18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009
http://mssanz.org.au/modsim09

Statistical downscaling method of regional climate model results for hydrological modelling

J. Sennikovs [1] **and U. Bethers** [1]

[1] *Laboratory for Mathematical Modelling of Environmental and Technological Processes, Faculty of Physics and Mathematics, University of Latvia, Riga, Latvia*
Email: jsenniko@latnet.lv

Abstract: Reasonable and consistent meteorological input data is a crucial factor for modelling the river runoff at the catchment scale. The regional climate models (RCMs) provide sufficient information for the hydrological modelling of impact of expected climate change on the river runoff. However, one must avoid direct usage of RCM data for the forcing of hydrological models without analysing RCM compliance with observations for the reference period.

The aim of this study was to provide reasonable meteorological input data for the hydrological models to predict the river runoff changes in the future.

We considered the calculations made by the European RCMs organised in a database at the Danish Meteorological Institute under European Commission research project “PRUDENCE” EVK2-CT2001-00132 (prudence.dmi.dk). The spatial resolution of the analysed models is approximately 50 km with temporal resolution of 1 day. The set of 21 model runs was analysed. Each considered model run contained at least calculation for the climatic reference period (1961-1990) and model predictions for climatic scenarios A2 and B2 (2071-2100). The RCMs provide meteorological parameters at each grid point, which we supposed to use as a forcing for hydrological models.

We analysed performance of different RCMs by statistically comparing air temperature and precipitation rate with the observations over the Eastern Baltic area for the same period. The penalty function describing the deviation of each of the RCMs from the meteorological observations was constructed, aiming at evaluation of model accuracy in terms of monthly average temperature and precipitation, their monthly and interannual variation, and spatial distribution. Generally, all models reasonably represent the seasonal cycle of temperature, though they overestimate winter precipitation and underestimate summer precipitation in the study area.

We proposed a method of RCM data correction, based on shifting the occurrence distribution of an individual daily output (temperature or precipitation). Two cumulative distribution functions – one of the observed data, and one of the RCM data – were constructed for each day-of-the-year, for each parameter in each observation station. The correction function was constructed in a way to have equal probabilities of particular daily parameter for both observed and corrected RCM data. The correction functions were spatially interpolated, giving the possibility to create modified RCM data both for the reference period and future climate scenarios.

We analysed the performance of the method by comparing monthly statistical parameters of observed data versus corrected RCM data at a selected station. We show that statistical moments of distribution of temperature and precipitation were corrected by the present method. Interannual variability and temperature/precipitation correlation properties, however, cannot be significantly improved.

The proposed approach of RCM data modification allows changing the modelled temperature and precipitation time series for the reference period in such a way that they preserve the characteristics on a small time-scale, and at the same time also having the statistical properties of the observed data. The time series for the future climatic scenarios were obtained assuming that the histogram modification algorithm is the same for present and future climate. The hydrological modelling with the modified meteorological forcing has not been carried out in the present study.

Keywords: *regional climate model, statistical downscaling, bias correction, histogram equalization*

1. INTRODUCTION

General circulation models (GCMs) provide means of estimating climate change in the future by providing a time series of climatic variables. If GCM outputs are directly used for regional/local impact studies, there arise two main problems: (1) they are biased with respect to observations of present climate, and (2) the spatial scale (e.g. 300 km) is usually too coarse. The dynamic downscaling methods employ regional climate models (RCMs) using the outputs of GCMs as forcing and boundary conditions. Regional climate models provide time series of climatic variables on the smaller scale, however, usually, their output deviates from observations similarly to GCMs. The projections of future climate comparing it to present climate usually follow the so-called delta-change approach, when only the differences between present and future climate are considered. If time series of RCMs are used as an input for external (e.g. hydrological) models with non-linear response to the climate signal, then delta-change approach may fail [see the discussion in, for example, Graham et al. (2007)].

The aim of this study was to provide daily time series of temperature and precipitation that would be used as a forcing for hydrological models to predict the river runoff changes under climatic change conditions in the Eastern Baltic region. It is widely accepted (e.g. Feddersen and Andersen, 2005) that outputs from GCMs and RCMs cannot be directly used to force hydrological models without removing the biases. There exist statistical downscaling methods to correct GCM predictions relative to observed climate. Quantile mapping has been used to correct bias of monthly precipitation (Wood et al., 2004). The corrected daily precipitation was constructed from GCM output by adjusting cumulative distribution function (CDF) separately for each of the 12 calendar months (Ines and Hansen, 2006). In the present study, we proposed adjusting CDFs at the grid of meteorological observation stations by constructing CDFs for each day of the year in a moving time interval. We used the interpolation of differences of observed and modelled CDFs to the particular point of interest to obtain the daily time series of precipitation and temperature.

We considered a set of regional climate model runs for Europe from the PRUDENCE project (Christensen et al., 2007). The typical spatial scale of the models is around 50 km.

The layout of the paper is as follows. In Section 2, we analyze the performance of several RCMs with respect to observations of temperature and precipitation in the control period (1961-1990). The methodology to rank the models according to their skill of representing the present climate is described and the need for the bias correction is also shown. We describe the statistical bias correction method in Section 3. Section 4 is devoted to analysis of results of the bias correction method applied to output of four RCMs.

2. ANALYSIS OF PERFORMANCE OF REGIONAL CLIMATE MODELS

The particular model run (PRUDENCE term - experiment) was identified by the applied RCM and its driving data from global climate model. The model runs for control period (1961-1990) were considered for comparison. We identify the model runs by abbreviations used in the list of PRUDENCE project experiments (http://prudence.dmi.dk) in the form organization/model/run. The 10 considered RCMs are: HIRHAM, HadCM3, CHRM, CLM, REMO, RCAO, RegCM, RACMO, and Arpège. They use driving data from HadAM3H, ECHAM4/OPYC, ECHAM5, HadAM3P, and HadCM3. More details on the models and their performance for the present climate for Europe can be seen in Jacob et al. (2007), and about their projections of future climate changes in Christensen and Christensen (2007).

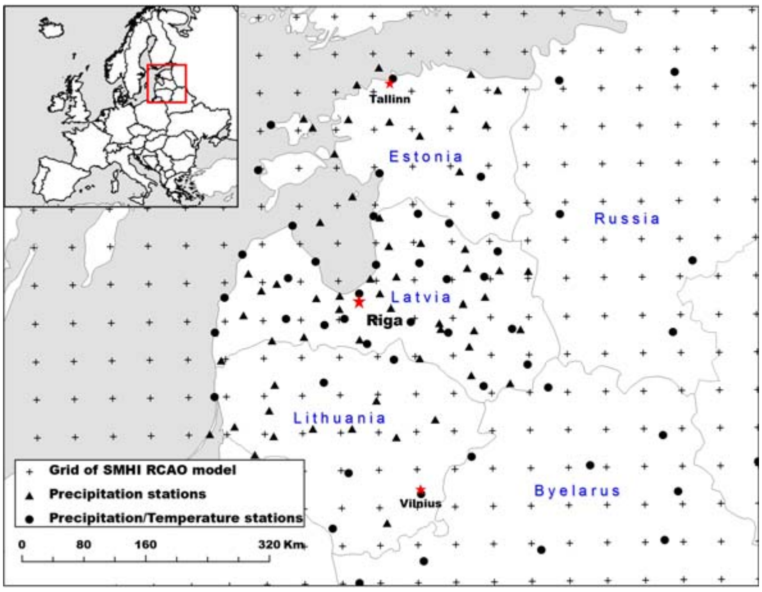

Figure 1. Location of considered region, observation stations and grid of RCAO model by SMHI.

Daily observations of temperature and precipitation for the time period 1961-1990 at 118 meteorological observation stations in the Eastern Baltic region were used for comparison between RCM control climate and observed climate. They were compiled mainly from two sources, namely, the Latvian Environmental, Geological and Meteorological Agency and the European Climate Assessment & Dataset. The geographical location of the considered region and positions of observation stations used for comparison are shown in Figure 1. We considered the region at the eastern coast of the Baltic Sea covering the territories of Latvia, Estonia, Lithuania and, partly, Russia and Belarus. The size of the considered domain is about 600 km x 600 km

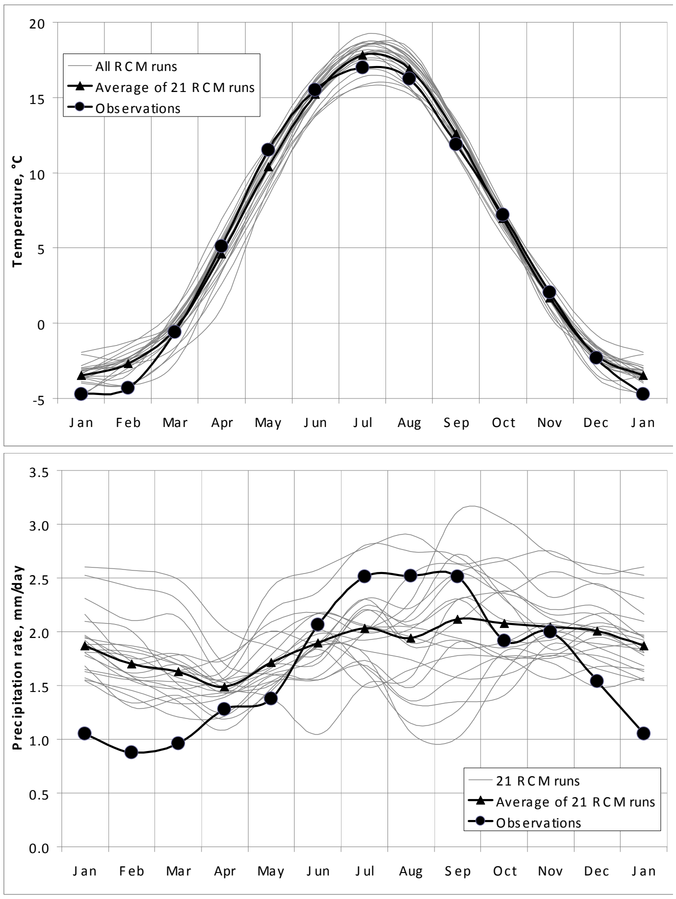

Figure 2. Monthly average temperature (top) and average mean precipitation (bottom) in Riga. Observations, calculations by RCMs, average values of all RCMs.

Figure 2 depicts the monthly average observed and calculated temperature and monthly average mean precipitation rate in a selected station (Riga). All of the models reasonably represent seasonal cycle of temperature, monthly average temperatures at Riga station deviate from the observed ones by -0.2 to 2.8ºC in January and from -1.2 to 2.2ºC in July. The root mean square error (RMSE) of monthly average temperature for different RCMs is between 0.9 and 1.6ºC, while maximum deviations of monthly average temperatures can reach 1.8 to 4.1ºC. RMSE of monthly mean precipitation rate at Riga station is between 0.45 and 0.95 mm/day that constitutes between 26% and 45% of yearly mean precipitation rate. Maximum deviations in a particular month are between 64% and 160% depending on the RCM. All models overestimate winter (December, January, February, and March) precipitation (in maximum by roughly 2 times). Most of them underestimate summer (July, August, September) precipitation by approximately 30%.

We used a penalty function to quantify the performance of the different models. This approach aimed at evaluating the control period performance of RCM in terms of monthly average temperature, precipitation, their monthly and interannual variation, and spatial distribution. It allows quantitative comparison and ranking of different models with respect to their ability to represent the observed control period climate.

Let i be the index of the particular RCM run, then the corresponding penalty function K_i is given by:

$$K_i = \frac{1}{4}\left(\frac{\Delta T_i}{\max\limits_i \Delta T_i} + \frac{\Delta p_i}{\max\limits_i \Delta p_i} + \frac{\Delta D_{T\,i}}{\max\limits_i \Delta D_{T\,i}} + \frac{\Delta CV_i}{\max\limits_i \Delta CV_i} \right).$$

The penalty function consists of four parts. Each part represents deviation of particular RCM run from observations with respect to (1) monthly mean temperature (ΔT_i), (2) monthly mean precipitation (Δp_i), (3) standard deviation of daily temperatures (ΔD_{Ti}), (4) coefficient of variation (CV - standard deviation of daily precipitation divided by mean precipitation) (ΔCV_i). Each of the parts was normalized to its maximum value among the model runs. We used the coefficient of variation of precipitation because standard deviation of precipitation has strong correlation with mean value, and, therefore would increase the weight of mean precipitation in the penalty function. The deviation of particular parameter (say T) took into account all months of the year (index m), all stations (index s) and were constructed as a sum of squares of differences between observed ($\overline{T}^*_{s,m}$) and calculated ($\overline{T}_{i,s,m}$) monthly average parameter as

$$\Delta T_i = \frac{1}{N}\sum_{s,m}\frac{\left(\overline{T}_{i,s,m} - \overline{T}^*_{s,m}\right)^2}{T_{\max}}, \quad T_{\max} = \max_{i,s,m}\left(\left(\overline{T}_{i,s,m} - \overline{T}^*_{s,m}\right)^2\right), \quad N = \sum_{s,m} 1 .$$

By constructing such a penalty function we measured the relative performance of a particular run among others. The weights of each of four parts were treated equal.

The values of $\overline{T}_{i,s,m}$ were obtained by spatially interpolating the daily time series of RCM output (given on the regular RCM grid) to the location of observation stations by geospatial kriging procedure and then calculating monthly averages. The deviations of each of the four parameters were normalized to their maximum values, therefore obtaining non-dimensional values that could be summed to form the penalty function K.

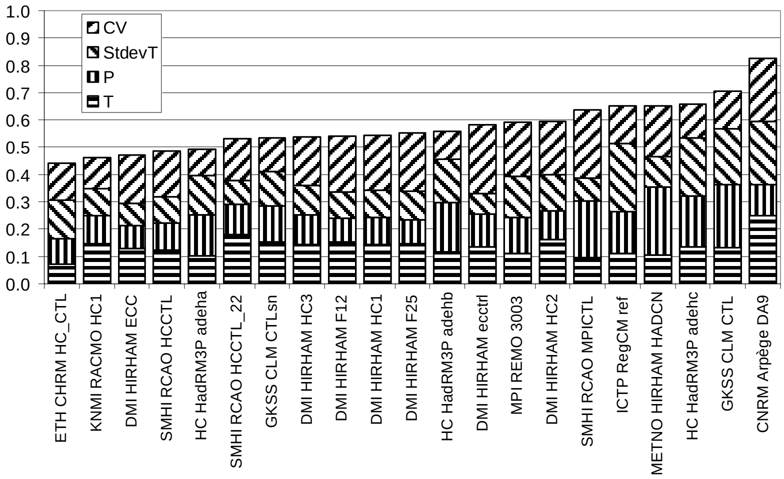

Figure 3. Penalty function and its components that characterize relative prediction skill of different RCM runs.

The values of penalty functions for all of the model runs are summarized in Figure 3. Correlation coefficients between model rank and the components of the penalty function – ΔT, ΔP, ΔD_T and ΔCV are, respectively, 0.23, 0.63, 0.59, and 0.33. This means that, on average, better skilled model runs predict also each of the components better, especially for mean precipitation and standard deviation of temperature, whilst regarding temperature and CV of precipitation, all of the model runs are more equal.

3. THE BIAS CORRECTION OF REGIONAL CLIMATE MODELS

The goal of bias correction was to obtain daily time series of modified temperature and precipitation at any point throughout the domain of our interest.

The bias correction method relies on changing of cumulative distribution function (CDF) of modelled data. We applied similar corrections separately for temperature and precipitation. Let x denotes the considered variable (temperature or precipitation), F(x) denotes the CDF of x. Then transformation that changes the particular daily value of RCM model run for control period (x_{ctl}) to modified (bias-corrected) value of it (x_{modctl}) at particular observation station is

$$x_{modctl} = F_{obs}^{-1}(F_{ctl}(x_{ctl}))$$

Here, F_{ctl} is CDF of x for RCM control period data (unmodified), and F_{obs}^{-1} is the inverse of observed CDF of x. Then at each observation station we constructed bias of x ($\Delta x(F) = x_{modctl}-x_{ctl}$) which is the function of cumulative probability F. Δx(F) can be spatially interpolated to any particular location. Therefore, to obtain modified value at any location we (1) interpolated the RCM time-series to this location and constructed CDF here, (2) interpolated values of Δx(F) to this location, and (3) transformed value of x_{ctl} as follows

$$x_{modctl} = x_{ctl} + \Delta x(F_{ctl}(x_{ctl}))$$

CDFs were constructed for each day of the year taking into account the values of x from this day and +/-5 days, e.g. from 330 values, thus obtaining CDFs in a moving time interval of 11 days. We expected smoother behaviour of the transformation functions by using overlapping intervals for the construction of CDFs, contrary to Ines and Hansen (2006). We sorted the array of values in ascending order and associated the probabilities accordingly. In the case of more than one equal value of x (e.g. daily precipitation equal to zero), we sorted equal values in the random order. The RCM data were produced on a 360 day year basis. We transformed it to the real dates by duplicating entries of each 72nd (normal year) or 60th (leap year) day of the year in the time series. If the value of modified precipitation after transformation was negative, we assumed it as zero.

The set of observation stations used for construction of CDFs are shown in Figure 1. Dense grid of precipitation measurements allows us to downscale outputs of RCM to smaller scale. The 11-day moving average precipitation is shown in Figure 4 for a selected location. The unmodified values differ significantly by their magnitude from both observations and modified values. On a scale smaller than 11 days, the character of the modified precipitation curve is similar to unmodified.

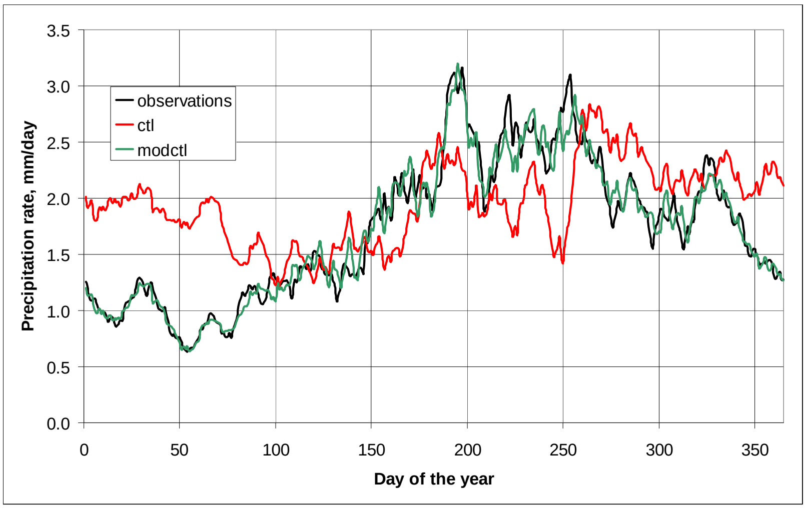

Figure 4. 11-day moving average precipitation rate. Observations, unmodified RCM calculations (SMHI RCAO HCTL) (ctl) and modified RCM calculations (modctl) are shown.

We applied transformation described above also to obtain the modified time series of temperature and precipitation in the future climate. The main assumption here was that the bias correction Δx(F) is the same for the same probability F both in the control period and in the future scenarios. Therefore, we used the following expression to obtain modified value of x ($x_{modscen}$) from RCM model value (x_{scen})

$$x_{modscen} = x_{scen} + \Delta x(F_{scen}(x_{scen}))$$

Here, the subscript *scen* denotes a particular future climate scenario.

4. ANALYSIS OF RESULTS OF BIAS CORRECTION

We analysed the performance of the bias correction procedure by comparing monthly average statistical parameters. The bias correction method, described above, modifies CDF of modelled climate, therefore, it is natural to expect that it will preserve all moments of the statistical distribution. We applied the method to correct output of the four most-skilled model runs, according to Section 2. In principle, this transformation can be used to correct output of any of the RCMs.

We checked the performance of bias correction by comparing monthly mean average temperatures, monthly mean precipitation rates, monthly standard deviations of daily average temperatures, and monthly coefficient of variation of daily precipitation rates. Interannual standard deviations of monthly average parameters were intended for evaluation of model representation of interannual variability of climate. Additional verified monthly parameters for precipitation are probability of dry days (precipitation rate smaller than 0.1 mm/day), average precipitation intensity (for days with precipitation), 95% and 98% percentiles.

Table 1. Performance parameters of bias correction for a particular observation station (Riga).

Parameter	Units	Abbreviation	KNMI	KNMI	DMI	DMI	ETH	ETH	SMHI	SMHI
			RACMO	RACMO	HIRHAM	HIRHAM	CHRM	CHRM	RCAO	RCAO
			HC1	HC1	ECC	ECC	HC_CTL	HC_CTL	HCCTL	HCCTL
			unmodified	modified	unmodified	modified	unmodified	modified	unmodified	modified
Temperature	°C	MBE	0.94	0.00	0.79	0.00	-0.43	0.00	0.61	-0.03
	°C	RMSE monthly average	1.35	0.04	1.03	0.06	1.24	0.03	0.94	0.10
	°C	RMSE STDEV	0.50	0.05	0.46	0.05	0.97	0.06	0.64	0.10
	°C	RMSE interannual	0.57	0.37	0.36	0.26	0.56	0.30	0.53	0.37
Precipitation	mm/day	MBE	0.00	0.00	-0.19	-0.01	0.16	-0.01	0.23	0.00
	mm/day	RMSE monthly average	0.45	0.06	0.51	0.04	0.52	0.06	0.54	0.05
	%	RMSE CV	36%	6%	52%	10%	38%	8%	44%	9%
	%	RMSE Interannual CV	15%	13%	12%	10%	12%	11%	9%	10%
	%	RMSE Pdry	17%	1%	23%	1%	21%	1%	22%	1%
	mm/day	RMSE intensity	1.08	0.12	1.82	0.09	1.16	0.13	1.19	0.13
	mm/day	RMSE 95% percentile	1.94	0.52	3.25	0.44	2.29	0.29	1.75	0.56
	mm/day	RMSE 98% percentile	3.08	0.81	5.64	0.81	3.33	0.46	3.08	0.74
	-	RMSE of T/P correl. coeff.	0.14	0.13	0.13	0.11	0.12	0.12	0.13	0.10

As a measure of agreement, we considered mean bias error (MBE) of monthly average temperature and precipitation, and RMSE of all the above-mentioned parameters. All RMS errors of particular parameter (x) were calculated as

$$RMSE_x = \left(\frac{1}{12} \sum_m (x_m^{obs} - x_m^{mod})^2 \right)^{0.5} .$$

As shown in Table 1, the MBE and RMSE of monthly average parameters were significantly reduced by the bias correction procedure. The errors in standard deviations (RMSE STDEV for temperature and RMSE CV for precipitation) are as well significantly smaller in modified model climate. The biases in probability of dry days, precipitation intensity and 95% and 98% percentiles of precipitation were reduced by the method. However, the errors in the interannual variability could not be significantly improved by this method, because this variability is inherent from the RCM.

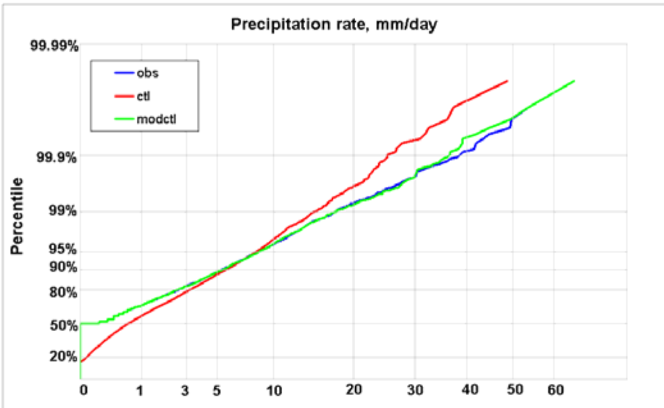

Figure 5. Yearly cumulative distribution functions of daily precipitation rate at Riga station representing observed (obs), unmodified (SMHI RCAO HCCTL) (ctl) and modified (modctl) climate.

The observed, unmodified and modified CDFs of precipitation are shown in Figure 5. There were too few dry days in unmodified climate, they were compensated by too many small intensity precipitation events (e.g. drizzle), the largest precipitation events (larger than approximately 8 mm/day) were underestimated. The CDF of modified RCM, as intended, were nearly equal to that of observed precipitation. There are some differences in histograms at very high percentiles (e.g. 99.9% and more).

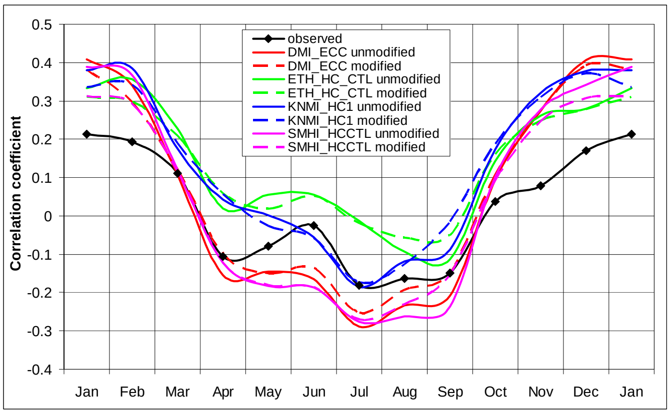

Figure 6. Correlation coefficients between temperature and precipitation. Results from the four most skilled models are depicted.

In the region considered, the snow accumulation in winter is one of the crucial factors determining the river runoff in spring. The hydrological models usually use a threshold temperature to recognize precipitation as either snow or rain. Therefore, correct temperature-precipitation correlation properties should be represented in the RCM output.

The monthly correlation coefficients between daily temperature and precipitation are depicted in Figure 6. Both modified and unmodified RCM output overestimated positive correlation between temperature and precipitation during autumn/winter (Nov-Feb) for all of the 4 considered models. Some of the models overestimated negative correlation during summer months. The agreement of temperature-precipitation correlation could not be significantly improved by the present

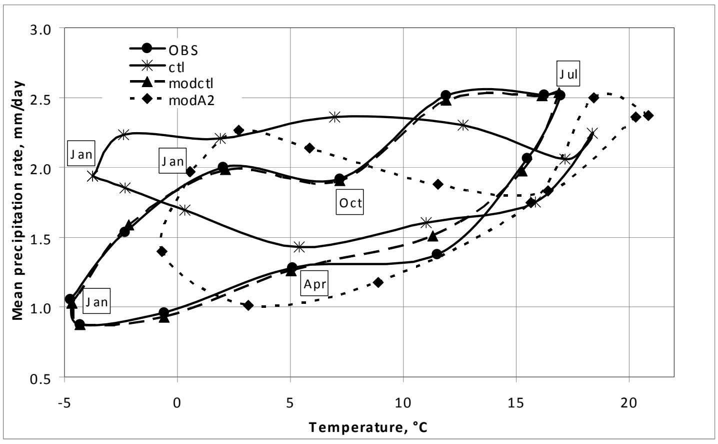

Figure 7. Temperature-precipitation diagrams for observed (obs), RCM control period (SMHI HCCTL, ctl), modified RCM control period (modctl), modified RCM A2 scenario (modA2).

method because the transformations of both of them are independent. Therefore, temperature-precipitation correlation properties were inherited from the RCM output. This means that, for instance, if large precipitation events in the RCM output are occurring mainly during the warmest periods in winter, then it will be the same in the bias corrected output, although average values of both temperature and precipitation will be corrected. The main consequence of this for the hydrological modelling in the area of study is underestimation of the snow percentage in precipitation during winter.

The monthly average temperature-precipitation diagrams representing observed climate, climate calculated by the selected RCM, modified RCM climate and modified climate of future A2 scenario are shown in Figure 7 for a selected location (Riga). Diagrams of modified control climate and observed climate nearly coincide. Diagram of unmodified climate differs significantly from observed one, especially with respect to precipitation. The future projected climate shows increase of temperature by around 4°C, the climate will become more continental in summer, more maritime during winter.

5. CONCLUSIONS

In the present paper, we described a method for ranking of the RCM model runs according to their relative performance. On average, better skilled models showed better performance on all of the considered components – temperature, precipitation, standard deviation of temperature, coefficient of variation of temperature.

The biases with respect to the present climate of even the best-performing model are too high for direct usage of their output to drive hydrological models.

The statistical bias correction method with daily cumulative distribution functions can remove biases, modifying the control period data obtained by the RCMs. We showed that statistical moments could be preserved during this procedure.

We used dense grid of observational stations to downscale the outputs of RCMs by interpolating differences of observed and calculated CDFs.

The interannual variability, as well as temperature-precipitation cross-correlation properties, cannot be significantly improved by such bias correction as they are inherited from the RCMs.

ACKNOWLEDGMENTS

The present work has been funded by the Latvian National Research Programme “Impact of the Climate Change on the Latvian Water Environment”.

The meteorological observation data in the Latvian meteorological stations used in this paper were provided by the Latvian Environmental, Geological and Meteorological Agency.

Regional climate model data have been provided through the PRUDENCE data archive, funded by the EU through contract EVK2-CT2001-00132.

REFERENCES

Christensen, J. H., Carter, T. R., Rummukainen, M. and Amanatidis, G. (2007), Evaluating the performance and utility of regional climate models: the PRUDENCE project. *Climatic Change* 81(Suppl.1), 1–6.

Christensen J.H., Christensen O.B. (2007), a summary of the PRUDENCE model projections of changes in European climate by the end of this century. *Climatic Change*, 81:7–30

Feddersen H., and U. Andersen (2005), A method for statistical downscaling of seasonal ensemble predictions. *Tellus*, 57A, 398-408

Graham LP, Hagemann S, Jaun S, Beniston M (2007) On interpreting hydrological change from regional climate models. *Climatic Change*, 81:97–122

Ines A. V. M., and J. W. Hansen (2006), Bias correction of daily GCM rainfall for crop simulation studies, *Agricultural and Forest Meteorology*, 138, 44-53.

Jacob D, Bärring L, Christensen OB, Christensen JH, de Castro M, Déqué M, Giorgi F, Hagemann S, Hirschi M, Jones R, Kjellström E, Lenderink G, Rockel B, Sánchez ES, Schär C, Seneviratne S, Somot S, van Ulden A, van den Hurk B (2007) An intercomparison of regional climate models for Europe: model performance in present-day climate. *Climatic Change*, 81:31-52

Wood, A. W., Leung, L. 5 R., Sridhar, V., and Lettenmaier, D. P (2004), Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs, *Climatic Change*, 62, 189–216.