The Precipitation and Temperature Regime Over Three European Sub-regions as a Result of Climate Change

H. Altinsoy, A. Yuncu, and M.L. Kurnaz

Abstract In this research, regional climate models' (RCMs) simulation data set of PRUDENCE project is used to analyze the uncertainties, direction and magnitude of the expected changes of precipitation, temperature, maximum and minimum air temperature with two different scenarios (B2 and A2) in the period 2070–2100 and 1960–1990 for three different sub-regions of Europe ($1.75°$–$15.75°$ East and $47.75°$–$55.75°$ North; $15.75°$–$25.75°$ East and $45.75°$–$55.75°$ North; $5.75°$–$13.75°$ East and $45.75°$–$47.75°$ North). SVM (support vector machine) is used to classify the models in order to choose the most appropriate ensembles with weighted coefficients from the multi model ensemble. The results are checked with ensemble-mean approach. In this study, we propose a new approach to multi-model ensemble. In general, the performance of our methodology is better than ensemble-mean approach. SVM can eliminate the models which have big amounts of absolute error and select the best ones that have small amounts of absolute error, with big coefficients. Moreover, our proposed methodology gives better results than ensemble-mean approach nearly in all cases.

1 Introduction

There are various statistical approaches to predict past and future temperature, precipitation and other climatological variables. Taylor Diagram (Gleckler et al. 2008), probability density functions (Ruosteenoja et al. 2007; Boberg et al. 2010), and weighted averages (Coppola and Giorgi 2010) are most common ones that have been enforced to find the best representative sample among multi-ensemble climatological data. SVM has been recently used in climatological studies for downscaling climatological variables (Tripathi et al. 2006; Chen et al. 2010; Anandhi et al. 2008),

---

H. Altinsoy (✉) • A. Yuncu • M.L. Kurnaz
Department of Physics, Bogazici University, Istanbul, Turkey
e-mail: hamza.altinsoy@boun.edu.tr

runoff modeling (Behzad et al. 2009), soil moisture data assimilation (Kashif Gill et al. 2007) etc. The genuine part of our work stems from the fact that SVM is used as a classification tool instead of other statistical methods (e.g. PDF, Taylor diagram, ensemble-mean, ANN etc.) and usage of SVM in other climatological studies. More detailed explanation and theory can be found in Vapnik (1995, 1998), Cortes and Vapnik (1995).

2 Data

In this study, PRUDENCE (Prediction of Regional scenarios and Uncertainties for Defining EuropeaN Climate change risks and Effects, EVK2-CT2001-00132), Fifth Framework European programme project (2002–2005), European research project (Christensen and Christensen 2007; Christensen et al. 2002) simulation results of 22 regional climate models and CRU data set are used to propose a new methodology for the multi-model ensemble researches. SVM (Support Vector Machine) is used with its classification option for three different sub-domains in Europe (AL:1.75°–15.75° East and 47.75°–55.75° North; EA:15.75°–25.75° East and 45.75°–55.75° North; ME:5.75°–13.75° East and 45.75°–47.75° North).

3 Methodology

There are two main procedures in the usage of SVM. In the first step, SVM determines the classification regions and the hyperplane (multidimensional linear decision surfaces) which separates the regions by using the train data. While it determines the discriminant function of the hyperplane, $f(x_i) = 0$, it assumes that one region satisfies $f(x_i) < 0$ and the other satisfies $f(x_i) > 0$. To simplify the constraint of the hyperplane we label regions with $y = +1$ and $y = -1$ so the constraint becomes $yf(\vec{x}_i) < 0$. Then, it calculates the perpendicular distance between hyperplane and the closest points of both regions and determines the most efficient $f(x_i)$ by maximizing the sum of these distances. In the second step, it uses the test data, chooses the suitable region and it determines the coefficients with a hyperplane function.

In our study, we used seasonal spatial averages of climatic variables (as shown in Table 1) of the CRU (Climatic Research Unit) for 1961–1990 period as train data. We use all the combinations of 40 and 60 percentile values as accepted region ($y = +1$) and all the combinations of 10 and 90 percentile values as rejected region ($y = -1$). We try various percentile values for accepted and rejected regions as shown in Table 1 and obtain three different hyperplanes which separate the space in two main regions. In second step, test data sets are used to determine the classes of each data points. Thirty year averages of the same period (1961–1990) of Regional Climate Model (RCM) results of PRUDENCE project are used as data points of the

Table 1 The definitions of SVM train data set

	Accepted region boundaries	Rejected region boundaries	Definition
SVM-1	49 and 51 percentiles	1 and 99 percentiles	Weighted averages of models in the accepted region with respect to their coefficients
SVM-5	45 and 55 percentiles	5 and 95 percentiles	
SVM-10	40 and 60 percentiles	10 and 90 percentiles	
SVM-BEST-1	49 and 51 percentiles	1 and 99 percentiles	Weighted averages of models that have coefficients bigger than 0.8
SVM-BEST-5	45 and 55 percentiles	5 and 95 percentiles	
SVM-BEST-10	40 and 60 percentiles	10 and 90 percentiles	
ENS-1	49 and 51 percentiles	1 and 99 percentiles	Equally weighted averages of models in the accepted region
ENS-5	45 and 55 percentiles	5 and 95 percentiles	
ENS-10	40 and 60 percentiles	10 and 90 percentiles	
ENS	–	–	Equally weighted averages of all models

Table 2 Biases of different methods over AL region for Spring (MAM) season

AL	ENS	ENS-1	ENS-5	ENS-10	SVM-1	SVM-5	SVM-10	SVM-BEST-1	SVM-BEST-5	SVM-BEST-10
PRE	−0.365	−0.333	−0.314	−0.660	0.469	0.557	0.434	0.487	−0.819	0.289
TMN	0.537	0.458	0.298	0.552	−0.553	−0.728	−0.507	0.260	0.250	0.153
TMP	0.981	1.310	1.185	0.675	−0.268	−0.978	0.016	−0.249	0.079	0.064
TMX	−0.714	0.213	0.417	1.087	−0.081	−0.914	0.250	−0.260	0.223	0.237

test data sets. The result of SVM shows the strength of each model by assigning positive and negative coefficients. These coefficients then determine the classes of each model. The models with positive or negative coefficients are respectively in the accepted or rejected region. We claim that models in the accepted region can represent the climatological specialties more accurately because their spatial averages are closer to observation (CRU) data set. Precipitation, air temperature, maximum and minimum air temperature are climatological variables that we observed for three selected regions (AL, EA and ME). By this way, we obtained the best representative models for each domain with respect to different variables. Moreover, the positive or negative coefficients of the models represent how correlated they are to the observation data set.

The main purpose of using SVM is to classify the models in two groups. After the classification, we try three simple methods by using the results and coefficients of the program. First of all, we calculate weighted averages of all models in the accepted region with respect to their coefficients (described in Table 1 as SVM-10, SVM-5, SVM-1). Secondly, equally weighted averages of these models are calculated. (Ens-10, Ens-5 and Ens-1) Then, we repeated the same process by using the models that have greater coefficients (greater than 0.8) and named them as SVM-Best-1 (SVM-Best-5 and SVM-Best-10) At last, we take the averages of all models in two regions to compare the efficiency of these new methodologies with a classical approach. We take the differences between the results that we obtained and the spatial averages of observation (CRU) data set for selected period and variable over each domain to determine the biases as shown in Table 2 for spring

Table 3 Absolute errors of precipitation (PRE) and max air temperature (TMX) of different methods over AL region for Spring (MAM) season

AL	ENS	ENS-1	ENS-5	ENS-10	SVM-1	SVM-5	SVM-10	SVM-BEST-1	SVM-BEST-5	SVM-BEST-10
PRE	0.693	0.672	0.712	0.673	0.678	0.656	0.659	0.650	1.007	0.616
TMX	0.691	0.760	0.959	0.672	0.832	0.671	0.694	1.012	1.216	1.279

season. Although the biases of the new methodologies are lower than the classical approach (ENS) results, this doesn't mean that using SVM gives more accurate results than the ENS method. The main reason behind this fact is taking spatial averages and positive and negative biases on different grids over a selected domain can cancel each other. Therefore, we calculated the absolute errors of each grid and took the spatial averages of the absolute errors.

As an example, SVM-10 and SVM-1 have the least mean biases of spring air temperature (e.g. 0.016) and maximum air temperature (e.g. −0.081) over AL domain, respectively. However, SVM-BEST-10 has the least absolute error of spring air temperature and absolute error of SVM-5 of maximum air temperature of the same season is the lowest one. The spatial means of absolute errors have to be small in contrast to the other methodologies to find the most appropriate method. Hence, the results in Table 3 are essential to claim the accuracy and success of SVM usage in this sort of studies. In most of the samples, the differences between the errors of proposed methods and ENS are not very significant. It stems from the fact that the distributions of absolute errors of RCM data in PRUDENCE project are very close to each other.

Finally, we choose the SVM method that has the minimum absolute error for each domain and then calculate its differences between the absolute errors of the ENS method. We can show only the spring season results of precipitation percentage and maximum air temperature results due to lack of space.

4   Results

We obtained significant correction in precipitation percentage amounts by using the best representative SVM as shown in the Fig. 1. The correction amount increases to 20–25% over Germany, Poland and East-Netherlands. In the northern part of the domain, ENS method can estimate the precipitation pattern over some parts such as Bosnia-Herzegovina. However, spatial average of the whole domain is positive (6.69%). The correction in the air temperature results and minimum air temperature results are not as much as the correction in maximum air temperature results. ENS method gives better results over the Western Poland, Northern Germany, Northern Slovenia and Netherlands. On the other hand, the most suitable SVM methodology (has minimum absolute error) corrects the maximum air temperature by 0–1.5°C over the rest of the domain.
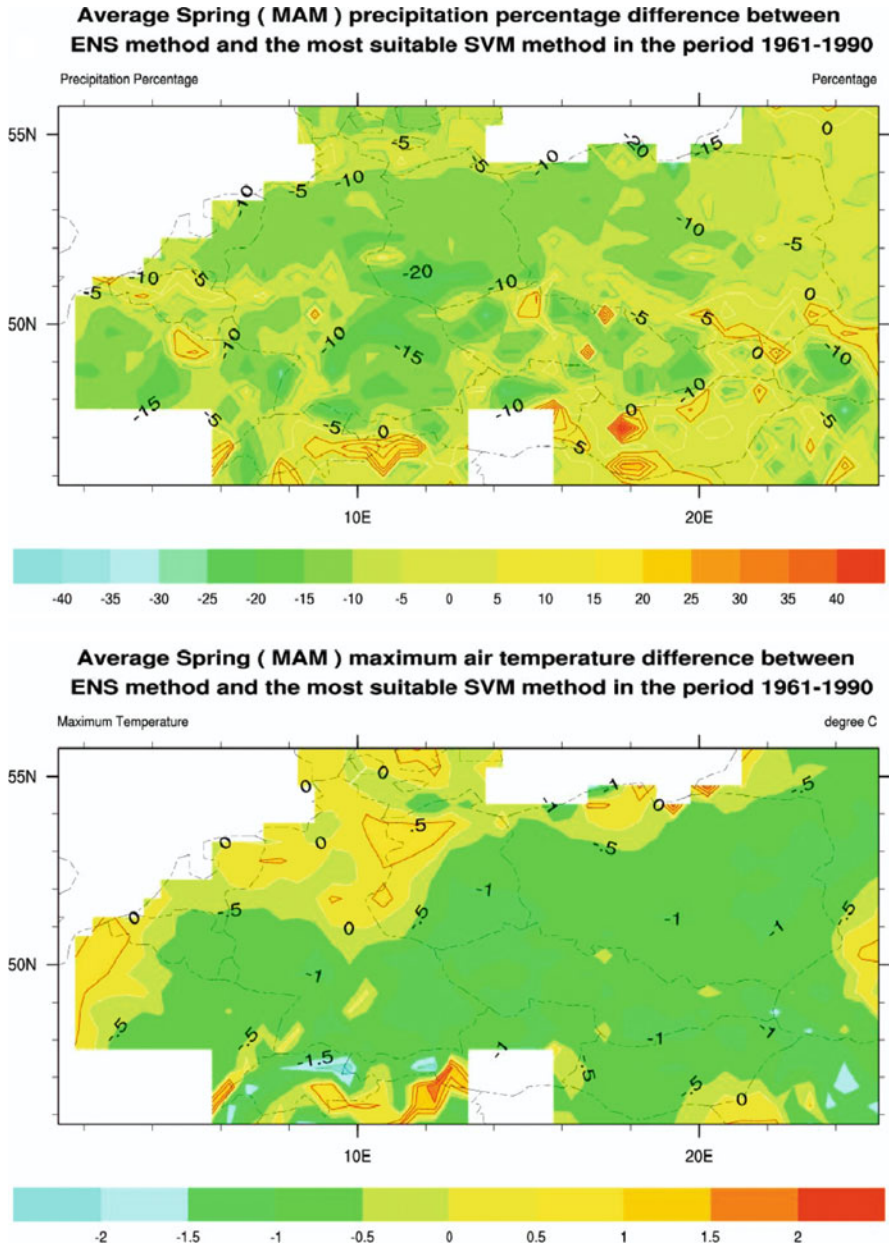
Fig. 1 The difference of absolute precipitation percentage and *max air temperature* error between ENS method and the SVM method over three (AL, EA, ME) domain

5 Conclusions

We have proposed a new methodology for representing the selected climatological variable over a specific domain in a more precise way. In general, the performance of our methodology is better than ENS approach. SVM effectively optimizes the model results by the usage of coefficients representing the level of correspondence with the observational data. With the analysis of the coefficients, the best model can easily be picked for different regions and/or seasons. The correction amount of absolute error changes in parallel with selected climatological variable, region and season. For some circumstances, there is a very small correction in selected region, because absolute errors of models are very close to each other. In order to increase the correction amounts, more Regional or Global Climate model results can be used. This type of methodology can be used to make further analysis of projections of the climatological variables.

References

Anandhi A, Srinivas VV, Nanjundiahb RS, Kumara DN (2008) Downscaling precipitation to river basin in India for IPCC SRES scenarios using support vector machine. Int J Climatol 28:401–420. doi:10.1002/joc.1529

Behzad M, Asghari K, Eazi M, Palhang M (2009) Generalization performance of support vector machines and neural networks in runoff modeling. Expert Syst Appl 36:7624–7629. doi:10.1016/j.eswa.2008.09.053

Boberg F, Berg P, Thejll P, Gutowski WJ, Jens HC (2010) Improved confidence in climate change projections of precipitation further evaluated using daily statistics from ENSEMBLES models. Clim Dyn 35:1509–1520. doi:10.1007/s00382-009-0683-8

Chen S-T, Yu P-S, Tang Y-H (2010) Statistical downscaling of daily precipitation using support vector machines and multivariate analysis. J Hydrol 385:13–22. doi:10.1016/j.jhydrol.2010.01.021

Christensen JH, Christensen OB (2007) A summary of the PRUDENCE model projections of changes in European climate by the end of this century. Clim Chang 81:7–30

Christensen JH, Carter TR, Giorgi F (2002) PRUDENCE employs NEW methods to assess European climate change. EOS 83:147

Coppola E, Giorgi F (2010) An assessment of temperature and precipitation change projections over Italy from recent global and regional climate model simulations. Int J Climatol 30:11–32

Cortes C, Vapnik V (1995) Support vector networks. Mach Learn 20:273–297

Gleckler PJ, Taylor KE, Doutriaux C (2008) Performance metrics for climate models. J Geophys Res 113:D06104. doi:10.1029/2007JD008972

Kashif Gill M, Kemblowski MW, McKee M (2007) Soil moisture data assimilation using support vector machines and ensemble Kalman filter. J Am Water Resour Assoc 43:1004–1015. doi:10.1111/j.1752-1688.2007.00082.x

Ruosteenoja K, Tuomenvirta H, Jylhä K (2007) GCM-based regional temperature and precipitation change estimates for Europe under four SRES scenarios applying a super-ensemble pattern-scaling method. Clim Chang 81:193–208. doi:10.1007/s10584-006-9222-3

Tripathi S, Srinivas VV, Nanjundiah RS (2006) Downscaling of precipitation for climate change scenarios: a support vector machine approach. J Hydrol 330:621–640. doi:10.1016/j.jhydrol.2006.04.030

Vapnik VN (1995) The nature of statistical learning theory. Springer, New York

Vapnik VN (1998) Statistical learning theory. Wiley, New York