Sensitivity of a Regional Atmospheric Model to a Sea State–Dependent Roughness and the Need for Ensemble Calculations

RALF WEISSE, HAUKE HEYEN, AND HANS VON STORCH

GKSS Forschungszentrum Geesthacht GmbH, Institut für Gewässerphysik, Geesthacht, Germany

(Manuscript received 16 August 1999, in final form 29 December 1999)

ABSTRACT

The sensitivity of an atmospheric high-resolution limited area model to a sea state–dependent roughness is examined. Two sets of Monte Carlo experiments are compared. In the first set the sea state was explicitly accounted for in the computation of the sea surface roughness. In the second set the roughness was parameterized by the standard Charnock relation. On climatic timescales of months and longer, the differences between the two sets are small. On the daily timescale large deviations between individual realizations of the two ensembles in the order of several hectopascals are occasionally found suggesting a considerable impact of the sea state–dependent roughness on the atmospheric circulation. It is shown, however, that the comparison of individual realizations, a frequently used approach in regional sensitivity studies, can be misleading. It is found here that the largest differences between the two ensembles occurred simultaneously with high inherent model variability. In these situations an eventually existing impact of the sea state–dependent roughness on the atmospheric circulation could therefore not be discriminated from the background variability and the null hypothesis that both ensembles stem from the same population could not be rejected at given risk. At times at which the internal model variability was small a statistically significant impact of the sea state–dependent roughness on the atmospheric circulation was found. However, the impact was small and it is concluded that compared with the sea state–dependent parameterization used in this study the Charnock relation represents a reasonable parameterization in regional atmospheric climate models.

1. Introduction

The presence of waves at the sea surface introduces an additional stress to the airflow. The waves extract momentum from the airflow, which is used in building up and maintaining the sea surface wave field. The wave-induced stress represents a considerable fraction of the total stress within the atmospheric surface layer (e.g., Janssen 1989).

In atmospheric general circulation models (AGCMs) the roughness of the sea surface z_0 is usually parameterized in terms of the Charnock relation (Charnock 1955) as

$$z_0 = \alpha_c u_*^2 g^{-1}, \tag{1}$$

where u_* represents the friction velocity and g the acceleration of gravity. The parameter α_c is a dimensionless constant called the Charnock parameter. Usually, the Charnock parameter is chosen in a way that it represents the best fit to a large number of available datasets; that is, it represents the mean value over a wide range of different sea states (Wu 1982). However, there is considerable scatter within the data around this mean value that may be attributed to a combination of several factors: an eventually existing sea state dependence of the Charnock parameter, observational uncertainties, and various local processes unaccounted for. Theories that explicitly account for the sea state dependence of the momentum flux have been developed and tested by, for example, Janssen (1989, 1991), Chalikov and Makin (1991), Makin et al. (1995), and most recently by Makin and Kudryavtsev (1999).

In atmospheric circulation models the Charnock parameter is usually treated as a constant. Several authors tested whether this assumption is a serious constraint for atmospheric modeling. By using coupled atmosphere–wave models they studied the impact of a parameterization that explicitly accounts for the sea state dependence of the momentum flux (hereinafter called sea state–dependent momentum flux) on the atmospheric circulation. Weber et al. (1993) and Weber (1994) first coupled a global AGCM with a global version of the wave model WAM (WAMDI Group 1988). They found that the changes in the surface fluxes in their model were too small to modify the climatological mean atmospheric circulation significantly. They concluded that in state-of-the-art climate models wave growth is

Corresponding author address: Ralf Weisse, GKSS-Forschungszentrum Geesthacht, Institut für Gewässerphysik, Max-Planck-Strasse 1, D-21502 Geesthacht, Germany.
E-mail: weisse@gkss.de

not a relevant process for the large-scale atmospheric circulation. This finding was consistent with an idealized experiment, in which an atmospheric GCM showed little sensitivity to an (unrealistically large) increase of surface roughness in the storm track of the Southern Ocean (Ulbrich et al. 1993). Although a constant increase of the sea surface roughness is not equivalent to modeling sea state–dependent effects, this sort of experiment may provide an estimate of the order of magnitude of the expected impact. Later, Janssen and Viterbo (1996) used a global coupled atmosphere–wave model with a higher spatial resolution than Weber et al. (1993). They found a small but statistically significant impact of wind waves on the mean atmospheric circulation and concluded that the atmospheric response found in the experiment of Weber et al. (1993) and Weber (1994) was too small because of the coarse spatial resolution of the models used by these authors.

Other studies focused more on the description of the modified physical processes at the air–sea interface due to the application of a sea state–dependent parameterization of the momentum flux. For instance, using the theory of Janssen (1989, 1991) for the computation of the sea state-dependent momentum flux at the air–sea interface Doyle (1995) and Lionello et al. (1998) coupled two different high-resolution atmospheric limited area models with a wave model and investigated the impact of the sea state dependence of the momentum flux on the development of a depression under idealized conditions in a situation of strong baroclinicity. Although, their results differ quantitatively they reached similar conclusions: compared to the Charnock parameterization the momentum flux between atmosphere and ocean is enhanced at the peak of a storm, leading to an enhanced surface roughness and causing a higher surface friction. In turn, the wind speeds and the wave heights are reduced and the depression becomes less intense. In both studies, the strength of the atmospheric response depends crucially on the strength of the depression itself. The changes in the atmospheric pressure field and the wind speed are restricted to the core region of the storm. These findings are consistent with the chain of effects described by Ulbrich et al. (1993) as atmospheric response to heavily enhanced surface roughness in a global model.

Janssen et al. (1997) showed a similar response of an atmospheric model for a realistic synoptic situation and model setup. However, using a sea state–dependent parameterization of the momentum flux they also found cases in which depressions were intensified compared to experiments in which the Charnock relation was used. They concluded that in these cases the effect described above is overcompensated for by enhanced heat fluxes that are due to enhanced coupling at the air–sea interface and lead to a further intensification of the depression. However, in agreement with the idealized case studies, Janssen et al. (1997) found the strongest atmospheric response in atmospheric situations that were characterized by strong baroclinic conditions.

In this paper, we describe two sets of Monte Carlo experiments using two given approaches for the parameterization of the sea surface roughness. In the first set, the atmosphere High-Resolution Limited Area Model (HIRLAM) was integrated using the standard Charnock relation for the computation of the sea surface roughness. In the second set, the HIRLAM model was coupled with the wave model WAM and a sea state–dependent roughness was computed using the wind over waves coupling theory of Janssen (1989, 1991). The objective of this study was twofold, namely to investigate to what extent differences between individual realizations of the two sets, a commonly used approach for assessing the sensitivity of regional climate models, do or do not represent the atmospheric response to the modified momentum flux parameterization and to assess whether an atmospheric response to the sea state–dependent momentum flux may be separated from the background atmospheric variability at a statistically significant level. The study is entirely focused on the statistical analysis of the problem and does not discuss the physics at the air–sea interface or possible drawbacks of the different parameterizations. Furthermore, we concentrate exclusively on the atmospheric circulation. We do not consider possible implications for wave and storm surge modeling.

In the following we will show that a large scatter between any two realizations from the two ensembles but also between any two realizations within each set of experiments does exist. The large differences between individual realizations of the two sets of experiments that occasionally occur on the daily timescale are found only when the variability within each ensemble is large, too. In these situations the wave impact on the atmospheric circulation can therefore not be separated from the background atmospheric variability. Nevertheless, in some cases a statistically significant response of the atmosphere model to the sea state–dependent momentum flux was found. In these cases, although statistically significant, the absolute values are small. We show further that on timescales of months and longer the difference between the two approaches is negligible and as a conclusion the Charnock relation appears to be a reasonable parameterization in climate models. Section 2 briefly describes the applied model and the Monte Carlo experiments performed. Our results are presented in section 3. A summary and discussion are given in section 4.

2. Brief description of model and experiments

a. The model

For this study we used the coupled limited area atmosphere–wave–ocean model ECAWOM (Weisse and Alvarez 1997). ECAWOM consists of the atmospheric

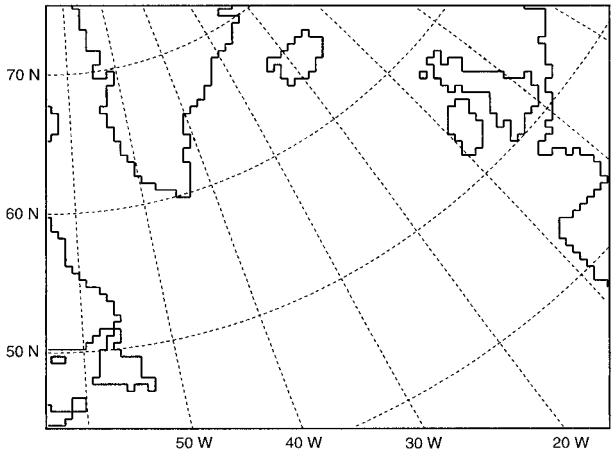

FIG. 1. The model domain. Dashed lines show regular spherical coordinates. The solid lines represent topography and show Newfoundland, Greenland, Iceland, the British Isles, and Europe (clockwise from the left).

model HIRLAM (Källén 1996), the wave model WAM (WAMDI Group 1988) and an ocean model. In order to be able to allocate the differences between our experiments either to internal dynamics of the atmosphere model or to the different formulations of the sea surface roughness, the ocean model was switched off in the present study.

The model operates on a rotated spherical grid with a $0.5^\circ \times 0.5^\circ$ horizontal resolution and a south pole at 58°S, 64°W. The integration area (Fig. 1) was selected such that it covers most of the North Atlantic storm track, that is, an area in which in principle a large atmospheric sensitivity to a sea state–dependent roughness can be expected according to the results of, for example, Doyle (1995) or Lionello et al. (1998). The time step is 5 min for the atmosphere and 20 min for the wave model. In a sponge zone (Davies 1976) along the lateral boundaries, the European Centre for Medium-Range Weather Forecasts reanalyses with a horizontal resolution of approximately $1.5^\circ \times 1.5^\circ$ (Gibson et al. 1996) are forced upon the regional atmospheric model. The wave model was initialized with a wind-independent spectrum.

b. The experiments

Two sets of experiments were performed. In the first set, hereafter referred to as control (CTR), the sea surface roughness is obtained from the standard Charnock relation. In the second set, the impact of waves is explicitly accounted for in the calculation of the sea surface roughness using the wind over waves coupling theory of Janssen (1989, 1991). In the following we refer to the second set as explicitly sea state dependent (ESD).

All experiments were carried out for the year 1993. In particular, the following realizations were computed:

TABLE 1. The quantiles q_x with x = 0.1%, 1%, 50%, 99%, and 99.9% for the spatial frequency distributions of the differences of the annual mean SLP ($\overline{p_i^{\mathrm{ESD}}} - \overline{p_i^{\mathrm{CTR}}}$) and the ratio of the annual SLP standard deviation ($\sigma_{p_i^{\mathrm{ESD}}}/\sigma_{p_i^{\mathrm{CTR}}}$) obtained from the two 1-yr realizations of the ESD and the CTR ensemble. For the instantaneous differences Δp_{it} the quantiles refer to a sample size of 5084 grid points × 1456 time steps. For definitions see text.

Variable	$q_{0.1}$	$q_{1.0}$	$q_{50.0}$	$q_{99.0}$	$q_{99.9}$
$\overline{p_i^{\mathrm{ESD}}} - \overline{p_i^{\mathrm{CTR}}}$, hPa	−0.069	−0.60	0.000	0.122	0.193
$\sigma_{p_i^{\mathrm{ESD}}}/\sigma_{p_i^{\mathrm{CTR}}}$	0.988	0.991	0.999	1.005	1.007
Δp_{it}, hPa	−2.930	−0.800	0.000	1.100	3.100

- For each set a 1-yr integration from 1 January to 30 December was performed.
- Five separate runs for January were made for each set with initial conditions for 2, 3, 4, 5, and 6 January. A sixth realization is given by the January of the 1-yr integration.
- Analogously, for each set five separate experiments for June were performed with initial conditions varying from 2 to 6 June. The June of the 1-yr integration can be interpreted as a sixth realization.

In all experiments the first three model days were discarded as spinup days since the wave model was initialized with a wind-independent wave spectrum. Since we are interested only in the response of the large-scale atmospheric circulation we investigated standard meteorological fields such as sea level pressure (SLP), 500- and 850-hPa geopotential height and temperature, or relative humidity at 700 hPa. Since the results for all these variables remain principally the same the following discussion focuses exemplarily on SLP. Some results for 500-hPa geopotential height are shown additionally.

3. Results

a. Sea state versus control in the 1-yr realizations

To elaborate the changes in the frequency distribution of SLP p between the two 1-yr realizations from the ESD and CTR ensemble the annual mean $\overline{p}$ and the annual standard deviation σ_p of SLP were computed locally at each grid point i ($\overline{p_i^{\mathrm{ESD}}}$, $\overline{p_i^{\mathrm{CTR}}}$, $\sigma_{p_i^{\mathrm{ESD}}}$, and $\sigma_{p_i^{\mathrm{CTR}}}$). Additionally, instantaneous differences Δp_{it} at each time t were calculated:

$$\Delta p_{it} = p_{it}^{\mathrm{ESD}} - p_{it}^{\mathrm{CTR}}. \tag{2}$$

The frequency and the size of these changes are illustrated in Table 1. The differences of the yearly averaged SLP, which provide an indication of an overall shift of the local frequency distribution to larger or smaller values, are relatively small. Only at 2% of the grid points did these changes exceed −0.06 or 0.12 hPa. The largest differences found were in the order of 0.2 hPa. Additionally, the annual standard deviations in the ESD and the CTR experiment are rather similar and the local ratio among both, which yields an indication of the change of the width of the local frequency distri-

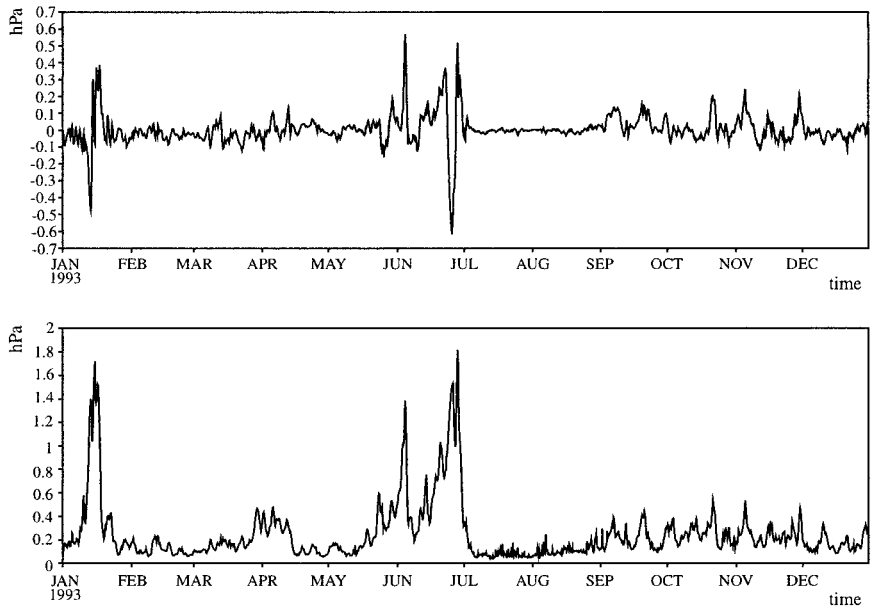

FIG. 2. (top) Spatial bias b and (bottom) rmsd for SLP in hPa between the two 1-yr realizations of the ESD and the CTR ensemble.

bution, is very close to one. We therefore conclude that in these experiments the changes in the temporal and spatial probability distribution of SLP due to the presence of interactive waves at the sea surface are minor. There is, however, on average a very small tendency for a slightly reduced SLP variability. This feature agrees qualitatively with the general findings of Doyle (1995), Lionello et al. (1998), and Ulbrich et al. (1993).

Additionally, we compared local instantaneous SLP differences (2) between both experiments. These differences provide an indication of the frequency and the size of large events (Table 1). It can be inferred that even locally and instantaneously it is most likely to find only small differences between both experiments that rarely exceed ± 1 hPa. Infrequently, however, positive and negative differences of several hectopascals may occur, as indicated by the 0.1% and 99.9% quantiles.

To elaborate the time development of the differences in more detail we calculated the time-dependent spatial bias b,

$$b_t = \frac{1}{M} \sum_{i=1}^{M} (p_{it}^{\mathrm{ESD}} - p_{it}^{\mathrm{CTR}}), \tag{3}$$

and the spatial root-mean-square difference (rmsd),

$$\mathrm{rmsd}_t = \sqrt{\frac{1}{M} \sum_{i=1}^{M} (p_{it}^{\mathrm{ESD}} - p_{it}^{\mathrm{CTR}})^2}, \tag{4}$$

between the SLP fields of both simulations. Here p_{it}^{ESD} and p_{it}^{CTR} denote the SLP at time t and grid point i in the ESD and the CTR simulation, respectively, and $M = 5084$ represents the total number of grid points in the model domain.

Figure 2 shows the bias and the rmsd between the ESD and the control simulation for 1993. Generally, deviations between both experiments are small. However, intermittently large biases and rmsd occur. At these times SLP differences in the order of ± 10 hPa were found among both simulations at some grid points. An example of such a situation is presented in Fig. 3. It can be inferred that the synoptic situation is simulated fairly similarly in both realizations. However, details differ. For instance, at 13 January the modeled depression in the ESD realization is about 4 hPa deeper than in the control realization while it is about 4 hPa shallower two days later. Additionally, the exact positions of the depressions differ slightly as well. Due to the large SLP gradients in this area both effects accumulate to large differences at a few grid points. Locally, the largest differences found were about ± 12 hPa at 13 and 15 January.

In the following we will show that these differences, although large, cannot be considered as the response of the atmosphere model to the sea state–dependent roughness. Instead, in these situations the variability due to the internal dynamics of the atmospheric model itself is high and prevents the separation of an eventually existing atmospheric response from the background variability.

b. Ensemble calculations

To investigate to what extend the large differences between the ESD and the control simulation in January and June can be considered as a response of the atmosphere model to the sea state–dependent parameter-

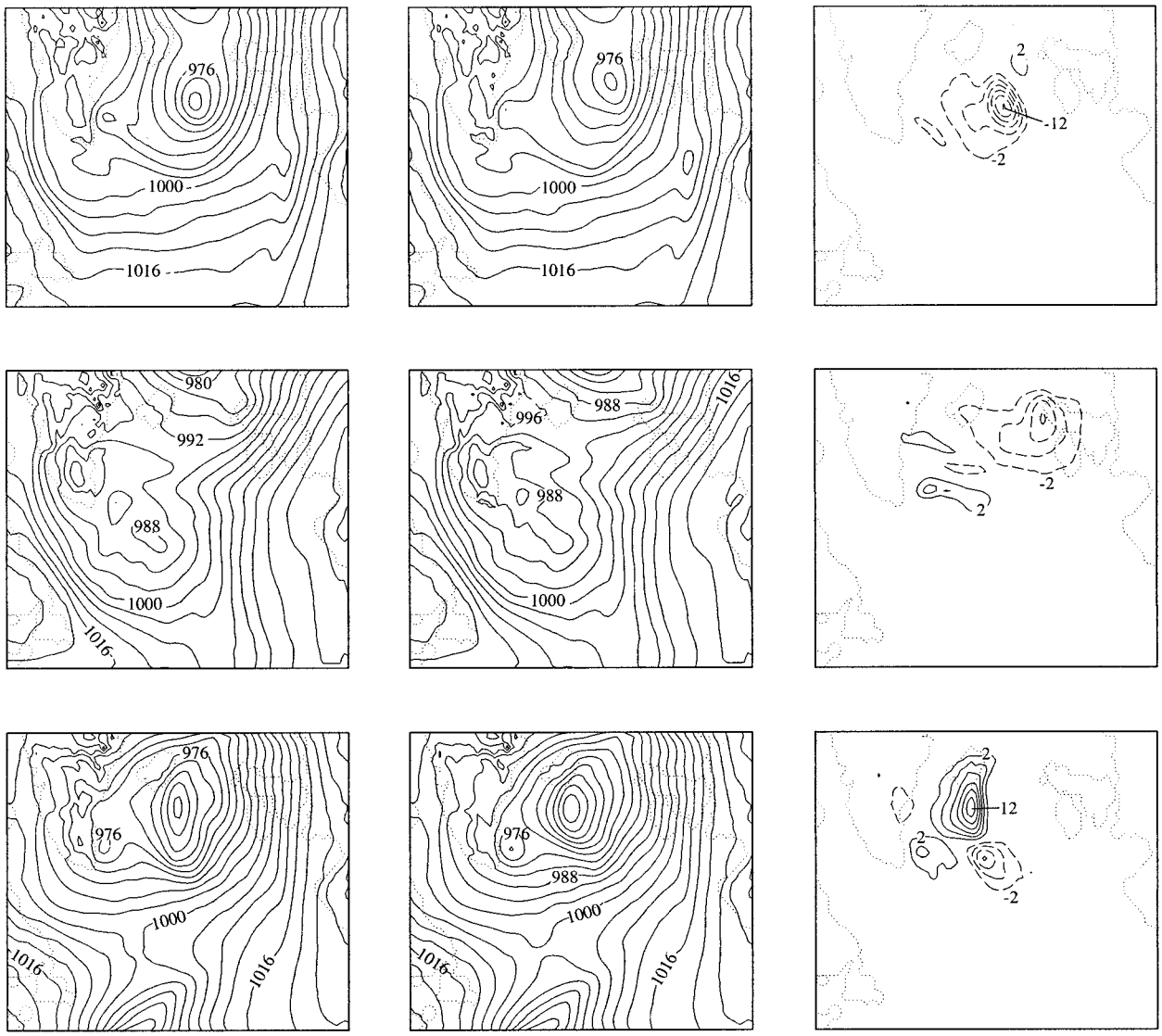

FIG. 3. SLP in hPa at 0600 UTC at 13, 14, and 15 Jan (from top to bottom) from (left) the 1-yr ESD and (middle) the 1-yr CTR simulation together with (right) the local SLP differences in hPa between both simulations.

ization of the momentum flux, a number of ensemble calculations for these months were carried out (cf. section 2b). The principal idea behind these computations is the following: differences between several realizations of the CTR experiment can only arise as a result of internal atmosphere model variability. On the other hand, the changes in the atmospheric circulation obtained in the ESD experiments generally consist of two constituents, namely the response of the atmosphere to the varying momentum flux due to the presence of waves at the sea surface (hereafter referred to as signal) and the inherent internal variability of the atmosphere model itself (hereafter referred to as noise). To separate between both effects, noise levels were estimated by means of ensemble calculations.

1) NOISE LEVEL

The view that in regional simulations ''noise'' would emerge is not widely acknowledged. In many sensitivity studies in which the impact of a ''treatment'' (e.g., a modified parameterization or boundary condition) is studied, the difference between a control and an experimental simulation is taken as being entirely related to this treatment (e.g., Schär et al. 1999). Also the abovementioned process studies by Doyle (1995) and Lionello et al. (1998) have adopted this view. Only recently have ensemble simulations been done to characterize the degree of inherent uncertainties originating from internal chaotic dynamical processes (e.g., Ji and Vernekar 1997; Rinke and Dethloff 2000). Indeed, when noise is un-

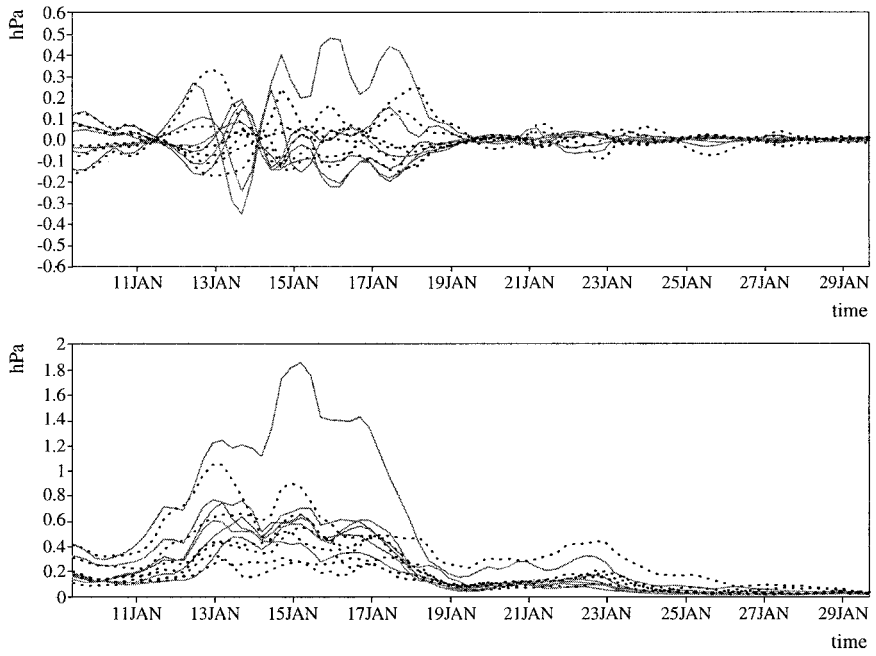

FIG. 4. (top) Bias and (bottom) rmsd of (solid) the six control simulations and (dotted) the six ESD simulations relative to their ensemble mean, respectively.

derstood as the integrated effect of many chaotic processes, such noise is generated on all spatial scales and not only in global atmospheric models. A difference to global simulations is, though, that the boundary control keeps the evolution in the regional integration domain somewhat bounded. We will later demonstrate (Fig. 4) that sometimes the trajectories strongly diverge while in other situations the trajectories converge or stay close together. Of course, such differences between different realizations will be smaller for smaller model domains, where the control by the lateral boundaries is more efficient (Rinke and Dethloff 2000). The same holds for spectral nudging forcing (von Storch et al. 1999; F. Feser 1999, personal communication).

A consequence of this view is that an analysis of regional sensitivity requires estimates of the noise level, exactly as has been pointed out earlier in the classical paper of Chervin and Schneider (1976) (for a review, refer to von Storch and Zwiers 1999). To check whether the differences are the result of random variations, uni- or multivariate tests are applied. In most cases, a conventional *t*-test is adopted.

When the noise level is stationary, that is, independent of time, it may be calculated from extended simulations. This is done for instance in global climate simulations, where stationarity of the noise is assumed for the specific seasons. In the case of regional models, which are generally run for few months or years only, the noise level cannot be regarded as stationary but depends on the synoptic situation enforced along the lateral boundaries. In that case, multiple integrations, or ensemble or Monte Carlo simulations, are needed.

In this study the latter approach was adopted. Ensemble integrations for both the CTR and the ESD experiments for January and June were carried out (cf. section 2b). These are the months for which the largest differences between the two 1-yr realizations were found. In the following we concentrate on the January simulations. The analysis for June yields essentially the same result.

For each time t and at each grid point i the ensemble mean for SLP $\langle p_{itk}^{\mathrm{CTR}} \rangle$ was calculated from the $k = 1, \ldots, 6$ control realizations where $\langle \cdots \rangle$ denotes ensemble averaging. Subsequently, a time-dependent spatial bias and a spatial root-mean-square difference were computed relative to this ensemble mean, analogously as in Eqs. (3) and (4). Compared to the ensemble mean the bias is indicative for the mean deviation of a realization while the rmsd measures the intensity of the local deviations from the ensemble average. The same calculations were repeated for the ESD experiments. The results of both computations for January are shown in Fig. 4.

It can be obtained that there is considerable variability between the two ensembles as well as within each ensemble. The latter is entirely related to the inherent internal variability of the model and was excited only by the modified initial conditions. The order of magnitude of the variability is the same for both ensembles.

To illustrate the variability inherent in the control

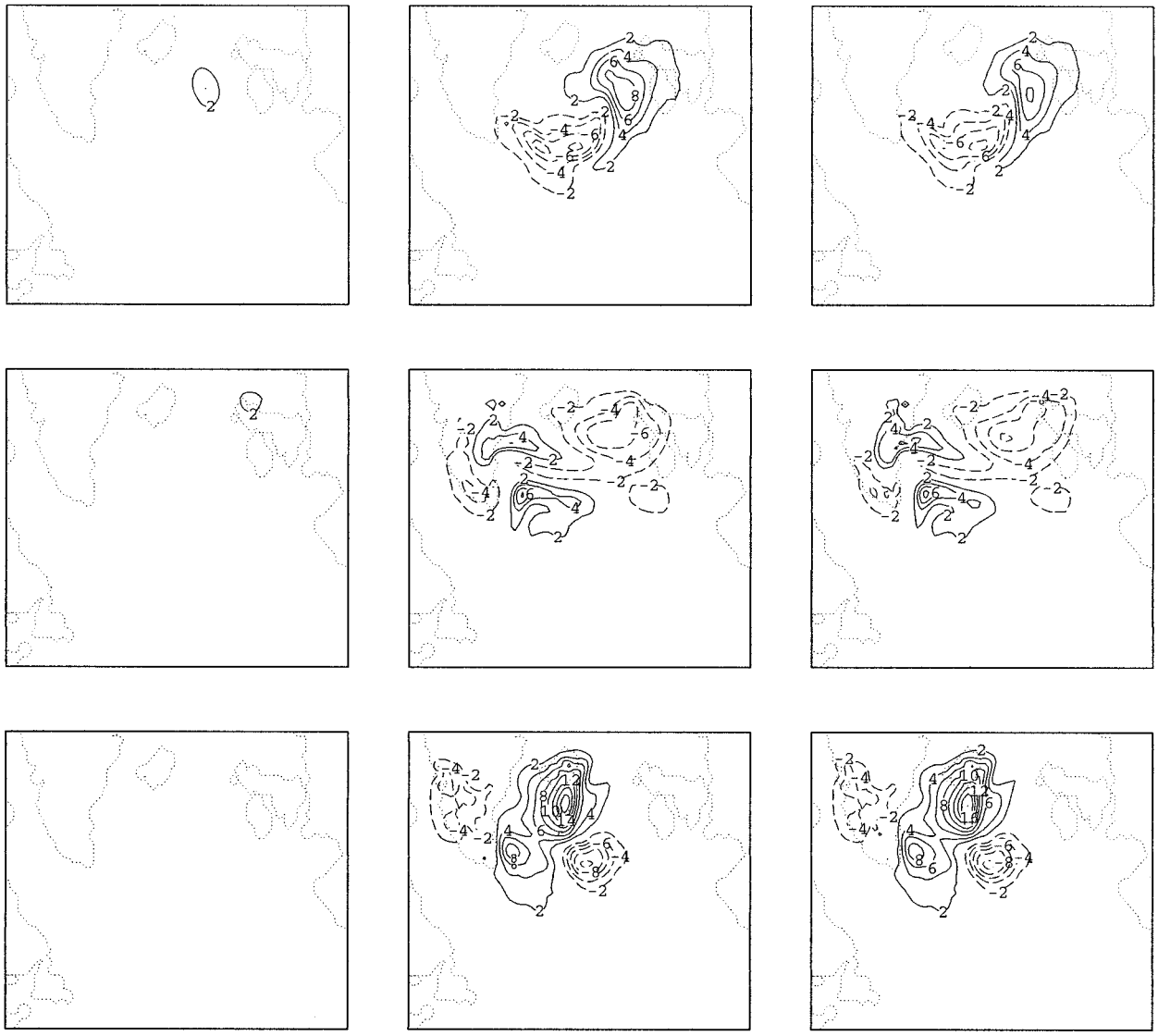

FIG. 5. Local SLP differences in hPa between three control realizations. Shown are the differences at 0600 UTC for 13, 14, and 15 Jan (from top to bottom) between the experiments with initial conditions from (left) 3 and 1 Jan, (middle) 6 and 1 Jan, and (right) 6 and 3 Jan.

simulations Fig. 5 shows the SLP differences at 0600 UTC at 13, 14, and 15 January between the three control realizations starting at 1, 3, and 6 January. It can be inferred that the differences within the control ensemble may have the same order of magnitude (e.g., up to 16 hPa between the realizations with initial conditions from 6 and 3 January) and a similar spatial structure as the SLP differences between the 1-yr ESD and the 1-yr CTR simulations (Fig. 3). Therefore, the large SLP differences obtained between the two 1-yr simulations in January and June reveal nothing about the sensitivity of the atmospheric circulation to a sea state–dependent momentum flux. They can be a result of both, sea state–dependent effects or modified initial conditions. An assessment on the basis of two individual realizations is not possible. In the following we therefore compare the signal to the noise level to assess to which degree differences are caused by model sensitivity or by random effects.

2) COMPARING SIGNAL WITH NOISE LEVEL

When comparing the two 1-yr integrations it was implicitly assumed that the noise level was stationary, that is, independent of time. However, the analyses of the previous section showed that this assumption is not valid in our case. In the following we therefore assess the significance of the signal by comparing it with the nonstationary noise level. Two different views are adopted. In the first we investigate whether a statistically significant effect exists in general, in the second the statistical

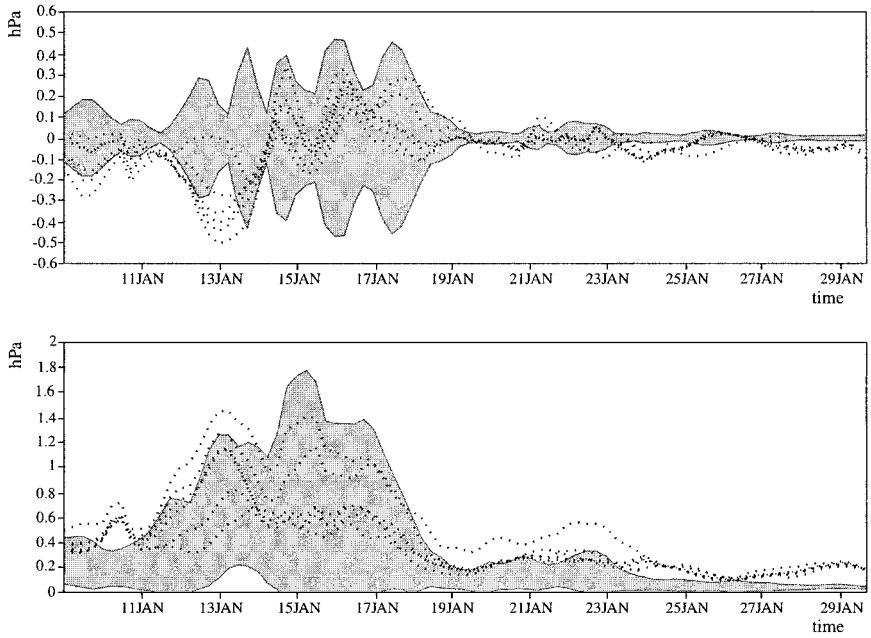

FIG. 6. (top) Spatial bias and (bottom) rmsd for SLP in hPa between (bottom) the ESD realizations and the ensemble mean of the control simulations. The gray-shaded areas represent the 95% confidence interval estimated from the control simulations.

significance of a signal at given time and/or location is assessed.

To find out whether there is a statistical significant effect in general we did the following. First, the 95% confidence intervals of the bias and the rmsd were derived from the control simulations. Then, the spatial biases and rmsd of the ESD realizations were computed relative to the ensemble mean of the control realizations. Finally, we counted how often the spatial biases and root-mean-square differences left the 95% confidence band obtained from the variability in the control realizations. For January, the result of this procedure is shown in Fig. 6. When the null hypothesis (both ensembles stem from the same population) cannot be rejected, the ESD realizations are expected to fall within the 95% confidence band 95% of the time. However, this is not the case (Fig. 6). For both, bias and rmsd only roughly 60% of the values fell within the 95% confidence band, indicating that on average both ensembles are significantly different in a statistical sense. However, using this method, no conclusions on the statistical significance of individual differences (for instance the 12 hPa between the 1-yr ESD and CTR realizations at 15 Jan) can be obtained. Furthermore, closer inspection of Fig. 6 reveals that the decision is mainly based on the second half of the examined period and that it is more likely to detect a significant response if the noise is small. However, in our experiments these cases usually came along with small signals too.

To assess the statistical significance of SLP differences at a particular time we used the local t test (e.g., von Storch and Zwiers 1999; Frankignoul 1995). In this test the null hypothesis that two univariate random variables x and y have equal means is tested. If both samples have the same sample size N the t statistic T is given by

$$T_t = \frac{\langle y_t \rangle - \langle x_t \rangle}{S_t \sqrt{\frac{2}{N}}}. \tag{5}$$

Note that the sample mean in (5) is not identical with the true mean that would be obtained from a sample with infinite sample size. This is accounted for by the denominator in (5) where S denotes the pooled estimate of the standard deviation (e.g., von Storch and Zwiers 1999):

$$S_t^2 = \frac{\sum_{k=1}^{N} (x_{tk} - \langle x_t \rangle)^2 + \sum_{k=1}^{N} (y_{tk} - \langle y_{tk} \rangle)^2}{2N - 2}. \tag{6}$$

The test statistic (5) has the form of a signal-to-noise ratio, with the unbiased estimate of the difference (signal) in the numerator, and an estimate of the expected random difference (noise) in the denominator. It can be inferred that a large T value does not necessarily imply a large signal, nor that a large signal automatically results in a large T value. We will show below that this point is crucial for assessing the results of our experiments.

The test statistic (5) was computed for the ensemble means of the spatial bias and the rmsd for each time step t in which both spatial bias and rmsd were computed relative to the respective ensemble mean of the control ensemble. Under the null hypothesis that the

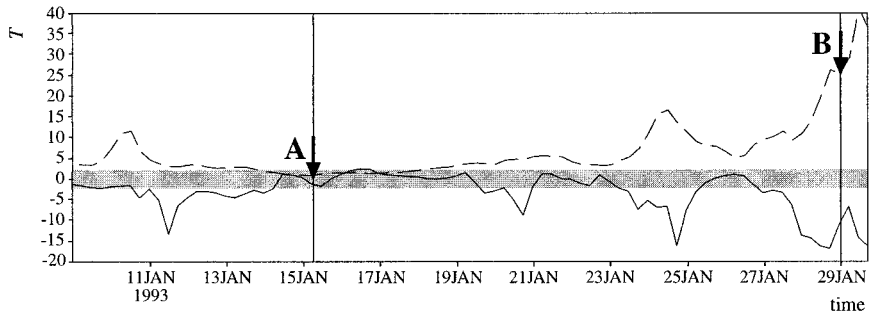

FIG. 7. The t statistic (5) for (solid) bias and (dashed) rmsd. The 95% confidence interval is indicated in gray. The points A and B are referred to in Fig. 8.

ensemble means of the bias and the rmsd of the control and the ESD simulations are identical, (5) has a t distribution with $2N - 2 = 10$ degrees of freedom. In the present analysis we adopt a 95% significance level so that the null hypothesis is rejected if $T > 2.23$ where 2.23 represents the 97.5% critical value of the t distribution with 10 degrees of freedom.

Figure 7 compares the t statistics for the SLP bias and root-mean-square difference against the critical value of 2.23. Most of the time both ensembles can be separated at the 95% confidence level. A striking result is that in mid-January the ensembles cannot be separated from each other, although the largest deviations between individual realizations of the ESD and the CTR ensemble were found in this period (cf. Figs. 3 and 6). On the other hand, the ensembles can be clearly separated at the end of January, where only small deviations between both ensembles occur (Fig. 6). This is due to the fact that large deviations between the ESD experiments and the control runs always correspond with times in which both the CTR and ESD runs show a high internal variability, too. Therefore, large deviations between individual members of the ensembles or even between the ensemble means do not necessarily indicate statistical significance of the results. On the contrary, it seems more likely to find a statistical significant response in periods in which the internal variability of the atmosphere is small. In our case this means, however, that the obtained atmospheric response, although statistically significant, is weak.

To further illustrate this point, two periods are compared in Fig. 8. The first period (0600 UTC 15 Jan) is characterized by local SLP differences of up to ± 4 hPa between the ensemble means of the ESD and the CTR experiments, large internal atmospheric variability (cf. Fig. 6), and a small signal-to-noise ratio (cf. Fig. 7). The second period (0000 UTC 29 Jan) is characterized by small local SLP differences of up to ± 0.75 hPa, small internal atmospheric variability, and a large signal-to-noise ratio. The different degrees of variability within and between the two ensembles on these days can also be obtained from the 500-hPa geopotential height field (Fig. 8). It can be inferred that, although statistically significant, the atmospheric response at 29 January is weak and may be considered as physically irrelevant while it is not clear whether the large SLP differences at 15 January are a result of the modified momentum flux parameterization or the internal model variability. In other words, in this case an eventually existing impact of the sea state–dependent momentum flux on the atmospheric circulation may not be discriminated from the background variability. Therefore, it cannot be concluded that these differences solely represent the model sensitivity due to the modified momentum flux parameterization. To further emphasize this point a local t-test analogous to (5) was applied for the ensemble means of the SLP in Fig. 8. The result of this test shows that the areas where the largest SLP differences were found do not coincide with those areas that are statistically significant at the 95% confidence level.

The fact that usually large differences between the ensembles fell together with small signal-to-noise ratios is elaborated in Fig. 9. The scatterplot shows that generally the largest values for the t statistics are found when the differences between the two ensembles are small. On the other hand, for the largest differences the t statistics are small since they usually occur simultaneously with high internal model variability. The declination of the scatter in the plots is due to the fact that the denominator in (5) is always nonzero and therefore T yields zero if the nominator is zero.

4. Summary and discussion

The impact of a sea state–dependent roughness on the atmospheric circulation in a high-resolution limited area model was investigated using Monte Carlo techniques. A number of ensemble runs with and without a sea state–dependent momentum flux was carried out. It was shown exemplarily for SLP that locally large differences on the order of ± 10 hPa can occur between individual members of the ensembles. However, the Monte Carlo experiments showed that these differences usually coincide with a large inherent internal variability of the model and can therefore not be a priori considered

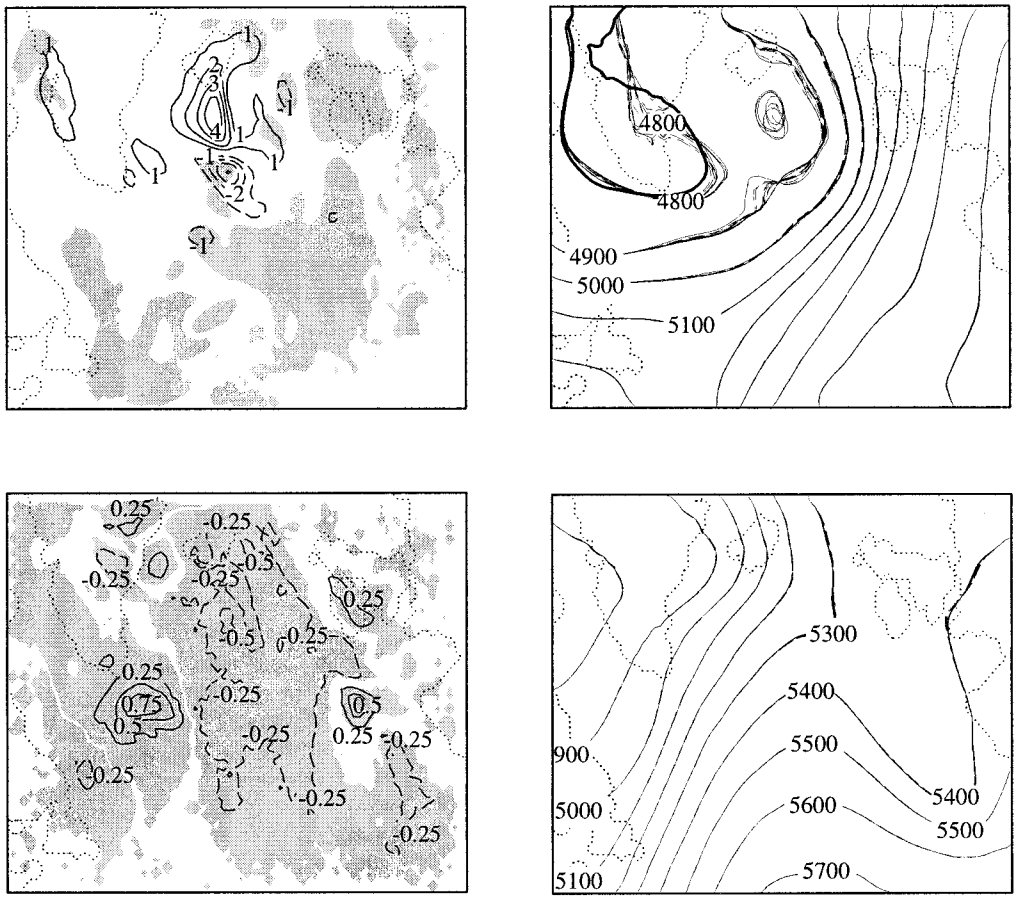

FIG. 8. (left) Local SLP differences in hPa between the ensemble means of the ESD and the CTR simulations; (right) 500-hPa geopotential height in gpm for (solid) the six ESD and (dotted) the six CTR realizations. (top) The situation at 0600 UTC 15 Jan, (bottom) the situation at 0000 UTC 29 Jan (cf. points A and B in Fig. 7). Areas for which the local *t* statistics exceed the value for the 95% confidence limit are indicated in gray. Note that the contour interval varies in the SLP plots.

as model sensitivity to the modified momentum flux parameterization.

We found that the largest SLP differences between the ensembles as well as between individual realizations within an ensemble usually occurred in particular synoptic situations that were characterized by strong baroclinicity and large horizontal SLP gradients. These are situations in which small random variability in the atmospheric model (e.g., small dislocations of depressions or differences in the absolute pressure) lead to large differences, if different realizations are compared with each other. Unfortunately, these situations coincide with those atmospheric conditions in which a strong impact of the sea state on the atmospheric circulation is usually expected (e.g., Doyle 1995; Lionello et al. 1998; Janssen et al. 1997). It is therefore hard to distinguish between the response of the atmosphere model to a modified parameterization of the momentum flux and the inherent model variability. We suggest that it is more likely that in these cases the true signal is obscured by internal atmospheric variability and may not be obtained from the direct intercomparison of two model simulations. Further sensitivity studies in this area therefore require an estimate of the level of internal model variability for assessing the statistical significance of the results obtained.

The application of Monte Carlo techniques to estimate the level of internal model variability is not a new approach. Monte Carlo techniques are a commonly used tool to assess the sensitivity of global climate models (e.g., Cubasch 1985; Cubasch et al. 1994) or the skill of numerical weather forecasts (e.g., Molteni et al. 1996). Janssen and Viterbo (1996) also used this approach to assess the sensitivity of a global coupled atmosphere–wave model. However, for regional atmospheric modeling it is widely believed that the model is to a large extent controlled by the prescribed lateral boundary conditions and that the sensitivity to the initial conditions can be considered as minor. For instance, Sass and Christensen (1995) pointed out that in the context of regional atmospheric modeling the fundamental question to be addressed is how strongly is the interior

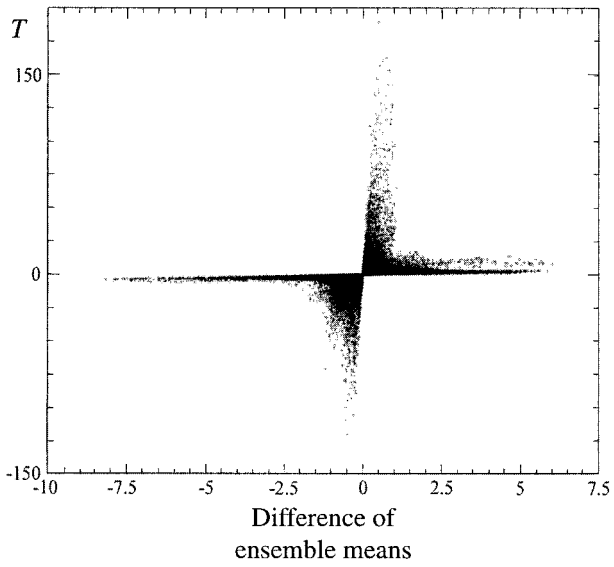

FIG. 9. Local *t* statistics as function of the the local difference of the ensemble means of the six ESD and the six CTR realizations for all grid points and all time steps in Jan.

solution affected by the quality of the boundary conditions. In our experiments the lateral boundaries are always identical. However, we found that under certain atmospheric conditions the interior solutions may also differ considerably as a result of the modified initial conditions. The importance of the boundary conditions for the model solution is nevertheless underlined by the fact that some time after a large scatter of the model solutions was observed, the solutions converged again and the scatter was reduced. Further examples of the need of ensemble calculations with regional atmospheric models in sensitivity studies have recently been provided by, for example, Ji and Vernekar (1997) or Rinke and Dethloff (2000). In atmospheric forecasting the Monte Carlo approach is already widely spread. A number of national weather forecast services use stochastic techniques in combination with regional models to account for the fact that a number of atmospheric states exist in which the atmosphere models behave less stable than in others. For instance, the German Weather Service uses a so-called MIX philosophy, in which the output from several different models is used as different realizations (Balzer 1998). In the context of this philosophy the results of F. Feser (1999, personal communication) are interesting and can be considered as another independent realization of the January 1993 period: Feser forced the regional atmosphere model REMO with the National Centers for Environmental Prediction–National Center for Atmospheric Research reanalysis data using a spectral nudging technique (von Storch et al. 1999) leading to enhanced control of the dynamical evolution by the prescribed large-scale evolution. She found that the REMO model also showed enhanced internal variability in the middle of January 1993.

In previous studies it was often argued that the atmospheric response to a sea state–dependent momentum flux may crucially depend on the model resolution used (e.g., Janssen and Viterbo 1996). In our study we therefore used a horizontal resolution of roughly 50 km $\times$ 50 km, which is comparable to the resolutions used in the studies by Doyle (1995) and Lionello et al. (1998) and even finer than the resolution used by Janssen and Viterbo (1996). However, running a global AGCM at this resolution is currently not feasible. As a compromise we decided to use a limited area model at a sufficiently high spatial resolution. The problem that the interior solution of such a model is to a large extent determined by the boundary conditions was solved by using the same lateral boundary conditions in all experiments such that the obtained response consists solely of the atmospheric response to the modified momentum flux at the sea surface and the internal model variability. The model was set up such that it covers most of the North Atlantic storm track. In this area wind and wave height variability is the largest worldwide and to a large fraction associated with the baroclinic instabilities (e.g., Bauer and Staabs 1998). According to theory, one would therefore expect a larger atmospheric response in this area compared to a region outside the storm track.

While the sensitivity of the atmospheric circulation to the modified momentum flux parameterizations was minor on timescales of months and longer we found that temporary large differences between individual realizations occurred. However, these differences occurred simultaneously with enhanced model variability such that in these situations the detection of an eventually existing sea state–dependent response of the atmospheric circulation was prevented. It should be emphasized that in our experiments the model was run in ''climate mode''; that is, the model was initialized once at the beginning of the simulation and then forced only by the boundary conditions. This is different from operational weather forecasts where the regional models are reinitialized periodically from global model simulations. Hence there remains a possibility that in the latter type of simulations the effect of a sea state–dependent parameterization on the atmospheric circulation may be larger if the model results depend more on the wave-dependent evolution of the initial conditions than on the wave-independent boundary forcing.

Waves play an important role in the air–sea interaction and they determine the stress at the air–sea interface to a large extent (e.g., Janssen 1989; Makin et al. 1995). The explicit account of waves in the calculation of the sea surface roughness may have implications for a number of research areas such as wave, storm surge, or atmospheric modeling. In this study we focused exclusively on the consequences for regional atmospheric modeling. Consequences for wave and storm surge modeling were not considered. The implications for the at-

mospheric circulation were assessed using a rather pragmatic approach. Rather than discussing physical aspects of the air–sea coupling we limit ourselves to a statistical analysis of the atmospheric response using two presently available state-of-the-art momentum flux parameterizations. We found that on longer timescales the frequency distribution of SLP does not change substantially due to the explicit account of waves in the calculation of the sea surface roughness. Because the latter makes the approach rather time consuming from a computational point of view we suggest that presently the Charnock relation provides a sufficiently reasonable parameterization of the momentum flux at the sea surface in regional atmospheric climate models.

Acknowledgments. We thank Vladimir Makin for many fruitful discussions and Beate Gardeike for preparing the figures. Ole Bøssing Christensen kindly provided the boundary data. This work was funded by the European Union under ENV4-CT97-0460 (ASPEN).

REFERENCES

Balzer, K., 1998: Aktuelle Herausforderungen und erste Antworten. *Wettervorhersage,* K. Balzer, W. Enke, and W. Wehry, Eds., Springer-Verlag, 71–92.

Bauer, E., and C. Staabs, 1998: Statistical properties of global significant wave heights and their use for validation. *J. Geophys. Res.,* **103,** 1153–1166.

Chalikov, D., and V. K. Makin, 1991: Models of the wave boundary layer. *Bound.-Layer Meteor.,* **56,** 83–99.

Charnock, H., 1955: Wind stress on a water surface. *Quart. J. Roy. Meteor. Soc.,* **81,** 639–640.

Chervin, R. M., and S. H. Schneider, 1976: On determining the statistical significance of climate experiments with general circulation models. *J. Atmos. Sci.,* **33,** 405–412.

Cubasch, U., 1985: The mean response of the ECMWF global model to the El Niño anomaly in extended range prediction experiments. *Atmos.–Ocean,* **23,** 46–66.

——, and Coauthors, 1994: Monte Carlo climate change forecasts with a global coupled ocean–atmosphere model. *Climate Dyn.,* **10,** 1–19.

Davies, H. C., 1976: A lateral boundary formulation for multi-level prediction models. *Quart. J. Roy. Meteor. Soc.,* **102,** 405–418.

Doyle, J., 1995: Coupled ocean wave–atmosphere mesoscale model simulations of cyclogenesis. *Tellus,* **47A,** 766–778.

Frankignoul, C., 1995: Statistical analysis of GCM output. *Analysis of Climate Variability,* H. von Storch and A. Navarra, Eds., Springer, 139–152.

Gibson, R., P. Kållberg, and S. Uppala, 1996: The ECMWF Re-Analysis (ERA) project. *ECMWF Newsletter,* Vol. 73, 7–17.

Janssen, P. A. E. M., 1989: Wave induced stress and the drag of of air flow over sea waves. *J. Phys. Oceanogr.,* **19,** 745–754.

——, 1991: Quasi-linear theory of wind wave generation applied to wave forecasting. *J. Phys. Oceanogr.,* **21,** 1631–1642.

——, and P. Viterbo, 1996: Ocean waves and the atmospheric climate. *J. Climate,* **9,** 1269–1287.

——, J. D. Doyle, J. Bidlot, B. Hanssen, L. Isaksen, and P. Viterbo, 1997: The impact of ocean waves on the atmosphere. *Proc. Seminar on Atmosphere–Surface Interaction,* ECMWF, Reading, United Kingdom, 85–111.

Ji, Y., and A. D. Vernekar, 1997: Simulation of the Asian summer monsoon of 1987 and 1988 with a regional model nested in a global GCM. *J. Climate,* **10,** 1965–1979.

Källén, E., Ed., 1996: HIRLAM documentation manual, level 2–5. SMHI, 215 pp. [Available from SMHI, S-601 75 Norrköping, Sweden.]

Lionello, P., P. Malguzzi, and A. Buzzi, 1998: On the coupling between the atmospheric circulation and the ocean wave field: An idealized case. *J. Phys. Oceaongr.,* **28,** 161–177.

Makin, V. K., and V. N. Kudryavtsev, 1999: Coupled sea surface–atmospheric model, Part 1, wind over wave coupling. *J. Geophys. Res.,* **104,** 7613–7624.

——, ——, and C. Mastenbroek, 1995: Drag of the sea surface. *Bound.-Layer Meteor.,* **73,** 159–182.

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF ensemble prediction system: Methodology and validation. Part A. *Quart. J. Roy. Meteor. Soc.,* **122,** 73–119.

Rinke, A., and K. Dethloff, 2000: On the sensitivity of a regional Arctic climate model to inital and boundary conditions. *Climate Res.,* **16,** 101–113.

Sass, B. H., and J. H. Christensen, 1995: A simple framework for testing the quality of atmospheric limited area models. *Mon. Wea. Rev.,* **123,** 444–459.

Schär, C., D. Lüthi, U. Beyerle, and E. Heise, 1999: The soil–precipitation feedback: A process study with a regional climate model. *J. Climate,* **12,** 722–741.

Ulbrich, U., G. Bürger, D. Schriever, H. von Storch, S. L. Weber, and G. Schmitz, 1993: The effect of a regional increase in ocean surface roughness on the tropospheric circulation: A GCM experiment. *Climate Dyn.,* **8,** 277–285.

von Storch, H., and F. W. Zwiers, 1999: *Statistical Analysis in Climate Research.* Cambridge University Press, 494 pp.

——, H. Langenberg, and F. Feser, 1999: Long-wave forcing for regional atmospheric modelling. GKSS Rep. 99/E/46, 29 pp. [Available from GKSS Forschungszentrum Geesthacht GmbH, Max-Planck-Str. 1, D-21502 Geesthacht, Germany.]

WAMDI Group, 1988: The WAM Model—A third generation ocean wave prediction model. *J. Phys. Oceanogr.,* **18,** 1776–1810.

Weber, S. L., 1994: Statistics of air–sea fluxes of momentum and mechanical energy in a coupled wave–atmosphere model. *J. Phys. Oceaongr.,* **24,** 1388–1398.

——, H. von Storch, P. Viterbo, and L. Zambresky, 1993: Coupling an ocean wave model to an atmospheric general circulation model. *Climate Dyn.,* **9,** 63–69.

Weisse, R., and E. F. Alvarez, 1997. The European Coupled Atmosphere–Wave–Ocean Model ECAWOM. MPI-Rep. 238, Max-Planck-Institut für Meteorologie, Hamburg, Germany, 34 pp.

Wu, J., 1982: Wind stress coefficient over sea surface from breeze to hurricane. *J. Geophys. Res.,* **87,** 9704–9706.