Uncertainties of Statistical Downscaling from Predictor Selection: Equifinality and Transferability

Submitted to: ***ATMOSPHERIC RESEARCH***

Guobin Fu[1], Stephen P Charles[1], Francis HS Chiew[2], Marie Ekström[2,3], Nick Potter[2]

1 CSIRO Land and Water, Private Bag 5, Wembley, Western Australia 6913

2 CSIRO Land and Water, GPO Box 1666, Canberra, Australian Capital Territory

3 School of Earth and Ocean Sciences, Cardiff University, Cardiff CF10 3AT

*Corresponding author:

Dr. Guobin Fu

Senior Research Scientist

CSIRO Land and Water, Private Bag 5, Wembley WA 6913, Australia

Phone: +61-8-9333 6342

Fax: +61-8-9333 6499

Email: Guobin.Fu@csiro.au

16 October 2017

Abstract:

The nonhomogeneous hidden Markov model (NHMM) statistical downscaling model, 38 catchments in southeast Australia and 19 general circulation models (GCMs) were used in this study to demonstrate statistical downscaling uncertainties caused by equifinality and transferability. That is to say, there could be multiple sets of predictors that give similar daily rainfall simulation results for both calibration and validation periods, but project different amounts (or even directions of change) of rainfall change in the future. Results indicated that two sets of predictors (Set 1 with predictors of sea level pressure north-south gradient, u-wind at 700hPa, v-wind at 700hPa, and specific humidity at 700hPa and Set 2 with predictors of sea level pressure north-south gradient, u-wind at 700hPa, v-wind at 700hPa, and dewpoint temperature depression at 850hPa) as inputs to the NHMM produced satisfactory results of seasonal rainfall in comparison with observations. For example, during the model calibration period, the relative errors across the 38 catchments ranged from 0.48 to 1.76% with a mean value of 1.09% for the predictor Set 1, and from 0.22 to 2.24% with a mean value of 1.16% for the predictor Set 2. However, the changes of future rainfall from NHMM projections based on 19 GCMs produced projections with a different sign for these two different sets of predictors: Set 1 predictors project an increase of future rainfall with magnitudes depending on future time periods and emission scenarios, but Set 2 predictors project a decline of future rainfall. Such divergent projections may present a significant challenge for applications of statistical downscaling as well as climate change impact studies, and could potentially imply caveats in many existing studies in the literature.

Key words: Statistical downscaling, Uncertainties, Equifinality, Transferability, Predictor selections, NHMM, GCMs

1. INTRODUCTION

General circulation models (GCMs) are the primary tools used to simulate present climate and project future climate, and their outputs are useful in understanding future global climatic changes for a given greenhouse gases emission scenario (IPCC, 2013). However, the present generation of global climate models are restricted in their usefulness for many sub-grid scale applications, such as hydrology and water resources, due to their coarse spatial resolution (Maraun et al., 2010; Zhang et al., 2016). Downscaling techniques have been developed to resolve the scale discrepancy between GCM climate change scenarios and the resolution required for climate impact assessments. Two approaches of downscaling are commonly used: *Dynamical downscaling* and *statistical downscaling* (Mehrotra and Shama 2006; Maraun et al., 2010).

Statistical downscaling is more widely used than the dynamical downscaling for climate change applications due to its pragmatic advantages (Fu et al., 2012). Compared to dynamical downscaling, statistical downscaling is simpler, uses minimal computing resources, can generate multiple stochastic realisations, and once developed can be relatively easily used to downscale from many GCMs to represent the full range of simulated uncertainty. In contrast, dynamical downscaling requires long computing run times, and because is not calibrated for rainfall characteristics it requires bias correction to produce rainfall that adequately matches observed rainfall characteristics (Corney et al., 2010). Ehret et al. (2012) even argued that bias correction technique is often used in an invalid way, i.e., it is added to the GCM/RCM model chain without sufficient proof that the consistency of the latter as well as the generality of its applicability increases.

The choice of predictor variable(s) is one of the most influential steps in the application and development of statistical downscaling schemes because this decision largely determines the characteristics of the downscaled scenario (Wilby et al. 2004). There are many studies in the literature on the selection of the best predictors. For example, Radanovics et al. (2013) proposed an extended version of the growing rectangular domain algorithm to provide an ensemble of near-optimum predictor domains for a statistical downscaling method. Sauter and Venema (2011) presented an approach for conditional airmass classification based on local precipitation rate distributions. It seeks, within the target region, three-dimensional atmospheric predictor domains with high impact on the local-scale phenomena, and concluded that predictor interactions played an important part in the modeling process and should be taken into account in the predictor screening. Phatak et al. (2011) described the use of a fast, sparse variable selection method, known as RaVE, for selecting atmospheric predictors, and illustrated its use on rainfall occurrence at stations in South Australia. Dayon et al. (2015) valued different combinations of predictors and found that a good temporal transferability is obtained only with a specific combination of predictors. Huth (2005) described a search for suitable predictors and predictands for downscaling of humidity variables. Hroton et al. (2017) have assessed multiple combinations of global optimization algorithms to select the best predictors and developed a new algorithms, i.e., the *chromosome of adaptive search radius.*

Despite the importance of predictor selection and the above-mentioned studies to explore the best predictors, there have been relatively few studies that specifically investigate the uncertainty due to choices of different downscaling predictors. For example, statistical downscaling of extreme precipitation for the Mediterranean region by Hertig et al (2014) found that “different predictor variables can lead to varying statistical downscaling results”. Lafaysse et al. (2014) have explored the uncertainties of future hydrometeorological projections and concluded that predictor selection is one of main uncertainties. Ben Daound et al. (2016) have reported that the downscaling model can be improved by introducing two additional predictors. Therefore, the choice of method/criteria often leads to the selection of different predictor variables. This could potentially result in an equifinality (Beven and Freer, 2001), as well as transferability problems for future rain projections (Dayon et al., 2015). That is to say, there are multiple sets of predictor variables and relationships that give similar overall simulation results when calibrating the statistical downscaling model. However, the different calibration parameter sets can give different projections of future rainfall due to both parameter equifinality and the temporal transferability of a statistical downscaling method in a climate change context. Regarding different spatial domains of predictors, it has been demonstrated by Radanovics et al. (2014).

Therefore, the objectives of this study are to demonstrate the equifinality and transferability for statistical downscaling by using the nonhomogeneous hidden Markov model (NHMM) stochastic daily downscaling model for 38 catchments in southeast Australia with two different sets of predictors, and to discuss a few guidelines as potential general principles to select more suitable predictors.

2. METHODS AND DATA

2.1. Equifinality and transferability

Equifinality, which was originally defined by Hans Driesch, a developmental biologist, is the principle that a given end state in open systems can be reached by many potential means (Cummings and Worley, 2005). It suggests that similar results can be achieved with different initial conditions and in many different ways. This term was introduced to hydrology by Beven (1993) in that an acceptable hydrological model can be achieved in many different ways, i.e., different model structure or parameter sets. Here we show that this also applies to statistical downscaling, i.e., a statistical downscaling model can be calibrated and validated well with different parameter sets.

This equifinality issue may present a challenge for future climate change projections, i.e., the temporal transferability of a statistical downscaling method in a climate change context (Dayon et al., 2015, Chaedon et al., 2014). That is to say, the current predictor-predictand relationship may not be transferred in future climate change conditions, and accordingly different calibration parameter sets could give different projections of future rainfall. This is defined as “Transferability” in this study. The two related issues, i.e., equifinality and transferability, are investigated in this study as they could be main sources of the uncertainty in statistical downscaling.

2.2. Study region

The study region comprises 38 catchments of the southern Murray-Darling Basin (MDB) in south-eastern Australia, i.e., Loddon, Avoca, Campaspe, and Goulburn rivers (Figure 1). This study region has been selected as: 1) The MDB is Australia’s most important agricultural region producing 38% of total Australian agricultural commodities, worth approximately $14 billion per year (Cai and Cowan, 2008) , thus future climate impact on the productivity of this region is of high national interest; 2) The MDB has experienced a decadal-long drought with unprecedented decline in the streamflow (Potter et al., 2010; Potter and Chiew, 2011), with a more significant impact of temperature change on streamflow than previously reported (Yu et al 2011); and 3) the nonhomogeneous hidden Markov model (NHMM) stochastic daily downscaling model has previously been shown to perform well in the study catchments in terms of various rainfall statistics (Chiew et al., 2010, Frost et al., 2011, Fu et al., 2013a;b). Note our study region is a very small part of the MDB, and consequently it is homogeneous in terms of rainfall processes.

2.3. Statistical Downscaling Model -- NHMM

The NHMM simulates multi-site patterns of daily rainfall occurrence and amounts conditional on a finite number of ‘hidden’ (i.e. unobserved) weather states (Hughes et al., 1999). The temporal evolution of these daily states is modelled as a first-order Markov process with state-to-state transition probabilities conditioned on a small number of synoptic-scale atmospheric predictors, such as mean sea-level pressure, geopotential heights, and measures of atmospheric moisture. When previously applied to these catchments, the NHMM showed good skill in reproducing characteristics of the observed daily rainfall distributions (Fu et al., 2013a). A detailed description of the current-generation NHMM, including its assumptions, mathematical parameterizations and estimation algorithms can be found in Hughes et al. (1999) and Kirshner (2005). Previously, the NHMM has been applied to many regions in the world, including Australia (Charles *et al*., 1999, Chiew et al., 2010, Frost et al., 2011, Fu et al., 2013a;b), Brazil (Robertson et al., 2004), China (Liu et al., 2011, 2013), India (Greene et al., 2011), and USA (Robertson et al., 2007).

In this study, data (rainfall and predictors) from 1981 to 2000 was used to calibrate the NHMM and the data from 1961 to 1980 is used to verify the model. Data over the more recent period (1981–2000) were chosen to calibrate the model because they have relatively higher quality than earlier periods.

This study focusses only on the April-October season, because it is the main rainfall season in the study region. Furthermore, rainfall statistical downscaling predictors usually vary from season to season, so annual rainfall is usually simulated by combining seasonal models.

2.4. General circulation models (GCMs)

The atmospheric predictors from 19 Coupled Model Intercomparison Project Phase 5 (CMIP5) GCMs (Table 1) were used in this study to drive the NHMM to project future rainfall. The CMIP5 GCM simulations were forced by plausible scenarios of greenhouse gas emissions and aerosols throughout the 21st century, referred to as Representative Concentration Pathways (RCPs) (Van Vuuren et al., 2011). This study used projections for the RCP4.5 scenario, an intermediate-emissions scenario, and for RCP8.5, a high-emissions scenario. The RCP numbers refer to the approximate enhanced radiative forcing levels by 2100, i.e. an additional

4.5 W/m^2 and 8.5 W/m^2 of radiative forcing at 2100, respectively, corresponding to equivalent CO_2 levels of ~650 ppm and ~1370 ppm by 2100. Detailed information for these models and RCP emission scenarios can be found at the website of the Program for Climate Model Diagnosis and Intercomparison (PCMDI) (http://www-pcmdi.llnl.gov/).

The CMIP5 GCM historical runs span 1961 to 2005 (45 years) and future climate projections 2006 to 2100 (95 years). To make an equal time length comparison, three different overlapping periods of 45 years, i.e., 2006–2050, 2031–2075 and 2056–2100, are used here to compare the future rainfall change with the historical period. The future projection rainfall is compared with each GCM's historical period (1961–2005) downscaled rainfall, not to observed rainfall. That is to say, the calibrated NHMM is applied to each GCM's historical and future periods, respectively. The difference between modelled rainfalls for the different time periods are then explored.

2.5. Predictor selections

The selections of NHMM predictors involved the following steps:

a) The gridded daily atmospheric variables (potential predictors) were extracted from the NCEP/NCAR Reanalysis 1 (NNR) (Kalnay et al., 1996) archive for the region shown in Figure 1. The size of domain is chosen to capture the relevant scale of the atmospheric predictor variables (10×8 = 80 2.5×2.5 degree longitude-latitude grids). The variables include mean sea-level pressure (MSLP), and for the 850, 700 and 500 hPa levels: temperature, specific humidity, U-wind (eastward wind speed component) and V-wind (northward wind speed component). This step produces 13 variables × 80 grids = 1040 potential predictors. The NCEP-NCAR reanalysis data (Kalnay et al., 1996) were chosen as it has been widely used in climate community. However, there are other reanalyses existing, which may perform better at some regional scale. A full comparison of these reanalysis products is beyond the scope of the current study. However, our recent study (Fu et al., 2016) shows that NCEP-NCAR and ERA-Interim produce similar results for potential predictors for the study region.

b) The dewpoint temperature depressions (DTD) at 850, 700 and 500 hPa levels were calculated from the temperature and specific humidity variables. This step produces 3 variables × 80 grids = 240 additional potential predictors.

c) For MSLP, each north-south, east-west, northeast-southwest, and northwest-southeast difference for adjacent grid-points were also calculated, because sea level pressure gradients typically have a high correlation with rainfall (Charles et al., 1999; Frost et al., 2011; Liu et al., 2011; Fu et al., 2013a,b; Liu et al., 2013). This step produced 70 (north-south) + 72 (east-west) + 2×63 (northeast-southwest and northwest-southeast) = 268 additional potential predictors.

d) The correlations between daily time-series of each variable (1548) and daily rainfall series for each catchment were calculated and then averaged over the 38 study catchments. This identified the sub-region for which each variable had the highest correlation with rainfall for the study catchment network. The candidate predictors were selected for the sub-regions of their highest correlation with rainfall. The independence of predictors are also considered in this step, because if the most highly correlated predictors are correlated, then they may be adding less information than other predictors with relatively lower correlations. We also define variable domains by merging several gird cells if they all have a high correlation with rainfall (Figure 2). This step reduced the potential predictors from 1548 to 14 (Table 2), including sea level pressure, east-west sea level pressure gradient (both predictors 2 and 3 as they are different grids), north-south sea level pressure gradient, u-wind at 700 and 850hPa, v-wind at 700 and 850 hPa, specific humidity at 500, 700 and 850hPa, and dew-point temperature depression at 500, 700 and 850hPa. Figure 2 shows the sub-regions of these 14 predictors.

e) The NHMM model was run with all possible four predictor combinations of these 14 candidate predictors. The calibrated NHMMs were assessed using a few criteria: 1) Bayes Information Criterion (BIC), as a measure of parameter parsimony, for the calibration period, and for both the calibration and validation periods; 2) mean bias of simulated rainfall calculated for each season and catchment and then averaged; 3) interannual correlation calculated for each catchment and then averaged; 4) a constraint that at least one of the four is a sea-level pressure predictor (i.e. predictor 1, 2, 3, or 4) and one is a moisture (humidity or DTD) predictor. This constraint is based on previous experience identifying the importance of surface circulation and atmospheric moisture for providing realistic projections (Charles et al., 1999). In addition, the pressure field and humidity predictors are used for almost all statistical downscaling methods, as well as for prefect model framework (Dixon et al., 2016).

f) There were a number of combinations of four predictors capable of reproducing rainfall characteristics that are close to the observed statistics. In this study, we simply pick two sets of predictors as a case study with the intention of demonstrating the equifinality issue and the implication on uncertainties of statistical downscaling influenced by the choice of predictors. Set 1 with predictors of sea level pressure north-south gradient (4), u-wind at 700hPa (5), v-wind at 700hPa(7), and specific humidity at 700hPa(10) and Set 2 with predictors of sea level pressure north-south gradient(4), u-wind at 700hPa(5), v-wind at 700hPa(7), and dewpoint temperature depression at 850hPa(14). The only difference is the inclusion of specific humidity at 700hPa (predictor 10) versus DTD at 850hPa (predictor 14). It needs pointing out that a full investigation of all relatively independent possibilities would be required to fully explore the relationship between predictor choice and equifinality/transferability but is out of scope of this current study.

g) The predicators from GCMs are standardized with each GCM's own historical period (1961–2005 in this case) to remove the biases of climatological performance at the present-day period.

3. RESULTS

3.1. Equifinality from NHMM calibration and validation results

The NHMM model calibration (1981–2000) and validation (1961–1980) results of winter-half (April-October) rainfall across 38 catchments, with two different predictor sets, are shown in Figure 3. Overall, NHMM produced pretty good results of seasonal rainfall in comparison with observations for both predictor sets for calibration period (Figure 3, first row): the relative errors across the 38 catchments range from 0.48 to 1.76% with a mean value of 1.09% for the predictor Set 1, and from 0.22 to 2.24% with a mean value of 1.16% for the predictor Set 2. The correlation coefficients between simulated and observed seasonal rainfall across 38 catchments for both sets of predictor are larger than 0.9999 (Figure 3).

There is slightly less agreement between observed and simulated seasonal rainfall during the validation period in comparison with calibrated period (Figure 3, second row), but they are still satisfactory results. For example, the relative errors across the 38 catchments range from -9.5 to 11.0% with a mean value of -1.74% for the predictor Set 1, and from -9.7 to 10.2% with a mean value of -2.23% for the predictor Set 2. The correlation coefficients between simulated and observed seasonal rainfall across 38 catchments for both sets of predictor are 0.991 (Figure 3).

In summary, both sets of predictors produced adequate seasonal rainfall, and their differences are too small to identify a better predictor set, because either can be recognized as the best depending on which model/criterion were used. It confirms that equifinality is indeed a challenging issue to statistical downscaling that may produce one of the main uncertainties of statistical downscaling.

It needs to point out that mean seasonal rainfall was used in this study to demonstrate the existence of equifinality and transferability issues because of its importance for climate change impact studies, as well as its being one of criteria to select the best models (Section 2.5e for details). In addition, our previous work (e.g., Fu et al., 2013a; b) showed that the best identified NHMM model usually reproduces a wide range of rainfall statistics, such as monthly distribution, winter/summer seasonal rainfall, daily maximum rainfall, 99th and 95th percentiles of daily rainfall, 99th percentiles of 3-day rainfall, numbers of rain days, maximum consecutive wet/dry days, etc.. Whether the two models are similarly able to simulate other rainfall statistics does not affect the conclusion that there is an equifinality issue inherent in statistical downscaling.

3.2. Transferability for future rainfall projections resulting from equifinality of predictors

Changes of future rainfall from the NHMM projections based on 19 GCMs are in different directions for the two different sets of predictors (Figure 4): one set of predictors projects an increase of future seasonal rainfall and another set a decline of future seasonal rainfall. It results from the transferability of calibrated models into a climate change context. The variations of the boxplot mainly comes from different GCMs, because the 38-catchment-averages of mean values of 50 NHMM simulations of each GCM was used to minimize the model uncertainty due to stochastic characteristics of statistical downscaling, and the averaging values across 38 catchment were used to minimize the spatial distribution and some deviations with modelling errors as shown in Figure 3 validation period.

Both projections seem rational: the magnitudes of rainfall changes become larger with time, i.e., larger magnitude of changes by the end of 21st century; RCP8.5 emission scenario usually result in a larger rainfall change than RCP4.5; The differences between RCP4.5 and RCP8.5 are very small for the near future (2006–2050), but significant by the end of 21st century (2056–

2100). This implies that the difference does not come from climate variability, rather it is coming from the trends of the predictors.

The predictor changes in the future (Figure 5) correspond to the future rainfall changes. Changes do not seem unreasonable from a physical processes perspective: rainfall generally is proportional to humidity, because more moisture implies a high probability of rainfall; but rainfall is inversely proportional to DTD, the difference between the temperature and dew-point temperature at a certain height in the atmosphere. The larger the DTD value (wider separation between air temperature and dew-point), the lower the relative humidity, and accordingly lower probability of rainfall.

It needs point out that it is the changes of 45-year mean of seasonal rainfall, and the interannual variability is not explored here. In reality, the rainfall variation would be much larger because climate variability overlaps with climate change signals.

4. DISCUSSION

This study provides evidence for the existence of equifinality and non-transferability issues in statistical downscaling. These may be one of the main uncertainties in statistical downscaling and could pose challenges for the validity of statistical downscaling as used in climate change impact studies. While there is not a simple method and/or criterion to determine which predictor set will provide reliable rainfall projections for a future climate, a few guidelines, but are not limited to, are discussed as below:

1) Regional rainfall physical processes should be considered. This is probably the most important guideline to select predictors, because the causality of the relationship between rainfall and the predictor variable and whether this relationship is still meaningful in the future projections are fundamental principles for statistical downscaling. Sometimes the best predictors identified with statistical criteria are not the best variables to associate with regional rainfall processes, and not completely adequate for climate change projection applications

(Wilby et al., 2004). For example, daily rainfall may be determined by geopotential heights in the extratropical areas, but changes of geopotential heights from GCMs in a warming climate might contain a non-dynamic signal, which will spuriously affect the estimation of rainfall changes (Wilby et al., 2004). This non-dynamic component should be corrected, as per suggestion of Burkhardt (1999), either by subtracting the mean changes of the geopotential heights in a sufficiently large area, or by using geopotential thickness, instead of geopotential heights, as predictors.

In our case study, both specific humidity and DTD represent moisture variables, which make sense as atmospheric moisture is an important factor influencing rainfall processes. However, DTD seems more suitable than specific humidity in this case, because it is a function of moisture and air temperature, given global and regional air temperature will increase in the future climate change scenarios. It is consistent with the conclusions of Timbal et al. (2008a) that “a relative humidity moisture predictor, rather than specific humidity, was needed for downscaled projections to be consistent with direct model output projections”. In fact, various choices of humidity predictors have been used in many studies in the literature to downscale precipitation, such as relative humidity and total column water (Radanovics et al., 2013, Chardon et al., 2014, Caillouet et al., 2015), soil moisture flux and specific humidity (Dayon et al., 2015), and soil moisture flux and relative humidity (Lafaysse et al., 2014).

Generally speaking, variables representing the absolute atmospheric moisture, e.g., specific humidity and total precipitable water do not account for the whole ‘climate signal’, including climate changes and variability (Mehrotra and Sharma, 2011). This is because, in a warming climate, the moisture-holding capacity of the atmosphere increases and as such additional information about atmospheric temperature is also required to know the amount of atmospheric moisture that can precipitate (Mehrotra and Sharma, 2011). Thus DTD, computed as a function of both atmospheric moisture and air temperature, appears to be a better choice.

2) Comparative studies could also be used to quantify the uncertainties and possible ranges of climate change projections, and accordingly to seek reliable predictors to make similar projections of future rainfall. A growing number of studies in the literature have compared different statistical downscaling methods or compared statistical downscaling with dynamical downscaling methods. For example, Wilby and Wigley (1997) and Wilby et al. (1998) compared six statistical downscaling approaches for multiple sites across the USA using observed and GCM data. Frost et al. (2011) have examined six downscaling methods (one scaling methods, four statistical downscaling and one dynamical downscaling) to simulate multi-site daily rainfall for 30 rain gauges in south-eastern Australia. Sunyer et al (2012) have compared five statistical downscaling methods based on a common change factor methodology from four different RCMs results, but reported a significant uncertainty in the downscaled projected changes of the mean, standard deviation, skewness and probability of dry days.

In our case study, the statistical downscaling results of NHMM with predictor Set 2 have a similar rainfall changing magnitude and direction with empirical scaling, statistical downscaling modelling (SDM) of Australian Bureau of Meteorology (Timbal et al., 2008b), and two dynamical downscaling methods (WRF and CCAM) for the June-July-August and September-October-November seasons (Potter et al., 2018). However, for the March-April-May season, the future rainfall projections from dynamical downscaling have different changing magnitudes and directions with statistical downscaling and empirical downscaling methods (Potter et al., 2018). This difference, not uncommonly reported in the literature, can drive users to further explore the reason in terms of plausible climate change signal. For example, Grose et al (2015) attributed the different climate projections for the southern Australia cool season rainfall from a statistical and dynamical downscaling comparison to a plausible relationship with topography and regional drivers that are not resolved by coarse global models.

However, a critical question for this kind of comparison study is that obtaining the same future rainfall changing directions does not guarantee it is “correct”.

3) Historical observed rainfall trend in a warming climate might imply the future change directions. In the last 50–60 years, almost all regions in the world have experienced a climate warming trend, and the associated rainfall changes in this warming climate in the past could be used as a reference for future rainfall changes. However, it is still a scientific debate whether the recent observed trend will necessarily continue into the future, especially for regions where there is little evidence linking observed rainfall trends to a warming climate.

4) The skill with which a climate variable (predictor) is simulated for the future, as measured by how consistent its simulations are across a range of GCMs, could also be used as a criteria to select better predictors for statistical downscaling. For example, Mehrotra and Sharma (2011) have used a variable convergence score method to evaluate the consistency associated with the prediction of 25 variables relevant for 19 GCMs. They found that mean sea-level pressure has the highest skill followed by the air temperature, geopotential heights and equivalent potential temperature. For the two different variables used in this case study, i.e., predictor 14 (DTD at 850 hPa) and predictor 10 (specific humidity at 700hPa), the former shows a slightly higher score than latter from the results of Mehrotra and Sharma (2011).

However, humidity variables are generally projected by GCMs to have larger changes in the future warming world. The future mean values (45 years) of 14 predictors used in this study across 19 GCMs are shown in Figure 6. Since it has been standardized by its own GCM historical runs (1961–2005), values of these predictors represent their changes in the future. The box plot variations come from 19 GCMs. It clearly demonstrates that humidity variables/predictors have a larger magnitude in the future than non-humidity ones. The changing magnitudes also vary with emission scenarios as well as time period, which is consistent with existing conclusions in the literature.

A critical question is whether we can only use non-humidity predictors because future changes of them have been observed in the historical periods and they have been simulated consistently across a range of GCMs, or we should combine both non-humidity and humidity predictors to have a better fitting model and physical explanation. In this case study, if only predictors 4, 5 and 7 are used, then the model errors are about 16–20% for seasonal rainfall for both calibration and validation periods, and accordingly project a slight decrease of future rainfall (Figure 7) — magnitudes being smaller than that with predictors 4, 5, 7 and 14. If we use only sea level pressure predictors (1, 2, 3 and 4, Table 2), given the surface circulation is probably the most important single predictor for rainfall physical processes, the model error is about 12–16% for seasonal rainfall for both calibration and validation periods (not shown).

5) Ensemble multi-models with different predictor sets can be used to quantify the ranges of future climate change projections. In most statistical downscaling studies, great efforts have been carried out in order to select predictors. However, this approach will accept all reasonable models (with different predictors), then quantify the range of future climate change projections and investigate how often a specific predictor is selected in the downscaling models in terms of absolute numbers and in relation to the total number of models (Hertig et al., 2014). Wu et al (2012) constructed sixty ensemble members for probabilistic estimates, instead of a single deterministic forecast, to statistically downscale climate forecast system for the Southeastern Mediterranean.

6) A perfect model framework can be used to isolate the uncertainties associated with the non-transferability stationarity assumption that are inherent to the future climate projections via statistical downscaling models (Dixon et al., 2016). Note that the name is not meant to imply that the model itself is perfectly free of errors. Rather, it is a name given to a model experiment approach in which model data is used as a substitute of observations or truth (Dixon et al., 2016). For example, outputs from a RCM is taken as predictands, both in the calibration

period and the future simulation period (Vrac et al., 2007, Beuchat et al., 2012, Dixon et al., 2016). Furthermore, using an ensemble of RCMs may help to take account of different RCM behaviors in regional changes (Dayon et al., 2015).

5. SUMMARY

Statistical downscaling is a useful tool and has been widely used for climate change impact studies. However, critical challenges and uncertainties remain, and selecting the most suitable and appropriate predictors is one of them, because it could potentially result in the equifinality and transferability issues highlighted here, i.e., there could be multiple sets of parameter values that give similar daily rainfall simulation results for both calibration and validation periods, but project different future rainfall change amount and directions in the global warming scenarios. For example, in this case study, two sets of predictors of NHMM model produce a pretty good of calibration and validation of seasonal rainfall, but project a different sign – one increasing and one decreasing – of future rainfall. Such divergent projections present a significant challenge for applications of statistical downscaling and could potentially imply caveats for many existing studies on climate change impacts studies using statistical downscaling.

There is not a simple method and/or criterion to determine which predictor set will provide a reliable rainfall projection for a future climate, because the future precipitation is unknown and we cannot verify or test the downscaled results. A few guidelines are discussed in this study as general principles, with the most important of them regional rainfall physical processes, because the causality of the relationship between rainfall and the predictor variable and whether this is still meaningful in the future projections are keystones of statistical downscaling.

In this case study, predictor set with DTD seems more reasonable than predictors with specific humidity, but it may still not the best predictor set for regional statistical downscaling as we simply selected two sets of predictors to demonstrate the equifinality and transferability issues of statistical downscaling. It needs further investigations, such as accepting all reasonable models with different predictors to quantify the range of future climate change projections (Hertig et al., 2014), or using a prefect model framework to isolate the uncertainties associated with the transferability stationarity assumption that are inherent to the future climate projections from statistical downscaling models (Dixon et al., 2016).

Acknowledgements

This research is partly funded by Victorian Climate Imitative (VicCI) project. We would like to thank Dr. Santosh Aryal and Dr Hongxing Zheng (CSIRO internal review), and two anonymous reviewers for their invaluable comments and constructive suggestions used to improve the quality of the manuscript.

Reference

Ben Daoud, A., Sauquet, E., Bontron, G., Obled, C., Lang, M., 2016. Daily quantitative precipitation forecasts based on the analogue method: Improvements and application to a French large river basin. Atmospheric Research, 169, Part A, 147-159. doi: 10.1016/j.atmosres.2015.09.015

Beuchat, X., Schaefli, B., Soutter, M. & Mermoud, A., 2012. A robust framework for probabilistic precipitations downscaling from an ensemble of climate predictions applied to Switzerland. Journal of Geophysical Research, 2012, 117, D03115. doi: 10.1029/2011JD016449

Beven, K., 1993. Prophecy, reality and uncertainty in distributed hydrological modelling. Advances in Water Resources, 16(1), 41-51. doi: 10.1016/0309-1708(93)90028-E

Beven, K.J. and Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems, Journal of Hydrology, 249, 11–29

Burkhardt, U. 1999. Alpine precipitation in a tripled CO2-climate. Tellus, A 51, 289-303.

Cai, W. and Cowan, T., 2008. Evidence of impacts from rising temperature on inflows to the Murray-Darling Basin. Geophys. Res. Lett., 35, L07701, doi:10.1029/2008GL033390.

Caillouet, L., Vidal, J.-P., Sauquet, E. & Graff, B., 2015. Probabilistic precipitation and temperature downscaling of the Twentieth Century Reanalysis over France. Climate of the Past, 12(3), 635-662. doi: 10.5194/cp-12-635-2016

Chardon, J., Hingray, B., Favre, A.-C., Autin, P., Gailhard, J., Zin, I. & Obled, C., 2014. Spatial similarity and transferability of analog dates for precipitation downscaling over France. Journal of Climate, 27(13), 5056-5074. doi: 10.1175/JCLI-D-13-00464.1

Charles SP, Bates BC, Whetton PH and Hughes JP (1999) Validation of downscaling models for changed climate conditions: case study of southwestern Australia. Climate Research 12, 1-14.

Chiew, F.H.S., Kirono, D.G.C., Kent, D.M., Frost, A.J., Charles, S.P., Timbal, B., Nguyen, K.C., Fu, G., 2010. Comparison of runoff modelled using rainfall from different downscaling methods for historical and future climates. Journal of Hydrology 387 (1–2), 10–23

Corney, S.P., Katzfey, J.J., McGregor, J.L., Grose, M.R., Bennett, J.C., White, C.J,, Holz, G.K., Gaynor, S.M. Bindoff, N.L. (2010), Climate Futures for Tasmania: climate modelling technical report, Antarctic Climate & Ecosystems Cooperative Research Centre, Hobart, Tasmania. Available online: <http://www.dpac.tas.gov.au/divisions/climatechange/adapting/climate_futures/climate_futures_for_tasmania_reports>.

Cummings and Worley, Organization Development and Change, Thomson, 2005, p. 87

Dayon G, Boé J, Martin E. 2015. Transferability in the future climate of a statistical downscaling method for precipitation in France. J. Geophys. Res. Atmos. 120*:* 1023–1043, *doi:* 10.1002/2014JD022236.

Dixon, K.W., Lanzante, J.R., Nath, M.J., Hayhoe, K., Stoner, A., Radhakrishnan, A., Balaji, V., Gaitán, C.F., 2016. Evaluating the stationarity assumption in statistically downscaled climate projections: is past performance an indicator of future results? Climatic Change, 135, 395–408.

Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K. & Liebert, J., 2012. HESS Opinions "Should we apply bias correction to global and regional climate model data?" Hydrology and Earth System Sciences, 16(9), 3391-3404. doi: 10.5194/hess-16-3391-2012

Frost, A. J., Charles, S. P., Timbal, B., Chiew, F. H. S., Mehrotra, R., Nguyen, K. C., Chandler, R. E., McGregor, J. L., Fu, G., Kirono, D. G. C., Fernandez, E. & Kent, D. M., 2011. A comparison of multi-site daily rainfall downscaling techniques under Australian conditions. Journal of Hydrology 408, 1-18

Fu, Guobin; Charles, Stephen P; Chiew, Francis; Teng, Jin and Zheng, Hongxing. Statistical downscaling for hydrological application: Advantages and challenges [online]. In: Hydrology and Water Resources Symposium 2012. Barton, ACT: Engineers Australia, 2012: 571-578. Availability: <http://search.informit.com.au/documentSummary;dn=917385523066423;res=IELENG> ISBN: 9781922107626.

Fu GB, SP Charles, FHS Chiew, J Teng, H Zheng, AJ Frost, W Liu, S Kirshner, 2013a, Modelling runoff with statistically downscaled daily site, gridded and catchment rainfall series, Journal of Hydrology, 492, 254–265, DOI: 10.1016/j.jhydrol.2013.03.041.

Fu GB, SP Charles, S Kirshner, 2013b, Daily rainfall projections from general circulation models with a downscaling nonhomogeneous hidden Markov model (NHMM) for south-eastern Australia, Hydrological Processes, 27, 3663–3673, doi: 10.1002/hyp.9483.

Fu GB, SP Charles, B Timbal, B Jovanovic, F Ouyang, 2016. Comparison of NCEP/NCAR and ERA-Interim over Australia. Int. J. Climatol. 36*:* 2345–2367

Greene, A. M., Robertson, A. W., Smyth, P. and Triglia, S., 2011. Downscaling projections of Indian monsoon rainfall using a non-homogeneous hidden Markov model. Q.J.R. Meteorol. Soc., 137: 347–359. doi:10.1002/qj.788

Grose MR, Moise AF, Timbal B, Katzfey JJ, Ekström M, Whetton PH (2015) Climate projections for southern Australian cool-season rainfall: insights from a downscaling comparison. Clim Res 62:251-265

Hertig E, Seubert S, Paxian A, Vogt G, Paeth H, Jacobeit J (2014) Statistical modelling of extreme precipitation indices for the Mediterranean area under future climate change. Int J Climatol 34, 1132-1156

Horton, P., Jaboyedoff, M. & Obled, C., 2017. Global optimization of an analog method by means of genetic algorithms. Monthly Weather Review, 145(4), 1275-1294. doi: 10.1175/MWR-D-16-0093.1

Hughes, J. P., Guttorp, P., Charles, S. P., 1999. A non-homogeneous hidden Markov model for precipitation occurrence. J. Royal Stat. Soc., Series C–Applied 48(1), 15–30.

Huth R. 2005. Downscaling of humidity variables: a search for suitable predictors and predictands. International Journal of Climatology 25*:* 243–250.

IPCC, 2013: 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, 1535 pp.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K.C., Ropelewski, C., Wang, J., Jenne, R., Joseph, D., 1996. The NCEP/NCAR 40-year reanalysis project. Bulletin of the American Meteorological Society 77 (3), 437–471.

Kirshner, S. 2005. Modeling of multivariate time series using hidden Markov models. Dissertation, University of California, Irvine, USA.

Lafaysse, M., B. Hingray, A. Mezghani,J. Gailhard, and L. Terray, 2014. Internal variability and model uncertainty components in future hydrometeorological projections: The Alpine Durance basin, Water Resour. Res. 50, 3317–3341, doi:10.1002/2013WR014897.

Liu Z, Xu Z, Charles SP, Fu G, Liu L. 2011. Evaluation of two statistical downscaling models for daily precipitation over an arid basin in China. International Journal of Climatology 31*:* 2006–2020

Liu WB, Fu GB, Liu CM, Charles SP. 2013. A comparison of three multi-site statistical downscaling models for daily rainfall in the North China Plain. Theoretical and Applied Climatology 111 (3): 585-600. DOI: 10.1007/s00704-012-0692-0

Maraun, D., et al. (2010) Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. Rev. Geophys. 48, RG3003, doi:10.1029/2009RG000314.

Mehrotra. R., Sharma, A. 2006. A nonparametric stochastic downscaling framework for daily rainfall at multiple locations, J. Geophys. Res., 111, D15101, doi:10.1029/2005JD006637.

Mehrotra R, Sharma A. 2011. Impact of atmospheric moisture in a rainfall downscaling framework for catchment-scale climate change impact assessment. International Journal of Climatology, 31*:*431–450.

Phatak, A., B. Bates, *and* S. Charles *(*2011*)*, Statistical downscaling of rainfall data using sparse variable selection methods, Environ. Modell. Software, 26*(*11*)*, 1363–1371.

Potter N. J., Chiew, F. H. S., Frost, A. J., 2010. An assessment of the severity of recent reductions in rainfall and runoff in the Murray-Darling Basin. Journal of Hydrology 381, 52–64.

Potter N. J., Chiew, F. H. S., 2011. An investigation into changes in climate characteristics causing the recent very low runoff in the southern Murray-Darling Basin using rainfall-runoff models. Water Resour. Res. 47, W00G10, doi:10.1029/2010WR010333.

Potter N.J., Ekström M., Chiew F.H.S., Zhang L., Fu G.B., Uncertainty from downscaling global climate model information for hydroclimate projections, *J. Hydrol.* (In review)

Radanovics S, JP Vidal, E Sauquet, A Ben Daoud, G Bontron. 2013. Optimising predictor domains for spatially coherent precipitation downscaling, Hydrol. Earth Syst. Sci., 17, 4189-4208

Robertson AW, Kirshner S, Smyth P. 2004. Downscaling of daily rainfall occurrence over northeast Brazil using a hidden Markov model. J. Climate 17, 4407 – 4424.

Robertson AW, Ines AVM, Hansen JW. 2007. Downscaling of seasonal precipitation for crop simulation. Journal of Applied Meteorology and Climatology 46*:* 677–693.

Sauter T, V Venema, 2011, Natural Three-Dimensional Predictor Domains for Statistical Precipitation Downscaling, Journal of Climate, 24, 6132-6145.

Sunyer M.A., Hadsen H. and Ang P.H., A comparison of different regional climate models and statistical downscaling methods for extreme rainfall estimation under climate change, *Atmos. Res.* **103**, 2012, 119–128.

Timbal B, Hope P, Charles S (2008a) Evaluating the consistency between statistically downscaled and global dynamical model climate change projections. J Climate 21: 6052–6059. DOI: 10.1175/2008JCLI2379.1

Timbal B, Li Z, Fernadez E (2008b) The Bureau of meteorology statistical downscaling model graphical user interface: user manual and software documentation. CAWCR Technical Report No. 004. Bureau of Meteorology, Australia

Van Vuuren DP, Edmonds J, Kainuma M, Riahi K, Thomson A, Hibbard K, Hurtt GC, Kram T, Krey V, Lamarque JF, Masui T, Meinshausen M, Nakicenovic N, Smith SJ and Rose SK (2011) The representative concentration pathways: an overview. Climatic Change 109(1–2), 5–31. DOI: 10.1007/s10584-011-0148-z.

Vrac, M., Stein, M. L., Hayhoe, K. & Liang, X.-Z., 2007. A general method for validating statistical downscaling methods under future climate change. Geophysical Research Letters, 34, L18701. Doi: 10.1029/2007GL030295

Wilby, R.L., Charles, S.P., Zorita, E., Timbal, B., Whetton, P., Mearns, L.O., 2004. Guidelines for Use of Climate Scenarios Developed from Statistical Downscaling Methods. IPCC Task Group on Data and Scenario Support for Impact and Climate Analysis (TGICA), http://www.ipcc-data.org/guidelines/dgm_no2_v1_09_2004.pdf.

Wilby, R.L., Wigley, T.M.L., Conway, D., Jones, P.D., Hewitson, B.C., Main, J. and Wilks, D.S. 1998. Statistical downscaling of general circulation model output: A comparison of methods. Water Resources Research, 34, 2995-3008.

Wilby, R.L. and Wigley, T.M.L. 1997. Downscaling general circulation model output: a review of methods and limitations. Progress in Physical Geography, 21, 530-548.

Wu, W., Liu, Y., Ge, M., Rostkier-Edelstein, D., Descombes, G., Kunin, P., Warner, T., Swerdlin, S., Givati, A., Hopson, T., Yates, D., 2012. Statistical downscaling of climate forecast system seasonal predictions for the Southeastern Mediterranean. Atmos. Res. 118 (15), 346–356. http://dx.doi.org/10.1016/j.atmosres.2012.07.019

Yu, J. J., Fu, G.B., Cai, W., Cowan, T. 2010. Impacts of precipitation and temperature changes on annual streamflow in the Murray-Darling Basin. Water International 35: 313–323.

Zhang, Y, Q. You, C. Chen, J. Ge, 2016, Impacts of climate change on streamflows under RCP scenarios: a case study in Xin River Basin, China, Atmos. Res., 178-179, 521–534

List of Figure Captions