Performance of Global Climate Models for Reproducing Present SLP Field over Eastern Asian Region in the CO₂ Transient Run

Hisashi KATO, Naoko OSHIMA, and Shinji KADOKURA

Central Research Institute of Electric Power Industry, Tokyo, JAPAN

(Manuscript received 28 May 2003, in final form 7 December 2004)

Abstract

In this study, we evaluate the performance of global climate models for reproducing the present SLP (sea level pressure) field in the CO₂ transient run, which will be used for the prediction of regional climate change. The outputs calculated using the model are projected in ‘PC-space’, with the component scores from the principal components (or EOF) of observed data as the axes, and their biases are evaluated quantitatively.

This evaluation method was applied to the previous runs (CSM125 and ACACIA) using NCAR-CSM. In the ACACIA run, the January SLP field is reproduced realistically and is improved from CSM125, although the winter monsoon is still weak. The reproduced pattern resembled that of December. On the other hand, the July SLP field is not reproduced well; the intensity of the North Pacific High is overestimated and its ridge is shifted to the north, which are both beyond the observed range. Through these analyses, the difference of bias pattern for each year between two runs which have similar mean distribution pattern is evaluated quantitatively.

Then the performances of the models are compared quantitatively to the other global climate models in the world through the application of this method. All the models tend to reproduce the Siberian High weakly, which is a common feature of the present models. For the summer SLP field, all the models show poor performance as well as two NCAR runs. A series of analyses revealed that, outputs from two NCAR runs in winter seem to be usable, with some care, for the prediction of regional climate changes. However, summer results are insufficient for such use, determined from the viewpoint of the performance of reproducing the present SLP field.

1. Introduction

IPCC (2001a, b, c) warns that global warming has clearly progressed during the last century. It is concerned that the climate change caused by global warming will have a large impact on the natural environment, ecosystem and human society (e.g., water resources and agriculture), and we must assess the effects and consider realistic adaptation programs as soon as possible. With this background, the authors have conducted a series of research studies to evaluate regional impacts of global warming through the development of regional climate models (Kato et al. 1999, 2001) and statistical downscaling (Oshima et al. 2002), which are necessary for impact assessment.

The accuracy of the regional climate change information depends not only on the performance of the downscaling technique, but also on the performance of the global climate model used. For example, Kato et al. (2001) and Oshima et al. (2002), who attempted downscaling for the eastern Asian region from the output of CO₂ transient runs using the NCAR (National Center for Atmospheric Research)-CSM (Climate System Model), regarded their regional
climate change information not as a prediction, but as a projection, or a test of their methods because the performance of the global climate model is not satisfactory in reproducing the present climate. This standpoint is common with other recent downscaling studies (Jones et al. 1997; Giorgi et al. 1997, 1998; Renwick et al. 1998; Deque et al. 1998; Laprise et al. 1998; Leung and Ghan 1999).

In the regional impact assessment for global warming, first of all, the performance of the global climate model in reproducing the present climate should be evaluated for an area that includes the target region. In this study, the performance of the global climate model is evaluated for the eastern Asian region, preceding the ongoing studies on predicting regional climate change using regional climate models and statistical downscaling.

In this study, we examine the sea level pressure (SLP) pattern as representative of the climate system, and the model biases (the difference between model output and observation) for the present climate are evaluated on a monthly basis. In the previous studies on evaluating the output of the global model for the eastern Asian region (IPCC, 2001a; Lal and Harasawa 2000), only the climatological means of air temperature and precipitation were treated. However, when we compare the bias only on the climatological mean basis, 1) large-scale spatial bias patterns may not be identified because of the local bias (e.g., Storch and Zwiers 1999), 2) bias pattern for each year is not clarified, and 3) it is difficult to compare the model features quantitatively. Therefore, for the evaluation, we consider ‘PC-space’ with the component scores of the first few principal components of observed SLP as the axes, and project the model outputs into this space. Figure 1 shows the schematic explanation of our method for a 2-dimensional plane (Z1–Z2 space). The bias of the model is evaluated independently whether it is statistically significant (i.e., the bias is meaningful) for each axis corresponding to a phenomenon or phenomena. This PC-space is a common measure for model outputs, and the bias of the model itself and the difference between models can be quantitatively examined statistically significantly.

The merit of this proposed method, is as follows: 1) a lot of information from n-dimensions (grid points) is integrated into a few independent principal component axes (i.e., local biases are removed), and is evaluated quantitatively, 2) the bias pattern for each year is also evaluated, (e.g., whether a bias pattern for a particular year is within the range of observed variation pattern), and 3) similarities among the models are evaluated quantitatively by a few indices. Furthermore, when we can match each principal component to a phenomenon (physical meaning), the biases on the axes are directly understood.

We establish a method of evaluating the performance of global models in this study. The method is explained in section 2, and the descriptions of the observed and model data used in the analyses are also given. In section 3, this method is applied to the previous data from two CO2 transient runs of the NCAR-CSM. Then the performance of the models is compared with other global climate models in the world. The results are summarized in section 4.

2. Data and evaluation method

2.1 Data

The observed data used in this evaluation is the NCAR SLP data set (ds010.1). First, this data set is examined carefully for consistent
availability through the time series analysis, and based on the information of its averaging technique (or its data source). Finally, we adopted the monthly data from 1963 to 1993 (31 years) and used them in the principal component analysis for each month. The target area in this study is eastern Asia, from 20 to 60 N and 90 to 175 E. The resolution of the data used in the analysis is 5 degrees for both latitude and longitude.

The data to which we applied the method are the outputs of two transient runs using NCAR-CSM. The first run (experiment), named CSM125 (Maruyama et al. 1997; Kato et al. 2001), was conducted using CSM version 1.0 under the condition of atmospheric CO2 concentration increasing at 1%/year. The second run, named ACACIA (Dai et al. 2001a, b, c), was conducted using an improved model (CSM version 1.4) under the BAU scenario, in which the atmospheric CO2 concentration in 2100 was set to be twice the present one. In the ACACIA run, the sulfate effect was also included, which was not considered in CSM125. Refer to the cited references for more detailed descriptions of the model and the experiments. In this study, the 30 years from 1990 are defined as the duration of the present climate because the transient run of CSM125 started with the CO2 concentration at 1990 (355 ppmv), and there is no base run data before the year. Then the results from the ACACIA run for 1990 to 2019 were also extracted and used for the present climate for the comparison with the results from the CSM125.

The outputs of other transient runs conducted throughout the world, which are compared with those of NCAR-CSM in this study, are from the IPCC-DDC (Data Distribution Centre). These runs were carried out by the Canadian Climate Center (CCC), Tokyo-University/NIES in Japan (CCSR), CSIRO in Australia (CSIRO), the Max-Planck Institute in Germany (ECHAM), and the Hadley Centre in U.K. (HADCM). The 30-year data from 1990 are also extracted for the present climate for comparison with the results obtained at NCAR. For the confirmation of the conclusion made in this study, we also conducted analyses using the data from 1970 to 1999; little qualitative difference was found between the final results of the two analyses.

2.2 Evaluation method

The evaluation method is based on a comparison between the model and observation in the multidimensional space, constructed of the principal component scores calculated from the observed data as the axes (PC-space). First, principal component analysis is conducted for the observed SLP data for both January and July, and the main variation patterns (distribution patterns) are extracted. Then the model output is projected to the PC-space constructed from the observed data, as follows.

When we consider \( Z_{obs}(i, t) \) as the \( i \)-th principal component score for the year \( t \), \( E_{obs}(i, j) \) as the vector component for the \( i \)-th principal component at the \( j \)-th grid point and \( X_{obs}(j, t) \) as the SLP value at the \( j \)-th grid point for the year \( t \),

\[
Z_{obs}(i, t) = \sum_j \left[ E_{obs}(i, j) \left( X_{obs}(j, t) - \overline{X_{obs}(j)} \right) \right] 
\]

(1).

Hence, \( X_{obs}(j) \) is the time-averaged SLP value at the \( j \)-th grid point. The notation ‘obs’ denotes that these are calculated from the observed data. The projected component score of each model, \( Z_{model}(i, t') \) is then calculated as

\[
Z_{model}(i, t') = \sum_j \left[ E_{obs}(i, j) \times ( X_{model}(j, t') - \overline{X_{obs}(j)} ) \right] 
\]

(2).

Hence, the model performance in reproducing the \( i \)-th SLP pattern can be statistically tested through the t-test of the time-averaged component scores between the observed (equal 0) and model results defined by Eq. (2). By this method, it is clarified whether the main SLP distribution patterns can be reproduced realistically by each model. Furthermore, the differences in the performance can be compared quantitatively based on this common scale in PC-space.

In this method, we assume that the reproduced model pattern resembles that of observed, i.e., bias components projected on the PC-space include the main part of total model bias, because the model tends to realistically simulate the observation pattern. In each analysis, we have calculated the contribution of bias components projected on the PC-space to the total bias, and confirmed the assumption. There is some possibility that, the bias of each
principal component is influenced by the regional mean bias (difference in regional mean between the model and the observation; hereafter Rm bias) in some cases. In this method, we regard the effect of Rm bias as important. For example, in a case with a positive Rm bias, which results from the overestimation of the strength of an anticyclone covering the region, the positive bias (i.e., the strong anticyclone) has important information for the simulation of a regional climate model for the region. However, some models have a positive or negative bias throughout the year for all grid points in the region, which may be inherent in the model. For the analysis of the model, the contribution of Rm bias, which is calculated as a product of Rm bias and the sum of vector components of the concerned principal component, is also examined.

In this study, not only January (July) data, but also pre- and post-month data, i.e., December and February (June and August) data, should also be analyzed because these biases may result from the fact that the model reproduced a feature in pre- or post-January. Because the extracted principal components are generally different from each other among months, individual principal components cannot simply be compared among months. Therefore, we used the same proposed method again, i.e., in the PC-space constructed for pre- and post-months (e.g., December and February results in the winter case), we projected the model data reproduced for the month of concern (January). This analysis clarifies whether each model reproduces the pre- or post-month SLP pattern for the month of concern, when the model has some biases for that month.

3. Results and discussions

3.1 SLP pattern in January

a. Principal component analysis of observed data

Figure 2 shows the geographical distributions of the factor loadings for the first three principal components, with the climatological mean and the standard deviation calculated from 31-year monthly SLP data in January. The mean SLP pattern in January is characterized by winter monsoon, with the Siberian High (anticyclone) to the west and the Aleutian Lows (cyclones) to the east. The standard deviation is large over the Aleutian Lows area, and to the north of the Siberian High. However, it is larger over the Aleutian Lows.

The first principal component of SLP (proportion: 33%) has positive factor loadings centered south of the Kamchatka Peninsula (southwest of the Aleutian Lows area), and negative factor loadings over the north of Mongolia to Russia. The second component (proportion: 30%) describes the so-called seesaw pattern of SLP variation between north and south of the Aleutian Lows area. Although these two patterns seem to correspond to the intensity (relative to Siberian High) and northern/southern shift of the Aleutian Lows, degeneration may occur for these two components, because their proportions are similar to each other. We do not examine here what phenomenon corresponds to each principal component. In order to clarify the cause of the bias of each principal component score, the contribution of the bias for each grid point is examined. Finally, the main model bias is clarified through the synthesized information. The third component has positive factor loadings over the continent to Japan, which seems to correspond to the expansion of the Siberian High to the southeast. The contribution of the third component is smaller (11%) than those of the first two components.

b. Model performance in reproducing SLP pattern

The evaluation method is applied to the output of both ACACIA and CSM125 runs. Because these two outputs have climatological means similar to each other, they are suitable for verifying whether the method distinguishes the differences in the performances of the models. The model-reproduced present SLP patterns (shown by the bias from the observation) for January (30-year mean) are shown in Figs. 3a and 3b, for ACACIA and CSM125, respectively. These two runs have similar biases in their reproduced patterns, i.e., the weaker Aleutian Lows in the southwestern part, and weaker Siberian High in the northern part. These model biases for each year are examined statistically in the PC-space constructed from the observed data (Fig. 4). These projected principal components are described as $Z_{1\text{jan}(ACACIA)}$, $Z_{3\text{jan}(CSM125)}$, and others.
Table 1 summarizes the mean model bias for each projected principal component with its statistical significance. It is clear that both \( Z_{1.jan}(ACACIA) \), and \( Z_{1.jan}(CSM125) \) have statistically significant positive biases. The contributions of mean bias at each grid point to the mean biases of \( Z_{1.jan}(ACACIA) \) and \( Z_{1.jan}(CSM125) \), i.e.,

\[
E_{\text{obs}}(1,j) \{X_{\text{model}}(j) - X_{\text{obs}}(j)\},
\]

are shown in Figs. 3c and 3d as a geographical map (hereafter, contribution map). These figures indicate that both the positive SLP biases in the southwestern part (weaker Aleutian Lows), and the positive SLP biases in the northern part (weaker Siberian High), cause the positive biases in \( Z_1 \). The small positive bias in \( Z_{1.jan}(ACACIA) \) compared with that of \( Z_{1.jan}(CSM125) \) results from the smaller positive SLP bias over the south of the Kamchatka
Fig. 3. Mean model bias (from the observation) of the reproduced January SLP in a) ACACIA run and b) CSM125 run (units are hPa). c)~h) Contribution of each model grid point to the mean bias of the projected first three principal component scores in the PC-space constructed using observed data. c), e), g) For ACACIA run, and d), f), h) for CSM125 run. Negative values are represented by dotted lines.
Peninsula in the ACACIA run, which seemed to be partly due to the improved sea-ice model in CSM, i.e., overestimated sea-ice area in the CSM125 (Kato et al. 2001) reduced in the ACACIA run. In Fig. 4, for the bias pattern for each year, both positive and negative biases exist for Z1.jan(ACACIA), which is realistic, while most of the biases are positive for Z1.jan(CSM125).

Both Z3.jan(ACACIA) and Z3.jan(CSM125) have statistically significant negative biases (Table 1), which means that these models tend to reproduce weaker expansion of the Siberian High to the east or southeast. This feature is also confirmed by Figs. 3g and 3h. From Fig. 4, we can see that ACACIA has a wider bias range in the third component, from positive to unrealistic negative values, while the mean bias of ACACIA is similar to that of CSM125. This suggests that the outputs of a global model should be examined carefully for each year when these data are used to evaluate the potential impact of climatic change, particularly an extreme event. For the second principal component, the bias is not so marked, and the bias for each year is within the observed range (Fig. 4), although both runs have mean negative biases.

Figure 4 shows that most of the model-reproduced patterns for each year are within the observed range (31 years) in the PC-space constructed from the first three principal components. With ACACIA, two-thirds of the reproduced January SLP patterns are within the observed range, while three-fourths are with CSM125. This seems to indicate that these...
models reproduce realistic January SLP patterns to some extent, although they still have significant mean biases.

c. Similarity to pre- and post-months patterns

In the previous section, it was clarified that some significant model biases exist in the January SLP field, both in ACACIA and CSM125 runs. It should be noted that these biases may result from the fact that the model reproduced the feature in pre- or post-January. Then the similarities of the reproduced January SLP pattern to observed pre- and post-months patterns are examined. Before the analyses, the differences in the observed SLP fields among the three winter months are clarified by the projection of December and February data to the PC-space constructed from January data (Fig. 5). Hereafter, the projected December data are described as $Z_{1,jan}(\text{Dec})$, $Z_{2,jan}(\text{Dec})$, ..., and so on.

In the figure, although $Z_{1,jan}(\text{Dec})$ for each year is within the distribution range of January data, it tends to show positive deviation. The mean deviation is +17.4, which is statistically significant at the 1% level. Figure 6c shows the contribution of each grid to this mean deviation. From Figs. 6a and 6c, it is evident that the Aleutian Lows are weak in the southern part, and the center of the low-pressure area shifts to the north in December, which mainly contributes to the positive deviation. On the other hand, in February, the Aleutian Lows are weaker in the north than in January (Fig. 6d), although $Z_{1,jan}(\text{Feb})$ also shows a mean posi-

Table 1. The mean bias of the model for each projected principal component score ($Z_1, Z_2, Z_3$) for January SLP. ** and * indicate that the bias is statistically significant at the 1% and 5% levels, respectively.

<table>
<thead>
<tr>
<th>Model</th>
<th>January</th>
<th>Z1</th>
<th>Z2</th>
<th>Z3</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACACIA</td>
<td>18.41**</td>
<td>-14.98*</td>
<td>-15.05**</td>
<td></td>
</tr>
<tr>
<td>CSM125</td>
<td>24.80**</td>
<td>-10.11</td>
<td>-14.93**</td>
<td></td>
</tr>
<tr>
<td>CCC</td>
<td>7.02</td>
<td>-50.85**</td>
<td>-17.09**</td>
<td></td>
</tr>
<tr>
<td>CCSR</td>
<td>-40.46**</td>
<td>-30.36**</td>
<td>-23.35**</td>
<td></td>
</tr>
<tr>
<td>CSIRO</td>
<td>-14.62*</td>
<td>-15.73*</td>
<td>-20.37*</td>
<td></td>
</tr>
<tr>
<td>ECHAM</td>
<td>0.83</td>
<td>5.85</td>
<td>-5.67</td>
<td></td>
</tr>
<tr>
<td>HADCM</td>
<td>22.06**</td>
<td>26.81**</td>
<td>-12.47**</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 5. Observed December and February SLP patterns for each year compared with observed January patterns in the PC-space constructed using the observed January first three principal component scores as the axes. All the data are plotted on both the $Z_1$–$Z_2$ surface (left figure) and the $Z_1$–$Z_3$ surface (right figure).
Fig. 6. Mean SLP deviation in a) December and b) February from that in January. c)–h) Contribution of each grid point to the mean deviation (December–January, February–January) of the projected first three principal component scores in the PC-space constructed using the January data, c), e), g) for December, and d), f), h) for February. Negative values are represented by dotted lines.
vative deviation (+11.2). These figures clarify that the winter SLP pattern is significantly different between December and January as to the location and intensity of the Aleutian Lows, although it is rather similar between January and February.

From these figures, some interesting results are obtained. In section 3.1b, it was pointed out that both ACACIA and CSM125 runs exhibit some unrealistic positive biases in the first principal component in January. However, a larger positive value exceeding the range of January data exists in February data, although in only one case (Fig. 5). This indicates that, for the first principal component, the reproduced January pattern may be acceptable (i.e., realistic) when the months are not strictly distinguished, but all three months are regarded collectively as winter. On the other hand, for the third component, both models show a negative bias in the January SLP pattern. Figure 5 shows that the negative scores of the third component prevail in December and February compared with January (Figs. 6g and 6h). The model-reproduced January SLP patterns are rather similar to those of December or February. However, note that the model-reproduced pattern sometimes exceeds this winter range; in other words, it unrealistically reproduces a very weak Siberian High. When we compare Figs. 4 and 5, it is noteworthy that the patterns reproduced by ACACIA for January are rather similar to those for December. This indicates that these reproduced patterns are, at least, representative of winter, although some of them are beyond the range of a realistic January pattern. From this viewpoint, the performance of the model is better than that based on a strict monthly definition.

These results indicate that this evaluation method is useful for the following points: 1) main bias integrated onto a few independent principal component axes is evaluated quantitatively, 2) bias pattern for each year is also evaluated, and 3) the model tendency of seasonally earlier, or later reproduction can be examined. As indicated in these analyses, we can clarify the difference in the performance of two models in detail (quantitatively), even when the mean performances are similar to each other.

3.2 SLP pattern in July

a. Principal component analysis of observed data

Figure 7 shows the geographical distribution of the factor loadings for the first three principal components, with the climatological mean and the standard deviation calculated from 31-year monthly SLP data in July. The mean SLP pattern in July is characterized by the North Pacific High (NPH) over the eastern ocean and low-pressure areas over the continent. The standard deviation is larger to the north and northwest of the NPH.

The proportion of the principal component for the observed SLP in July is about 20% for each of the first three components and equally weighted. This means that the degeneration may occur for these three principal components. Then, similar to the winter case, we do not examine here what phenomenon corresponds to each principal component. However, the cause of the bias of each principal component score is examined from the contribution map, and the main model bias is clarified through synthetic information. The first component is characterized by negative factor loadings for the area from China to the eastern ocean to Japan (Fig. 7c). The second component has positive factor loadings over the entire area, except the wedge-shaped area of negative values north of the NPH (Fig. 7d). The third component is characterized by the positive factor loadings centered in the Okhotsk Sea (Fig. 7e).

b. Model performance in reproducing SLP pattern

The present model-reproduced SLP patterns (mean model bias from the observation) for July (30-year mean) are shown in Figs. 8a and 8b for ACACIA and CSM125, respectively. These two runs have similar biases in their reproduced patterns, i.e., the stronger intensity of the NPH over its northern part, and the northern shift of its ridge as well as the higher SLP over the continent (i.e., weaker cyclonic activities). Although the overestimated intensity of the NPH revealed in the CSM125 run is slightly improved in the ACACIA run, the cyclonic activities over the continent seem to become weak. As a result, the east-west pressure gradient is reproduced as being weaker
than the observation in the ACACIA run, while it is stronger in the CSM125 run.

These model biases for each year are examined statistically in the PC-space constructed from the observed data (Fig. 9). Table 2 summarizes the mean model bias for each projected principal component, with its statistical significance. The negative bias for the first component of the model is statistically significant at the 1% level for both ACACIA and CSM125 runs. The contribution map (Figs. 8c and 8d) shows that both the stronger NPH, and weaker cyclonic activities over the continent increase the negative bias of the first component. The mean negative bias is about the same for both runs. This result indicates that neither model reproduces well the cyclonic activities over the Bai-u front, or over the continent in terms of their intensity (frequency) or their locations (pass), in other words, the location or intensity of the NPH is not reproduced well. More than half of the reproduced first component scores for all years are outside the observed range (Fig. 9).

For the second component, the bias is small in the CSM125 run, with a weak positive
contribution over the entire area, except the eastern ocean area (Fig. 8f), and all the projected second components are within the observed range (Fig. 9). On the other hand, in the ACACIA run, the positive bias is statistically significant (Table 2) because of the higher SLP over the continent (weaker cyclonic activities), and the stronger NPH over the ocean area to...
the southeast of Japan (Fig. 8e). About one-third of the projected component scores are outside the observed range in the run (Fig. 9). Also for the third component, the negative bias is statistically significant at the 1% level in the ACACIA run, while the bias is small in the CSM125 run (Table 2). The notable bias in the ACACIA run is due to the stronger NPH over the ocean to the southeast of Japan, as well as to the comparatively weak Okhotsk High. In the ACACIA run, about half of the projected third component scores are outside the observed range.

From Fig. 9, it is evident that most of the model-reproduced SLP pattern in July is outside the range in the observed PC-space constructed with the first three principal component scores, i.e., only two data, and one-third of all the data (year), are within the range in the ACACIA and CSM125 runs, respectively. Then, for July, the model does not reproduce the SLP pattern well, not only for the climatological

Table 2. The mean bias of the model for each projected principal component score (Z1, Z2, Z3) for July SLP.

<table>
<thead>
<tr>
<th>Model-Observation</th>
<th>July</th>
<th>Z1</th>
<th>Z2</th>
<th>Z3</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACACIA</td>
<td>-30.30**</td>
<td>14.66**</td>
<td>-9.85**</td>
<td></td>
</tr>
<tr>
<td>CSM125</td>
<td>-25.85**</td>
<td>0.23</td>
<td>4.64</td>
<td></td>
</tr>
<tr>
<td>CCC</td>
<td>19.33**</td>
<td>-57.37**</td>
<td>10.08**</td>
<td></td>
</tr>
<tr>
<td>CCSR</td>
<td>24.33**</td>
<td>-55.02**</td>
<td>6.11</td>
<td></td>
</tr>
<tr>
<td>CSIRO</td>
<td>-13.94**</td>
<td>9.03**</td>
<td>21.97**</td>
<td></td>
</tr>
<tr>
<td>ECHAM</td>
<td>-22.80**</td>
<td>14.08**</td>
<td>14.06**</td>
<td></td>
</tr>
<tr>
<td>HADCM</td>
<td>7.21*</td>
<td>-18.39**</td>
<td>15.09**</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 9. Same as Fig. 4, except for July.
mean pattern, but also the pattern for each year.

c. Similarity to pre- and post-month patterns

It was clarified that significant model biases exist in the July SLP field in both ACACIA and CSM125 runs. Similar to the winter case, these biases may result from the fact that the model reproduced earlier or later features for July. Then the similarities of the reproduced July SLP pattern to observed pre- and post-month patterns are examined. First, the differences in the observed SLP fields for the three summer months are clarified by the projection of June and August data to the PC-space constructed from July data (Fig. 10). Please note that the scales of the abscissa are different between Fig. 9 and Fig. 10. From the figure, it is evident that Z1.jul(Jun) for each year is almost within the range of July data, and the difference in mean values is not significant. On the other hand, the mean value of Z1.jul(Aug) is –8.2, which is significantly different from the July mean (i.e., zero) at the 1% level, because the NPH is stronger, and the cyclonic activities over the continent, and over Japan (Bai-u front), are weaker in August (Figs. 11b and 11d). However, the August pattern for each year is within the range of the July data for the first component (Fig. 10).

For the second component, the mean values of Z2.jul(Jun) and Z2.jul(Aug) are +13.6 and +8.2, respectively, which are significantly different from the July mean at the 1% level. Because the contributions of the difference between the June and July SLP values at each grid point to the mean positive deviation (Fig. 11e) in Z2.jul(Jun) are geographically similar to the factor loading of the second component (Fig. 7d), excluding the eastern part over the NPH ridge, the positive second component scores in the principal component analysis of July data, may indicate that the SLP pattern in June extends to July. On the other hand, the positive deviation of Z2.jul(Aug) is due to the higher SLP over the continent (Figs. 11b and 11f). For both June and August patterns, some projected second components exceed the maximum value for July (Fig. 10). There are no significant differences between June and July, and between August and July, for the mean values of the third component.

From the comparison between Figs. 9 and 10, it is clarified that the reproduced first components for July, in both the ACACIA and CSM125 runs, tend to exceed the observed minimum in the three summer months. Then these negative biases do not result from the seasonally earlier, or later, reproduction of the SLP pattern within summer. Both models re-
produced the weak cyclonic activities over the Bai-u front, and/or stronger NPH unrealistically.

In the ACACIA run, although some reproduced second and third component scores for July are outside the observed range, these are within the observed range for June. Furthermore, the contribution maps (Figs. 8e and 8g)
resemble those of June (Figs. 11e and 11g) from the continent to the southwestern ocean region. Therefore it seems that the SLP pattern for July in the ACACIA run reproduced a seasonally earlier (i.e., June) pattern. However, this should not be concluded prematurely, because the model bias from the Okhotsk Sea to the ocean southeast of the Kamchatka Peninsula (Fig. 8a), is different from the deviation of the June SLP from the July SLP (Fig. 11a), even in its sign. On the other hand, for the CSM125 run, some reproduced third components exceed the maximum of the three summer months (Figs. 9 and 10), which means the model performance of reproducing the third SLP pattern in July is not good. This large bias results from the unrealistically strong NPH, and its northern shift (Figs. 8b and 8h). Judging from the fact that the unrealistically reproduced phenomena (positive biases) are reduced in the ACACIA run, it seems to be improved from the CSM125 run. These results for summer analyses indicate again that this evaluation method is useful to clarify the difference in the performance between the models in detail.

3.3 Seasonal change

In the previous sections, reproduced January and July SLP patterns were examined. It was pointed out that some biases may be due to the seasonally earlier reproduction in the models. In this section, further examinations are conducted of the performance of the models in reproducing the seasonal change of SLP patterns for the three months of both winter and summer. Here, we concentrate on the total squared biases rather than the bias for each principal component mentioned previously. One index of model performance, i.e., total squared biases for all the grid points of each model for a reproduced SLP pattern, \( \sum_j (X_{\text{model}}(j) - X_{\text{obs}}(j))^2 \), equals \( L^2 \), the squared distance between the original point to the mean point projected from the model into PC-space with all the principal component scores as the axes, because it also equals \( \sum_i (Z_{\text{model}}(i) - Z_{\text{obs}}(i))^2 \) from the definition of the principal component. Hence, \( X(j) \) and \( Z(i) \) denote the \( j \)-th grid point value, and \( i \)-th principal component score. In the following analyses, we define the distance \( L \) as the ‘main bias’ not for all the principal components, but for the first three components. This \( L \) in the 3-dimensional PC-space includes the absolute value of the bias only for the main SLP bias pattern, and excludes the higher order or local small phenomena and the errors in the data. It is an index to evaluate the model performance to reproduce notable SLP biases. The squared main bias \( L^2 \) was about half or greater of the total squared bias in most models analyzed in this study, in both January and July.

For the evaluation of the performance of each model in reproducing seasonal changes, we conducted other principal component analyses. For winter, we constructed a PC-space for both the observed December and February data. Then the model-reproduced SLP patterns for January were projected to both PC-spaces, and the main biases of the model in both PC-spaces were calculated. The same analyses were also carried out for the model-reproduced December data (to observed November and January PC-spaces), and February data (to January and March PC-spaces). We also examined the summer data in the same manner. These results are summarized in Fig. 12.

In the case of the ACACIA run in winter, the main bias of the reproduced January SLP pattern in the observed PC-space for December is smaller than that in the PC-space in January, which suggests that the reproduced January SLP pattern rather resembles the observed December pattern. On the other hand, the reproduced patterns for December and February have minimum main bias in the PC-space only for the respective month. We also examined the total squared biases in the same manner, and confirmed that the results are similar to Fig. 12, except that the total squared bias of the reproduced January SLP pattern in the observed PC-space for January, is about the same as that in the PC-space in February. These results indicate that the strong winter monsoon in mid-winter cannot be reproduced in the ACACIA run, although the weak winter monsoon appearing in the early and late winter season can be reproduced by the model; in other words, the weak winter monsoon pattern continues without realistic seasonal change. Such a consistent feature is not found in the results of the CSM125 run.

For summer, the model bias is slightly smaller in the PC-space for the pre-month, i.e.,
the model tends to reproduce the earlier seasonal pattern, for June and July in the ACACIA run. Such a feature is not found in the results of the CSM125 run. The reproduced SLP pattern in July also resembles the observed August pattern, because the reproduced cyclonic activities over the continent are weak, which is representative of August.

3.4 Comparison of model performances

In this section, the model performances are compared with other global climate models, using the proposed method. In the analyses, it is clarified whether the model biases found in the ACACIA and CSM125 runs are inherent to these models, or common to other models.

Figure 13 summarizes the main bias calculated for each model for each month. In the figure, the main bias is smaller for all the winter months in the ACACIA run compared with the CSM125 run, although that for summer is not. In the figure, the main bias is about the same value (order) between summer and winter. However, it must be noted that the model bias is comparatively large in summer from the statistically significant point of view, because the variance of the summer data is only one-fifth of that of the winter data. For the quantitative evaluation, the threshold for the statistical significance of the bias is shown as a gray line in the figure. Most models, except ECHAM whose total squared bias is also small, have main bias somewhat greater than the threshold in winter. On the other hand, all models have a significant main bias in summer. The main bias of both ACACIA and CSM125 runs is about the same order with that of other models. Although the CCSR, and CCC have a larger main bias, it is mainly due to the contribution of Rm bias mentioned in section 2.2. For the analyzed region, the CCSR and CCC tend to have a negative Rm bias throughout the year, and for the warm season, respectively. Then, after removing these contributions, the main biases of the models are about the same order as those of other models.

The details of the performance of each model in reproducing January and July SLP patterns,
are summarized in Tables 1 and 2 for each principal component. For January, the models exhibit common features with a negative third component score, i.e., the Siberian High is reproduced as being weaker than the observed pattern. It seems to be the common weak point of the models at this stage, because they cannot reproduce the dam effect for cold air over the continent using the unrealistic coarse-resolution topography of the model (Manabe and Terpstra 1974). On the other hand, positive biases in the first component, i.e., the weakly reproduced Aleutian Lows appearing in ACACIA and CSM125, are not a common feature of the models. In the CCSR and CSIRO runs, the Aleutian Lows are stronger than those of the observation, which is confirmed in the bias maps (not shown). Then this bias may partly depend on the performance of the ocean, or sea-ice model in each global model.

Table 2 shows that all the models, except ACACIA, have positive bias in the third component in July. This positive bias results from the stronger Okhotsk High than the mean observed pattern. However, the reproduced SLP pattern is different among models. In the CSIRO, ECHAM and HADCM runs, the SLP over the Okhotsk sea is higher than the observed mean, while the SLP is comparatively high there in CCC and CCSR, because of the lower SLP than the observed mean over the surrounding region, particularly over the continent. The CSIRO run is similar to the ECHAM run. The ACACIA run rather resembles these two runs, except for the negative bias of the third component. The common feature of these three models is the notable positive SLP bias, extending from east to west, as clarified in the previous sections.

Table 2 indicates that the HADCM run yields a reversed pattern to that of the ACACIA run. However, the stronger NPH in the eastern ocean region is common in both models. In the CCC and CCSR model runs, a relatively strong NPH is reproduced by the notable negative bias over both north and south of the NPH, although no positive biases are reproduced over the NPH. This is due to the Rm bias previously mentioned. Therefore, it is concluded that all the models reproduced a stronger summer NPH than that observed, although its extent or the location of its ridge differed in each model.

From these results, outputs from the two NCAR runs in winter may be usable for the regional climate simulations with some care. On the other hand, it should be concluded that summer results are insufficient for such use, determined from the viewpoint of the performance of reproducing the present SLP field.

4. Conclusions

The performance of the global climate models in reproducing the present SLP field as a representative of a climate system was evaluated for the east Asia region, preceding the ongoing studies on predicting regional climate change caused by global warming using regional climate models and statistical downscaling. The evaluation concerned not only the reproduction of climate means, but also the bias pattern for each year. We constructed a space with a few principal component scores of the observed SLP as the axes for the quantitative evaluation of the model performance. That space to which model outputs were projected (integrated) was a common measure for the model outputs. The bias of the model itself, and the difference between the models, were quantitatively examined in the space from the statistically significant point of view.

The proposed method was applied to the outputs of two CO2 transient runs (ACACIA and CSM125) conducted using NCAR-CSM. For the January pattern, the models reproduced a weaker monsoon pattern, which resembled the observed December pattern. Although some SLP biases were found in the mean, most of the bias patterns reproduced for each year were within the range of the observations for, at least, the three winter months. It can be concluded that the winter SLP pattern is reproduced realistically by the models. On the other hand, it was concluded that the July SLP pattern was not reproduced well at this stage, because the North Pacific High and the cyclonic activities over the continent were not realistically reproduced. The main bias is smaller for all the winter months in the ACACIA run compared with the CSM125 run, although that for summer is not. We can clarify the difference in the performance between two models in detail (quantitatively), even when the mean performances are similar to each other.

Five model outputs compiled at IPCC-DDC
were also applied in this method, and the performances of ACACIA and CSM125 were compared with those of all the models quantitatively. Most models, except ECHAM, have a small main bias in winter. On the other hand, all models have a significant main bias in summer. The main biases of the two NCAR runs are about the same order as those of other models. From the analyses, it was clarified that the bias of a weak Siberian High in winter appearing in both ACACIA and CSM125 runs, is common for all models, although the weaker Aleutian Lows are an inherent bias of the former two models. In summer, the models tend to reproduce a stronger NPH, although the location of its ridge differs in each model. The intensity of the Okhotsk High is stronger in all models, except ACACIA. We conclude that the outputs from these two runs for winter seem to be usable, with some care for the prediction of regional climate change. However, summer results are insufficient for the same use, determined from the viewpoint of the performance in reproducing the present SLP field.

Acknowledgements

We thank NCAR scientists and staff for allowing us to use ACACIA data, and observed SLP dataset. We also acknowledge the efforts of all the staff who made it possible to obtain the model-generated data of the Data Distribution Centre of IPCC. We thank anonymous reviewers and the editor for their useful comments.

References


Leung, L.R. and S.J. Ghan, 1999: Pacific Northwest


