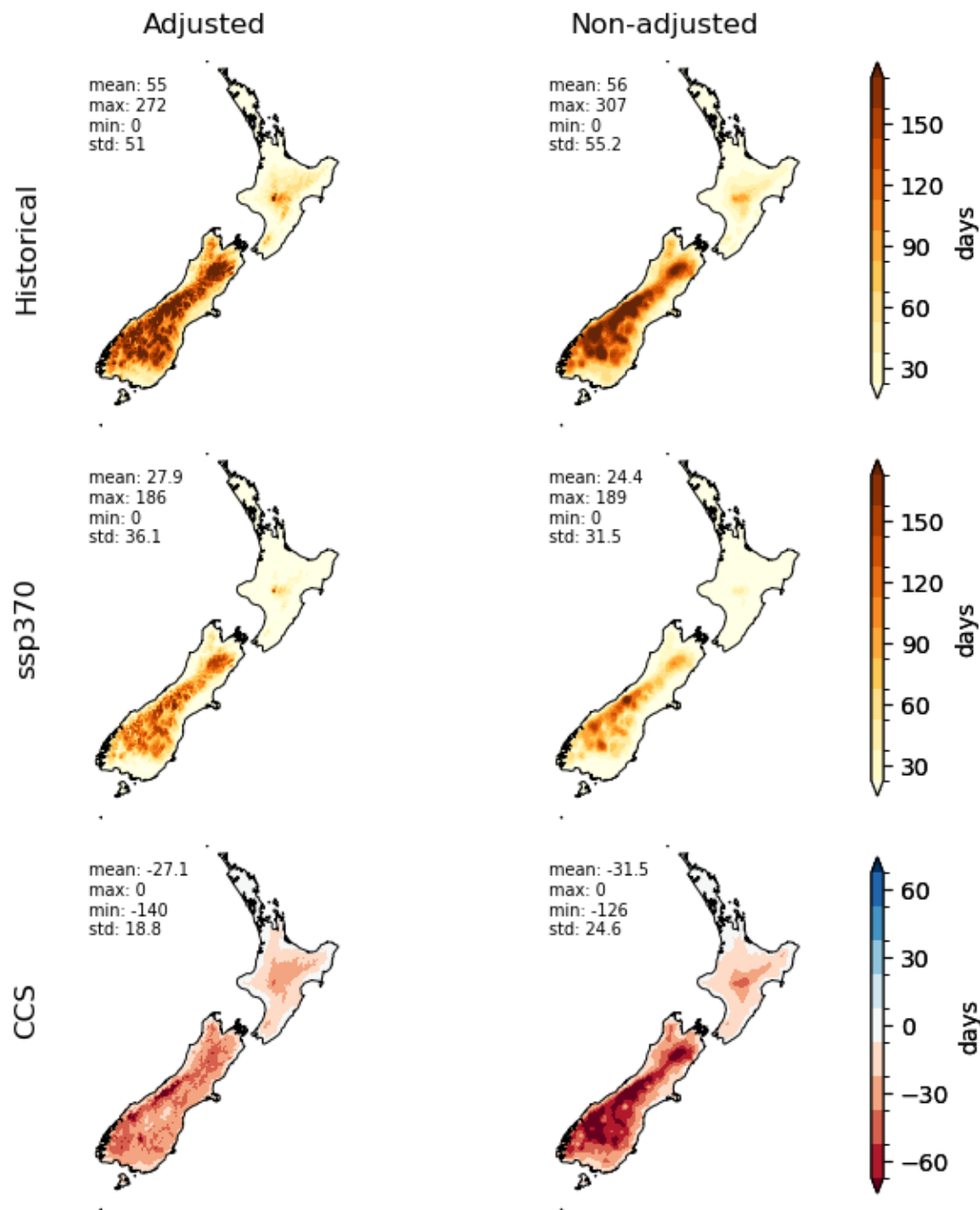


**Figure 3-8: Climate change signal between the historical (1985-2014) and SSP3-7.0 (2070-2099) experiments for summer (DJF) and winter (JJA) for accumulated precipitation over those seasons.** The top row shows the CCS for bias corrected GFDL-ESM4—CCAM output; the middle row shows the corresponding raw model output; and the bottom row indicates the difference between the two.

The CCS for tasmin and tasmax were almost completely unchanged between the raw and corrected CCAM output (not shown), indicating the trend was well-preserved, whereas there was greater alteration to the CCS for precipitation. Figure 3-8 provides an example figure for the GFDL-ESM4 SST-driven run, with some areas seeing a CCS difference exceeding  $\pm 10\%$ . This mostly comes in regions where there is little precipitation to begin with, so the absolute changes are much smaller. However, there are regions where there are considerable absolute and relative changes, such as over the Southern Alps in winter. As mentioned previously, this is likely related to the higher sensitivity to



error of multiplicative QM. Also, the reference period for this CCS was 2070-2099, whereas the whole future period (2015-2099) was used to calculate the eCDF. If the end-of-century period differed considerably from the rest of the century, whether due to the high temporal variability characteristic of rainfall, or a linear trend, this could impact the CCS preservation.



**Figure 3-9: Annual climatologies number of frost days for the historical (1985-2014) and SSP3-7.0 (2070-2099) experiments and the CCS.** The left column shows the bias corrected GFDL-ESM4—CCAM output, and the corresponding raw output in the right column.

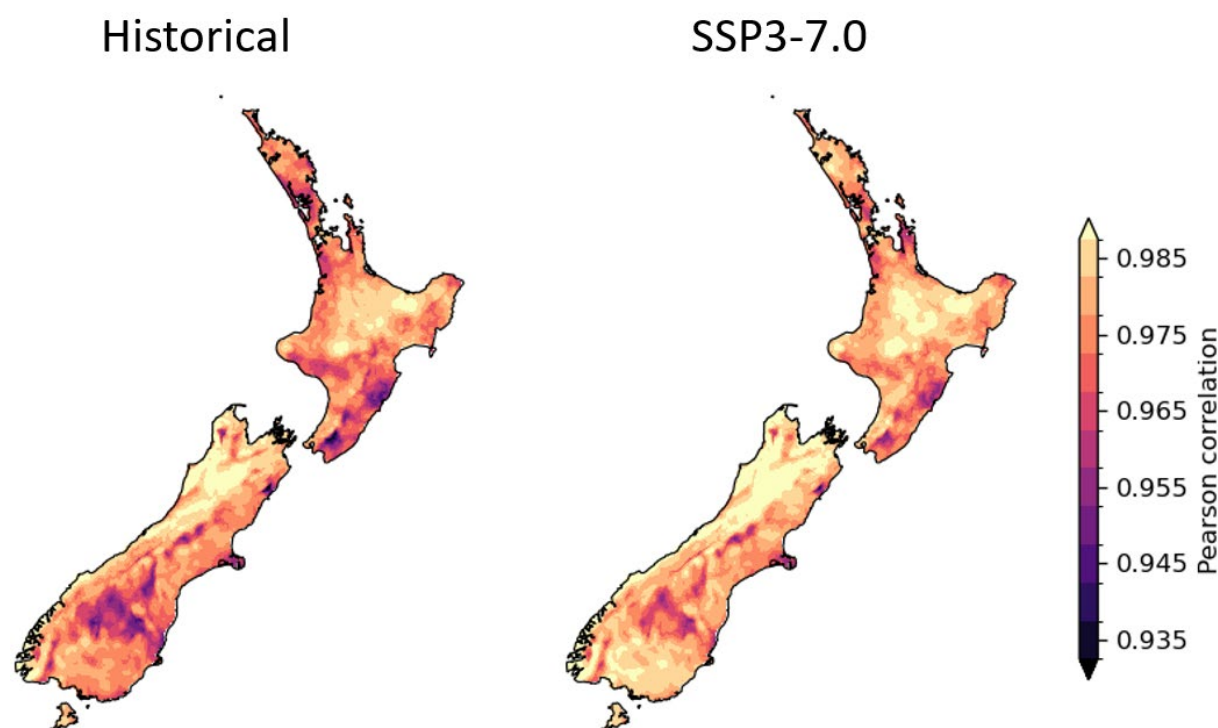
We also explored the trend preservation for the ETCCDI indices and provide one pertinent example in the change in frost days. Generally, the trend preservation was considered to perform well across variables and seasons. The clearest difference was the number of frost days (FD) across high-

elevation regions, where the bias corrected CCAM output shows a smaller decrease in FD relative to the raw CCAM output (Figure 3-9). This is possibly related to the snow albedo effect: with the reduction of snow in a warmer world, the cooling effect of the increased albedo is lost. The QM cannot be expected to capture this change which introduces some differences in the CCS in high elevation regions. Indices with a strong threshold dependence are likely subject to the greatest differences, where a small change in a continuous variable (e.g. air temperature) can induce relatively large differences in a categorical variable (e.g. frost days).

### 3.5 Temporal correlation

An important assumption of the QM method is the ability to retain statistical properties related to the underlying physical processes in CCAM. We use temporal correlation between the bias corrected and raw model output as a measure of the retention of these properties.

The bias corrected daily output is still highly temporally correlated with the raw daily output (e.g., Figure 3-10), suggesting that important temporal properties have not been degraded through bias correction. For precipitation, the reasons for the smaller correlations over the Wairarapa and Central Otago cannot be clearly linked to any one cause, although do appear to roughly coincide with the strong modifications to the wet spell lengths described earlier. Similarly, the smaller correlations for temperature variables generally appear larger in high-elevation regions (not shown).



**Figure 3-10:** Temporal correlation between bias corrected ACCESS-CM2—CCAM output and the corresponding raw model output for daily accumulated precipitation in the historical (1960-2014) and SSP3-7.0 (2015-2099) experiments.

## 4 Summary and conclusions

We have statistically downscaled and bias corrected CMIP6 climate projections from the CCAM dynamical model over the NZ region. The bias correction procedure targeted four daily variables: precipitation, maximum and minimum near-surface air temperature, and potential

evapotranspiration. An empirical technique known as quantile mapping was used to bias correct the modelled CCAM distributions against VCSN.

This report has described the applied methodology and thoroughly evaluated the output. Biases and climatological spatial patterns show substantial improvements for all four key climate variables. The seasonal cycle, especially for precipitation, also shows large improvements. Furthermore, the climate change signal and daily temporal correlations for these key variables are shown to be well preserved. However, no bias correction is perfect, and here we have highlighted some remaining challenges. While the climatological biases for extreme events (e.g. rainfall and temperature extremes) are also substantially reduced relative to the raw CCAM output, some regional biases for these indices remain after bias correction. Similarly, bias correction tends to slightly reduce the length of wet spells compared to the raw CCAM output. Another challenge is in preserving the climate change signal for indices that are strongly threshold dependent (e.g. frost days). We have provided guidance on the limitations of the bias corrected data, and how it can be used most appropriately.

## 5 Acknowledgements

We would like to acknowledge the contributions of our colleagues to this work and the broader CMIP6 downscaling project: Nicolas Fauchereau, Hamish Lewis, Stephen Stuart, Ashley Broadbent, Abha Sood, Olaf Morgenstern, Daithí Stone, Andrew Tait, and Jonny Williams.

## 6 Glossary of abbreviations and terms

AF	Adjustment Factor
BCSD	Bias Correction-Statistical Downscaling
CCS	Climate Change Signal
CMIP	Coupled Model Intercomparison Project. CMIP6 refers to the sixth generation and CMIP5 to the fifth generation of this project.
CV	Cross validation
eCDFs	Empirical Cumulative Distribution Function
EQM	Empirical Quantile Mapping
ETCCDI	Expert Team on Climate Change Detection and Indices
GCM	Global Climate Model
LOYO	Leave One Year Out. A method for cross-validation where one year of data is left out of the bias correction model to see how well the model generalises to out-of-sample data.
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
PCORR	Pattern correlation
RCM	Regional Climate Model
RMSE	Root Mean Squared Error
SSP	Shared-socio-economic pathway. Assumed emissions scenarios for the future period, combining societal developments with radiative forcings. Replaces Representative Concentration Pathways (RCPs) from CMIP5.
tasmax	Daily maximum near-surface air temperature
tasmin	Daily minimum near-surface air temperature
VCSN	Virtual Climate Station Network

## 7 References

- Alavoine, M., Grenier, P. (2023) The distinct problems of physical inconsistency and of multivariate bias involved in the statistical adjustment of climate simulations. *International Journal of Climatology*, 43(3): 1211–1233. <https://doi.org/10.1002/joc.7878>
- Boé, J., Terray, L., Habets, F., Martin, E. (2007) Statistical and dynamical downscaling of the Seine basin climate for hydro-meteorological studies. *International Journal of Climatology*, 27(12): 1643–1655. <https://doi.org/10.1002/joc.1602>
- Bourgault, P., Huard, D., Smith, T. J., Logan, T., Aoun, A., Lavoie, J., Dupuis, É., Rondeau-Genesse, G., Alegre, R., Barnes, C., Laperrière, A. B., Biner, S., Caron, D., Ehbrecht, C., Fyke, J., Keel, T., Labonté, M.-P., Lierhammer, L., Low, J.-F., ... Whelan, C. (2023) xclim: Xarray-based climate data analytics. *Journal of Open Source Software*, 8(85): 5415. <https://doi.org/10.21105/joss.05415>
- Campbell, I., Gibson, P. B., Stuart, S., Broadbent, A., Sood, A., S. Pirooz, A. A., & Rampal, N. (2024, under review) Comparison of three reanalysis-driven regional climate models over New Zealand: Climatology and extreme events. *International Journal of Climatology*.
- Cannon, A.J. (2018) Multivariate quantile mapping bias correction: An N-dimensional probability density function transform for climate model simulations of multiple variables. *Climate Dynamics*, 50(1): 31–49. <https://doi.org/10.1007/s00382-017-3580-6>
- Cannon, A.J., Sobie, S.R., Murdock, T.Q. (2015) Bias Correction of GCM Precipitation by Quantile Mapping: How Well Do Methods Preserve Changes in Quantiles and Extremes? *Journal of Climate*, 28(17): 6938–6959. <https://doi.org/10.1175/JCLI-D-14-00754.1>
- Casanueva, A., Herrera, S., Iturbide, M., Lange, S., Jury, M., Dosio, A., Maraun, D., Gutiérrez, J.M. (2020) Testing bias adjustment methods for regional climate change applications under observational uncertainty and resolution mismatch. *Atmospheric Science Letters*, 21(7): e978. <https://doi.org/10.1002/asl.978>
- Gibson, P.B., Rampal, N., Dean, S.M., Morgenstern, O. (2024a) Storylines for Future Projections of Precipitation Over New Zealand in CMIP6 Models. *Journal of Geophysical Research: Atmospheres*, 129(5): e2023JD039664. <https://doi.org/10.1029/2023JD039664>
- Gibson, P.B., Stuart, S., Sood, A., Stone, D.A., Rampal, N., Lewis, H., Broadbent, A., Thatcher, M., Morgenstern, O. (2024b, under review) Dynamical downscaling CMIP6 models over New Zealand: Added value of climatology and extremes. *Climate Dynamics*.
- Gudmundsson, L., Bremnes, J.B., Haugen, J.E., Engen-Skaugen, T. (2012) Downscaling RCM precipitation to the station scale using statistical transformations – a comparison of methods. *Hydrology and Earth System Sciences*, 16(9): 3383–3390. <https://doi.org/10.5194/hess-16-3383-2012>

- Gutjahr, O., Heinemann, G. (2013) Comparing precipitation bias correction methods for high-resolution regional climate simulations using COSMO-CLM. *Theoretical and Applied Climatology*, 114(3): 511–529. <https://doi.org/10.1007/s00704-013-0834-z>
- IPCC (2021) *The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. IPCC, Geneva, Switzerland.
- Ivanov, M.A., Kotlarski, S. (2017) Assessing distribution-based climate model bias correction methods over an alpine domain: Added value and limitations. *International Journal of Climatology*, 37(5): 2633–2653. <https://doi.org/10.1002/joc.4870>
- Kim, K.B., Kwon, H.-H., Han, D. (2016) Precipitation ensembles conforming to natural variations derived from a regional climate model using a new bias correction scheme. *Hydrology and Earth System Sciences*, 20(5): 2019–2034. <https://doi.org/10.5194/hess-20-2019-2016>
- Maraun, D. (2016) Bias Correcting Climate Change Simulations—A Critical Review. *Current Climate Change Reports*, 2(4): 211–220. <https://doi.org/10.1007/s40641-016-0050-x>
- Maraun, D., Shepherd, T.G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J.M., Hagemann, S., Richter, I., Soares, P.M.M., Hall, A., Mearns, L.O. (2017) Towards process-informed bias correction of climate change simulations. *Nature Climate Change*, 7(11): 764–773. <https://doi.org/10.1038/nclimate3418>
- Maurer, E.P., Pierce, D.W. (2014) Bias correction can modify climate model simulated precipitation changes without adverse effect on the ensemble mean. *Hydrology and Earth System Sciences*, 18(3): 915–925. <https://doi.org/10.5194/hess-18-915-2014>
- Met Office (2018) *UKCP18 Guidance: Bias correction*. <https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/research/ukcp/ukcp18-guidance---how-to-bias-correct.pdf>
- Mullan, B., Sood, A., Stuart, S., Carey-Smith, T. (2018) *Climate Change Projections for New Zealand: Atmosphere Projections Based on Simulations from the IPCC Fifth Assessment*. Ministry for the Environment.
- Penman, H.L., Keen, B.A. (1997) Natural evaporation from open water, bare soil and grass. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 193(1032): 120–145. <https://doi.org/10.1098/rspa.1948.0037>
- Perkins, S.E., Pitman, A.J., Holbrook, N.J., McAneney, J. (2007) Evaluation of the AR4 Climate Models' Simulated Daily Maximum Temperature, Minimum Temperature, and Precipitation over Australia Using Probability Density Functions. *Journal of Climate*, 20(17): 4356–4376. <https://doi.org/10.1175/JCLI4253.1>
- Pierce, D.W., Cayan, D.R., Maurer, E.P., Abatzoglou, J.T., Hegewisch, K.C. (2015) Improved Bias Correction Techniques for Hydrological Simulations of Climate Change. *Journal of Hydrometeorology*, 16(6): 2421–2442. <https://doi.org/10.1175/JHM-D-14-0236.1>

- Tait, A., Macara, G. (2014) Evaluation of interpolated daily temperature data for high elevation areas in New Zealand. *Weather and Climate*, 34: 36–49.  
<https://doi.org/10.2307/26169743>
- Tait, A., Sturman, J., Clark, M. (2012) An assessment of the accuracy of interpolated daily rainfall for New Zealand. *Journal of Hydrology (New Zealand)*, 51(1): 25–44.
- Thiemeßl, M.J., Gobiet, A., Heinrich, G. (2012) Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal. *Climatic Change*, 112(2): 449–468. <https://doi.org/10.1007/s10584-011-0224-4>
- Vrac, M., Noël, T., Vautard, R. (2016) Bias correction of precipitation through Singularity Stochastic Removal: Because occurrences matter. *Journal of Geophysical Research: Atmospheres*, 121(10): 5237–5258. <https://doi.org/10.1002/2015JD024511>
- Vremec, M., Collenteur, R.A., Birk, S. (2023) Technical note: Improved handling of potential evapotranspiration in hydrological studies with `PyEt`. *Hydrology and Earth System Sciences Discussions*, 1–23. <https://doi.org/10.5194/hess-2022-417>
- Zhang, X., Alexander, L., Hegerl, G.C., Jones, P., Tank, A.K., Peterson, T.C., Trewin, B., Zwiers, F.W. (2011) Indices for monitoring changes in extremes based on daily temperature and precipitation data. *WIREs Climate Change*, 2(6): 851–870.  
<https://doi.org/10.1002/wcc.147>



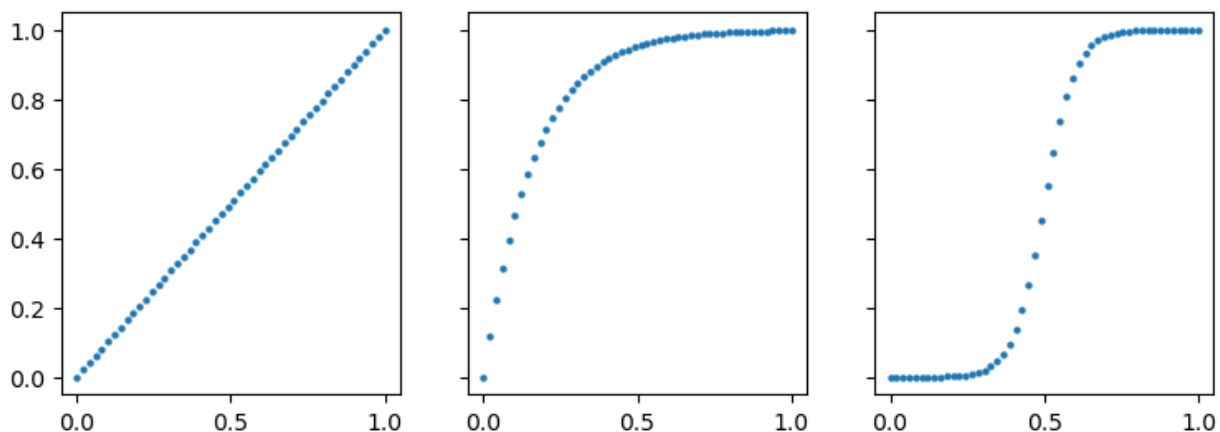
## Appendix A Details of Quantiles Mapping Configuration

We use the `xclim` Python package (Bourgault et al., 2023) for applying the QM algorithm (<https://xclim.readthedocs.io/en/stable/index.html>). `xclim` is a weather and climate focused library created by Ouranos (Quebec, Canada; <https://www.ouranos.ca/en>), written with the intention of manipulating and documenting NetCDF files in line with CF conventions. `xclim` is built on top of `Xarray` and `dask` libraries, utilising the extensive vectorised computation and parallelisation of those libraries. The following figures and discussion are for illustrative purposes only.

### Quantile settings

The number of quantiles and spacing between those quantiles should be chosen to achieve the best fit without over-fitting to the training data. A high number of quantiles will more precisely estimate the CDF for the calibration period but can also result in over-fitting. Likewise, some distributions have specific features that make linear spacing between quantiles less suitable, such as the discontinuity at zero for precipitation. We tested a range of configurations of quantile density and spacing for both temperature and precipitation variables to select the best configuration (see Figure A-1).

For temperature, a continuous Gaussian curve is most representative of the distribution. The dynamic range is not interrupted by any discontinuities.



**Figure A-1: Quantile spacing for linearly spaced nodes (left), log-spaced nodes (centre) and sigmoid spaced nodes (right).** Note that the unevenly spaced nodes were found to over fit the training data, therefore the linearly spaced nodes were preferred here.

We tested a range of quantiles, including log-spaced quantiles for precipitation and sigmoid quantiles for tasmax and tasmin (Figure A-1). We found that out-of-sample performance did not improve greatly above 50 quantiles (not shown). Also, we found variable spacings between the quantile could improve within-sample performance, but over-fitting leads to poorer performance out-of-sample. The tails of the distribution, where rare events occur, tend to be poorly represented by models, and can suffer over-fitting.

### Method for calculating adjustment factors - the `kind` argument

Consider  $y$  to be an array of quantiles of the reference dataset, and  $x$  to be an equivalent array of the simulated historical quantiles. The adjustment factors (AFs) are then calculated by:

- For multiplicative kind:  $AFs = y / x$
- For additive kind:  $AFs = y - x$

This returns an array containing the *AFs* of equal length to *x* and *y*. Any value that falls within the range of two quantiles are corrected with equivalent *AF*.

As an example, assuming the reference grid cell took values of  $x = (10, 6, 2)$  and the modelled grid cell took values of  $x = (20, 12, 4)$ , the adjustment factors would be  $AFs = (0.5, 0.5, 0.5)$  for the multiplicative kind, and  $AFs = (-10, -6, -2)$  for the additive kind.

Typically, the multiplicative kind is used for precipitation, and additive for temperature fields. Multiplicative kind imposes a natural limit on the transformation of the variable, as the *AFs* cannot correct a variable to below zero unless there are such negative values in the reference dataset.

It is possible to change precipitation into an additive variable by performing a log transform. This is advocated for by (Alavoine & Grenier, 2023) because it avoids "Type 1" physical inconsistencies, i.e., instances where a correction to a variable produces values beyond their dynamic range, e.g., negative precipitation.

We tested this method with varying numbers of quantiles to assess whether there is any improvement in the performance for precipitation QM. We find no real improvement in performance (not shown), and type 1 physical inconsistencies can also be avoided by calculating *AFs* multiplicatively.

### Accounting for seasonality

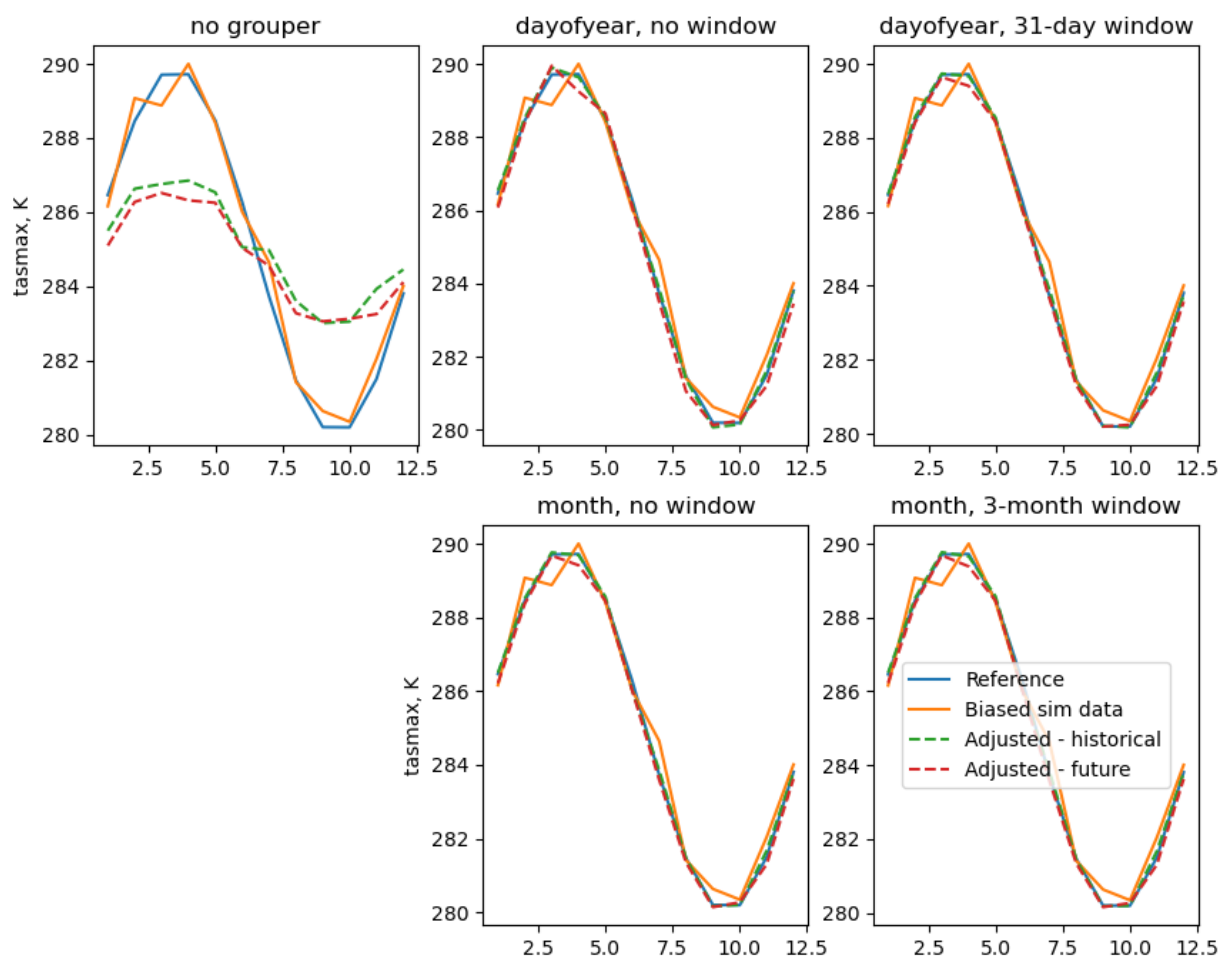
Temperature fields undergo a strong seasonal cycle (so too precipitation in some locations). To account for this, we can disaggregate the data into sub-samples depending on the time of year, to which we apply QM independently.

`xclim` provides several adaptable settings, including grouping by month or by day-of-year, and providing a window. The window option will include data from either side of the group of interest. For example, for the month of June, data from May and July will be included in the derivation of transfer functions.

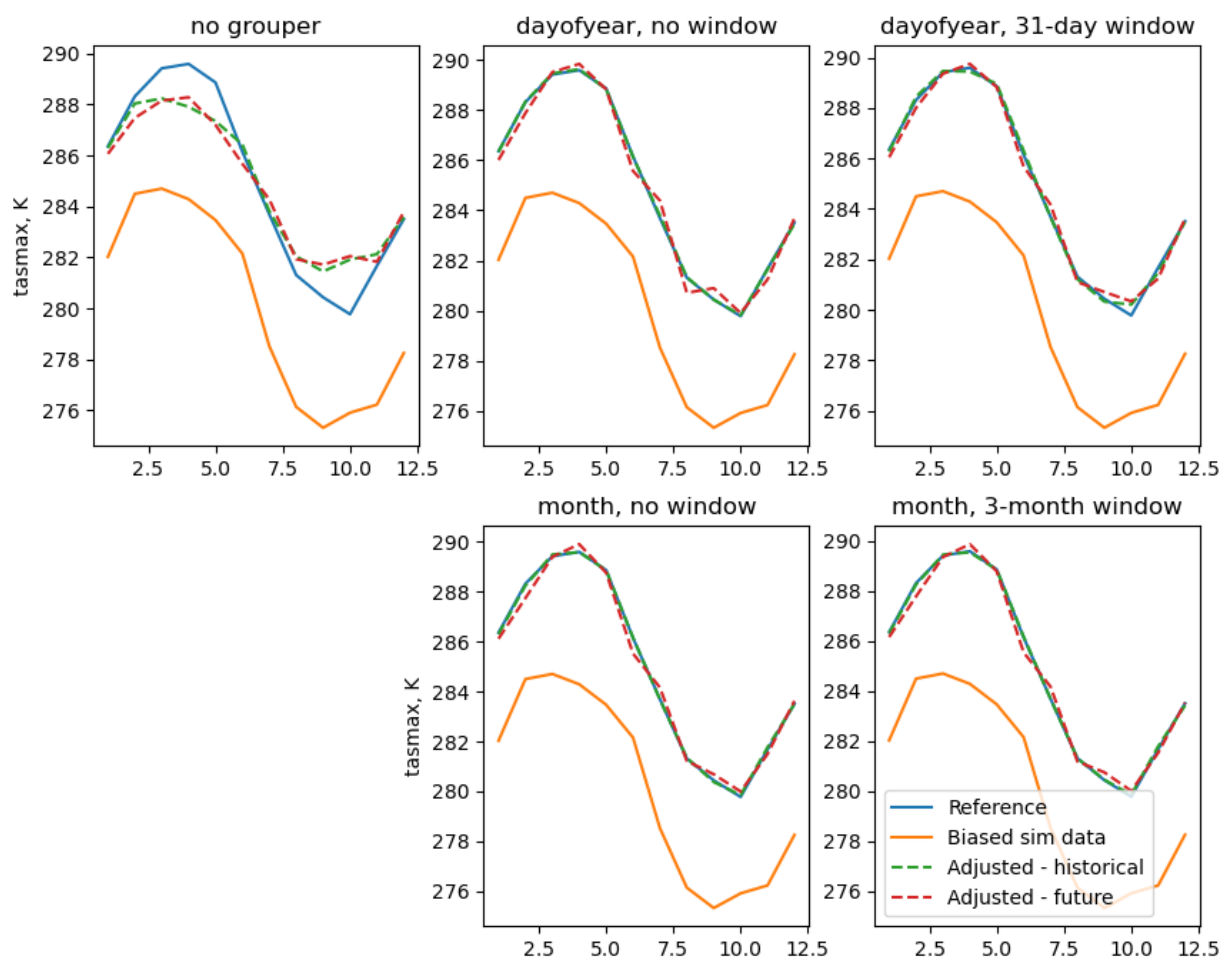
We tested five grouping methods:

- No grouping.
- Group by month with no window.
- Group by month with a 3-month window.
- Group by day-of-year with no window.
- Group by day-of-year with a 31-day window.

As can be seen in the monthly means in Figure A-2 and Figure A-3, without a grouper the seasonality is significantly smoothed out. Providing a grouper by month or day-of-year greatly improves performance.

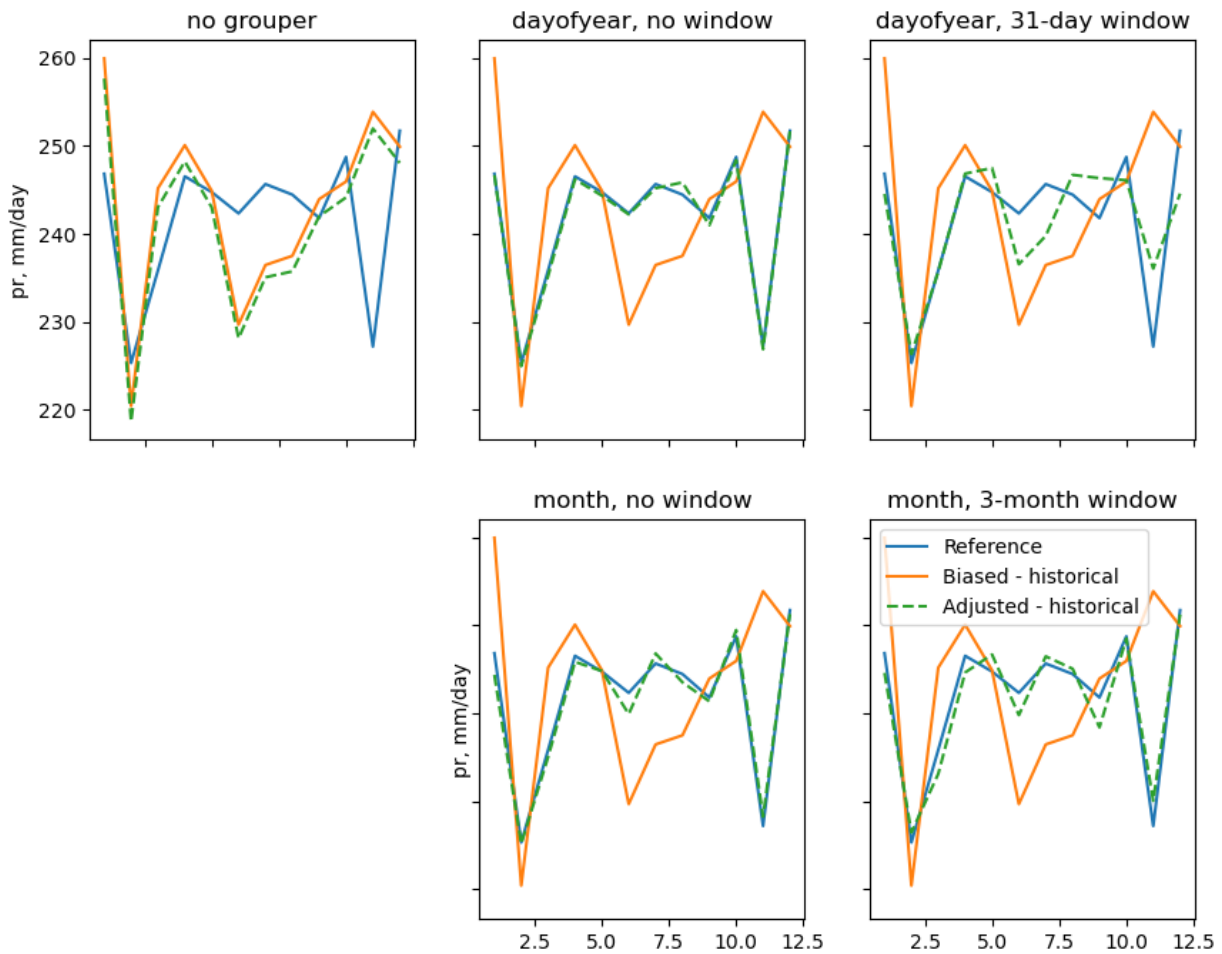


**Figure A-2:** Dummy data, reference and simulated data drawn from distribution with the same mean and high variance.



**Figure A-3: Dummy data, reference and simulated data drawn from distribution with the low mean and high variance.**

Because of the nature of the statistics that these figures show (long-term monthly means), they hide the discontinuities in *AFs* that can occur between months within a time series. Using a day-of-year grouper with a 31-day window decreases this effect, by applying a greater smoothing between days (e.g., Figure A-4). This limits the sample size from which to calculate *AFs*, however, the moving window provides additional data creating a suitably large sample size.



**Figure A-4: Monthly means for multiplicative precipitation dummy data (Gamma distribution) with groupers.**

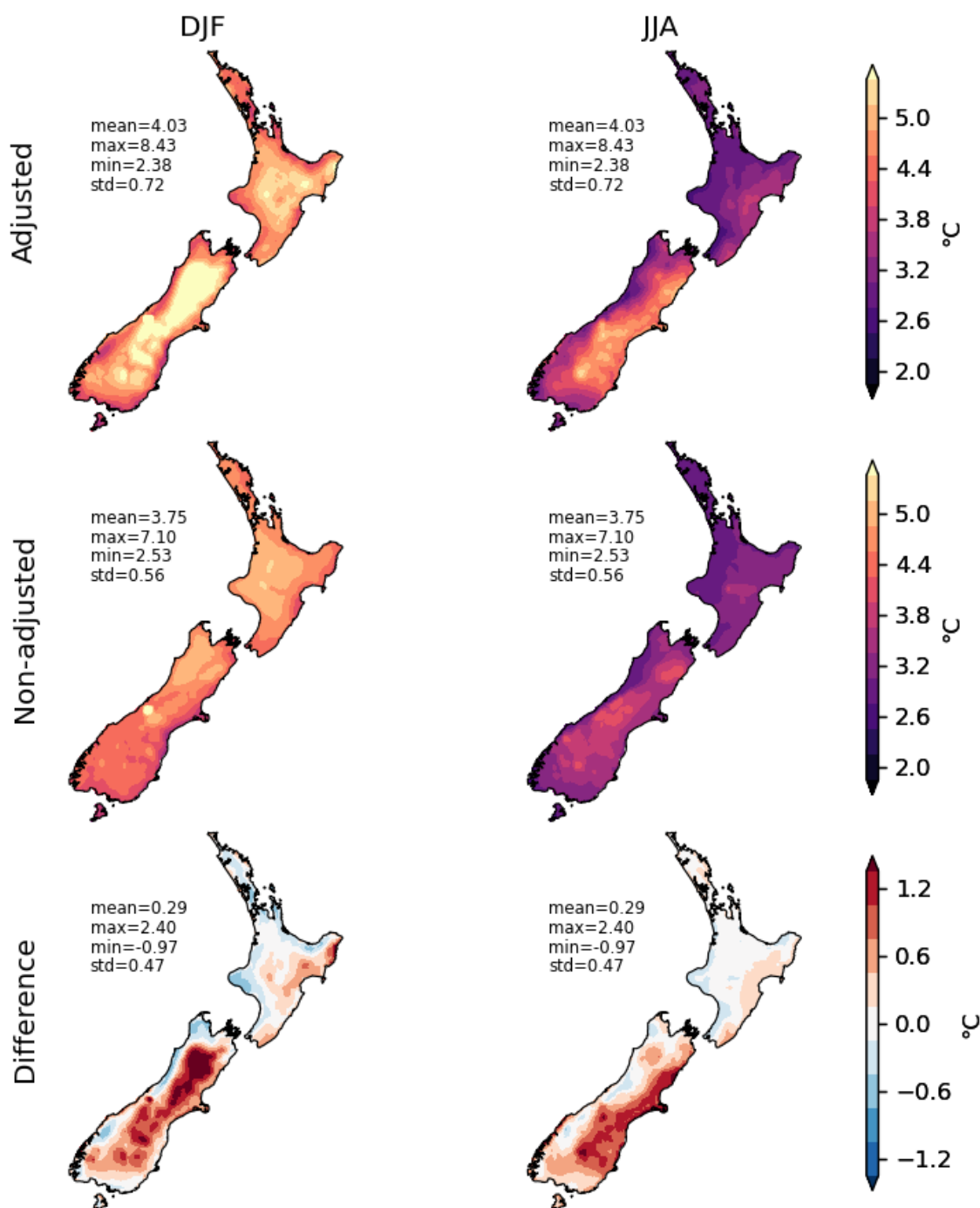
For any given value, the quantile to which it belongs and a subsequent interpolation to the corresponding AF is needed. Three different methods can be specified in the bias correction phase – nearest neighbour, linear and cubic interpolation – which progressively decreases the granularity of the interpolation, increasing the smoothness between interpolated points. For transfer functions that have been stratified into monthly or day-of-year groups, the interpolation method will be 2D. However, linear and cubic interpolation introduces missing values as out-of-sample data cannot be interpolated if they fall outside the calculated quantile range.

To balance these issues, we use a day-of-year grouper with a window of 31 days, and nearest-neighbour interpolation. This maintains seasonality, whilst reducing the discontinuities in AFs between days-of-year and avoid over-fitting to the data.

### Effect of non-stationarity (or the risk of *artificial inflation*)

Variables such as temperature have well-known trends over time. This complicates the bias correction of simulated futures because the distribution can be significantly different to that used in the calibration (for which we have observations). EQM can have significant issues by introducing an artefact known as ‘inflation’, as documented in e.g., (Alavoine & Grenier, 2023; Casanueva et al., 2020; Maraun, 2016).

# tasmax, ACCESS-CM2 Climate Change Signal (ssp370)

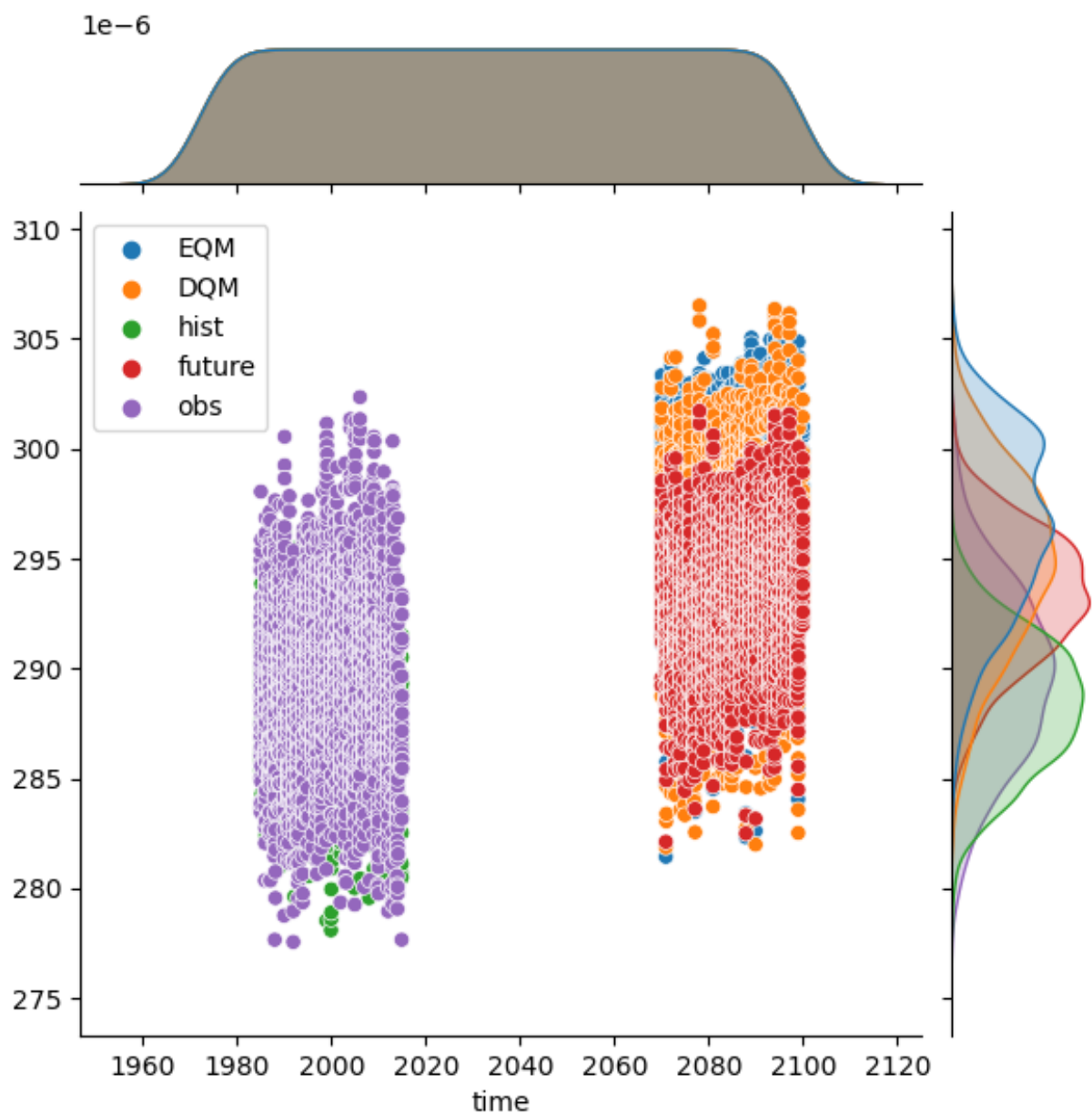


**Figure A-5: Summer (DJF) and winter (JJA) climate change signals of ACCESS-CM2-CCAM under the ssp370 scenario.** The difference between bias corrected and raw signals demonstrates the modification of trends by the EQM method. Note that the EQM method was not used here for this reason.

An example of inflation occurs over the central South Island, where the CCS is different between the bias corrected and raw data, particularly in summer (see Figure A-5). Inflation can occur when there is a mismatch in standard deviation between observations and the model training data. The transfer function will be trained to stretch or compress the model distribution to match the observations. However, when applying the transfer function to projections expressing a linear trend, this can lead to inflation (deflation) when the standard deviation of the training data is too small (large).

Specifically, only a portion of the transfer function is used for out-of-sample data since it is applied by mapping the *AFs* to the calculated historical quantiles. These quantiles are no longer representative of the projected distribution, since a linear trend results in a concerted shift. For example, for training data with a median (50<sup>th</sup> percentile) of 15°C, the shifted distribution now 15°C now corresponds with the 65<sup>th</sup> percentile in the future period. This means that

To demonstrate how this can occur using EQM, we use a single grid cell (Figure A-6; co-located with Mahanga EWS). Here, we see the EQM method stretches the PDF of the future period, exaggerating the double-peak structure found in the training (historical) data. Whereas, DQM preserves the CCS of the mean temperature, whilst still accounting for the lower standard deviation in the historical period. Therefore, the use of detrended quantile mapping (DQM) is more appropriate for variables with known linear trends.



**Figure A-6: Climate change signal over the Mahanga station under the ssp370 scenario, highlighting the inflation effect of EQM on trends, without explicit trend preservation.** Note that the EQM method as not used here for this reason.

## Extrapolation behaviour outside the historical quantiles

Any data points lying outside the historical quantiles is extrapolated with a constant factor corresponding to the nearest quantile (Boé et al. 2007). For example, if the upper quantile limit calculated from the training period is 500mm for precipitation, any value in the future period above this limit will be adjusted by the same factor, whether 700mm or 1000mm. The only alternative option provided by `xclim` is to replace values outside the quantiles with missing values, which is unsuitable for many applications.

Note that this behaviour is not necessary for QDM, where the quantiles are recalculated for the future period. By design, no data can lie outside these quantiles.