Evaluation of regional climate models performance over the area of the Czech Republic

Eva Holtanová, Martin Motl, Jaroslava Kalvová

Department of Meteorology and Environment Protection, Faculty of Mathematics and Physics, Charles University in Prague, Czech Republic (eva.holtanova@matfyz.cz)

1. Objectives

The new climate change scenarios for the Czech Republic (being created in the frame of the project VaV SP/1A6/108/07) will be based on the outputs of the regional climate model ALADIN-CLIMATE/CZ (CHMI). Therefore it is necessary to evaluate the model performance in simulating observed climate characteristics. Model simulations driven by reanalysis provide a possibility to evaluate the RCM itself, without the influence of driving GCM. The comparison to other RCMs can provide an estimate of the uncertainties in climate change scenario based on one RCM. Several criteria based on monthly mean air temperature and precipitation has been chosen for this analysis (see boxes 3-5).

2. Data

The regional climate models used in this study are described in table 1. We used model runs driven by ERA40 in the period 1961-1990 in horizontal resolution of 25 km (except of the model CLM/GKSS with resolution of 50 km). The area used for the analysis is shown in fig. 2. The E-OBS gridded dataset (Haylock et al., 2008) was chosen to serve as the observations, mainly because we took advantage of the common grid. However, the orography of this dataset differs from the real terrain in the orographically complex area of the Czech Republic. Largest discrepancies occur over the mountains in country's border regions. We are aware of the fact that the results of presented analysis might be influenced by the choice of the dataset.

Model	Institute	Country
ALADIN-CLIMATE/CZ	CHMI	Czech Republic
RCA3/C4I	C4I	Ireland
ALADIN/CNRM	CNRM	France
HIRHAM5	DMI	Denmark
CLM/ETHZ	ETHZ	Switzerland
CLM/GKSS	GKSS	Germany
RACMO	KNMI	Netherlands
REMO5.7	MPI	Germany
CRCM	OURANOS	Canada
RCA3/SMHI	SMHI	Sweden
PROMES	UCLM	Spain

Tab. 1: The regional climate models used in this study.

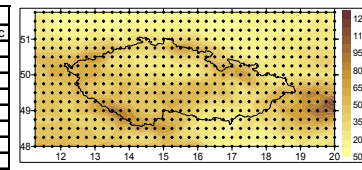

Fig. 1: The grid points used for the analysis (diamonds). The area is between 11.125° and 19.875° of longitude. and 48.125° and 51.875° of latitude. Color scale illustrates the orography of the E-OBS dataset. The borders of the Czech Republic are also shown.

3. Model performance criterion

(Adopted from Reliability Ensemble Averaging (REA) (Giorgi and Mearns, 2002))

Model performance weights R_{Bi} for individual models are based on comparison of model bias B_i in simulating present-day (1961-1990) average monthly mean air temperature and precipitation to the measure of natural variability ε. This parameter is estimated as the difference between minimum and maximum of 30-year moving averages of monthly mean air temperature (precipitation) during the 1890-1990 period.

The value of R_{Bi} is fixed to 1 when B_i is smaller then ε: $R_{Bi} = \left\{ \left[\frac{\varepsilon}{abs(B_i)} \right] \right\}$, REA rmsd: $\delta = \left[\frac{\sum R_{Bi} (m_i - \widetilde{m})^2}{\sum R_{Bi}} \right]^{1/2}$

The REA weighted multi-model mean: $\widetilde{m} = \frac{\sum R_{Bi} m_i}{\sum R_{Bi}}$

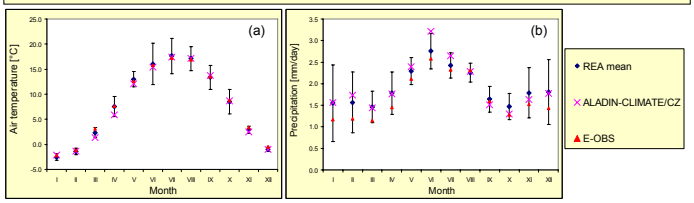

Fig. 2: REA weighted multi-model mean (blue diamonds) of monthly mean air temperatures (a) and precipitation (b). Red triangles indicate the corresponding values derived from E-OBS dataset and pink crosses show the simulated values from ALADIN-CLIMATE/CZ. Vertical lines indicate the REA uncertainty ranges given by the REA rmsd.

Temperature	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	Mean
ALADIN-CLIMATE/CZ	1,0	1,0	0,7	0,5	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	0,9
RCA3/C4I	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
ALADIN/CNRM	1,0	0,9	1,0	1,0	0,5	0,3	0,6	1,0	1,0	0,5	1,0	1,0	0,8
HIRHAM5	1,0	1,0	1,0	1,0	1,0	0,9	1,0	1,0	1,0	1,0	1,0	1,0	1,0
CLM/ETHZ	1,0	1,0	1,0	1,0	1,0	0,7	0,7	1,0	1,0	1,0	1,0	1,0	1,0
CLM/GKSS	1,0	1,0	0,8	1,0	1,0	0,5	0,6	1,0	1,0	0,7	1,0	1,0	0,9
RACMO	1,0	1,0	1,0	1,0	0,8	0,5	0,7	0,9	1,0	1,0	1,0	1,0	0,9
REMO5.7	1,0	1,0	1,0	0,9	0,9	0,5	0,5	0,4	0,5	0,3	1,0	1,0	0,7
CRCM	0,4	0,3	0,2	0,3	1,0	1,0	1,0	0,8	0,6	0,3	0,4	0,3	0,6
RCA3/SMHI	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
PROMES	1,0	1,0	0,5	0,4	1,0	1,0	1,0	1,0	0,8	0,4	0,9	1,0	0,8
Precipitation	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	Mean
ALADIN-CLIMATE/CZ	0,4	0,4	1,0	1,0	1,0	0,7	1,0	1,0	1,0	1,0	1,0	0,8	0,9
RCA3/C4I	0,2	0,4	0,5	0,5	1,0	1,0	1,0	1,0	1,0	0,9	0,4	0,4	0,7
ALADIN/CNRM	1,0	1,0	1,0	1,0	1,0	1,0	0,9	1,0	0,7	0,8	0,4	1,0	0,9
HIRHAM5	0,2	0,2	0,5	0,5	1,0	1,0	1,0	1,0	1,0	1,0	0,4	0,3	0,7
CLM/ETHZ	0,2	0,3	0,5	0,4	1,0	1,0	1,0	1,0	1,0	1,0	0,6	0,4	0,7
CLM/GKSS	0,3	0,5	0,5	0,6	1,0	1,0	1,0	1,0	1,0	1,0	0,5	0,6	0,7
RACMO	0,3	0,4	0,7	1,0	1,0	0,9	1,0	1,0	1,0	1,0	0,8	0,6	0,8
REMO5.7	0,3	0,4	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	0,6	0,9
CRCM	0,4	0,6	1,0	0,6	0,9	0,6	1,0	1,0	1,0	1,0	1,0	0,6	0,8
RCA3/SMHI	0,3	0,4	0,6	0,5	1,0	0,9	1,0	1,0	1,0	0,8	0,4	0,4	0,7
PROMES	0,5	1,0	1,0	1,0	0,8	1,0	1,0	1,0	1,0	1,0	0,8	0,6	0,9

Tab. 2: The model performance weights for monthly mean air temperature and precipitation. In the last column there are the weights averaged over all months. Richer blue color indicates values closer to 1 (best result possible).

4. Skill score based on MSE

- a method introduced by Pierce et al. (2009), who used spatial mean square error
- here we use this skill score in two versions – original (a) and modified (b)
- Skill score: $SS = 1 - \frac{MSE(m,o)}{MSE(o,o)}$, $MSE(m,o) = \frac{1}{N} \sum_{i=1}^{N} (m_i - o_i)$
- where m_i (o_i) indicates
  - (a) 30-year seasonal averages of simulated (observed) value in grid point *i* (spatial MSE)
  - (b) 30-year monthly averages of simulated (observed) values in *i*-th month averaged over the whole area (MSE of annual cycle)
- ΔSS is the Euclidian distance from the model's skill score point to perfect skill point (1, 1, 1, . . . , 1) in *n*-dimensional space, where *n* is the number of metrics used. Lower values of ΔSS indicate better matches to observations.

Skill score		ALADIN-CLIMATE/CZ	RCA3/C4I	ALADIN/CNRM	HIRHAM5	CLM/ETHZ	CLM/GKSS	RACMO	REMO5.7	CRCM	RCA3/SMHI	PROMES
Precave	MAM	-0,5	-3,0	0,4	-4,1	-1,7	-0,9	0,2	0,0	-0,9	-2,3	0,0
	JJA	0,1	-0,2	0,2	-0,3	0,4	0,3	0,0	0,2	-0,1	-0,4	0,4
	SON	0,0	-3,5	-3,5	-4,2	-1,3	-0,1	0,4	0,2	-0,1	-3,2	-1,0
	DJF	-2,6	-3,9	0,1	-10,1	-4,0	-2,1	-1,9	-2,1	-1,2	-2,6	-0,8
Preciad	MAM	-0,7	-14,3	-1,3	-15,6	-3,3	-1,1	-4,2	-1,9	-0,5	-7,2	-2,3
	JJA	0,4	-0,2	-0,9	-1,8	-0,3	0,2	-0,1	-0,1	-1,2	-0,4	-0,9
	SON	-1,5	-0,4	-5,3	-3,3	-1,3	-0,2	-0,3	-1,0	-2,9	-0,3	-3,6
	DJF	-2,8	-0,2	-0,2	-4,4	-0,8	-0,1	-0,5	-0,6	-0,3	0,3	-0,8
Precsdev	MAM	-0,3	-5,3	-0,2	-7,4	-1,2	-0,3	-0,6	-0,3	-0,2	-2,9	-0,3
	JJA	0,3	0,1	-0,2	-1,2	-0,4	0,1	0,2	-0,3	-0,6	-0,2	-0,6
	SON	-2,2	-0,1	-1,4	-2,1	-1,6	0,2	-0,6	-0,6	-0,5	0,0	-1,7
	DJF	-2,2	-0,1	0,3	-4,4	-0,9	0,2	-0,2	-0,5	-0,2	0,3	-0,3
Tave	MAM	-0,7	0,7	-0,2	0,9	0,9	0,6	0,8	0,5	-4,9	0,8	-1,8
	JJA	0,8	0,9	-0,6	0,6	0,2	-0,4	-0,2	-1,2	0,5	0,8	0,7
	SON	0,5	0,8	-0,2	0,8	0,8	0,4	0,8	-1,6	-3,5	0,7	-0,8
	DJF	0,7	0,8	-0,5	-0,1	0,8	-0,3	0,8	0,7	-20,9	0,5	0,7
Tiad	MAM	-0,3	0,2	-4,8	-26,0	-10,9	-15,6	-15,5	-10,6	-55,2	-0,2	-9,2
	JJA	-20,5	-1,8	-32,3	-9,1	-42,4	-47,8	-8,1	-35,3	-34,2	-2,6	-80,1
	SON	-0,9	-1,0	-0,2	-11,6	-2,2	-0,3	-0,2	-2,2	-9,1	-1,4	-28,9
	DJF	-9,9	0,0	-2,3	-3,1	0,0	0,4	-1,2	-2,2	-1,1	0,4	0,2
Tsdev	MAM	-8,2	-0,3	-17,0	-30,3	-4,2	-7,1	-12,6	-20,3	-47,3	-0,4	-46,9
	JJA	-18,8	-2,6	-39,7	-14,8	-64,0	-47,5	-7,1	-45,5	-69,4	-2,8	-68,6
	SON	-1,5	-1,4	-1,3	-11,2	-5,9	-3,9	-2,1	-7,4	-18,7	-0,8	-7,6
	DJF	-21,0	-0,3	-14,3	-2,9	0,3	0,5	-1,3	-1,0	-1,0	0,4	-0,5

Tab. 3: Skill score based on spatial MSE of seasonal mean air temperature (Tave) and the annual cycle of its standard deviation (Tsdev) and interannual variability (Tiad). Similarly for precipitation (Precave, Precsdev, Preciad). Color scales indicate values between -1 and +1, the richer the color, the better skill score. White color corresponds to values smaller than -1.

All metrics	ΔSS	Only means	ΔSS
RCA3/SMHI	13,2	RACMO	3,5
RCA3/C4I	19,4	CLM/GKSS	4,4
RACMO	26,0	PROMES	4,5
ALADIN-CLIMATE/CZ	40,2	ALADIN-CLIMATE/CZ	4,5
HIRHAM5	55,0	REMO5.7	4,9
ALADIN/CNRM	58,8	ALADIN/CNRM	5,5
REMO5.7	64,9	CLM/ETHZ	6,2
CLM/GKSS	71,6	RCA3/SMHI	6,6
CLM/ETHZ	80,1	RCA3/C4I	7,8
CRCM	113,0	HIRHAM5	13,4
PROMES	121,9	CRCM	23,4

Tab. 4. The values of ΔSS based on all metrics in tab. 3 (all metrics), and the values of ΔSS based on Tave and Precave only (only means). The models are arranged according to the values of ΔSS. It is obvious that most of the models get different ranks when the characteristics of variability are accounted for, and the values of ΔSS are larger than in case we consider only the mean values of air temperature and precipitation.

Skill score	Tave	Tsdev	Tiad	Precave	Precsdev	Preciad	ΔSS	Model rank
ALADIN-CLIMATE/CZ	0,99	0,21	0,30	0,53	0,41	0,56	1,37	9
RCA3/C4I	1,00	0,90	0,85	-0,13	0,57	0,50	1,32	8
ALADIN/CNRM	0,97	0,10	0,14	0,26	0,27	0,13	1,84	11
HIRHAM5	0,99	0,76	0,58	-0,54	-0,57	-0,57	2,74	12
CLM/ETHZ	0,99	0,67	0,62	0,06	0,21	0,39	1,46	10
CLM/GKSS	0,98	0,74	0,62	0,29	0,80	0,74	0,91	4
RACMO	0,99	0,84	0,80	0,35	0,80	0,73	0,77	2
REMO5.7	0,97	0,76	0,65	0,60	0,51	0,74	0,80	3
CRCM	0,82	0,66	0,41	0,37	0,42	0,39	1,27	7
RCA3/SMHI	1,00	0,95	0,91	-0,15	0,76	0,73	1,21	6
PROMES	0,97	0,53	0,46	0,59	0,79	0,68	0,91	5

Tab. 5: Skill score based on the MSE of the annual cycle of monthly mean air temperature (T_{ave}) and the annual cycle of its standard deviation (T_{sdev}) and interannual variability (T_{iad}). Similarly for precipitation ($Prec_{ave}$, $Prec_{sdev}$, $Prec_{iad}$). The model rank according to the value of ΔSS is in the last column of the table. The red (blue) color indicates values lower (higher) than 0.5.

5. RMSE and Relative Error

- a method introduced by Gleckler et al. (2008) and in the Czech Republic used for evaluation of AR4 GCM performance (Kalvová et al., 2009).
- Root mean square error of a model: $RMSE = \sqrt{MSE} = \sqrt{\frac{1}{\sum w_i} \sum w_i (S_i - P_i)^2}$

where S_i denotes simulated mean value in month *i* averaged over the whole area in fig. 1, P_i is corresponding observed value and w_i is number of days in month *i*.

- Relative error: $RE = \frac{RMSE - \overline{RMSE}}{RMSE}$, where $\overline{RMSE}$ is median of values of models' RMSE

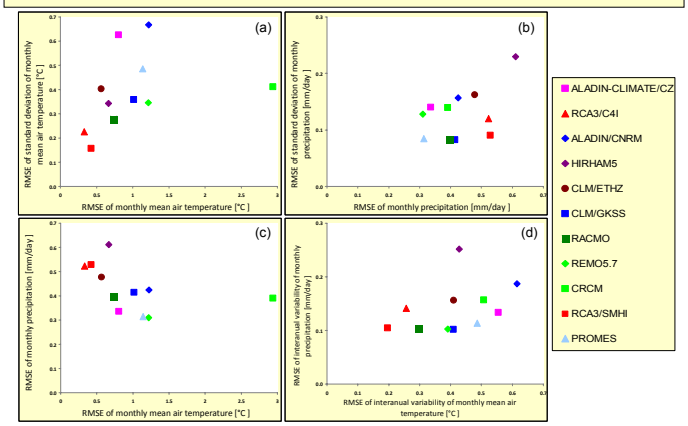

Fig. 3: a) RMSE of monthly mean air temperatures and standard deviation of monthly mean air temperatures. b) RMSE of monthly mean precipitation and standard deviation of monthly mean precipitation. c) RMSE of monthly mean air temperatures and monthly mean precipitation. d) RMSE of inter-annual variability of monthly mean precipitation and inter-annual variability of monthly mean air temperatures.

RE	Tave	Tsdev	Tiad	Precave	Precsdev	Preciad	Average rank
ALADIN-CLIMATE/CZ	0,00	0,75	0,36	-0,19	0,10	0,00	7
RCA3/C4I	-0,59	-0,37	-0,37	0,26	-0,06	0,06	4
ALADIN/CNRM	0,52	0,86	0,50	0,03	0,23	0,40	10
HIRHAM5	-0,18	-0,04	0,04	0,48	0,79	0,88	8
CLM/ETHZ	-0,30	0,12	0,00	0,15	0,27	0,17	7
CLM/GKSS	0,26	0,00	0,00	0,00	-0,36	-0,24	5
RACMO	-0,09	-0,22	-0,27	-0,04	-0,36	-0,22	3
REMO5.7	0,51	-0,04	-0,05	-0,25	0,00	-0,23	5
CRCM	2,65	0,15	0,24	-0,06	0,09	0,18	8
RCA3/SMHI	-0,47	-0,56	-0,52	0,28	-0,29	-0,22	4
PROMES	0,41	0,36	0,19	-0,24	-0,34	-0,15	6

Tab. 6: Relative model errors for monthly mean air temperature (T_{ave}) and the annual cycle of its standard deviation (T_{sdev}) and interannual variability (T_{iad}). Similarly for precipitation ($Prec_{ave}$, $Prec_{sdev}$, $Prec_{iad}$).

6. Summary and conclusions

• Model performance criterion

In general, the model performance is worse for precipitation than for air temperature, except for models CRCM and PROMES. Precipitation is overestimated by most of the models. The models' rmsd (a measure of the variance of multi-model ensemble) is larger in summer (winter) for temperature (precipitation). ALADIN-CLIMATE/CZ is quite close to the weighted multi-model mean except for March and April in case of air temperature, and June and July in case of precipitation.

• Skill score based on MSE

(a) Generally, the skill scores based on spatial MSE are quite good for seasonal mean air temperature and standard deviation of seasonal precipitation, and worst for characteristics of air temperature variability (Tsdev and Tiad). ALADIN-CLIMATE/CZ is ranked as fourth according to ΔSS values when considering whatever combination of metrics combination.

(b) The skill scores are best in case of annual cycle of air temperature. Surprisingly, the worst results are for the annual cycle of monthly precipitation. We would expect the worst results for the characteristics of variability.

• RMSE and Relative Error

ALADIN-CLIMATE/CZ has relatively small RMSE of monthly mean air temperature and precipitation, but the RMSE of interannual variability and standard deviation of air temperature are one of the largest among other models. According to the average rank based on relative model error, the model RACMO is the best and model Aladin/CNRM the worse. ALADIN-CLIMATE/CZ is on 7th place.

• Conclusions

- Using various metrics gives different results of model performance.
- ALADIN-CLIMATE/CZ performs well in simulating annual cycle of monthly mean air temperature and precipitation. However, it seems quite problematic when considering the annual cycle of standard deviation and interannual variability of monthly mean air temperature. According to the most of metrics used here, ALADIN-CLIMATE/CZ occurred in the top third of model ranking.

Acknowledgements

We acknowledge the E-OBS dataset from the EU-FP6 project ENSEMBLES (http://www.ensembles-eu.org) and the data providers in the ECA&D project (http://eca.knmi.nl)

The ENSEMBLES data used in this work was funded by the EU FP6 Integrated Project ENSEMBLES (Contract number 505539) whose support is gratefully acknowledged.

This work has been supported by the project VaV SP/1a6/108/07 funded by the Ministry of Environment of the Czech Republic.

References

Haylock, M.R., N. Hofstra, A.M.G. Klein Tank, E.J. Klok, P.D. Jones, M. New. 2008: A European daily high-resolution gridded dataset of surface temperature and precipitation. J. Geophys. Res (Atmospheres), 113, D20119, doi:10.1029/2008JD10201

Gleckler, P.J., Taylor, K.E., Doutriaux, C., 2008: Performance metrics for climate models, Journal of Geophysical Research, roč. 113.

Giorgi F. and Mearns L. O., 2002. Calculation of average, uncertainty range and reliability of regional climate changes from AOGCM simulation via the Reliability Ensemble Averaging (REA) method. *J. Climate*, 15, 11141-1157.

David W. Pierce, Tim P. Barnett, Benjamin D. Santer, and Peter J. Gleckler, 2009: Selecting global climate models for regional climate change studies. Proceedings of the National Academy of Sciences of the United States of America, Vol. 106, No. 21, 8441-8446.

Kalvová, J., Holtanová, E., Mikšovsky, J., Motl, M., Pišoft, P., Raidl, A., Farda, A., Kliegrová, S., Metelka, L., 2009: Výběr globálních klimatických modelů pro posouzení neurčitostí odhadů budoucích změn klimatu v oblasti ČR. Meteorological Bulletin, 62, 97-106. (in czech)