Annals of Agrarian Science

Journal homepage: http://journals.org.ge/index.php

Evaluation of multi-model hindcasts for land surface air temperature over Georgia

T.Davitashvili[a], N.Kutaladze[b*], L.Megrelidze[b], G.Mikuchadze[b], G.Gogichaishvili[b], I.Samkharadze[c], R.Qvatadze[d]

[a]Ilia Vekua Institute of Applied Mathematics of Ivane Javakhishvili Tbilisi State University; 2, University Str., Tbilisi, 0186, Georgia

[b]National Environmental Agency of Ministry of Environmental Protection and Agriculture of Georgia; 150, David Agmashenebeli Ave., Tbilisi, 0112, Georgia

[c]Ivane Javakhishvili Tbilisi State University; 1, Chavchavadze Ave., Tbilisi, 0179, Georgia

[d]Georgian Research and Educational Networking Association GRENA; 10, Chovelidze Str., Tbilisi, 0108, Georgia

Received: 17 November 2019; accepted: 15 December 2019

A B S T R A C T

Present study evaluates hindcast over the Caucasus Region of the multi-model system, comprising from 4 ERA-Interim-driven regional climate models (RCM) and the high resolution GCM-MRI-AGCM3 of Meteorological Research Institute (MRI). In total, five climate models simulations were assessed against the CRU observational database. Present work focuses on the mean surface air temperature. The study shows the performance of the members of ensembles in representing the basic spatiotemporal patterns of the climate over the territory of Georgia for the period of 1991–2003. Different metrics covering from monthly and seasonal to annual time scales are analyzed over the region of interest: spatial patterns of seasonal mean, annual cycle of temperature, as well monthly mean temperature bias and inter annual variation. The results confirm the distinct capabilities of climate models in capturing the local features of the climatic conditions of the Caucasus Region. This work is in favor to select models with reasonable performance over the study region, based on which a high-resolution bias-adjusted climatic database can be established for future risk assessment and impact studies.

Keywords: Regional climate models (RCMs), Reanalysis, Domain, Hindcast, Gridded data, Boundary conditions.

*Corresponding author: Nato Kutaladze: E-mail address: cwlam08@gmail.com

Introduction

Assessing the impacts of the anticipated climate variations and change on regionally important sectors is growing in their importance. The primary tool for projecting climate are global climate models (GCMs) that are typically run at course horizontal resolutions, because of their massive computational and data storage requirements. GCMs' output cannot be directly used for impact study and requires downscaling on a finer scale, for which regional climate models (RCM) are developed. RCM data are essential for assessing the impact of climate change on different socio-economic sectors.

Climate model evaluation is a fundamental step in estimating the uncertainty in future climate projections. Because of the crucial role of climate models in this process, it is essential to characterize their strengths, weaknesses and uncertainties. Impact assessments typically start with climate data as inputs, and involve a series of calculations using multiple models through which information propagates hierarchically. Climate model evaluation is used not only for model development and improvements but also for assessing and correcting model biases. Model evaluations are also used to weight individual models in multi-model ensembles, alleviate the effect of model error on assessment models, and estimate the range of uncertainty in projected impacts.

Uncertainties in climate projections come from multiple sources including incomplete model formulations, future emissions scenarios, and how these are folded together with elements of human behavior, technological advancements and social/government structures. Among these, model errors are probably the most viable to characterize, and potentially remedy to reduce the uncertainty. Model evaluations are typically performed by comparing model outputs against reference data from observations or reanalysis using suitable metrics [1-3], and can be further used to guide model improvement and/or for bias correction [4]. Measuring model performance objectively is of a particular importance in the practice of applying climate model outputs to climate change impact assessments which employs bias correction and/or multi-model ensemble [5-9]. Previous studies attempt to identify a single parameter representative of overall model performance [10-12], that can be applicable to objective multi-model ensemble and/or bias correction. Caution must be exercised in such attempts and application of their results, however, because spatial and temporal variations in model performance can introduce a substantial amount of uncertainties in calculating such indices. GCM evaluations have been well-established [13-17], but collective and systematic evaluation of RCMs is much less mature [18-21]. Considering the importance of RCMs in studying climate change and assessing its impacts, it is critical to apply as much observational scrutiny as possible to RCMs.

Systematic multi-model RCM experimentations and observation-based evaluations are much less mature than those for GCM studies. The long history of GCM analyses for assessments and other climate variability issues has resulted in a mature process of model experimentation and evaluation [22-25]. Evaluation of the fidelity in simulating the present-day climate of multiple GCMs that have contributed to the archives of phase 3 of the Coupled Model Intercomparison Project (CMIP3) used for the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4). The CORDEX program was established as the first activity of the Task Force on Regional Climate Downscaling established by World Climate Research Program (WCRP). Common experimental designs in CORDEX are advantageous for many practical purposes including model evaluations, uncertainty assessments, and constructing multi-model ensemble (ENS).

This study evaluates the 4 RCM and a high resolution global model simulations over the Georgia using Regional Climate Model Evaluation System (RCMES). We include in this research 2 RCMs' (RegCM v 4.7.0 and WRF-ARW v3.9.1.1) simulations over the domain centered to Georgia performed by us and 2 simulations over different domains from the CORDEX program. Namely RCA4 over MENA (Middle East and North Africa) and HadRM3P over CAS (central Asian domain). Such a choice has resulted in the fact that only these two domains overlaps our target area and the evolutionary simulations are available only for these 2 models on ESGF (Earth System Grid Federation)-CORDEX archive. High resolution GCM - MRI-AGCM3.2 output was provided from the Meteorological Research Institute of Japan Meteorology agency.

Section 1 provides details of the experimental design including the evaluation domain, RCMs, reference datasets, section 2 – climate description of Country, section 3 provides details of the Regional Climate Model Evaluation System (RCMES) used in the model evaluation. Section 4 presents the evaluation of RCM skill in simulating the targeted variables and examines the uncertainties in model evaluation related with reference data. Results are summarized in Sect. 5.

Main part

1. Data. In this paper, we used several data archives, most of them are available from the federative ESGF infrastructure, including Coordinated Regional climate Downscaling Experiment (CORDEX). We downloaded CORDEX simulations over central Asia (CAS) and the Middle East and North Africa (MENA) domains, covering South Caucasus territory.

Hindcast from the high-resolution atmospheric general circulation model of the Meteorological Research Institute (MRI) and two our simulations from two regional models RegCM4 and WRF with different configuration over the same domain (centered to South Caucasus territory) and with the same resolution have also been evaluated.

As for observations, we used global gridded observations and reanalysis of 2-meter air temperature data.

1.1. Observational data. For validation of individual models, also for ensemble the gridded global data set of the Climate Research Unit (CRU) was used. The monthly CRU data represent one of the most comprehensive observational data sets available and have been widely used for former studies. These observational data sets are based on statistical interpolation methods, which are a gridded time-series dataset. We used latest version TS 4.03, released 15 May 2019, covering the period 1901-2018 Coverage: All land areas (excluding Antarctica) at 0.5° resolution for six Variables [26].

Despite the existing studies reveals that there are clear biases between the observed surface air temperature and the (GHCN + CAMS) data sets, they vary in space and seasons. We decide to compare this two gridded temperature dataset over Georgia. This data set at 0.5 X 0.5 latitude-longitude resolution is different from some existing surface air temperature data sets in: (1) using a combination of two large individual data sets of station observations collected from the Global Historical Climatology Network version 2 and the Climate Anomaly Monitoring System (GHCN + CAMS) [27].

1.2. Reanalysis data. As all Regional climate models are forced with ERA-Interim - reanalysis of the global atmosphere dataset, compatibility between reanalysis and observations (CRU) was also established.

The main objectives of the ERA-Interim project were to improve on certain key aspects of ERA-40 (previous version of ERA-Interim), such as the representation of the hydrological cycle, the quality of the stratospheric circulation, and the handling of biases and changes in the observing system. These objectives have been largely achieved as a result of combination of factors, including many model improvements, the use of 4-dimensional variation analysis, a revised humidity analysis, the use of variation bias correction for satellite data and other improvements in data handling.

The ERA-Interim atmospheric model and reanalysis system uses cycle 31r2 of ECMWF's Integrated Forecast System (IFS), configured for the spatial resolution - T255 spherical-harmonic representation for the basic dynamical fields and a reduced Gaussian grid with approximately uniform 79 km spacing for surface and other grid-point fields [28-29].

2. Climate models used in the study. Regional climate models (RCMs) are useful tools for the projection of climate change on regional scales. Unlike GCMs, the model domain of an RCM does not cover the entire globe. It is restricted to a certain area of regional scale.

This restriction allows for long-term simulations with higher resolutions. On the other hand, this implies that information about the lateral and lower boundary conditions (LBCs) has to be provided. These LBCs can be derived from GCM simulations or from observational data sets (usually reanalysis products).

RCA4. Since 1997 the Rossby Centre has developed an international standing in the field of regional climate modelling with the development of the atmospheric model RCA, at SMHI.

RCA is based upon the numerical weather prediction (NWP) model HIRLAM. The RCA4 dynamical core is a two time-level, semi-Lagrangian, semi-implicit scheme with six-order horizontal diffusion applied to the prognostic variables. Grid boxes in RCA4 can include fractions of sea (with fractional ice cover) or lake (with ice or not) and land. The land fraction can be further subdivided into forest and open land, where both can be partly snow covered. Each sub-grid scale tile has a separate energy balance equation and individual prognostic surface temperatures.

The model was driven by European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim reanalysis data to run the CORDEX Evaluation experiment, representative of the period from 1981 to 2010, over the Middle East and North Africa (MENA) domain called CORDEX-MENA [30-31].

HadRM3P is Hadley's Regional limited-area regional climate model widely used worldwide as part of the PRECIS (Providing Regional Climates for Impacts Studies) system, which was developed at the Hadley Centre of the United Kingdom Met Office. HadAM3P is a grid-point model which solves equations of motion, radiative transfer and dynamics explicitly on the same scale as the grid. The atmospheric equations are a quasi-hydrostatic version of the primitive equations with full representation of the Coriolis force. Other, mostly thermodynamic, processes that occur at the sub-grid-scale are represented by physical parametrizations. Model has 0.44 x 0.44 degrees' resolution with a rotated pole to achieve approx. 50 km x 50 km resolution on 19 levels (used for the EU region, South Asia, planned for East Asia). Also used is a double resolution variant at 0.22 x 0.22 degrees (Western US, planned for EU and Africa).

The model was driven by European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim reanalysis data to run the CORDEX Evaluation experiment, representative of the period from 1990 to 2011, over Central Asia domain with 0.44-degree resolution (CAS-44 domain) [32, 33].

RegCM v 4.7.0. Regional Climate Model version 4 - RegCM model is limited-area regional climate model. It uses the radiation scheme of the NCAR CCM3, the cloud scattering and absorption parameterization follow that of *Slingo* (1989), whereby the optical properties of the cloud droplets (extinction optical depth, single scattering albedo, and asymmetry parameter) are expressed in terms of the cloud liquid water content and an effective droplet radius. The soil hydrology calculations include predictive equations for the water content of the soil layers.

Compared to previous versions, RegCM4 includes new land surface, planetary boundary layer, and air-sea flux schemes, a mixed convection and tropical band configuration, modifications to the pre-existing radiative transfer and boundary layer schemes, and a full upgrade of the model code towards improved flexibility, portability and user friendliness. The model can be interactively coupled to a 1D lake model, a simplified aerosol scheme (including organic carbon, black carbon, SO4, dust and sea spray) and a gas phase chemistry module (CBM-Z) [34-36].

WRF-ARW v3.9.1.1 - Weather Research and forecasting model (Grell scheme). The WRF model is a state-of-the-art, next-generation mesoscale numerical weather prediction system designed to serve both operational forecasting and atmospheric research needs (http://www.wrf-model.org). It is a non-hydrostatic model, with several available dynamic cores as well as many different choices for physical parameterizations suitable for a broad spectrum of applications across scales ranging from meters to thousands of kilometers. The physics package includes microphysics, cumulus parameterization, planetary boundary layer (PBL), land surface models (LSM), longwave and shortwave radiation.

The WRF model is a state-of-the-art, next-generation mesoscale numerical weather prediction system designed to serve both operational forecasting and atmospheric research needs (http://www.wrf-model.org). It is a non-hydrostatic model, with several available dynamic cores as well as many different choices for physical parameterizations suitable for a broad spectrum of applications across scales ranging from meters to thousands of kilometers. The dynamic cores in WRF include a fully mass- and scalar-conserving flux form mass coordinate version. The soil scheme solves the thermal diffusivity equation using five soil layers and the energy budget includes radiation, sensible, and latent heat fluxes. It treats the snow-cover, soil moisture as fixed quantities with a land use and season-dependent constant value. The terrain, land use and soil data are interpolated to the model grids from USGS global elevation, 24 category USGS vegetation data and 17 category FAO soil data with suitable spatial resolution (arc 5 minutes) to define the lower boundary conditions [37-38].

MRI-AGCM3. A new version of the atmospheric general circulation model of the Meteorological Research Institute (MRI), with a horizontal grid size of about 20 km, has been developed. The previous version of the 20-km model, MRIAGCM3, which was developed from an operational numerical weather-prediction model, provided information on possible climate change induced by global warming, including future changes in tropical cyclones, the East Asian monsoon, extreme events, and blockings. For the new version, MRI-AGCM3.2, various new parameterization schemes have been introduced that improve the model climate. Using the new model, a present-day climate experiment has been performed using observed sea surface temperature. The model shows improvements in simulating heavy monthly-mean precipitation around the tropical Western Pacific, the global distribution of tropical cyclones, the seasonal march of East Asian summer monsoon, and blockings in the Pacific [39, 40].

2. Country climatology. Georgia's location on the northern edge of the subtropical zone between the Black and Caspian seas, on the one hand and on the other hand, complexity of its special topography determines the variety of climate conditions. Local climate creates Black Sea and the Caucasus. The last protects Georgia from direct invasion of cold air masses from the North and the Black Sea makes moderate temperature fluctuations and contributes large amount of precipitation, especially in western Country. Very important is the Likhi Range, running from the North to the South and dividing the Country into its eastern and western parts, with quite different climatic pictures.

Temperature regime of the territory, as a climate in whole, is characterized by a range of peculiarities stipulated mainly by the geographic location of Georgia, complex relief of the occupied territory, radiation pattern and prevailing atmospheric circulation processes. As orography on the territory varies from 0÷5068 m correspondently mean annual temperature fluctuate in the range of -5, +15^{0}C approximately.

Table 1. *Summary of data used in this study*

Data set/version	Time range	Resolution	References
Observation & reanalysis			
CRUvTS 4.03	1901-2018	0.5^0 X 0.5^0	Climate Research Unit (CRU)
GHCN + CAMS	1948- to near present	0.5^0 X 0.5^0	NOAA/OAR/ESRL PSD
ERA-Interim	1979- to near present	79 km	ECMWF
Climate models			
SMHI-RCA4	1981-2010	0.22^0 X 0.22^0	SMHI/Rossby Centre
HadRM3P	1990-2011	0.44^0 X 0.44^0	Met Office
RegCM v 4.7.0	1985-2015	15 km	ICTP
WRF-ARWv3.9.1.1	1985-2015	15 km	NCAR/NCEP
MRI-AGCM3	1979 -2003	20 km	MRI -JMA

Due to the Country climate regime, territory was divided in 8 sub-regions to examine the simulation performance across the experiments on different sub-regions. These regions mostly cover Georgia's territory but also include some other parts, according to the factors of local climate formation. On Fig.1. location and names of sub-regions are presented, where R01, R02 and R03 are, respectively, Western, Central and Eastern parts of Greater Caucasus mountains, R04 - Kolkheti Lowlands, R05 - Central part of Georgia including Likhi range, R06 – Adjara Black Sea coastal zone with adjacent mountains, R07 - Lesser Caucasus mountains, R07 - Eastern Country plane territory.

3. The Regional Climate Model Evaluation System (RCMES). RCMES is an open, publicly accessible process enabled by leveraging the Apache Software Foundation's OSS library, Apache Open Climate Workbench (OCW). RCMES provides datasets and tools to assess the quantitative strengths and weakness of climate models, typically under present climate conditions for which we have observations for comparison, which then forms a basis to quantify our understanding of model uncertainties in future projections.

RCMES is composed of two main components, the Regional Climate Model Evaluation Database (RCMED) and the Regional Climate Model Evaluation Toolkit (RCMET). RCMED can reside on a single server or be distributed on multiple servers to allow efficient data management and sharing while reducing the hardware and software burdens for handling the data storage and traffic. Observed data are fundamental to model evaluation. The demands from higher resolution and multivariate evaluation make the scientific and logistical process of model evaluation ever more challenging. RCMED bringing together massive amounts of observational and model data, but also dealing with the wide variety of sources and formats of data, necessitating significant investments in computer and personnel resources to transfer, decode, (re)format, (re)archive, and analyze the data. Such steps can make the process of performing robust model evaluations extremely difficult and time consuming even for highly trained scientists.

RCMET includes a software suite for calculating statistical metrics popularly used in model evaluations and visualizations. Model-evaluation metrics and visualization generally vary widely according to users and targets; RCMET includes the capability to incorporate user-defined metrics as well as pathways to extract partially processed data (e.g., both model and reference data regridded onto a common grid) so that users can do their own specific data processing and visualizations.

4. Results. The spatial distribution and annual cycle of mean monthly temperature along with the bias averaged over the entire analysis region and eight sub-regions (Fig.1), selected to examine climate models skills across varied geographical landscape are presented in this work. The analysis focuses on how the model simulates surface climate (temperature) in response to the large-scale forcing imposed by the ERA-Interim reanalysis and by local topographical features. All the analysis presented here is carried out over the interior domain to eliminate the buffer zone where the direct effect of the lateral boundary conditions is maximum.

4.1. Evaluation metrics. Different metrics have been used in order to represent the performance of climate models in simulating climatic conditions. Besides computing the mean bias and root mean square error (RMSE), the degree of statistical similarity between two climatic fields was quantified in the form of normalized Taylor diagrams that can be considered as the combination of different measures such as the centered (or bias removed) RMSE, spatial standard deviation (STD), and spatial correlation. The Taylor diagrams reported in the present study are based on 13-yr annual and seasonal means in grid points.

4.2. Uncertainties assessment. The accuracy of reference data is among the most important concerns in model evaluation. All observations and/or analyses include errors of unknown/estimated magnitudes; e.g., analyses based on surface station data are directly affected by local station density. This especially true for the Caucasus region in which station density varies substantially according to regions. Uncertainties in model evaluation originating from reference data are examined using three different reference datasets. In addition to CRU, ERA-Interim (ECMWF Re-Analysis) and the GHCN_CAMS (Global Historical Climatology Network v2 and the Climate Anomaly Monitoring System) are selected for the same period as models evaluation was performed. All CMs yield higher spatial correlations with the CRU than GHCN_CAMS and ERA-Interim. The standardized deviations and RMSE are smaller against CRU and GHCN_CAMS, i.e. the spatial variability of the ERA-Interim data is larger than other two datasets (Fig.2).

Fig.3 shows the spatial distribution of mean annual temperature biases averaged over the entire 13-yr period compared to the CRU, GHCN and ERAINT datasets. In all seasons (not shown) the temperature bias against CRU data ranges between -4° and 4°C over the most of domain, except in winter and for annual means, when the bias mostly ranges between -3° and 3°C. Differences between models and ERAINT and GHCN_CAMS data range between -4° and 4°C in summer and for annual biases, but in the rest of seasons it is increased up to -6° and 6° C. As for GHCN_CAMS dataset, all models deviate from observation in the range -4° and 4°C, increasing to -6° and 6°C in winter season.

Best fit was revealed with CRU, biases of models are the smallest, but there are some systematically occurred features in the spatial distribution of these differences, i.e. relative to CRU and GHCN_CAMS it is noticeable cold bias over lowlands and plain territory and warm bias is the most evident in summer and winter over Caucasus. As for ERAINT, in contrary, for all simulations cold bias is occurred over mountainous areas, especially over Greater Caucasus range, warm bias – over lowlands and plain territory, that is the most evident in summer. Although evaluation of models against all three observation datasets demonstrates the spatial features of temperature biases and bias pattern is comparable with the terrain profile. It must be noticed, that moving towards the originally higher resolution information, the finer the details are in the spatial distribution of the seasonal temperature fields and the spatial features of deviations of mean seasonal temperature fields of ensemble simulations even reveal the ranges of Likhi Mountains (dividing west and east Country), which is in fact is especially clearly seen in winter season for all simulations.

The differences between the temperature evaluations based on the three observation datasets, may have resulted from the difference in the observational platform and methodologies. This examination shows that, quality control and cross-examination of reference datasets are important for model evaluations.

4.3. Evaluation results. In this study the baseline evaluation of the mean surface air temperature is presented against CRU dataset. As it was already mentioned, the most noticeable feature is the general warm bias over the Greater and Lesser Caucasus mountains and cold bias in the lowlands and plain territory (Fig.3). The spatial patterns of cold biases for all simulations except AGCM3 are similar, with the largest magnitudes being located in west Georgia lowlands. However, the cold biases in the WRFC and RegCM4 simulations are generally larger and extended over east Georgia plain territory. The warm bias is found in all simulations except RCA4. RCA4 is an outlier among these five CMs in the sense that it generates general cold biases over almost the entire study area. Another difference between the five simulations is that in the HaDRM3P simulation warm biases of more than 2°C are largely confined to Greater and Lesser Caucasus highlands, reaching its maximum magnitude up to 3°C in DJF season, when this warm bias clearly depicted in other three simulations (except RCA4) as well. On the contrary, the only GCM (AGCM3) (Fig.3) generates overall warm biases. Unlike in the AGCM3 simulation, annual bias is the smallest (±1°C) but without clear dependence on topography, only in winter months there is evident positive (up to 3°C) deviation area covering mountainous part of the Country.

As it seems all RCMs except RCA4, overestimate the surface temperatures over the high elevation regions and underestimate low elevations resulting the least deviated ENS results regarding to observation in the range ±1.5°C.

Overall, all models simulate the spatial variations in the annual mean temperatures over Georgia with the spatial pattern correlation coefficients between 0.95 and 0.99 and standardized deviations (the spatial standard deviation of the simulated surface air temperature normalized by that of the observed data) of 0.8–1.15 with respect to the CRU data, except WRFC with much lower STD of 0.65 (Fig.2). Fig.2 also shows that the multi-model ensemble mean (ENS), along with AGCM3, yield the smallest RMSE.

Comparison of the simulated annual cycle against the CRU analysis for the sub-regions shows that the multi-model ensemble is generally in well agreement with observed climatology in these regions and all five simulations have almost identical annual cycle and a similar range in monthly mean temperatures averaged over sub-regions, with differences up to 5°C between separate models. However, despite the reasonable performances, model biases vary noticeably according to regions and seasons (Fig.4).

Fig.4 shows some time dependency of model deviations, as temperature biases are not constant in time. They have a more or less clear annual cycle: there is one of five CMs (RCA4) with a constant negative temperature bias through the entire year, for other four models temperature is generally overestimated in winter (exceeding 4°C in January and February), whilst underestimated to a varying extent in the rest of the year resulting ensemble simulations negative bias. Therefore, the seasonal variation in the magnitude of the bias in area-average temperature means that the ENS simulation has a less extreme annual cycle than the annual cycle of the observations. In the transient seasons (spring, autumn), all regions of the study territory have a cold bias. This appears to be largest over the west Georgia lowlands. Warm bias in area-mean temperature is greatest during winter. In this season, warm biases extend over entire mountainous regions including the Greater and Lesser Caucasus. This may be related to the simulation of cold-season snowpack in the high-elevation regions and/or the lack of resolutions both in model simulations and the CRU data, suitable for representing the large orographic variations and associated variations in surface temperature in the mountainous region.

Fig.5 presents the normalized biases and interannual variability in terms of the percentage of the temporal standard deviations of the CRU data over the 13-yr period, of the simulated surface air temperatures in the eight sub-regions during each season. The temporal standard variations are adopted as the measure of the interannual variability. The scaled model bias shows that the warm bias over the Caucasus mountains is common for nearly all models (except RCA4) only in winter; RCA4, which generates quite strong (by 50%–150% of the observed interannual variability) cold biases over the region in winter, is the only exception. As for cold bias on the intermountain low elevation area negative bias is systematic regardless of models and seasons for west Georgia and more evident in transient seasons for east Country plains.

Models skill in simulating the interannual variability of the seasonal temperature is further examined using RMSE and the temporal correlation coefficients between the simulated and CRU data over the 13-yr period. The resulting RMSE (Fig.5) generally exceeds the interannual variability of the CRU data (i.e., normalized RMSE>100%), especially during autumn. In spring and summer clear overestimation is also revealed. For winter, the RMSE varies according to models in most regions; the normalized RMSE for RegCM4, as well as the multi-model ensembles, is less than 100% while that for AGCM3 is well above 150% for all sub-regions. As for ENS, it yields the smallest RMSE in summer and winter due to opposite signs of deviations for different models, but for annual means and in transient seasons because of mostly underestimation is evident, ENS RMSE is greater than for separate models, that are AGCM3 or HaDRM3P having positive bias.

The spread of bias fields mostly ranges between –3°C and +3°C, only AGCM3, HaDRM3P (overestimation), and WRFC (underestimation) models are slightly exceeding these limits in summer. WRFC and RCA4 typically show a strong cold bias when compared to the CRU observational dataset. In general, RegCM4 and AGCM3, having higher initial (before regridding) resolution, performs among the best climate models: i.e., producing close to zero mean annual bias due to the least biased performance during the period of March-August. Again, higher resolution simulations (RegCM4, AGCM3) are not expected to decrease the mean bias fields, and actually the standard deviation of bias averaged over the region of interest in each season is larger in case of RegCM4 and AGCM3 compared to the ensemble (Fig.5). The wide range of the spread in seasonal biases can be directly attributed to the different topography and parameterizations implemented in the evaluated climate models simulations.

Overall, models show consistently better skill in simulating the monthly-mean surface air temperature in the cold period (September-February) than for warm period (March-August) of the year (Fig.6).

The model biases also vary systematically according to regions. For spring, the most noticeable systematic biases are the cold bias in the entire western part including and no systematic warm bias revealed in this season. For autumn, the most systematic biases are the cold bias in the central mountainous regions including Likhi Range and South Caucasus highlands. In winter warm bias is evident relative to other seasons and this bias varies closely with orography as shown in Fig.5. This feature of orography dependence bias is noticeable during whole year but most evident in winter. The evaluation of the temporal standard deviation, a surrogate for the interannual variability, shows that all models perform reasonably well in simulating the interannual variability of winter and summer temperatures for all sub-regions. Most of CMs overestimate the interannual variability of the transient season's temperatures; overestimation is greatest for RCA4 in spring and AGCM3 in autumn. For all seasons, ensemble simulations have the least STD.

The correlation coefficients between the simulated and CRU time series (Fig.5) also shows that climate models examined in this study generally perform better in simulating the phase of the interannual variation in the surface air temperatures during winter than in other seasons, the poorest correlation was found in spring. In contrary with annual correlations, overview of seasonal means revealed that AGCM3, the only GCM in this study, almost not correlated with observation, whilst all RCMs has a quite high score as for annual, as well for seasonal means.

5. Discussion and Conclusion. In the present study, five climate models have been evaluated over a 13-yr reference period (1991−2003) against the CRU observational dataset. Overarching aim of the present study is to provide useful information on general capabilities of given models in reproducing climatic conditions over the Caucasus Region. By and large, the annual temperature cycle averaged over the study region is well represented by ensembles simulation. According to the spatial distribution of seasonal temperature, models performing well for annual temperature do not necessarily perform well in separate seasons and model performance varies widely and, often systematically, according to regions and seasons. These characteristics in model errors make it difficult to design a set of model weightings that can be universally applied to the construction of multi-model ensemble.

According to the findings reported in the present work, the following considerations can be made: (1) there is not a single model outperforming the other ones in all aspects, but it is also important to note that all models have their strength and weaknesses; (2) higher resolution simulations may adequately resolve the temperature variations in the region; (3) due to the amplification of biases or the increased internal variability on small scales induced by strong local surface heterogeneities within the regional domain, higher resolution simulations not necessarily reduce the uncertainties; (4) model performances are also influenced by observational uncertainties.

We assess that the model can provide useful information on variables that are important for the assessment of climate change impacts. We therefore plan to use this model configuration in simulations of other essential climate variables and construction future climate scenarios for Caucasus region.

Appendix:

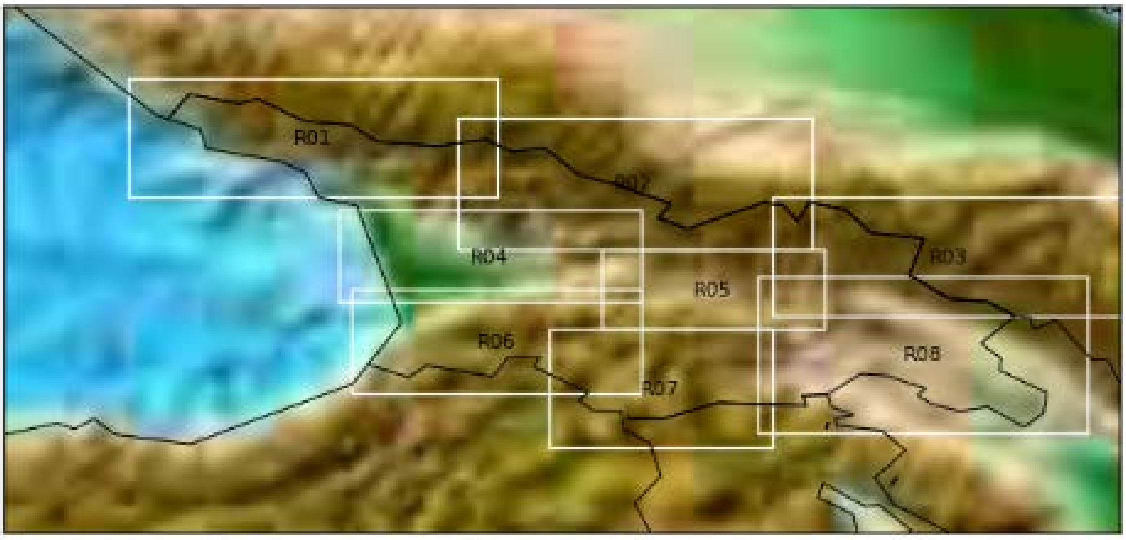

Fig. 1. *Study (Caucasus) domain. The color contours represent the terrain elevation. The numbered boxes with white boundaries indicate the eight sub-regions in which the area-mean time series are evaluated.*

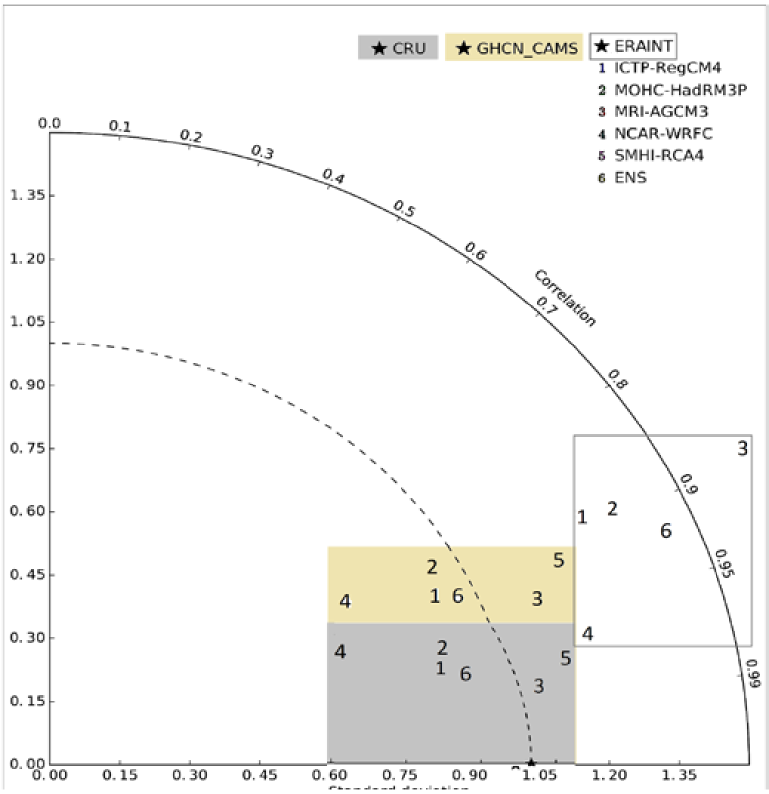

Fig. 2. *Evaluation of the simulated temperature climatology over the land using three different reference datasets. The dots on grey (CRU), yellow (GHCN_CAMS) and squared (ERAINT) backgrounds, respectively, indicate the model ensemble evaluated against different reference data.*

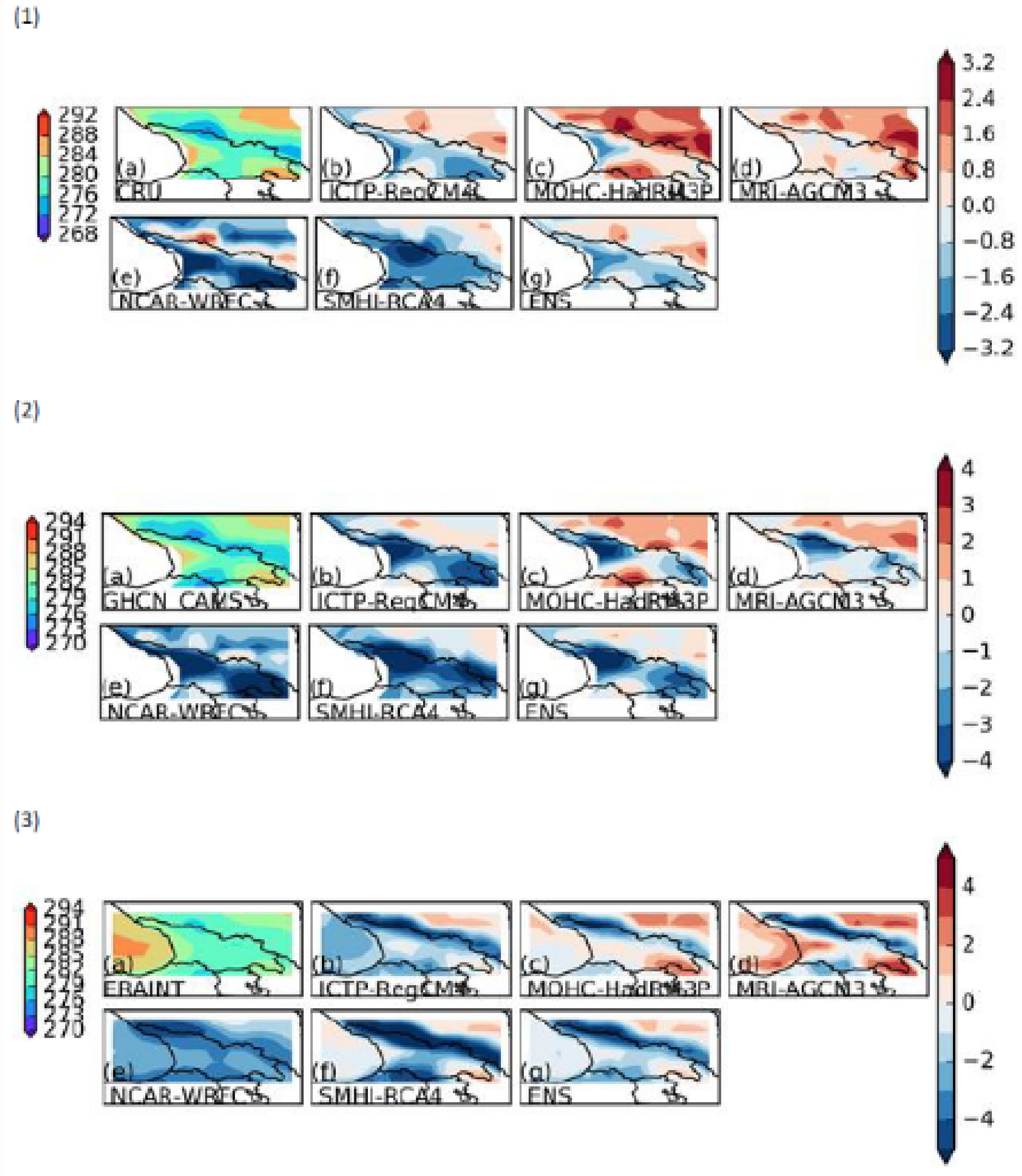

Fig. 3. *Annual-mean surface air temperature (°C) from the CRU (1), GHCN_CAMS (2) and ERAINT (3) analysis. The biases (°C) from the reference data for (b)–(f) the individual models and (g) the multi-model ensemble (ENS).*

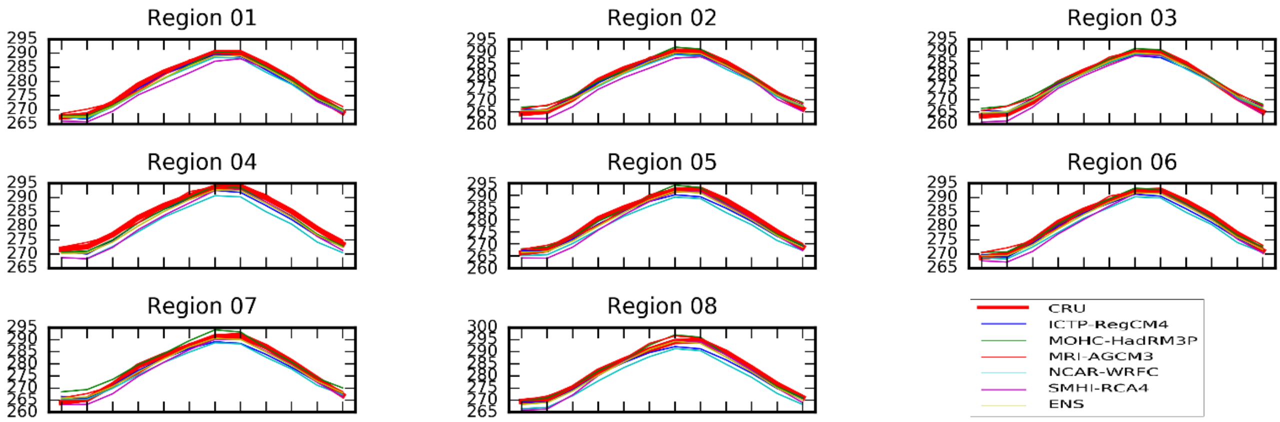

Fig. 4. *Simulated and observed (CRU, thick red) temperature annual cycle (°C) for the eight sub-regions. The thin yellow line indicates the multi-model ensemble temperature.*

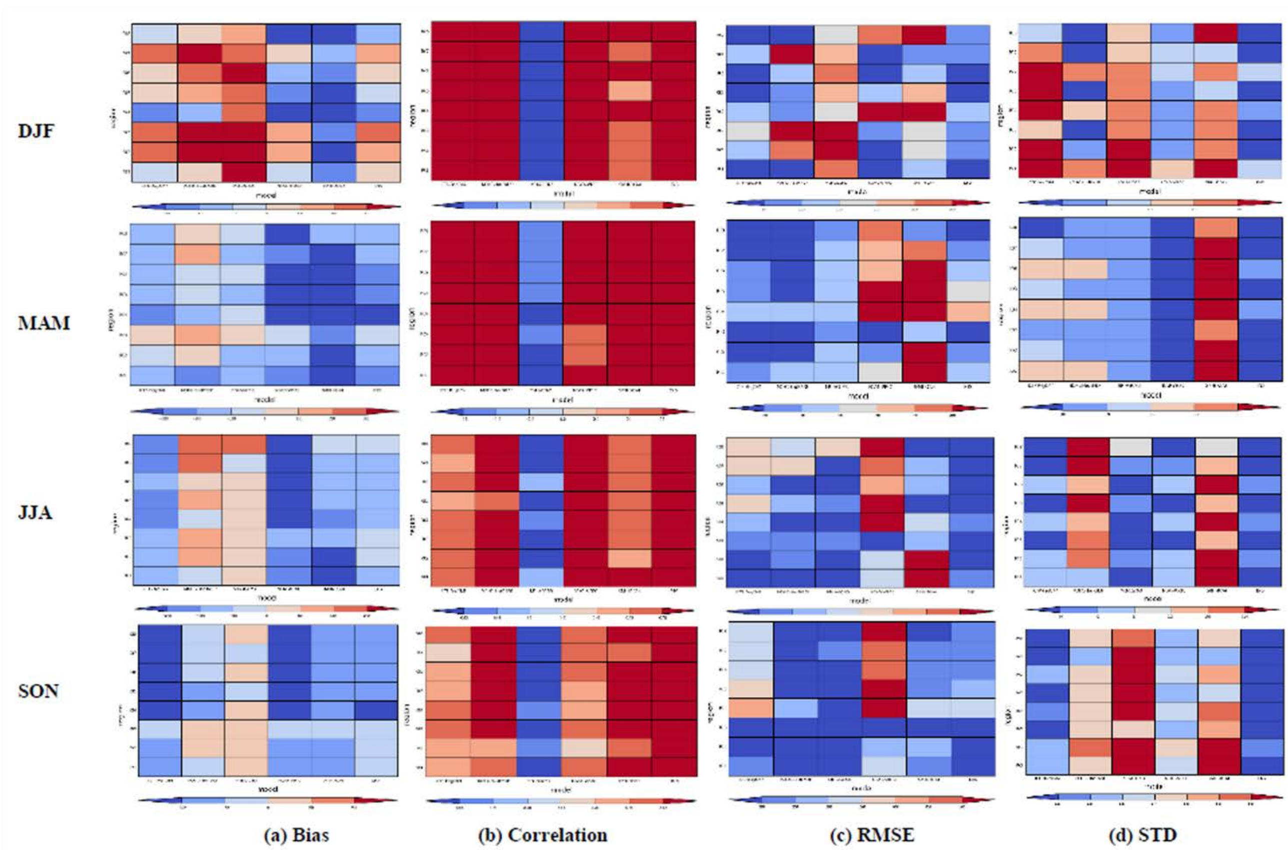

Fig. 5. *Regional (a) bias, (b) temporal correlation coefficients, (c) root mean square error and (d) temporal standard deviation of simulated average seasonal air temperatures relative to CRU observations. Seasons are defined as follows: winter-DJF (December–February), spring-MAM (March–May), summer-JJA (June–August) and autumn-SON (September–November). The bias, standard deviation, and RMSE are normalized by the standard deviation of the CRU data.*

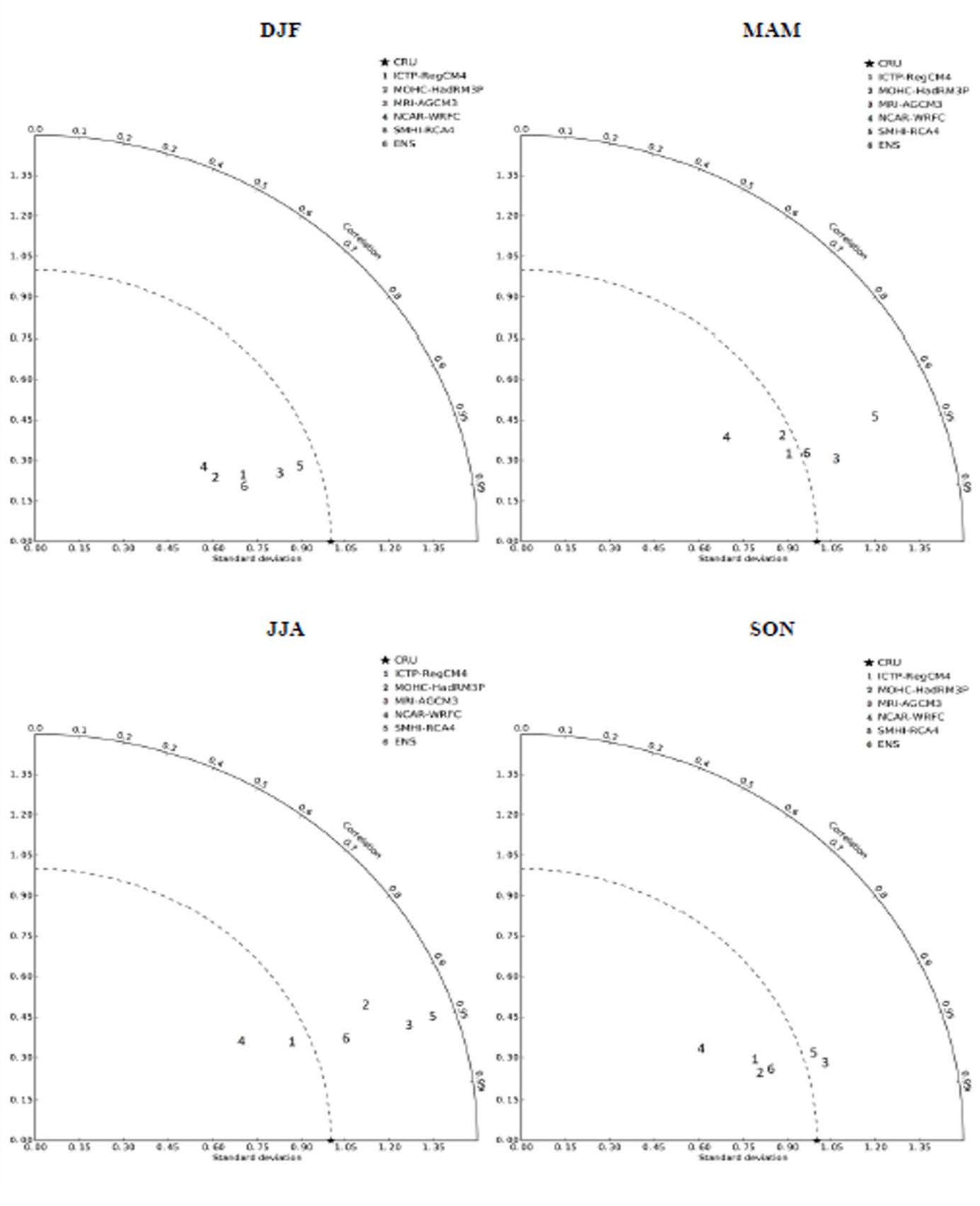

Fig. 6. *Standardized deviations and spatial pattern correlations of temperature between the CRU data and the individual model results for separate seasons over the land surface. Seasons are defined as follows: winter-DJF (December–February), spring-MAM (March–May), summer-JJA (June–August) and autumn-SON (September–November).*

Acknowledgements

The research was funded by the Shota Rustaveli National Scientific Foundation Grant #FR17_548.

We are thankful to our colleagues from Meteorological Research Institute (MRI) of Japan Meteorological Agency who provided AGCM3.2 model outputs that gretly assisted the research.

References

[1] J. Kim, J. E. Lee, Amulti-year regional climate hindcast for the western U.S. using the Mesoscale Atmospheric Simulation (MAS) model. J. Hydrometeorol.4, (2003) 878–890.

[2] G. Lenderink, Exploring metrics of extreme daily precipitation in a large ensemble of regional climate model simulations. Clim Res., 44 (2010) 151–166.

[3] J. Overpeck, G. Meehl, S. Bony, D. Easterling, Climate data challenges in the 21st century. Science, 331,2011. doi: 10.1126/science1197869.

[4] A. Dosio, P.Paruolo, Bias correction of the ENSEMBLES highresolution climate change projections for use by impact models: evaluation on the present climate. J. Geophys Res.,116 2011: D16106. doi:10.1029/2011JD015934.

[5] O. Christensen, M. Drews, J. Christensen., The HIRHAM regional climate model version 5. DMI Tech Rep., 22 (2006), 06–17.

[6] J. Christensen, E. Kjellstrom, F. Giorgi, G. Lenderink,M.Rummukainen, Weight assignment in regional climate models. Clim Res.,44 (2010)179–194.

[7] E. Coppola, F. Giorgi, S. Rauscher, C. Piani, Model weighting based on mesoscale structures in precipitation and temperature in an ensemble of regional climate models. Clim Res., 44 (2010) 121–134.

[8] T.Krishnamurti, C. Kishtawal, D. Shin, C. Williford, Multimodelsuperensemble forecasts for weather and seasonal climate. J.Clim., 13 (2000) 4196–4216.

[9] W. Yoon, L. Stefanova, A. Mitra, T. Vijaya, V. Kumar, W. Dewar, T. Krishnamurti, A multi-model superensemble algorithm for seasonal climate prediction using DEMETER forecasts. Tellus A. 57 (2005) 280–289. doi:10 .1111/j.1600-0870.2005.00131.

[10] P. Gleckler, K. Taylor, C. Doutriaux, Performance metrics for climate models. J.Geophys. Res., 113 (2008) doi:10.1029/2007JD008972.

[11] T. Reichler, J. Kim, Haw well do coupled models simulate today's climate? Bull. Am.Meteorol. Soc., 89 (2008) 303–311.

[12] T. Reichler, J. Kim, Uncertainties in the climate mean state of global observations, reanalysesandthe GFDL climate model. J.Geophys. Res., 113 (2008):D05106. doi:10.1029/ 2007JD009278.

[13] IPCC, Climate change, The science of climate change. IPCC, WMO, 1995.

[14] IPCC, Climate change, The scientific basis. IPCC, WMO, 2001 881.

[15] IPCC, Climate change—synthesis report, Intergovernmental Panel on Climate Change, WMO, 2007.

[16] M.Baldauf, A. Seifert, J. Forstner, D. Majewski, M. Raschendorfer, T. Reinhardt, Operational convective-scale numerical weather prediction with the COSMO model: description and sensitivities. Mon. W. Rev. (2011): doi:10.1175/MWR-D-10-05013.

[17] G. Meehl, C. Covey, T. Delworth, R. Stouffer, M. Latif, B. Mcavaney, J.Mitchell, The WCRP CMIP3 multi-model dataset: a new era in climate change research. Bull. Am.Meteorol. Soc., 88 (2007):1383–1394.

[18] G.Nikulin,C. Jones, P. Samuelsson, F. Giorgi, M. Sylla, G. Asrar, M. Buchner,R.Cerezo-Mota, Precipitation climatology in an ensemble of CORDEX-Africa regional climate simulations. J.Clim. (2012) doi:10.1174/JCLI-D-11-00375.

[19] M. Mearns, R. Arritt, S. Biner, M. Bukovsky, S. Mcginnis et al., The North American Regional Climate Change Assessment Program: overview of phase I results. Bull. Am.Meteorol. Soc. (submitted), (2012).

[20] F. Tapiador, E. Sa´nchez, Changes in the European precipitation climatologies (2070–2100) as derived by eight regional climate models. J. Clim.21 (2008) 2540–2557.

[21] F. Tapiador, A joint estimate of the precipitation climate signal in Europe using eight regional models and five observational datasets. J. Clim, 23 (2010) 1719–1738.

[22] J. T. Houghton,L. G. Meirafilho, B. A. Callander, N. Harris, A. Kattenberg, and K. Maskell (Eds.),Climate Change 1995: The Science of Climate Change. Cambridge University Press 19962.

[23] J. T. Houghton, Y.Ding,D. J.Griggs,M.Noguer, P. J. van der Linden,X.Dai, K. Maskell, and C. A. Johnson (Eds.), Climate Change 2001: The

Scientific Basis. Cambridge University Press 2001.

[24] S.Solomon, D. Qin, M. Manning, M. Marquis, K. Averyt, M. M. B. Tignor, H. L. Miller J, and Z. Chen (Eds.), Climate Change 2007: The Physical Science Basis. Cambridge University Press, 2007.

[25] D. C. Bader, C. Covey, W. J. Gutkowski, I. M. Held, K. E. Kunkel, R. L. Miller, R. T. Tokmakian, and M. H. Zhang, Climate models: An assessment of strengths and limitations. U.S. Climate Change Science Program Synthesis and Assessment Product 3.1, Department of Energy, Office of Biological and Environmental Research, 2008, 124.

[26] I. Harris et al., Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset, 2014. doi:10.1002/joc.3711.

[27] Y. Fan and H. van den Dool,, A global monthly land surface air temperature analysis for 1948–present, J. Geophys. Res, 113, (2008), D01103, doi:10.1029/2007JD008470.

[28] http://www.ecmwf.int/research.

[29] http://onlinelibrary.wiley.com/doi/10.1002/qj.828/abstract.

[30] H. W. Barker, A parameterization for computing grid-averaged solar fluxes for inhomogeneous marine boundary layer clouds. Part I: Methodology and Homogeneous Biases. J. Atmos. Sci. 53, (1996) 2289– 2303.

[31] S. Bergstrom, L. P. Graham, On the scale problem in hydrological modelling, J. Hydrology 211 (1998) 253–265.

[32] F. Giorgi, C. Jones, G. Asrar, Addressing climate information needs at the regional level: The CORDEX framework. WMO Bull., 58 (2009) 175– 181.

[33] K. Thorpe, M. Zwarenstein, A. Oxman, S. Treweek, C. Furberget. al., A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers J. of Clinical Epidemiology 62 (5) (2009), 464-475.

[34] F. Giorgi., On the simulation of regional climate using a limited area model nested in a general circulation model, J. Clim, 3 (1990) 941-963.

[35] F. Giorgi and G.T. Bates, On the climatological skill of a regional model over complex terrain, Mon. Wea. Rev., 117(1989) 2325-2347.

[36] F. Giorgi, X. Bi and Y. Qian, Y., Direct radiative forcing and regional climatic effects of anthropogenic aerosols over East Asia: A regional coupled climate-chemistry/aerosol model study, J. Geophys. Res. 107 (D20) (2002) 107, 4439.

[37] J. M. Done, G.J. Holland, C.L. Bruyère, L.R. Leung, and A. Suzuki-Parker., Modeling High-Impact Weather and Climate: Lessons from a Tropical Cyclone Perspective, Climatic Change, 2013. doi:10.1007/s10584-013-0954-6.

[38] C. L. Bruyère, J.M. Done, G.J. Holland, and S. Fredrick, Bias Corrections of Global Models for Regional Climate Simulations of High-Impact Weather. Climate Dynamics, 43 (2014) 1847-1856. doi:10.1007/s00382-013-2011-

[39] L. Feng, T. Zhou, B. Wu, T. Li, J. Luo. Projection of future precipitation change over China with a high-resolution global atmospheric model. Adv. Atmos. Sci., 28 (2) (2011) 464–476. doi: 10.1007/s00376-010-0016-1;

[40] C. He and T. Zhou., The two interannual variability modes of the Western North Pacific Subtropical High simulated by 28 CMIP5-AMIP models. Clim.Dyn., 43 (2014) 2455–2469. doi: 10.1007/s00382-014-2068-x.