Instrument: **SP1 Cooperation**

Thematic Priority: **Priority Area 1.1.6.3 "Global Change and Ecosystems"**

FP7 Collaborative Project – Grant Agreement 212492

CLARIS LPB
A Europe-South America Network for Climate Change Assessment and Impact Studies in La Plata Basin

DELIVERABLES

D7.2: Recommendations on methods and datasets for calibration and combination of multi-model ensemble simulations

Due date of deliverable: Month 18

Start date of project: **01/10/2008** Duration: 4 **years**

Organisation name of lead contractor for this deliverable: P-INPE

Deliverable No	Deliverable title	WP	Lead beneficiary	Estimated person-months	Real person-months	Nature	Dissemination level	Delivery date
D7.2	Recommendations on methods and datasets for calibration and combination of multi-model ensemble simulations	WP 7	P9-INPE	12,00	12,84	O	CO	18

D7.2: Recommendations on methods and datasets for calibration and combination of multi-model ensemble simulations

Partner involved: INPE

1. Introduction

The production process of future climate simulations for climate change studies using climate models involves a considerable amount of uncertainties. These uncertainties arise, for example, from: a) assumptions about the future atmospheric constitution in terms of gases; b) our inability to precisely observe the present climate; and c) a number of assumptions that need to be made in climate models in order to simplistically represent the real climate.

Recognising these sources of uncertainties, approaches have been developed for sampling some of these uncertainties in order to make estimates of future climate conditions. These approaches involve the generation of a number of climate simulations with a single climate model with different choices for various model parameters – the so called perturbed physics ensemble, or the generation of a number of climate simulations with different climate models in order to generate the so called multi-model ensemble of future climate conditions. In this procedure, a large number of future climate simulations is generated and used to estimate possible future ranges of values for a particular variable of interest (e.g. temperature). The motivation for the use of multi-model ensembles as a complement to single model perturbed physics ensembles lies on the fact that the use of different models potentially allows a more comprehensive quantification of uncertainty aspects involved in the future climate projection problem.

This deliverable aims to review and recommend methods for dealing with large ensemble simulations in order to address uncertainties in future climate change simulations for producing combined and calibrated multi-model ensemble simulations for impact studies in the La Plata Basin (LPB). Recommendations on climate datasets for dealing with the future climate projection problem are also provided.

2. Methods for calibration and combination of multi-model ensemble simulations

Raisanen and Palmer (2001) can be considered the pionner study applying a probabilistic view of climate change projections on the basis of a multi-model ensemble composed by 17 ocean-atmosphere general circulation models participating in the Coupled Model Intercomparison Project phase 2 (CMIP2). Based on these models, probabilities of threshold events such as ‘the warming at the time of doubled CO2 will be greater than 18ºC’ were computed as the fraction of models that simulated such an event, therefore assigning equal weight to each model when counting frequencies of exceedance.

The next important step towards objective probabilistic projections was published in articles by Giorgi and Mearns (2002, 2003) introducing the reliability ensemble average (REA) approach. This approach assumes a different perspective when considering the contribution of each model to the final multi-model ensemble estimate. Two criteria are chosen for defining the contribution of each model: a) model performance in replicating current climate and b) inter-model agreement in the projections of future change. In other words, models with small bias and projections that agree with the ensemble ‘consensus’ are rewarded while models that perform poorly in replicating observed climate and that appear as outliers are penalised. The REA method proposes an estimation of model weights through which ‘bias’ and

'convergence' criteria are quantified. The weights for each contributing model *i* for the ensemble are defined as

$$R_i = \left[(R_{B,i})^m.(R_{D,i})^n\right]^{1/(m.n)} = \left\{\left[\frac{\varepsilon_T}{|B_{T,i}|}\right]^m.\left[\frac{\varepsilon_T}{|D_{T,i}|}\right]^m\right\}^{[1/(m.n)]}$$

and the final weighted multi-model ensemble average is computed as

$$\tilde{\Delta T} = \frac{\sum_i R_i \Delta T_i}{\sum_i R_i}$$

where ΔT_i is the projected change of each model. The weight R_i for an individual model is defined as the product of two terms ($R_{B,i}$ and $R_{D,i}$), one inversely proportional to the absolute bias $B_{T,i}$, and the other to the absolute distance $D_{T,i}$ between the model projected change and the final weighted ensemble average. At the numerator, ε_T is a measure of natural variability in 30-yr average temperature and precipitation, and ensures that models whose bias and deviation are not large relative to natural fluctuations are not unfairly penalised. For estimating ε_T, time series of observed temperature and precipitation for the twentieth century are selected. Next 30-yr moving averages of the series after linearly detrending the data (to remove century-scale trends) are computed and finally ε_T is estimated as the difference between the maximum and minimum values of these 30-yr moving averages. The exponents *m* and *n* are designed to modulate the relative importance of the two terms in the weighted average, but are set equal to 1.

The REA studies by Giorgi and Mearns (2002, 2003) motivated the work by Tebaldi et al. (2004, 2005) and Smith et al. (2009). The latter authors proposed a Bayesian implementation of the REA approach making assumptions on the statistical distribution of model outputs and observations in order to determine the so-called likelihood distribution, which is then combined through Bayes theorem with the prior distributions to derive posterior distributions of uncertain quantities of interest (e.g. the climate change signal). Gaussian (normal) assumptions are made for the current (X_i) and future (Y_i) model projections, centred around the mean climate signals, μ and ν, respectively, with model-specific variances:

$$X_i \sim N(\mu, \lambda_i^{-1})$$
$$Y_i \sim N(\nu, (\theta\lambda_i)^{-1})$$

where the notation $N(\mu, \lambda^{-1})$ stands for a Gaussian distribution with mean μ and variance $1/\lambda$. Similarly, the observed current climate X_0, is modelled as a Gaussian distribution centred around the same current climate signal μ, with variance estimated through the observed record

$$X_0 \sim N(\mu, \lambda_0^{-1})$$

The posterior (i.e. multi-model or combined) distribution for the climate signal is derived using Bayes theorem and numerical Markov Chain Monte Carlo methods, and translated into a probability distribution for climate change defined as $\nu-\mu$. The form of the posterior means for μ and ν is approximately

$$\tilde{\mu} = \frac{\sum_i \lambda_i X_i}{\sum_i \lambda_i}$$

and

$$\tilde{\nu} = \frac{\sum_i \lambda_i Y_i}{\sum_i \lambda_i}$$

where the model-specific λ_i resamble the REA weights, being estimated as

$$\tilde{\lambda}_i = \frac{a+1}{b + \frac{1}{2}\left[(X_i - \tilde{\mu})^2 + \theta(Y_i - \tilde{\nu})^2\right]}$$

The first term of the denominator in the equation above is a measure of bias, being the distance of the present climate average X_i simulated by model *i* from the optimal estimate of the current climate. The second term is a measure of convergence, computing a distance between the model's future projection Y_i from the future climate signal's posterior mean. The terms a and b are parameters chosen as orders of magnitude smaller than the remaining terms to avoid significant impact on the final estimates. As in Giorgi and Mearns (2002), in this Bayesian implementation of the REA approach, models with large bias and too far from the multi-model ensemble receive less weight. Sharp criticisms have been raised against the validity of the convergence criterion when analysing a set of models that are by design 'best guesses' rather than attempting to sample a wide range of uncertainties, and whose agreement may be a consequence of inbreeding rather than reciprocal validation of individual tendencies. In particular, it has been often argued that there may exist common weaknesses in the representation of certain processes in a majority of models, and consequently outliers may not appear at random. In response to these concerns, the authors proposed a variant of the analysis in which the models considered as outliers are not heavily penalized (Tebaldi et al. 2004). This is achieved by a priori assigning a large probability to the models being less 'precise' in their future projections compared with their skill in current projections. This formally translates into a prior distribution for the parameter θ in the equation above that is concentrated on values less than 1. Another related consequence of assuming independence among GCM projections (which is implicitly or explicitly the case for all the methods described so far) is that any statistical analysis will produce increasingly more precise estimates (e.g. narrower posterior distributions of climate change signals) as the number of models in the ensemble increases.

Another objective procedure was published by Greene et al. (2006), who have combined multi-model ensembles with a method commonly used in seasonal and interannual forecasting. In this method, a Bayesian hierarchical linear model is fitted to an observational dataset of regionally aggregated seasonal and annual temperatures, where the predictors are similarly aggregated GCM projections. The observed temperature (Y_{ik}) for region *i* and time *k* are modelled as centred around a mean value, with a Gaussian error, as

$$Y_{ik} \sim N(\mu_{ik}, \sigma_k^2)$$

with the mean value modelled as a linear combination of GCM outputs

$$\mu_{ik} = \beta_{ok} + \sum_j \beta_{jk} X_{ijk}$$

where X_{ijk} indicates the simulated temperature in region *i* at time *k* by model *j*. This procedure is similar to performing model calibration using retrospective forecasts and past observations. The calibrated ensemble is used to derive climate change projections, and given the random nature of the parameters in this hierarchical Bayesian model, the posterior distribution translates into a probability distribution for the climate change projection. The main assumption of this approach is that of stationarity of the relationship between observed and simulated climate, estimated in the training period of the twentieth century and applied to future simulations. This strong assumption causes obvious differences between the simple average projections from the GCMs and the projections synthesized from the calibrated ensemble, in many cases resulting in distributions over a range of values significantly shifted, more often towards lower values.

Using a spatial Bayesian model Furrer et al. (2007) investigated GCM outputs at the grid point level, rather than at the aggregated level of large subcontinental regions as performed in all methods presented so far. The central idea of the approach in this spatial model is to model each GCM field of temperature or precipitation change as a random process on the sphere. The field is composed of two additive components: a large-scale climate signal and a small-scale error field, representing both model bias and internal variability. Thus, modelling the field of change, denoted as D_i for the *i*th GCM, and defined as the difference, grid point by grid point, of the future mean projection minus the current mean projections

$$D_i = Y_i - X_i$$

the statistical model is given by

$$D_i = M\theta_i + \varepsilon_i.$$

The large-scale signal, represented as the first additive term in the equation above, is modelled as a linear combination of a set of truncated basis functions, filling the columns of the matrix M. The basis functions are spherical harmonics, apt to represent spatial structure on a sphere, plus a set of additional vectors modelling the expected geographical patterns like, for example, a land/ocean mask. Observations are also used as one of the additional columns in the linear combination, in the hope that they will help explain some of the effect of the physical processes that create climate on Earth but are not easily represented through statistical modelling. In this study, there is no direct use of either a bias or convergence criterion in the spatial model. The coefficients of the linear combination are the components of the vector θ_i. The small-scale residual field ε_i is modelled as a realization of a stationary Gaussian random field of mean zero. Both the linear coefficients θ_i and the scale parameters of the covariance function in the Gaussian process ε_i are model-specific, to account for the different GCMs' characteristics in replicating the true climate signal. The vectors θ_i of linear coefficients are samples from a distribution whose mean are the 'true' coefficients. The goal of the Bayesian analysis is to estimate the posterior distribution of the true coefficients. Once recombined with the basis functions, the posterior distribution for the true coefficients will translate into a multidimensional probability distribution of the large-scale signal of climate change, jointly quantifying the uncertainty over the global grid. This remains the only published method to represent the uncertainty over a global map, using spatial statistics to model

geographical patterns of varying degrees of smoothness (e.g. temperature change fields rather than precipitation change fields) as a function of the spatial correlation between locations.

Figure 1 shows a comparision of probability density functions (pdfs) of temperature change produced by the methods in Tebaldi et al. (2004), Greene et al. (2006) and Furrer et al. (2007) using the Intergovernmental Panel on Climate Change (IPCC) fourth assessment (AR4) model experiments based on Special Report on Emissions Scenario (SRES) A1B scenario. The pdfs represent projections of temperature change in boreal winter (DJF, left panels) and summer (JJA, right panels) for the end of the twenty-first century. The figure also shows the empirical distributions of GCM projections (shaded histograms). Four out of the 22 regions introduced by Giorgi and Francisco (2000), which have become standard for regional climate change analysis in GCMs, have been selected: Western North America (WNA, first row), the Mediterranean basin (MED, second row), Northern Asia (NAS, third row) and Southeast Asia (SEA, fourth row). The empirical distribution represents the method by Raisanen and Palmer (2001) and Palmer and Raisanen (2002). The figures highlights that different methods produce different curves. The methods by Tebaldi et al. (2004) and Furrer et al. (2007) produce similar curves in terms of location of the central (most likely) estimate, and in most cases also present similar width. These two methods have narrower distributions than the empirical GCM distribution. The method by Greene et al. (2006) produces wider pdfs, probably becasuse of the large degree of uncertainty in the estimation of the calibration coefficients. This methos also tends to show for some regions shifted curves with respect to the empirical distribution. This effect is most likely due to the calibration coefficients being significantly different from those of a simple average of all available GCM projections.

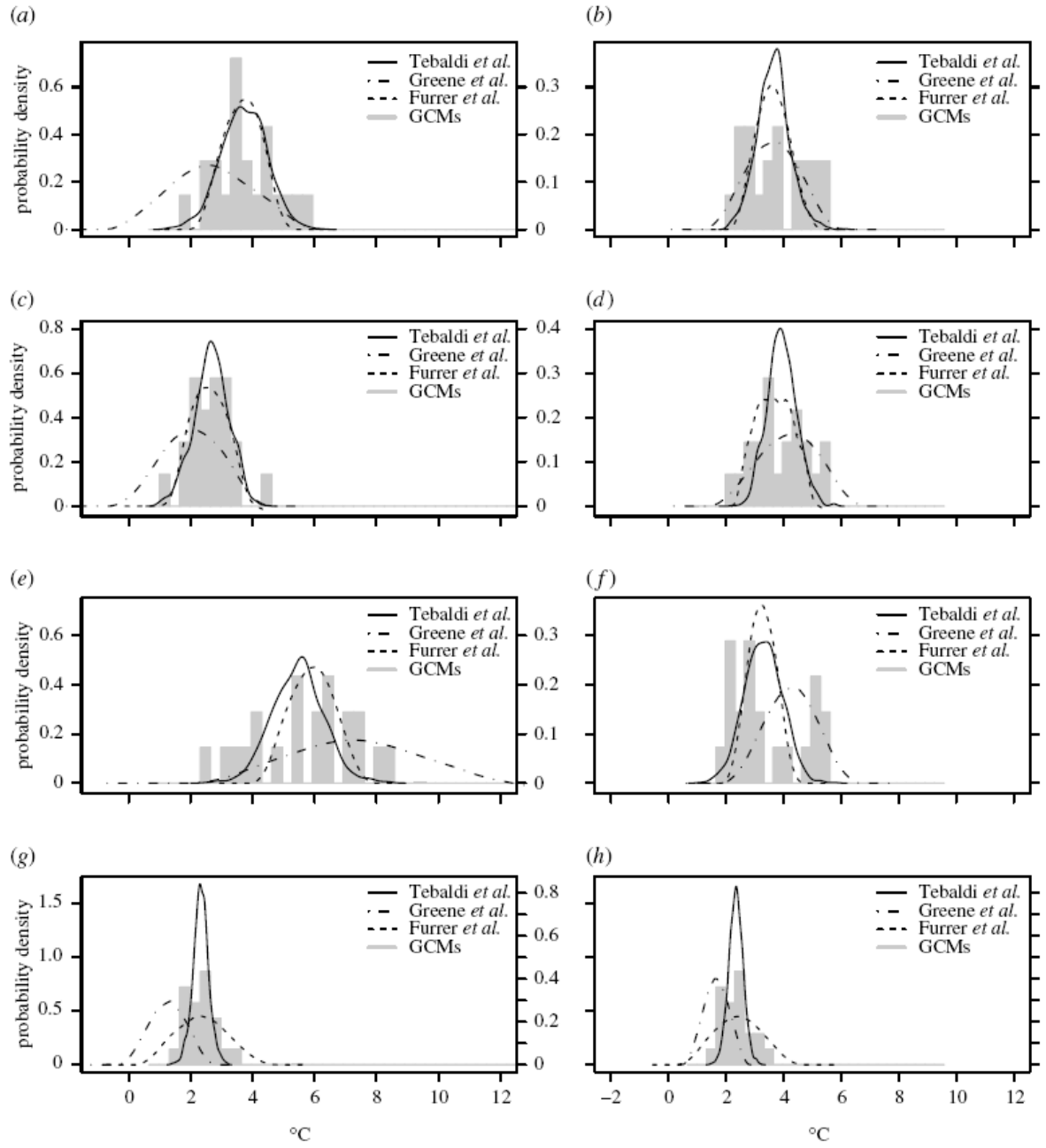
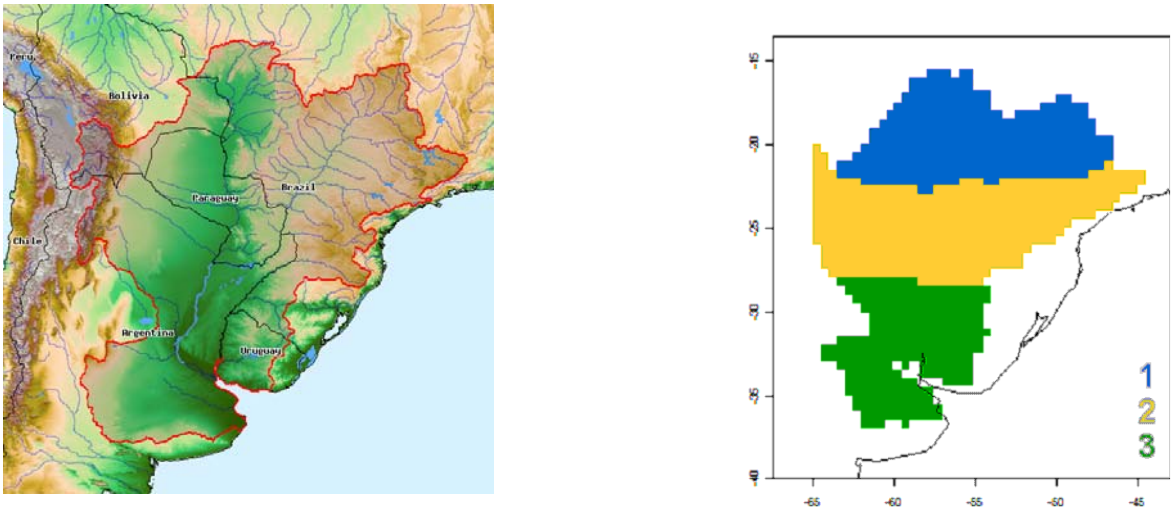


Figure 1. Comparison of pdfs of temperature change under SRES A1B scenario produced with the methods of Tebaldi et al. (2004), Greene et al. (2006) and Furrer et al. (2007) for boreal winter (DJF, left panels) and summer (JJA, right panels) for four regions: Western North America (WNA, first row), the Mediterranean basin (MED, second row), Northern Asia (NAS, third row) and Southeast Asia (SEA, fourth row). A total of 20 IPCC AR4 model projections is used to produce the pdfs. Grey histograms represent empirical GCM projected change distributions obtained with the method by Raisanen and Palmer (2001) and Palmer and Raisanen (2002). The temperature changes are computed as the difference between two 20-year averages, 2080–2099 versus 1980–1999. Source: Tebaldi and Knutti (2007).

3. Calibration and combination of temperature scenario simulations over LPB with linear and fuzzy regression

In this section two methods are investigated for the calibration and combination of annual mean temperature scenario simulations over the La Plata Basin (LPB), which is illustrated in Figure 2a. The two investigated methods are: a) multiple linear regression (Greene et al., 2006), and b) fuzzy regression (Bisserier et al., 2010). First, cluster analysis was performed on 20th century temperature observations (Mitchell and Jones, 2005) to identify climatic homogeneous regions within LPB. Three main climatic homogeneous regions have been identified and are illustrated in Figure 2b, representing the meridional variation of temperature (warmer temperatures towards the north and colder temperatures towards the south). Next, temperature projections produced by five IPCC AR4 models under A1B scenario (see table 1 for the list of models used in the investigation) have been extracted and averaged over each homogeneous region in order to produce a time series (i.e. an index) representative of the region. The observed temperatures of the 20th century have also been averaged over homogeneous regions. Then, the observed temperatures and the climate model temperature projections for the 20th century for each region have been used to build both the multiple linear regression and the fuzzy regression equations for the temperature indices of each homogeneous region, which have then been used for producing combined temperature projections for the 21st century for each homogeneous region in LPB.



(a) **(b)**

Figure 2: a) Boundary of LPB (red line) in Southeastern South America. b) Three homogeneous regions in terms of annual mean temperature according to a k-means cluster analysis.

Table 1: Observations (first row) and IPCC AR4 models (remaining rows) used in the investigation.

Acronym	Instituiton	Country	Ensemble members	Spatial resolution
CRU[1]	University of East Anglia	U.K	-	0.5° x 0.5°
GFDL[2]	US Dept. of Commerce / NOAA / Geophysical Fluid Dynamics Laboratory	USA	4	2.5° x 1.5°
GISS[3]	NASA / Goddard Institute for Space Studies	USA	9	5° x 5°
IPSL[4]	Institut Pierre Simon Laplace	France	3	3.75° x 3.75°
ECHAM5[5]	Max Planck Institute for Meteorology	Germany	8	1.875° x 1.875°
HADCM3[6]	Hadley Centre for Climate Prediction and Research / Met Office	U.K.	3	3.75° x 3.75°

1: Mitchell and Jones (2005), 2: Delworth et al., (2006), 3: Schmidt et al., (2005), 4: Hourdin et al., (2005), 5: Roeckner et al., (2003), 6: Johns et al., (2003).

Figure 3 shows the observed annual mean temperature during the 20th century (solid red lines in panels a, c and e) and individual model projections (solid black lines in panels a, c and e) for the five IPCC AR4 models for homogeneous regions 1, 2 and 3, respectively. The right panels of Figure 3 show combined and calibrated annual mean temperature projections for the 21st century with multiple linear regression (solid red lines in panels b, d and f) and individual model projections (solid black lines in panels b, d and f) for the five IPCC AR4 models for homogeneous regions 1, 2 and 3, respectively. Model projections for the end of the 21st century suggest an increase in temperature of the order of 1°C for the three homogenous regions.

Figure 4 shows combined and calibrated annual mean temperature projections for the 20th and 21st centuries with multiple linear regression (solid thick line) and 90% confidence intervals (red band) for homogeneous regions 1, 2 and 3. The uncertainty in temperature projections for the 21st century is larger than for the 20th century for the three homogenous regions, as indicated by the larger confidence intervals of the 21st century when compared to the confidence intervals of the 20th century.

Figure 5 shows combined and calibrated annual mean temperature projections for the 20th and 21st century with fuzzy regression. The blue band represents the 90% confidence interval and the green band the 40% confidence interval for homogeneous regions 1, 2 and 3. The cental grey line in panels a, c and e is the observed annual mean temperature during the 20th century, and the upper and lower grey lines are the observed 95th and 5th percentiles, respectively. The uncertainty in temperature projections for the 21st century is generaly larger when estimated with fuzzy regression compared to when estimated with multiple linear regression for the three homogenous regions, as indicated by the larger 90% confidence

intervals in Figure 5 when compared to Figure 4. In accordance with the results from multiple linear regression, fussy regression also suggests an increase in temperature of the order of 1°C for the end of the 21st century for the three homogenous regions. These results are also supported by Figure 6, which shows a comparison of decadal mean temperature projections for the 21st century with multivariate linear and fuzzy regression.

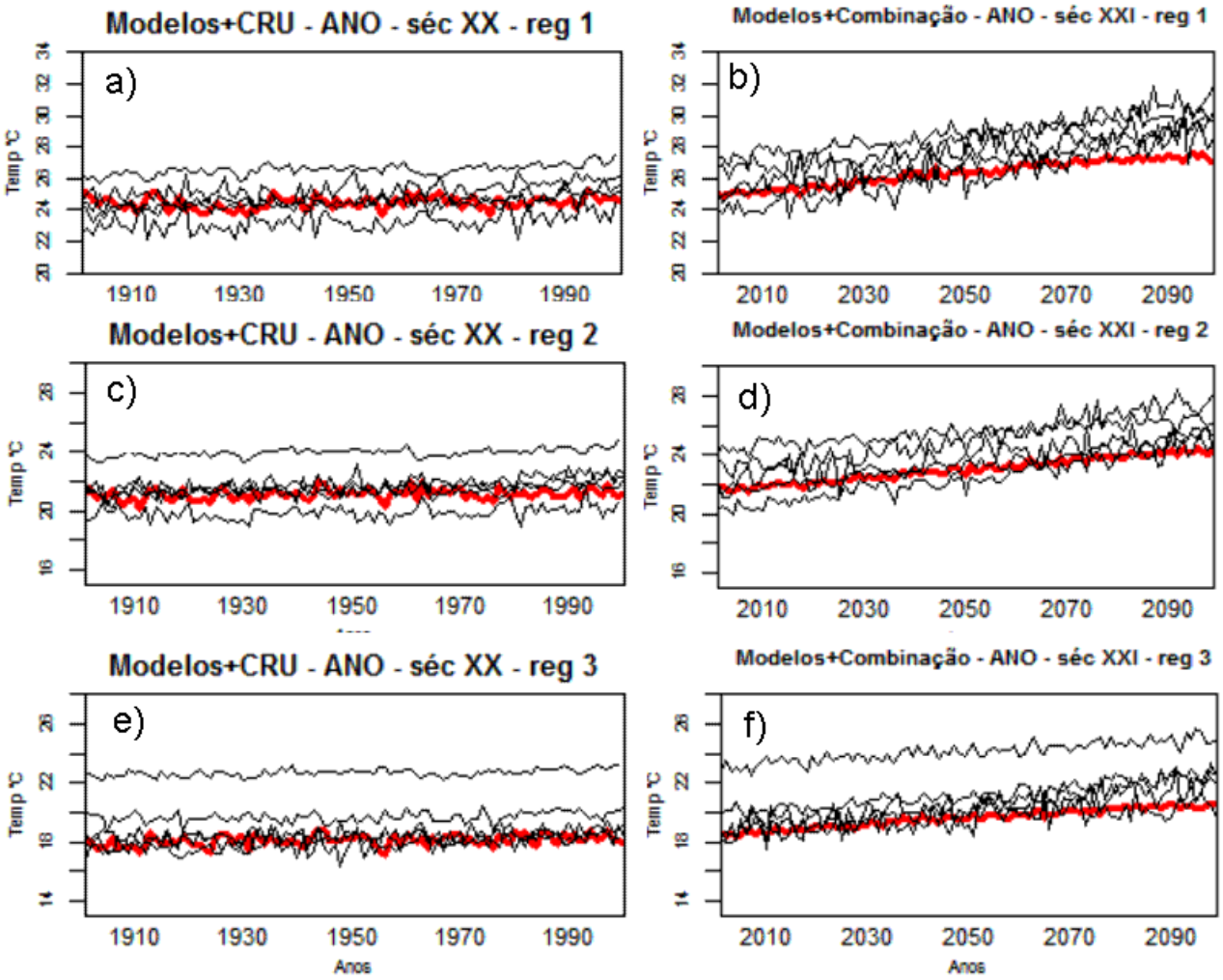


Figure 3: Left panels, observed annual mean temperature during the 20th century (solid red line) and individual model projections (solid black lines) for the five IPCC AR4 models for homogeneous regions 1 (panel a), 2 (panel c) and 3 (panel e). Right panels, combined and calibrated annual mean temperature projection for the 21st century with multiple linear regression (solid red line) and individual model projections (solid black lines) for the five IPCC AR4 models for homogeneous regions 1 (panel b), 2 (panel d) and 3 (panel f).

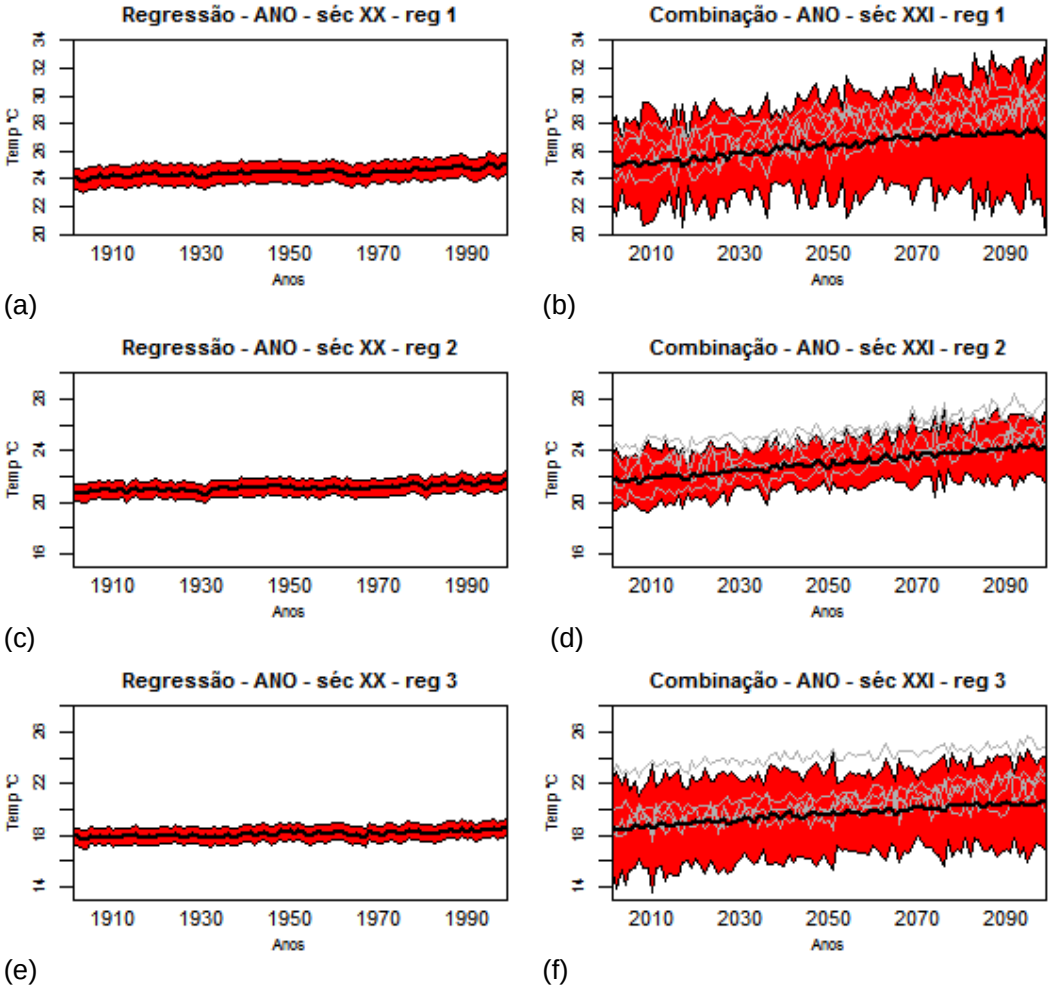


Figure 4: Left panels, combined and calibrated annual mean temperature projection for the 20th century with multiple linear regression (solid thick line) and 90% confidence interval (red band) for homogeneous regions 1 (panel a), 2 (panel c) and 3 (panel e). Right panels, individual model projections for the 21st century for the five IPCC AR4 models (solid grey lines), combined and calibrated annual mean temperature projections for the 21st century with multiple linear regression (solid thick line) and 90% confidence interval (red band) for homogeneous regions 1 (panel b), 2 (panel d) and 3 (panel f).

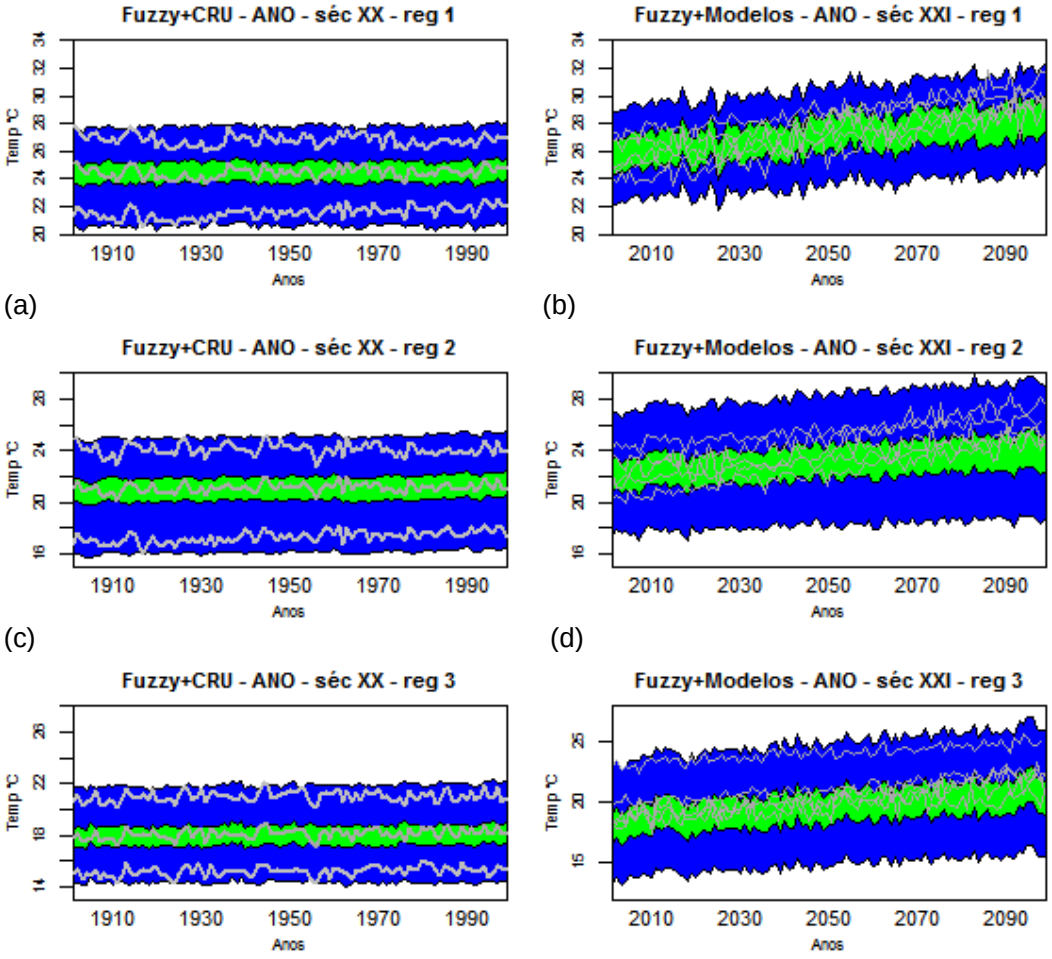


Figure 5: Left panels, combined and calibrated annual mean temperature projection for the 20th century with fuzzy regression (90% confidence interval is given by the blue band and 40% confidence interval is given by the green band) for homogeneous regions 1 (panel a), 2 (panel c) and 3 (panel e). The central grey line is the observed annual mean temperature. The upper grey line is the observed 95th percentile. The lower grey line is the observed 5th percentile. Right panels, individual model projections for the 21st century for the five IPCC AR4 models (solid grey lines), combined and calibrated annual mean temperature projections for the 21st century with fuzzy (90% confidence interval is given by the blue band and 40% confidence interval is given by the green band) for homogeneous regions 1 (panel b), 2 (panel d) and 3 (panel f).

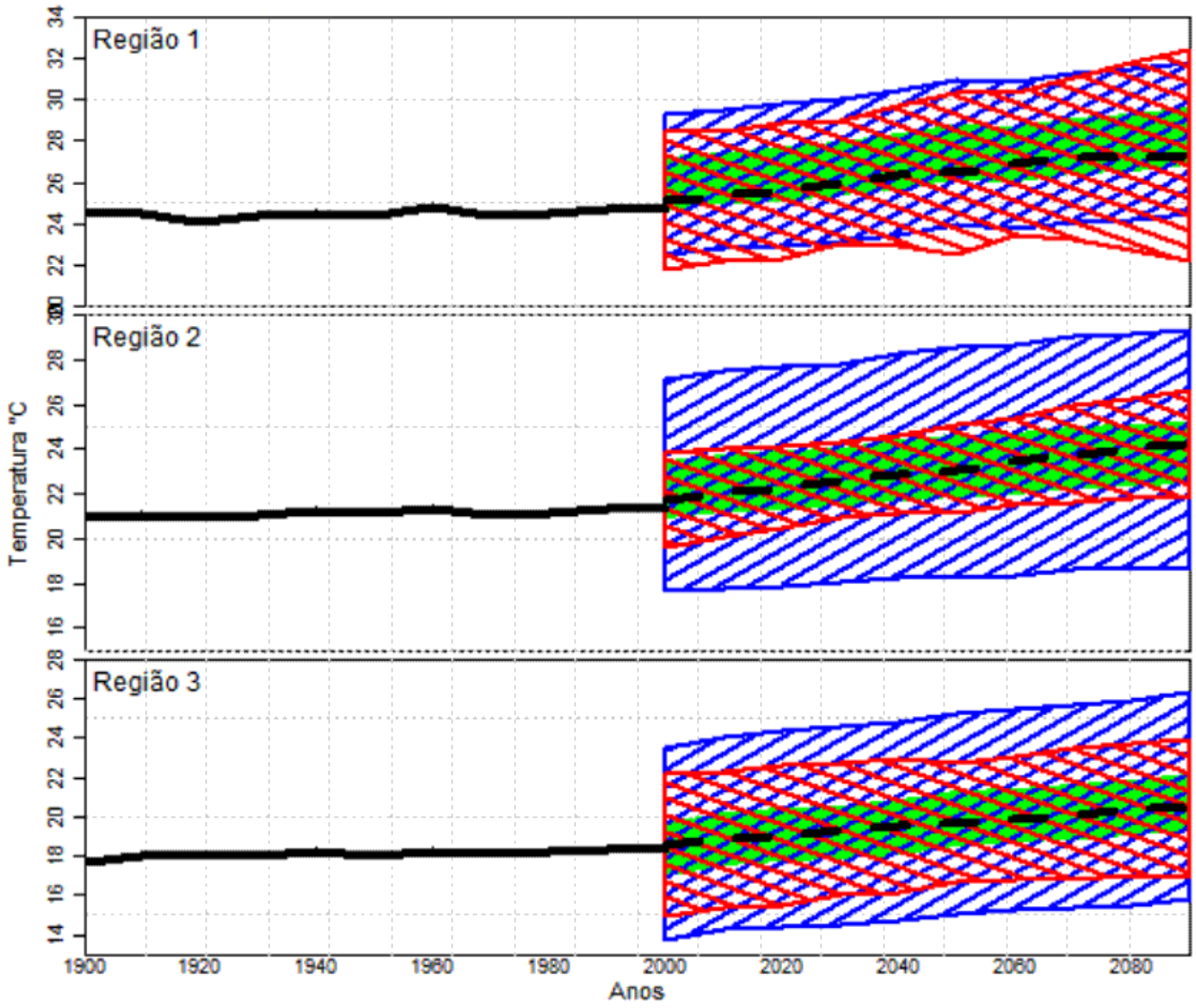


Figure 6: Observed decadal mean temperatures during the 20th century (solid thick line) for homogeneous region 1(top panel), 2 (middle panel) and 3 (bottom panel). Combined and calibrated decadal mean temperature projection for the 21st century with multiple linear regression (dashed thick line). The red hashed band is the 21st century projection 90% confidence interval for multiple linear regression. The blue hashed band is the 21st century projection 90% confidence interval for fuzzy regression. The green band is the 21st century projection 40% confidence interval for fuzzy regression.

4. Spatial calibration of El Niño-induced drought pattern in climate change projections

Climate change models are capable of simulating El Niño precipitation variability during the 20th century and present robust precipitation projections for the 21st century (Coelho and Goddard 2009). Given the ability of climate change models to simulate El Niño precipitation variability during the 20th century and the robustness of the projected precipitation for the 21st century one can use spatial calibration techniques to produce improved projections of this variability, and thus better assess increasing or decreasing drought vulnerability, for the 21st century. To illustrate how climate change projections can be spatially calibrated the IPCC AR4 HADCM3 model is used. This is one of the models that best simulates El Niño-induced spatial patterns during the autral summer (DJF) in the 20th century (see Table 2 of Coelho and

Goddard 2009). The illustration is performed using the first member of HADCM3 simulations for both 20th and 21st centuries.

For the spatial calibration a simple Bayesian approach known as forecast assimilation (Stephenson *et al.* 2005) has been used. This approach is currently used in seasonal forecasting (Coelho *et al.* 2006). The calibration is performed when modeling a so-called likelihood function, where 20th century simulations are regressed on 20th century observations for obtaining a calibration equation. Because of spatial dependence of neighboring grid points and the large dimensionality of the datasets involved in the calibration procedure, the regression is performed on the leading mode of the canonical correlation analysis between the modeled and observed 20th century DJF precipitation anomalies, which has been found to represent ENSO variability. Additional and more detailed information about the calibration procedure is given in Stephenson *et al.* (2005) and Coelho *et al.* (2006).

The illustration is performed in order to calibrate precipitation anomaly fields and produce a composite field (mean of a few El Niño events) representative of El Niño conditions. Prior to performing the canonical correlation analysis the following procedure is adopted to overcome the lack of temporal consistency of the climate change models in simulating the observed ENSO variability. First, both modeled and observed 20th century Niño-3.4 indices have been sorted in ascending order. Next, both modeled and observed 20th century precipitation have been paired according to the sorted Niño-3.4 index. Finally, the paired precipitation has been put in chronological order according to the observations. Figure 7 shows the leading mode of the canonical correlation analysis between modeled and observed 20th century precipitation, which represents ENSO variability. Figure 7a shows the time series (expansion coefficients) of modeled (solid line) and observed (dashed line) precipitation. The two time series are highly correlated (correlation coefficient of 0.92) indicating that HADCM3 has a good representation of ENSO variability. Large positive values indicate El Niño years. Large negative values indicate La Niña years. Figures 7b-c show modeled and observed spatial precipitation patterns, respectively. The spatial pattern of Fig. 7b is given by the correlation of the observed time series (i.e. dashed line in Fig. 7a) and the modeled precipitation time series at each grid point. The spatial pattern of Fig. 7c is given by the correlation of the modeled time series (i.e. solid line in Fig. 7a) and the observed precipitation time series at each grid point.

Figure 7b shows that HADCM3 has a good response to ENSO in the equatorial Pacific and simulates positive precipitation anomalies over the region of positive sea surface temperature anomalies during El Niño years, as indicated by the positive correlation in the equatorial Pacific. The comparison of Figs. 7b and 7c also reveals that the model is able to simulate the observed negative precipitation anomalies over northern South America. Over southern Africa, the model produces weaker precipitation anomalies than observed. Over part of Indonesia, the model simulates the opposite precipitation signal to the observed. The patterns of Figs. 7b and 7c are similar to the canonical correlation analysis loading patterns used in the calibration procedure to project the 21st century model simulated patterns for each El Niño towards the observed pattern. For the 20th century simulations the calibration procedure for each year is performed using a cross-validation (leave one year out) procedure. After calibration is performed for each year in the 20th and 21st century, El Niño composites are produced for the calibrated precipitation anomalies.

Figures 8a shows the raw (i.e. uncalibrated) and Fig. 8b shows the calibrated HADCM3 El Niño precipitation anomaly composites for the 21st century. The calibrated composite (Fig. 8b) is in much better agreement with the observed composite (Fig. 9) than the raw composite (Fig. 8a). The calibration resulted in important regional changes in drought risk. For example, both northeast South America and northeast Australia appear as drought prone regions in the raw composite (Fig. 8a), while in the

calibrated composite (Fig. 8b) this drought feature is mostly replaced by normal conditions. Some regions in southern Africa appear as less prone to droughts in the raw composite (Fig. 8a), while in the calibrated composite (Fig. 8b) these regions now appear as more drought prone. Over Eastern Indonesia in the raw composite drought risk appears reduced during DJF El Niño conditions (Fig. 8a), while in the calibrated composite (Fig. 8b) this region now appeard as more at risk of drought during El Niño.

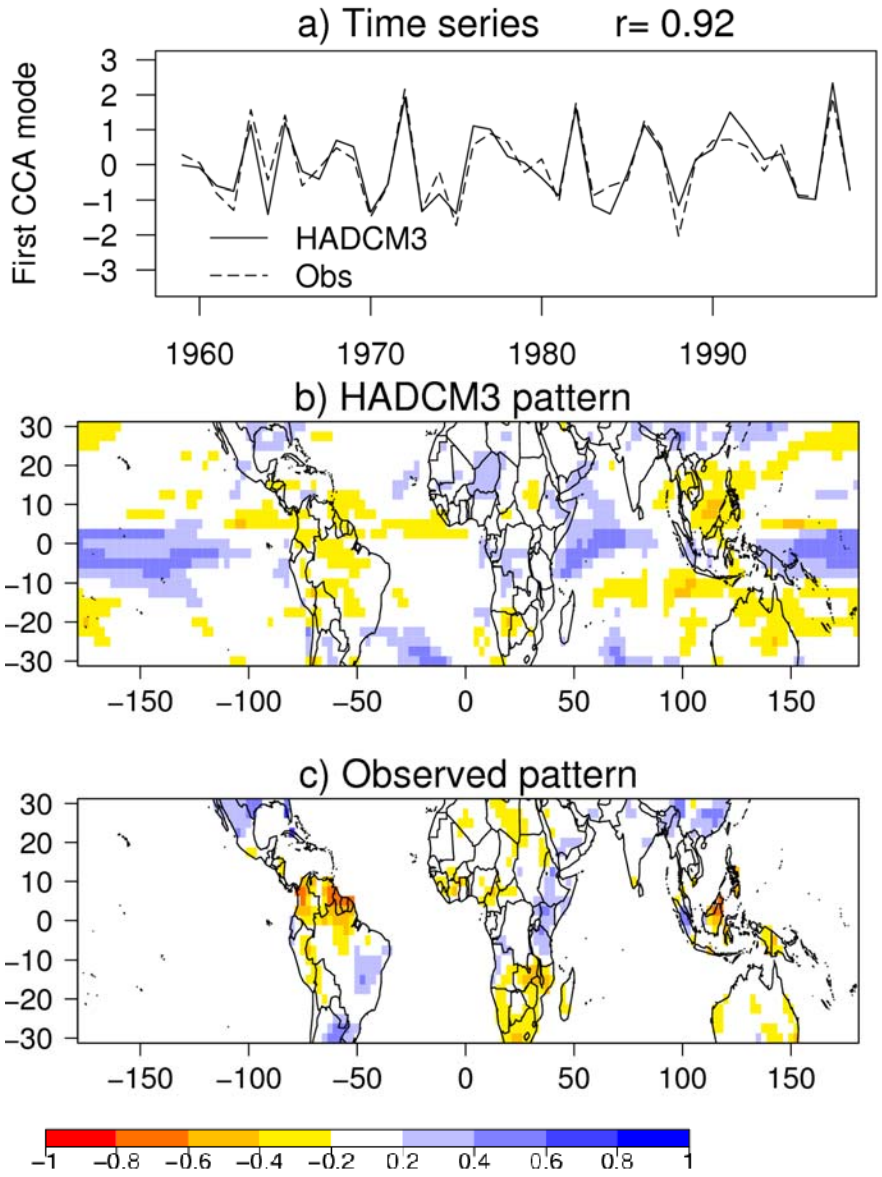


Figure 7: Leading canonical correlation analysis mode between DJF HADCM3 simulations and observed precipitation anomalies in the 20th century. a) time series (expansion coefficient). b) HADCM3 pattern. c) observed pattern. See text for additional explanation.

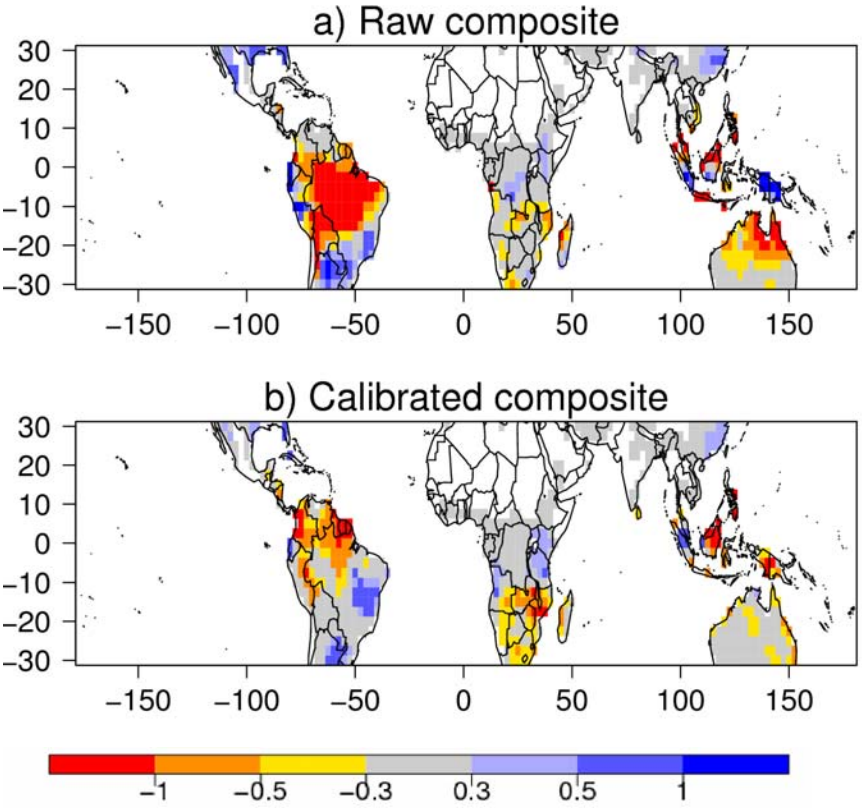


Figure 8: HADCM3 El Niño precipitation anomaly composites. a) Raw composite for the 21st century. b) Calibrated composite for the 21st century. Precipitation anomalies are expressed in mm.day^{-1}.

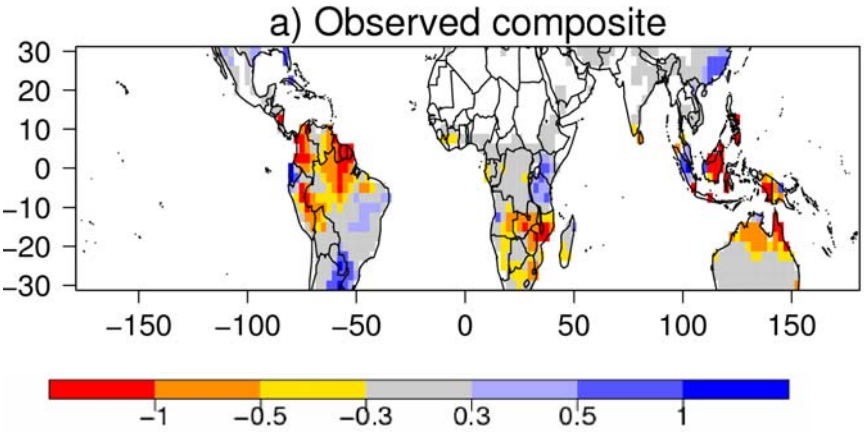


Figure 9: Observed DJF precipitation anomaly El Niño composite. The composite is given by the mean of the observed DJF standardized precipitation anomalies for the eleven years (1963/64, 1965/66, 1968/69, 1972/73, 1977/78, 1982/83, 1986/87, 1987/88, 1991/92, 1994/95 and 1997/98) classified as El Niño years. El Niño classification is determined by the Niño-3.4 index (SST anomaly in the equatorial Pacific region between 5°N, 5°S, 120°W,170°W) falling in the upper quartile of the observed distribution.

5. Datasets for calibration and combination of multi-model ensemble simulations

Several observational datasets are currently available for climate research and are appropriate to be used for the calibration and combination of multi-model ensemble simulations. Below is a list of the most commonly used datasets:

- The Climate Reseach Unit-University of East Anglia (CRU-UEA TS2.1) dataset (Mitchell and Jones, 2005) contains daily mean, minimum and maximum temperature, diurnal temperature range, precipitation, wet day frequency, frost day frequency, vapour pressure and cloud cover at a spatial global resolution of 0.5° x 0.5° degrees in latitude and longitude for the period 1901-2002. This dataset and the correpondent documentation is available at http://www.cru.uea.ac.uk/cru/data/hrg/timm/grid/CRU_TS_2_1.html

- The Global Precipitation Climatology Project (GPCP V2.1) dataset (Adler et al., 2003) contains monthly combined satellite-gauge global precipitation estimates at a spatial resolution of 2.5° x 2.5° degrees in latitude and longitude for the period 1979-present. Additional information is provided at ftp://precip.gsfc.nasa.gov/pub/gpcp-v2.1/doc/V2.1_doc.pdf
and the dataset is available at ftp://precip.gsfc.nasa.gov/pub/gpcp-v2.1/

- The Hadley Centre Climatic Research Unit Temperature version 2 (HadCRUT2v) dataset (Jones and Moberg 2003; Rayner et al. 2003) contains combined land and marine gridded monthly surface temperature at a spatial resolution of 5° x 5° degrees in latitude and longitude for the period 1870-2005. This dataset is available online at http://www.cru.uea.ac.uk/cru/data/tem2/. This is one of the best datasets with long time coverage (136 yr) available for climate research.

- The National Centers for Environmental Predictions/National Center for Atmospheric Research (NCEP/NCAR) Reanalysis (Kalnay et al., 1996; Kanamitsu et al., 2002) dataset contains global daily and monthly analysis of a comprehensive list of climate variables at a spatial resolution of 2.5° x 2.5° degrees in latitude and longitude for the period 1948-present. This dataset and correspondent documentation is available at http://www.esrl.noaa.gov/psd/data/reanalysis/reanalysis.shtml

- The European Centre for Medium-Range Weather Forecasts (ECMWF) 40-year Reanalysis (ERA-40) dataset (Uppala et al., 2005) contains global daily and monthly analysis of a comprehensive list of climate variables at a spatial resolution of 2.5° x 2.5° degrees in latitude and longitude for the period 1957-2002. This dataset and correspondent documentation is available at
http://data-portal.ecmwf.int/data/d/era40_moda/
http://data-portal.ecmwf.int/data/d/era40_daily/
http://www.ecmwf.int/research/era/do/get/era-40

- The European Centre for Medium-Range Weather Forecasts (ECMWF) Interim Reanalysis (ERA-Interim) dataset contains global daily and monthly analysis of a comprehensive list of climate variables at a spatial resolution of 1.5° x 1.5° degrees in latitude and longitude for the period 1989-present. This dataset and correspondent documentation is available at
http://data-portal.ecmwf.int/data/d/interim_moda/
http://data-portal.ecmwf.int/data/d/interim_daily/
http://www.ecmwf.int/research/era/do/get/era-interim

- The National Centers for Environmental Predictions/Climate Forecast System (NCEP/CFS) Reanalysis (Saha et al., 2010) is the first reanalysis dataset produced with a coupled atmosphere-

ocean-land surface-sea ice system and contains global daily and monthly analysis of a comprehensive list of climate variables at a spatial resolution of 0.5° x 0.5° degree in latitude and longitude for the period 1979-2010. This dataset and correspondent documentation is available at http://cfs.ncep.noaa.gov/cfsr/

- The Japanese 25-year Reanalysis (JRA-25) dataset (Onogi et al., 2007) produced by the Japan Meteorological Agency (JMA) and the Central Research Institute of Electric Power Industry (CRIEPI) contains global daily and monthly analysis of a comprehensive list of climate variables at a spatial resolution of 1° x 1° degree in latitude and longitude for the period 1979-2004. This dataset and correspondent documentation is available at http://jra.kishou.go.jp/JRA-25/index_en.html

6. Recommendations

Considerable research has been devoted to methods for combining and calibrating multi-model ensemble climate simulations (e.g., Giorgi and Mearns 2002, 2003; Tebaldi et al. 2004, 2005; Greene et al. 2006; Furrer et al. 2007; Smith et al. 2009). This deliverable reviewed these methods and explored the application of regression based methods (Greene et al. 2006; Bisserier et al., 2010) for producing combined and calibrated annual mean temperature projections for LPB. The deliverable also illustrated how climate change projections can be spatially calibrated. For the spatial calibration a simple Bayesian approach known as forecast assimilation (Stephenson et al. 2005) was used. This approach is currently used in seasonal forecasting (Coelho et al. 2006). The following recommendations are provided:

- As illustrated in Figure 1, different methods produce different projected climate change distributions. In order to increase understanding and confidence on currently available methods it is recommended the comparison of the empirical distribution as obtained in Raisanen and Palmer (2001) and Palmer and Raisanen (2002) with the distribution of each method described in section 2.
- Given the lack of robustness among currently available methods as evident by the significant disagreement among the distributions produced by the different methods, and the open questions and issues associated with the interpretation of multi-model ensembles for climate projections, the results of these multi-model ensemble analyses should be considered experimental, altough all developed methods have enough theoretical and statistical basis to justify their use.
- Simple regression approaches for the calibration and combination of multi-model ensemble climate simulations are atractive because of their straightforward implementation. However, these approaches assume stationarity of the relationship between observed and simulated climate, estimated in the training period of the twentieth century and applied to future simulations. This is a strong assumption that should be kept in mind when interpreting future climate projections obtained with these approaches.

In terms of datasets for the calibration and combination of multi-model ensemble climate simulations, in addition to the datasets listed in section 5, the following datsets are recommeded:

- CLARIS LPB data server (http://eolo.cima.fcen.uba.ar/DS/menu.html), developed by WP7, contains the archive of IPCC AR4 climate model simulations and will also have the arquive of regional climate model simulations over LPB produced by WP5.
- CLARIS LPB station dataset, compiled by WP3, containing daily maximun and minimun temperature and precipitation over LPB, which is available from http://wp32.at.fcen.uba.ar/

- CLARIS LPB gridded dataset, produced by WP3 using the station dataset above, which will be made available at the CLARIS LPB data server at http://eolo.cima.fcen.uba.ar/DS/menu.html
- The Climate explorer (http://climexp.knmi.nl/) contains a comprehensive list of datasets available on-line for climate analysis.
- The IRI data library (http://iridl.ldeo.columbia.edu/) contains a comprehensive list of datasets available on-line for climate analysis.

7. References

Adler, R.F., G.J. Huffman, A. Chang, R. Ferraro, P. Xie, J. Janowiak, B. Rudolf, U. Schneider,S. Curtis, D. Bolvin, A. Gruber, J. Susskind, P. Arkin, E.J. Nelkin, 2003: The Version 2.1 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979 - Present). *J. Hydrometeor.*, **4**(6), 1147-1167.

Bisserier A., R. Boukezzoula and S. Galichet, 2010: Linear Fuzzy Regression Using Trapezoidal Fuzzy Intervals. *Journal of Uncertain Systems*. 4(1), 59-72.

Coelho C.A.S. and L. Goddard, 2009: El Nino-induced tropical droughts in climate change projections. *J. Climate* . Vol. 22, No. 23, 6456-6476.

Coelho C.A.S., D. B. Stephenson, M. Balmaseda, F. J. Doblas-Reyes and G. J. van Oldenborgh, 2006: Towards an integrated seasonal forecasting system for South America. *J. Climate*. Vol. 19, No. 15, 3704-3721.

Delworth and co-authors, 2006: GFDL's CM2 global coupled climate models - Part 1: Formulation and simulation characteristics, *J. Climate*, 19(5), 643-674.

Furrer, R., Sain, S., Nychka, D. and Meehl, G., 2007: Multivariate Bayesian analysis of atmosphere–ocean general circulation models. *Environ. Ecol. Stat.*, 14(3), 249-266.

Giorgi, F. and Francisco, R., 2000: Uncertainties in regional climate change predictions. A regional analysis of ensemble simulations with the HADCM2 GCM. *Clim. Dynam*. 16, 169–182.

Giorgi, F. and Mearns, L., 2002: Calculation of average, uncertainty range and reliability of regional climate changes from AOGCM simulations via the 'reliability ensemble averaging' (REA) method. *J. Climate*. 15, 1141–1158.

Giorgi, F. and Mearns, L., 2003: Probability of regional climate change calculated using the reliability ensemble average (REA) method. *Geophys. Res. Lett.* 30, 1629–1632.

Greene, A., Goddard, L. and Lall, U., 2006: Probabilistic multimodel regional temperature change projections. *J. Climate*. 19, 4326–4343.

Hourdin, F. and co-authors, 2005: The LMDZ4 general circulation model: climate performance and sensitivity to parametrized physics with emphasis on tropical convection. *Climate Dynamics*, 19, 3445-3482.

Johns T.C. and co-authors, 2003: Anthropogenic climate change for 1860 to 2100 simulated with the HadCM3 model under updated emissions scenarios. *Climate Dynamics*, 20, 583-612.

Jones, P. D., and A. Moberg, 2003: Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001. *J. Climate*, 16, 206–223.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C.; Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., Joseph, D., 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteorol. Soc.* , v.77, p. 437-471.

Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S-K., Hnilo, J. J., Fiorino, M., Potter, G. L., 2002: NCEP-DEO AMIP-II Reanalysis (R-2). *Bul. Amer. Meteorol. Soc.*, v. 83, p. 1631-1643.

Mitchell, T. D. and P. D. Jones, 2005: An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int. J. Climatol.*, 25, 693-712.

Onogi, K., J. Tsutsui, H. Koide, M. Sakamoto, S. Kobayashi, H. Hatsushika, T. Matsumoto, N. Yamazaki, H. Kamahori, K. Takahashi, S. Kadokura, K. Wada, K. Kato, R. Oyama, T. Ose, N. Mannoji and R. Taira (2007) : The JRA-25 Reanalysis. J. Meteor. Soc. Japan, 85, 369-432.

Palmer, T. N. and Raisanen, J., 2002: Quantifying the risk of extreme seasonal precipitation events in a changing climate. *Nature*. 415, 512–514.

Raisanen, J. and Palmer, T. N., 2001: A probability and decision-model analysis of a multimodel ensemble of climate change simulations. *J. Climate*. 14, 3212–3226.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Globally complete analyses of sea surface temperature, sea ice and night marine air temperature, 1871-2000. *J. Geophys.Res.*, 108, 4407.

Roeckner, E. and co-authors, 2003: The atmospheric general circulation model ECHAM5. Part I: Model description, *Max Planck Institute for Meteorology Report Number 349*. 127pp.

Saha, S. and co-authors, 2010: The NCEP Climate Forecast System Reanalysis. BAMS. Submitted.

Schmidt, G. A. and co-authors, 2005: Present day atmospheric simulations using GISS Model E: Comparison to in-situ, satellite and reanalysis data. *J. Climate*, 19, 153-192.

Smith, R., Tebaldi, C., Nychka, D. and Mearns, L., 2009: Bayesian modeling of uncertainty in ensembles of climate models. *Journal of the American Statistical Association*. 104(485), 97-116.

Stephenson D. B., C.A.S. Coelho, F. J. Doblas-Reyes and M. Balmaseda, 2005: Forecast Assimilation: A unified framework for the combination of multi-model weather and climate predictions. *Tellus A* . Vol. 57, 253-264.

Tebaldi, C. and Knutti, R., 2007: The use of the multi-model ensemble in probabilistic climate projections. *Phil. Trans. R. Soc. A*, 365, 2053–2075.

Tebaldi, C., Mearns, L., Nychka, D. and Smith, R., 2004: Regional probabilities of precipitation change: a Bayesian analysis of multimodel simulations. *Geophys. Res. Lett.* 31, L24 213.

Tebaldi, C., Smith, R., Nychka, D. and Mearns, L., 2005: Quantifying uncertainty in projections of regional climate change: a Bayesian approach to the analysis of multi-model ensembles. *J. Climate*. 18, 1524–1540.

Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., Da Costa Bechtold, V., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A.; Beljaars, A. C. M., van de Berg, L., Bidlot, J, Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., Mcnally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., Woollen, J., 2005: The ERA-40 re-analysis. *Quart. J. R. Meteorol. Soc.*, v.131, p. 2961-3012.