Application of Scale-Selective Data Assimilation to Regional Climate Modeling and Prediction

SHIQIU PENG
Department of Marine, Earth, and Atmospheric Sciences, North Carolina State University, Raleigh, North Carolina, and Laboratory for Tropical Marine Environmental Dynamics, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou, China

LIAN XIE, BIN LIU, AND FREDRICK SEMAZZI
Department of Marine, Earth, and Atmospheric Sciences, North Carolina State University, Raleigh, North Carolina

(Manuscript received 19 February 2009, in final form 24 September 2009)

ABSTRACT

A method referred to as scale-selective data assimilation (SSDA) is designed to inject the large-scale components of the atmospheric circulation from a global model into a regional model to improve regional climate simulations and predictions. The SSDA is implemented through the following procedure: 1) using a low-pass filter to extract the large-scale components of the atmospheric circulation from global analysis or model forecasts; 2) applying the filter to extract the regional-scale and the large-scale components of the atmospheric circulation from the regional model simulations or forecasts; 3) assimilating the large-scale circulation obtained from the global model into the corresponding component simulated by the regional model using the method of three-dimensional variational data assimilation (3DVAR) while maintaining the small-scale components from the regional model during the assimilation cycle; 4) combining the small-scale and the assimilated large-scale components as the adjusted forecasts by the regional climate model and allowing the two components to mutually adjust outside the data assimilation cycle. A case study of summer 2005 seasonal climate hindcasting for the regions of the Atlantic and the eastern United States indicates that the large-scale components from the Global Forecast System (GFS) analysis can be effectively assimilated into the regional model using the scale-selective data assimilation method devised in this study, resulting in an improvement in the overall results from the regional climate model.

1. Introduction

Since the first successful numerical weather forecasts by Charney et al. (1950), a variety of numerical models have been developed to meet different needs. These models can be classified into two major categories according to spatial coverage (i.e., global versus regional models). A global model is suitable for simulating or predicting large-scale meteorological phenomena such as ENSO, global warming, etc., while a regional model is mainly developed for the prediction of local and regional weather and climate. In general, a regional model uses a higher spatial resolution to resolve much more details of the local forcing than a global model, but relies on the global model to supply lateral boundary conditions through which effects of teleconnection and remote forcing can be passed into the regional domain. The effectiveness of prescribing lateral boundary conditions, however, may be limited, especially in the interior of the regional model domain, and inconsistencies between the regional model solution and the global model solution along the boundaries may produce undesirable noise and result in instabilities due to wave reflection (Davies 1983). Thus, systematic large-scale errors could develop within the regional domain (Kanamaru and Kanamitsu 2007), resulting in a distortion of the large-scale information in the regional model (known as “climate drift”) after a long-period integration. Similar problems exist in dynamical downscaling in which regional models are used to produce high-resolution analysis of the atmosphere that cannot be resolved in the global analysis. In a study of dynamical downscaling, Waldron et al. (1996) incorporated...
large-scale data from a larger domain into a smaller inner domain using a spectral nudging technique. Similar techniques were also employed in other studies to nudge the long waves of the model variables (such as the zonal/meridional wind components, temperature, and perturbation Exner function) in the regional model to those of the driving fields (von Storch et al. 2000; Miguez-Macho et al. 2004, 2005; Knutson et al. 2007). Castro et al. (2005) used a grid nudging technique in which all scales were nudged into the regional model, and a comparison between grid nudging and spectral nudging was given in the study of Rockel et al. (2008). Hoyer (1987) proposed a method called "perturbation filter," which decomposes a full forecast model field of the regional domain into the base field and the regional perturbation field. In this method, the spectral truncation filters out the waves longer than the width of the regional domain from the perturbation tendencies so that any scales longer than the width of the regional domain cannot be modified by the regional model during the model integration. A more practical application of this method was described by Juang and Kanamitsu (1994). In a more recent study, Kanamaru and Kanamitsu (2007) proposed a scale-selective scheme that, in many aspects, is similar in function to the spectral nudging technique of von Storch et al. (2000). This scale-selective scheme aims to reduce the systematic large-scale errors within the regional domain in the regional downscaling procedure using a combination of spectral tendency damping and area average correction of temperature, humidity and surface pressure in the Regional Spectral Model (RSM). In addition to the above nudging techniques, Qian et al. (2003) proposed a technique of continuous reinitialization to reduce the errors between the driving reanalysis and the regional climate model in their regional climate downscaling study.

While efforts have been made to reduce the errors in the large-scale component of the atmospheric circulation in the context of dynamical downscaling, the benefit of scale-selective data assimilation in regional climate modeling and prediction has yet to be assessed. The large-scale circulation, which is determined largely by global climatic forcing, can only be properly predicted by a global model. Therefore, large-scale circulation from a global model should be taken as the "target state" and be used to correct the large-scale bias in the regional model. On the other hand, while correcting the large-scale bias,
the regional-scale details in the regional model should be preserved since they are strongly influenced by the regional forcing and are only resolved by the regional model. Therefore, in this study, we propose a data assimilation scheme, referred to as scale-selective data assimilation (SSDA), to incorporate the large-scale circulation from global model forecasts or analysis into regional models while keeping the regional-scale details unchanged during the assimilation cycle. This data assimilation scheme involves a low-pass filter and a three-dimensional variational data assimilation (3DVAR) technique, thus providing a dynamically more consistent method than the spectral nudging technique used in past scale-selective data assimilation studies. In this study, a filter was used to perform the scale separation of forecast fields from both the global model and the regional model. The large-scale components of the global model fields were incorporated into the corresponding components of the regional model fields using 3DVAR. Such a procedure was applied periodically during the integration of the regional model over a period of 4 months, in a way that mimics a 4-month-long seasonal climate prediction. Thus, it extends from the application of scale-selective data assimilation in climatic data mapping (downscaling) to a more realistic seasonal forecasting. For the large-scale components of the atmosphere, the pressure field is adapted to the wind field, so to demonstrate the effectiveness of scale-selective data assimilation in seasonal climate forecasting, we selected the wind field as the only variable used in the assimilation procedure.

The rest of the paper is organized as follows. Section 2 gives a description of the data, the regional model (including its 3DVAR system), and the low-pass filter used in this study. The design of numerical experiments is described in section 3. The results and corresponding analysis are presented in section 4. A summary and discussion are given in the last section.

2. Data, model, and methodology

The Final (FNL) global analysis on 1.0° × 1.0° global grids every 6 h was used to provide the initial conditions and boundary conditions as well as the large-scale information of the atmosphere for the regional model. These data were prepared operationally by the National Centers for Environmental Prediction (NCEP), interpolated from their Global Forecast System (GFS) model. The analyses are available on the surface, 26 mandatory

![Fig. 3. Evolution of the gradient norm of cost function as a function of the number of iterations.](image)

![Fig. 4. Vertical profiles of the mean analysis increments of wind components and temperature for the 3DVAR cycle at 0000 UTC 29 Sep 2005, averaged over all grids in the domain.](image)
(and other pressure) levels from 1000 to 10 hPa. Parameters include surface pressure, sea level pressure, geopotential height, temperature, sea surface temperature, soil parameters, ice cover, relative humidity, $u$ and $v$ winds, vertical motion, vorticity, and ozone. To evaluate the impacts of assimilating the large-scale information from the global analysis on the regional weather/climate forecasting, the NCEP North American Regional Reanalysis (NARR) was used as the benchmark. The NARR dataset is a high-resolution reanalysis of the North American region produced by the high-resolution NCEP Eta Model (32 km/45 layers) together with the Regional Data Assimilation System (RDAS), which, significantly, assimilates precipitation along with other variables. This dataset covers the whole North America with 349 × 277 grids, which is approximately 0.3° (32 km) resolution at the lowest latitude and 29 vertical pressure levels from 1000 to 100 hPa. It is available at 3 hourly, daily, and monthly from 1979 to 2008, and is continuously updated.

The regional model used is the Weather Research and Forecasting (WRF) model developed mainly at the National Center for Atmospheric Research (NCAR). WRF is a next-generation mesoscale numerical weather prediction system designed to serve both operational forecasting and atmospheric research needs (Michalakes et al. 1998). It features multiple dynamical cores, a 3DVAR data assimilation system, and a software architecture allowing for computational parallelism and system extensibility. WRF is suitable for a broad range of applications with scales from meters to thousands of kilometers. The version of the model employed in this study is the Advanced Research WRF (ARW) whose control equations are fully compressible, Eulerian, and nonhydrostatic with a run-time hydrostatic option. It uses a terrain-following, hydrostatic-pressure vertical coordinate with the top of the model being a constant pressure surface. The horizontal grid is the Arakawa-C grid. The time integration scheme in this study uses the third-order Runge–Kutta scheme, and the spatial discretization employs fifth-order schemes. The model supports both idealized and real-data applications with various lateral boundary condition options (Michalakes et al. 1998).

<table>
<thead>
<tr>
<th>Iterations</th>
<th>Jt</th>
<th>Jb</th>
<th>Jo</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>128 200.415</td>
<td>0.000</td>
<td>128 200.415</td>
</tr>
<tr>
<td>48</td>
<td>2710.097</td>
<td>396.797</td>
<td>2313.301</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Iterations</th>
<th>Jt</th>
<th>Jb</th>
<th>Jo</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>128 200.415</td>
<td>0.000</td>
<td>128 200.415</td>
</tr>
<tr>
<td>48</td>
<td>2710.097</td>
<td>396.797</td>
<td>2313.301</td>
</tr>
</tbody>
</table>

Table 1. The variation of the values of each term of the cost function with the iterations. Jt represents the total cost function, Jb the background term, and Jo the observation term.

![Image](https://example.com/image1)

**FIG. 5.** Vertical profiles of monthly-mean RMSEs for the large-scale wind components against GFS analysis for control run (thin solid line), SSDA5 (heavy solid line), and SSDA2 (dash solid line) experiments: (left) $u$ and (right) $v$ component. They were averaged over all grids at sigma levels $\sim (12–24)$ and every 12 h between 0000 UTC 1 Sep and 0000 UTC 1 Oct 2005 (m s$^{-1}$).

![Image](https://example.com/image2)

**FIG. 6.** As in Fig. 5, but for cross correlation.

Table 2. Diagnostics (analysis minus observation) of 3DVAR performance compared with the OI for the last 3DVAR cycle at 0000 UTC 29 Sep 2005.

<table>
<thead>
<tr>
<th>3DVAR</th>
<th>OI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u$ (m s$^{-1}$)</td>
<td>$v$ (m s$^{-1}$)</td>
</tr>
<tr>
<td>Min</td>
<td>$-6.3916$</td>
</tr>
<tr>
<td>Max</td>
<td>$13.1616$</td>
</tr>
<tr>
<td>Avg</td>
<td>$0.0363$</td>
</tr>
<tr>
<td>RMSE</td>
<td>$0.8500$</td>
</tr>
</tbody>
</table>
The WRF 3DVAR data assimilation system is based on incremental variational data assimilation technique, in which the conjugate gradient method is utilized to minimize the cost function in analysis control variable space. The horizontal component of the background (first guess) error is represented via the recursive filter (for regional) or the power spectrum (for global). The vertical component is applied through projections on climatologically averaged eigenvectors and its corresponding eigenvalues. Both the horizontal and vertical background errors are nonseparable, with each eigenvector having its own horizontal length scale determined climatologically. A preconditioning is performed on the background term of the cost function through control variable transform $U^T$ (defined as $B = UU^T$, where $U^T$ is the conjugate transpose of $U$). It can ingest various types of observations, such as those from ground stations, radar, sounding, satellites, etc., to improve the regional weather forecasts.

The low-pass filter employed in this study is a 1D filter that consists of a detrending preprocessing and a discrete fast Fourier transform (FFT). A detrending preprocessing is used to remove the distorted spectrum in high frequency caused by the aperiodicity in the limited-area domain (Errico 1985). As shown by Denis et al. (2002), the possible spurious Gibbs waves caused by extreme large slopes (derivatives) of the field at lateral boundaries can be removed by applying a simple smoothing on the field at the boundaries. In this study, the wind component at grids of each zonal line was taken as a “data series,” that is, $U(i, j) (i = 1, 2, \ldots, M)$ was taken as the $j$th data series, where $i$ varies from west to east and $j$ from south to north. Suppose $N$ is the maximum number of $j$, then there are $N$ data series with the length of $M$. The low-pass filter was applied to each of these data series. Since the eastern–western boundaries of the model domain in this study mainly cross the ocean and flat land, no extreme large slopes of wind field were found and no Gibbs phenomena were apparent, thus no smoothing was applied in the boundaries. More details such as the setup of cutoff wavenumber will be addressed in the next section.

3. Design of experiments

To test the effectiveness of the scale-selective data assimilation scheme in improving regional climate forecasts, a set of experiments were designed and performed. We chose the summer season of 2005 as the time period and the regions covering the eastern United States and most of the Atlantic Ocean as the model domain. WRF was used to perform a 4-month simulation from 0000 UTC 1 June 2005 to 0000 UTC 1 October 2005. The horizontal resolution is 30 km with $275 \times 155$ grids. There are 27 sigma levels with a model top of 50 hPa. For comparison, a control run (denoted as CTRL)

FIG. 7. The large-scale $v$ component at 200 hPa for (a) GFS analysis, (b) control run, and (c) SSDA5 at 0000 UTC 24 Sep 2005 (at the end of the data assimilation cycle; m s$^{-1}$).
was first made. Two data assimilation experiments with the same SSDA scheme but different intervals of data assimilation cycles 5- and 2-day (denoted as SSDA5 and SSDA2, respectively) were performed, in which the large-scale components of wind fields were assimilated into the model every 5 days and every 2 days during the 4-month period. Figure 1 gives a schematic depiction of the SSDA5 experiment. It should be noted that it is not necessary for the positions of 3DVAR cycles to be the same as shown in Fig. 1, and it can be shifted forward or backward in the same interval. When shifting the positions of 3DVAR cycles, the overall effects of the data assimilation on the regional model in a certain period might not change significantly, though the effects on a specific day or week could be different (i.e., the effects might be more evident near the positions of data assimilation cycles).

The low-pass filter was applied to the wind components from both the GFS analysis and the WRF model simulation at an interval of every 5 or 2 days to decompose the wind fields into the large-scale components and small-scale ones. The cutoff wavenumber was set to 3 (i.e., only wavenumbers 0–3 were allowed to pass through the filter), corresponding to the mean flow (wavenumber 0) and the long waves with wavelengths of about 8220, 4110, and 2740 km in the model domain. Figure 2 shows an example of the decomposition of the $v$ component of the wind field at 200 hPa at 0000 UTC 25 September 2005 using the filter. For consistency between the GFS analysis and the WRF simulation, the wind components from GFS analysis were mapped on the 30-km-resolution grids of WRF domain before the decomposition. At every 5 or 2 days, the large-scale wind components with wavenumbers 0–3 from the GFS analysis were assimilated into the WRF-simulated wind fields to correct the simulated large-scale wind components using 3DVAR, while the smaller-scale components (with wavenumbers greater than 3) remained unchanged in the WRF simulation. Considering that the mid- to upper levels of the atmosphere are dominated by large-scale circulation, and the small-scale features are mainly developed or originated near the surface and lower atmosphere (Giorgi and Mearns 1999), only the large-scale wind components of the GFS analysis between sigma levels 13 and 24 (corresponding to pressure levels 500–100 hPa) were assimilated into the regional model. The choice of limiting the SSDA procedure to the mid- and upper troposphere increases the computational efficiency (fast convergence) while allowing the small-scale features to develop in the lower atmosphere in the regional model. This practice of confining the fitting to mid- and upper levels of the troposphere was commonly used in the studies of regional climate model simulation using nudging techniques (von Storch et al. 2000; Miguez-Macho et al.)
2004, 2005; Castro et al. 2005). After the data assimilation, WRF resumes its integration (warm start) until the next data assimilation cycle (Fig. 1).

4. Results and discussion

During the SSDA experiments, each 3DVAR cycle was completed normally, with the gradient of cost function with respect to the model variables decreasing by two to three orders. The values of the observation term in the cost function are generally reduced by one to two orders within 100 inner iterations. As an example, we present some diagnostic results for the last 3DVAR cycle (0000 UTC 29 September 2005) of SSDA5 in Figs. 3–4 and Tables 1–2. Figure 3 shows the evolution of the gradient norm with the inner iterations. The gradient norm decreased sharply during the first 10 iterations. The variation of the values of each term of the cost function with the iterations is given in Table 1. The additional computational expense of SSDA scheme depends on the frequency of the 3DVAR cycles. For SSDA5 experiment, the SSDA scheme increased the total CPU running time by approximately 7%.

The 3DVAR procedure was able to effectively fit the model simulated values to the “observations.” Table 2 presents a comparison of the performance of 3DVAR with optimal interpolation (OI). The root-mean-square errors (RMSEs) of the $u$ and $v$ components are about 0.86 and 0.79 m s$^{-1}$ for 3DVAR, compared to values of 5.69 and 6.86 m s$^{-1}$ for OI. It is obvious that the 3DVAR outperformed OI significantly. Figure 4 shows the vertical profiles of the mean analysis increments for wind components ($u$ and $v$) and temperatures which were averaged over all grid points in the domain. Although only the mid- to upper layers (500–100 hPa) of the large-scale flow were assimilated into the model, the maximum increments of the $u$ component are found around 700 hPa while those of $v$ component are found near the surface and around 400 hPa. For the temperature, the maximum increments are seen at around 300 hPa.

The RMSEs and cross correlation of the large-scale wind components against the GFS analysis were calculated over all grids of sigma levels $(13–24)(500–100 \text{ hPa})$ at every 12 h between 0000 UTC 1 September and 0000 UTC 1 October 2005. For both SSDA5 and SSDA2 experiments, data immediately after the assimilation cycle were not used in the comparison. At the time steps when 3DVAR was performed, the model simulated values before the start of the assimilation cycle were used to calculate the monthly-mean RMSEs. Figure 5 shows the vertical profiles of the monthly-mean RMSEs for the control run, SSDA5 experiment, and SSDA2 experiment. The RMSEs of both $u$ and $v$ components were reduced substantially from the mid- to upper levels (about 500–200 hPa) for both SSDA5 and SSDA2. It is interesting to note that the RMSEs in
SSDA5 were generally smaller than those in SSDA2 at all levels. This indicates that higher frequency of data assimilation cycles does not guarantee that it fits the “true” large-scale flow more closely than the less frequent data assimilation cycles as measured in monthly-averaged RMSEs at each model level from 500 to 100 hPa. Figure 6 shows the vertical distribution of the correlation coefficients between the simulated wind components from the control experiment, SSDA5, SSDA2, and those from the GFS analysis. For the $u$ component, SSDA5 shows the largest correlation with the GFS analysis at all levels except at the upper boundary region (Fig. 6a). However, for the $v$ component, SSDA2 shows the largest correlation above 300 hPa, whereas SSDA5 shows the largest correlation below that (Fig. 6b). The difference between SSDA2 and SSDA5 might be caused by the difference in the large-scale flow patterns at the time of the assimilation cycle between the two experiments. SSDA2 and SSDA5 assimilated different large-scale winds into the model except once every 10 days (the least common multiple of 2 and 5). Nevertheless, both SSDA5 and SSDA2 are more closely correlated with the GFS analysis than the control experiment. Additional experiments and analysis are needed to address the issues of the causes for the differences and the relative advantages between different intervals of the assimilation cycles, which are beyond the scope of this paper.

Consider now the effect of SSDA on the regional climate simulations from the two-dimensional perspective using the 200-hPa wind field from SSDA5 as an illustration. Figures 7–8 display the large-scale $v$ components at 200 hPa for observations (i.e., GFS analysis), control run, and SSDA5 experiment at 0000 UTC 24 September and 0000 UTC 29 September 2005 (right after the last two 3DVAR cycles), respectively. The magnitudes of the large-scale $v$-component winds at 200 hPa from the control run (Figs. 7b and 8b) showed large difference compared to the GFS analysis (Figs. 7a and 8a). After the assimilation of the large-scale wind components from the GFS analysis, the difference was reduced substantially, as seen from Figs. 7c and 8c. Figures 9–10 show the large-scale wind components at 0000 UTC 25 September and 0000 UTC 30 September 2005 (24 h after the data assimilation cycles), respectively. It is evident that the patterns of the large-scale wind components in the SSDA experiment (Figs. 9c and 10c) are still closer to the observations (Figs. 9a and 10a) than those from the control run (Figs. 9b and 10b). These features can be seen more clearly in the wind vector fields (Figs. 11–12). Compared to the GFS analysis (Figs. 11a and 12a), the control run (Figs. 11b and 12b) missed two vortices over the Gulf of Mexico and the southeastern Atlantic Ocean, whereas the SSDA experiment was able to reproduce them (Figs. 11c and 12c). The cyclone over the western Atlantic Ocean...
and the anticyclone over the central Atlantic Ocean were also well captured by the SSDA experiment but they were poorly simulated by the control run in both location and intensity.

To assess the impacts of adjustments of the large-scale fields on the regional-scale fields, we validate the simulated total wind field (large scale plus small scale) from the regional model using the “independent” NARR dataset, which has a comparable resolution to the regional model. The RMSEs and cross correlation were calculated, over the area where the model domain and NARR data coverage overlap, at every 12 h during the whole month of September. Figures 13–14 display the vertical profiles of the monthly-mean grid-averaged RMSEs and cross correlation, respectively. For SSDA5, a RMSE decrease accompanied by an increase of cross correlation is found at all vertical levels for the $u$ component, and at upper levels and near the surface for the $v$ component; for SSDA2, RMSE decrease is found at the mid- to upper levels with an increase of cross correlation for the $u$ component, and at the lower and upper levels for the $v$ components.

Precipitation forecasting has been one of the most challenging tasks in weather/climate forecasts and one of the most concerned issues of the public. Thus, improvement in precipitation forecasting is one of the most important criterions in testing the performance of a regional model. One of the most widely used measurements for the precipitation forecasting skill is the threat score, defined as

$$TS = \frac{N_h}{N_h + N_m + N_f},$$

where $N_h$, $N_m$, and $N_f$ denote the numbers of grid points of “hitting,” “missing,” and “false alarming” by the model forecasting. We calculated the TS of 24-h accumulated precipitation forecasting for September for different thresholds for CTRL, SSDA5, and SSDA2. The results are presented in Fig. 15. For both SSDA5 and SSDA2, although there is a deterioration of threat score during the period of 25–28 September, the magnitude of the deterioration during this period is smaller than the improvements occurred during the periods of 4–10 and 18–23 September. Therefore, the overall monthly-mean impact of SSDA is positive, which is indicated by the percentage of monthly-mean improvements of the threat scores, as shown in Table 3. The monthly-mean threat score improvements by SSDA for different threshold precipitation values varied from 11% to 24%, with slightly higher threat scores for SSDA2. Such an improvement of precipitation forecast by SSDA experiments was also found for the whole month of August (not shown here).
5. Summary and conclusions

In this study, a scale-selective data assimilation (SSDA) scheme for correcting the large-scale biases in the regional climate models was proposed and tested using WRF model and GFS/NARR analysis. The procedure of this SSDA scheme includes scale separation using a low-pass filter and data assimilation using 3DVAR technique. The initial test of this SSDA scheme on a 4-month simulation from 1 June to 1 October 2005 indicates that the scheme was able to incorporate the large-scale features from the global analysis into the regional model effectively and the accumulated effect of the assimilated large-scale circulation led to remarkable improvements in the regional model. By verifying the performance of the regional model against the high-resolution NARR data, it is found that assimilating large-scale information is beneficial to the regional weather forecasting. The effectiveness of scale-selective data assimilation in improving the mid- and upper-level large-scale circulation in the regional climate model provides optimism for improving the skills of dynamic forecast of seasonal tropical cyclone activity and precipitation forecasting. In fact, the seasonal simulation using the SSDA scheme reproduced the vortices that were missing in the control run, although the relatively low resolution used in the present study prevented the test case from actually simulating tropical cyclones at the right intensity.

It is also found that increasing the frequency of assimilating cycle from once every 5 days to once every 2 days does not necessarily lead to improvements in the performance of the regional model. The relative

![Fig. 13. Vertical profiles of monthly-mean RMSEs of the full wind fields against the high-resolution NARR dataset for control run (thin solid line), SSDA5 (heavy solid line), and SSDA2 (dash line) experiments, which were averaged over all overlapped grids between the model and NARR data of each pressure level and every 12 h between 0000 UTC 1 Sep and 0000 UTC 1 Oct 2005 (m s⁻¹): (left) u and (right) v component.](image-url)
advantages of SSDA2 and SSDA5 are somewhat mixed. SSDA5 generally showed a smaller RSME in the wind fields than SSDA2, but SSDA2 showed slightly higher threat scores of precipitation forecast than SSDA5 for the month of September 2005. Therefore, more experiments and analyses need to be carried out to determine the optimum frequency of SSDA.

It should be noted that, although the results of testing the SSDA scheme shown here are from hindcasting, the scheme can be applied to forecast regional climate. In the later case, the large-scale information should be obtained from the forecasts by global climate models instead of global analysis.

The results shown here demonstrate the feasibility of using 3DVAR in scale-selective data assimilation. However, some issues remained unsolved. For instance, in addition to the wind fields, is it beneficial to assimilate the large-scale information of other model variables (such as temperature, moisture, and pressure) into the regional model? Is the effect of the SSDA scheme sensitive to domain size? What is the “optimal” cutoff wavenumber for defining the large scales for a specific domain and the optimal frequencies of data assimilation cycles for different model resolution? Further analyses and experiments are needed to comprehensively study the benefits of the SSDA scheme in regional downscaling and climate prediction. These issues are beyond the scope of the present paper and will be addressed in the future. Furthermore, a comparison between the SSDA

![Fig. 14. As in Fig. 13, but for cross correlation.](image)

![Fig. 15. Threat scores of 24-h accumulated precipitation between 0000 UTC 1 Sep and 0000 UTC 1 Oct 2005 for control run (thin solid line), SSDA5 (heavy solid line), and SSDA2 (dash solid line) experiments. (top to bottom) Different thresholds of 1, 5, 10, and 25 mm.](image)
The method described here and other downscaling techniques (such as grid or spectral nudging) should also be made in future studies.

Acknowledgments. This study is jointly supported by U.S. Department of Energy Grant DE-FG02-07ER6448, the Chinese Academy of Sciences Project KZCX2-YW-Q11-02, and the National Oceanic and Atmospheric Administration (NOAA) ISET Program via Grant NA06OAR4810187. The Changjiang Scholar Program from the Ministry of Education of China provided partial support.

REFERENCES


Table 3. The percentage of monthly-mean improvements of threat scores of precipitation for different thresholds (1, 3, 5, 10, 15, and 25 mm) by SSDA experiments compared to control run for September 2005.

<table>
<thead>
<tr>
<th>Method</th>
<th>1 mm</th>
<th>3 mm</th>
<th>5 mm</th>
<th>10 mm</th>
<th>15 mm</th>
<th>25 mm</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSDA5</td>
<td>10.8</td>
<td>13.7</td>
<td>15.7</td>
<td>19.9</td>
<td>20.1</td>
<td>22.3</td>
</tr>
<tr>
<td>SSDA2</td>
<td>11.3</td>
<td>13.8</td>
<td>14.3</td>
<td>19.2</td>
<td>21.2</td>
<td>24.0</td>
</tr>
</tbody>
</table>

