A Markov Model for Seasonal Forecast of Antarctic Sea Ice

DAKE CHEN AND XIAOJUN YUAN
Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York

(Manuscript received 1 August 2003, in final form 9 February 2004)

ABSTRACT

A linear Markov model has been developed to simulate and predict the short-term climate change in the Antarctic, with particular emphasis on sea ice variability. Seven atmospheric variables along with sea ice were chosen to define the state of the Antarctic climate, and the multivariate empirical orthogonal functions of these variables were used as the building blocks of the model. The predictive skill of the model was evaluated in a cross-validated fashion, and a series of sensitivity experiments was carried out. In both hindcast and forecast experiments, the model showed considerable skill in predicting the anomalous Antarctic sea ice concentration up to 1 yr in advance, especially in austral winter and in the Antarctic dipole regions. The success of the model is attributed to the domination of the Antarctic climate variability by a few distinctive modes in the coupled air–sea–ice system and to the model's ability to detect these modes. This model is presently being used for the experimental seasonal forecasting of Antarctic sea ice, and a current prediction example is presented.

1. Introduction

Antarctic sea ice, with its large seasonal and interannual variabilities, greatly affects the energy balances in the atmosphere and ocean by changing surface albedo, salt injection, and the insolation at the air–sea interface. Long-range forecasts of Antarctic sea ice are very much in demand, not only because of the potential importance of sea ice in the global climate, but also for the practical purpose of exploring the Antarctic continent. Unfortunately, such forecasts are not yet feasible with any state-of-the-art general circulation models (GCMs) because the complex air–sea–ice interaction processes on long time scales are still not well understood and are by no means well simulated by these models. An alternative is to apply statistical methods to Antarctic sea ice prediction. The linear Markov model described here represents one of the first attempts in this direction.

Since satellite observations became available in the 1970s, tremendous efforts have been devoted to investigating Antarctic sea ice variability and its role in the global climate system (e.g., Zwally et al. 1983; Gloersen et al. 1992; Parkinson 1992, 1994; Jacobs and Comiso 1997). It has been shown that to some extent the Antarctic sea ice fields covary linearly with the tropical Pacific El Niño–Southern Oscillation (ENSO; Chiu 1983; Carleton 1989; Simmonds and Jacka 1995; Ledley and Huang 1997; Harangozo 2000; Yuan and Martinson 2000, 2001; Kidson and Renwick 2002; Kwok and Comiso 2002). Various mechanisms have been proposed for the ENSO–high-latitude teleconnection (e.g., Mo and Ghil 1987; Karoly 1989; Mo and Higgins 1998; Renwick and Revell 1999; Rind et al. 2001; Liu et al. 2002). In any case, the low-latitude influence is generally manifested in a few prominent southern high-latitude climate modes, such as the progressive Antarctic circumpolar wave (ACW; White and Peterson 1996) and the standing Antarctic dipole (ADP; Yuan and Martinson 2001).

The ocean, atmosphere, and cryosphere are highly coupled in the southern polar region. Recent studies have revealed strong links among various components of the air–sea–ice system (White and Peterson 1996; Yuan et al. 1999; Venegas and Drinkwater 2001). Anomalous signals in the coupled system have been detected in different polar basins from synoptic to interannual time scales. The sea ice field is clearly driven by the atmosphere through both dynamic and thermodynamic processes (Yuan et al. 1999; Venegas et al. 2001), and the ocean, with its huge heat content, also regulates the ice field. In some cases, there are positive feedbacks from sea ice to the atmosphere and ocean, which reinforces the anomalies in the coupled system and produces persistent and predictable patterns. For example, on interannual time scales, an anomalous poleward heat flux through the atmosphere could reduce sea ice concentration, which means an increased open water area and enhanced heat flux from the ocean to the atmosphere. The heat loss could destabilize the upper
ocean and cause intense overturning. This could bring warm deep water to the surface, allowing more heat to be vented to the atmosphere and more ice to be melted (Martinson 1990).

Therefore, the Antarctic sea ice variability is likely to be controlled by both remote and local processes. The atmospheric anomalies from low latitudes could excite certain modes of the Antarctic climate system, which
could then be amplified and sustained by the local air–sea–ice interaction. Ideally, in order to predict the Antarctic sea ice variation, we should use a sophisticated global GCM that realistically simulates the feedbacks among the ocean, atmosphere, and cryosphere. Since such a model is presently not available, it would be desirable to develop a statistical model to meet the immediate need for Antarctic sea ice prediction. There are at least two possible approaches here: one is to build a regression model that uses relevant equatorial variables as predictors and the other is to construct a self-evolving Markov model based on local variables at high latitudes. In this study, we explore the possibility of the latter using a technique combining multivariate empirical orthogonal function (MEOF) analysis and linear Markov prediction. Our model results indicate that the dominant modes of the Antarctic climate variability are indeed predictable up to 1 yr in advance and that our simple statistical model can serve as a useful tool for Antarctic sea ice forecasting until better dynamical models come along.

2. Model construction

Our model is constructed in the MEOF space. In other words, the base functions of the model’s spatial dependence consist of the MEOFs of several variables that are chosen to define the state of the Antarctic climate, while the temporal evolution of the model is a Markov process with its transition functions determined from the corresponding principal components (PCs). By retaining only a few leading modes of the MEOFs, we can greatly reduce the model space and, more importantly, filter out incoherent small-scale features that are basically unpredictable. This kind of model has been used in some ENSO predictability studies (Blumenthal 1991; Xue et al. 1994; Canizares et al. 2001), in which the outputs from various versions of a dynamic model (Zebiak and Cane 1987; Chen et al. 1995, 2000) were used to compute MEOF bases. Recently, Xue et al. (2000) built a Markov model using observational data and successfully applied it to ENSO forecasting. Our model here follows the same spirit of Xue et al. (2000),
but differs in model variables and in the procedure used to calculate MEOFs.

We chose to define the coupled Antarctic climate system with eight variables: sea ice concentration, surface air temperature (SAT), sea level pressure (SLP), zonal and meridional surface winds, 300-hPa geopotential height, and zonal and meridional winds at the 300-hPa level. Other variables have also been included in some test cases, but their effects were found to be insignificant. The sea ice data were obtained from the National Snow and Ice Data Center and then binned into $0.5^\circ \times 2^\circ$ latitude–longitude grids and monthly intervals. All of the other datasets came from the reanalysis product of the National Centers for Environmental Prediction.
Fig. 4. Model hindcasts at different lead times for anomalous sea ice concentration and associated atmospheric conditions in JJA of 1992.

(NCEP), which are monthly data on a 2.5° × 2.5° grid. The model domain covers the southern polar region (50°–90°S). Twenty-two years (1979–2000) of observational and reanalysis data were used in this study. Since our focus is on climate variability, the climatological seasonal cycle based on this 22-yr period was subtracted to obtain the monthly anomalies for the eight model variables. By normalizing these anomalies and combining them into a single matrix $V$, we can decompose it into MEOFs (spatial patterns) $E$ and their corresponding PCs (time series) $P$:

$$V = EP^T,$$

where the columns of $E$ are orthogonal and the columns
of $\mathbf{P}$ are orthonormal; the superscript $T$ denotes matrix transposition. Reduction in space is achieved by truncating (1) to the first few modes. The Markov model is computed using the single-step correlation matrix, that is, a transition matrix $\mathbf{A}$ that satisfies the linear relation,

$$\mathbf{P}_{i+1} = \mathbf{A}\mathbf{P}_i + e_i,$$

where $i$ denotes the $i$th month and $e$ is the error in the model fit. Multiplying (2) with $\mathbf{P}_i^T$ and averaging over time, we have

$$\langle \mathbf{P}_{i+1}\mathbf{P}_i^T \rangle = \mathbf{A}\langle \mathbf{P}_i\mathbf{P}_i^T \rangle + \langle e_i\mathbf{P}_i^T \rangle,$$

where angle brackets indicate time averaging. There should be no correlation between $e_i$ and $\mathbf{P}_i^T$ for the best model fit, so $\mathbf{A}$ can be calculated as

$$\mathbf{A} = \langle \mathbf{P}_{i+1}\mathbf{P}_i^T \rangle (\mathbf{P}_i\mathbf{P}_i^T)^{-1}.$$
3. MEOF analyses

Before proceeding with model experiments, we first examine some salient features of the Antarctic climate variability by analyzing the leading MEOF modes of the model variables. Since our emphasis is on sea ice, it is weighted 2 times more than other variables in calculating MEOFs. Figure 1 shows the first three MEOF modes of all model variables, which account for about 15%, 8%, and 6% of the total variance, respectively. The first mode is clearly dominated by a dipole pattern in sea ice and SAT, with one pole centered in the central polar Pacific and the other opposite pole in the central polar Atlantic. In association with this large-scale seesaw (the ADP; Yuan and Martinson 2000), the pressure fields are characterized by a Pacific–South American (PSA)–like pattern, as evident in the first MEOF mode of 300-mb geopotential height and, to a lesser extent, in that of SLP. The center of this pressure pattern straddles between the two poles of the ADP, and the associated wind field brings warm air poleward on one side

Here, \( A \) is considered season dependent; thus (4) is actually applied to 12 subsets of PCs to obtain different transition matrices for each of the 12 calendar months. Such a seasonal Markov model has been shown to be more useful than a nonseasonal one in ENSO forecasting (e.g., Xue et al. 2000). Once the MEOF bases and the monthly transition matrices are determined, the model construction is completed. In practice, the procedure of model prediction consists of three steps: first, the PCs corresponding to the initial anomalous climate conditions are calculated by projecting observations to the MEOFs; second, the predictions of the PCs are made at increasing lead times by successively applying the transition matrices; and finally, the predicted PCs are combined with the respective MEOFs to give forecasts of the selected climate variables. Note that this is an anomaly model, and the seasonally varying climatology is predetermined based on the observations from the 1979–2000 period. The forecast of the total field, if desired, is simply the sum of the model-predicted anomaly field and the climatology.
15 AUGUST 2004

Figure 8. Cross-validated model skills in predicting the sea ice anomalies averaged in the DP1 (60°–70°S, 130°–150°W) and DP2 (58°–68°S, 20°–40°W) regions, and in predicting the difference between the two.

Fig. 8. Cross-validated model skills in predicting the sea ice anomalies averaged in the DP1 (60°–70°S, 130°–150°W) and DP2 (58°–68°S, 20°–40°W) regions, and in predicting the difference between the two.

and cold air equatorward on the other. The anomalous heating and cooling caused by these airflows are probably responsible for the formation of the dipole in SAT and sea ice.

There are some indications of wavenumber-3 and wavenumber-2 circumpolar waves in the first and second MEOF modes. The slowly progressive wavenumber-2 wave is the ACW (White and Peterson 1996). Yuan and Martinson (2001) suggested that the ADP and the ACW are actually related, with the latter being excited by the former. The sea ice anomalies can propagate eastward out of the ADP area, probably due to the advection by the Antarctic Circumpolar Current and/or air–sea–ice interaction, but their magnitude decreases away from the source regions and becomes barely detectable in the Eastern Hemisphere. The second MEOF mode looks like the first one being rotated eastward by a quarter of a wavelength. Yuan and Martinson (2001) also noted that the first two modes are in quadrature, meaning that the second mode represents the transition between the two phases of the ADP. The third MEOF mode is basically a summer mode, which is evident in the proximity of the sea ice anomalies to the continent. This mode appears in austral summer and fall when there is usually very little sea ice, while the dipole mode generally occurs in winter and spring when the sea ice coverage is most extensive.

A practical issue in building a Markov model with MEOF bases is the number of leading modes to retain. Using too few of them may miss predictable signals and too many may contaminate the model with noise. The right number is usually achieved by trial and error, and we will show in the next section that seven is the best number of modes in our case. Figure 2 compares the contributions from different numbers of MEOF modes to the sea ice concentration anomalies averaged in the 58°–68°S zonal band. When only the first mode is present, the sea ice variability is limited almost entirely to the standing dipole. When the first three modes are retained, some eastward propagation takes place and more anomalies appear in austral summer and fall. When the first seven modes are included, all the major anomalous signals in the Western Hemisphere are captured, and the differences between this case and the total sea ice field are basically random noise. Although the first seven MEOF modes only account for about 50% of the total variance (it appears to be more in Fig. 2 because of meridional averaging), they seem to contain most of the predictable signals in the sea ice field.

4. Model experiments

In this section, we first evaluate the model with a 22-yr hindcast experiment, then further test the skill and sensitivity of the model in a cross-validated fashion, and finally show some real forecasts, including the latest predictions from our forecast web site.

a. Hindcasts

Hindcast experiments were performed for the period from January 1979 to December 2000. The Markov model was initialized with observational and reanalysis data in each month and predictions were made for up to 12 months for all model variables. Figure 3 shows the correlations between model-predicted and observed sea ice concentration anomalies at different lead times and for different seasons. The hindcast scores are quite remarkable in the ADP regions, especially in austral winter [June–July–August (JJA)]. The model has significant skills in predicting certain aspects of the Antarctic sea ice variability even at a 9-month lead time. As an example of individual predictions, Fig. 4 displays the model hindcasts at different lead times for the winter season of 1992, when a typical dipole pattern occurred in response to the 1991/92 El Niño. The top row (0-month lead) is simply the observations represented by the first seven MEOF modes and thus can be considered as the target. The model did a fairly nice job predicting all of the main features in the observed sea ice and atmospheric variables, though the prediction made 9-months ahead is a bit weak in magnitude. Another example is shown in Fig. 5 for the winter conditions in
2000, when the dipole was in the opposite phase in response to the 1999/2000 La Niña. Again, the model was able to predict the event at least 9 months in advance.

**b. Cross validation and sensitivity tests**

There may be some artificial skills in the hindcasts shown above since the model was trained with the same sets of data as those used to initialize and verify the model. A more convincing way to evaluate model skill is to use a cross-validation scheme (Barnston and Ropelewski 1992), in which the data used to verify the model hindcasts are not used for model training. Thus we built a somewhat different Markov model for each month with a 1-yr moving window of data removed, and then used this window of data to evaluate each model. Figure 6 shows the cross-validated model skills in predicting the average sea ice anomaly at DP1 (60°–70°S, 130°–150°W; the center of the dipole in Pacific). The model beats the persistence prediction by a large amount in terms of both anomaly correlation and root-mean-square error (rmse). Among the four model cases with different numbers of MEOF modes included, the one with seven modes has the highest overall score. In this case, the anomaly correlation is above 0.6 and the rmse is below 9% for nearly all lead times up to almost 1 yr. It is worth noting that the model is not particularly sensitive to the number of modes retained, which is a desirable property.

The contributions of different model components are evaluated in Fig. 7, where cross-validated model skills at DP1 are shown for five cases with different variables included in the model. It is clear that a multivariate model is always better than a sea ice–only model. The SAT, pressure, and wind fields all contribute to improving the model, with the winds being most effective. The best model performance is achieved when all of these variables are included. In order to test the sensitivity of the model to the geographic coverage of the seven atmospheric variables, we successively extended the northern boundary of their domain to 30°S, 10°S, and 10°N. It turns out that the model skill is not much affected by the choice of model domain for the atmospheric variables (not shown). So far we have used the average sea ice anomaly at DP1 as an index to measure the model performance because DP1 is the region where the largest interannual variability is found. Figure 8 compares the model skill at DP1 with that at DP2 (60°–70°S, 40°–60°W; the other pole of the ADP in the At-
The model’s correlation score is lower at DP2, but its rmse is smaller there. This is consistent with the weaker predictable signal and thus the lower signal-to-noise ratio in that region. The difference between DP1 and DP2, which is a measure of the ADP strength, is well predicted by the model. The skill for this index is better than that for either DP1 or DP2 in terms of both anomaly correlation and rmse.

c. Forecasts

We started real-time seasonal forecasting of Antarctic sea ice in the beginning of 2003. Since then, we have been providing forecasts on a monthly basis in our experimental sea ice prediction web site (http://rainbow.ldeo.columbia.edu/forecasts/sea_ice.html) at the Lamont-Doherty Earth Observatory (LDEO), Columbia University. By now our model forecasts can be verified with several months of observational data. An example of such verification is shown in Fig. 9, where the 3- and 6-month lead ensemble forecasts initialized in October–November–December (OND) of 2002 are compared with the observations averaged in January–February–March (JFM) and April–May–June (AMJ) of 2003, respectively. The gross features of the observed sea ice anomaly fields were well predicted while some mismatches remain in details. The model tends to emphasize the sea ice anomalies in the Pacific part of the dipole, which may be statistically correct but is not quite the case for this year. Another problem is that the model underpredicted the large ice increase in the Ross Sea during the early months of 2003. This weakness can be traced back to the initial conditions in OND of 2002; even then, the model was not able to pick up the observed large anomaly in the Ross Sea. This is because such an event rarely occurred in the past two decades and thus was not represented by the leading MEOFs of the model.

Finally, the latest seasonal predictions of Antarctic sea ice concentration as shown on our forecast web site are displayed in Fig. 10. These are forecasts for the next four seasons, starting from the observed sea ice and atmospheric conditions in OND of 2003. Both anomalous (Fig. 10a) and total (Fig. 10b) sea ice concentration predictions are given, with the latter being simply the sum of the former and the climatology based on the 1979–2000 period. According to our model forecasts, ice concentration would be below normal in the Pacific sector (Amundsen Sea) while above normal in the Atlantic sector (Weddell Sea) in the austral summer. As the ice extent increases during the following autumn and winter seasons, the ice anomalies would intensify and become concentrated along the ice edge, with a slow eastward propagation. They would then weaken considerably toward the end of the year. It remains to be seen whether or not such a sequence will actually take place in the coming months.

5. Summary and discussion

We have developed a low-order linear Markov model to simulate and predict the short-term climate change in the Antarctic, with particular emphasis on sea ice variability. Seven atmospheric variables along with sea ice were chosen to define the state of the Antarctic climate, and the multivariate empirical orthogonal functions of these variables were used as the building blocks of the model. The predictive skill of the model was evaluated in a cross-validated fashion, and a series of sensitivity experiments were carried out. In both hindcast and forecast experiments, the model showed considerable skill in predicting the Antarctic sea ice anomalies a few seasons in advance, especially in austral winter and in the Antarctic dipole regions. We are presently using this model for experimental seasonal forecasting of Antarctic sea ice, and our predictions are updated on a monthly basis in the sea ice forecast web site of the Lamont-Doherty Earth Observatory. Such forecasts are expected to be useful for planning Antarctic field expeditions.

It is somewhat surprising that such a simple statistical model could have fared so well in predicting the interannual variations of the Antarctic sea ice field, a task that has rarely been attempted before. In fact, the forecast skill of this model in predicting the Antarctic sea ice anomalies is comparable to that of the state-of-the-art ENSO forecast models in predicting the sea surface temperature anomalies in the tropical Pacific Ocean. It is truly remarkable for a statistical model to have cross-validated anomaly correlations above 0.6 for lead times up to almost 1 yr, based on more than two decades of monthly samples. Admittedly, this kind of score is only achieved for certain indices such as DP1 or DP1 − DP2, but so is the present skill in ENSO forecasting; little predictability is found for sea surface temperature anomalies away from the eastern and central equatorial Pacific (i.e., aside from the commonly used Niño-3 and Niño-3.4 indices).

We draw a specific analogy to ENSO here because it is the largest and perhaps the most predictable interannual signal in the earth’s climate system. The predictability of ENSO largely relies on the strong ocean–atmosphere interaction in the tropical Pacific and on the low-dimensional nature of this tightly coupled system. Likewise, the predictability of the Antarctic interannual variations demonstrated here can be attributed to the domination of the coupled air–sea–ice system by a few distinctive, slowly changing modes such as the ADP and the ACW, although an ENSO-like self-sustained low-frequency oscillation is not likely to exist in the Antarctic. Our understanding of the physical processes operating in the Antarctic climate system is still rather limited compared with that of ENSO, but we do have a good grasp of the statistical characteristics of the system, and a well-constructed statistical model can be a
useful forecast tool as long as it contains those dominant climate modes.

The performance of this Markov model is quite robust. It is not particularly sensitive to the number of MEOF modes retained, nor is it sensitive to the size of the model domain for the atmospheric variables. Because of the high coherency of these model variables and the redundant information contained in them, we could even drop some of the variables without too much skill reduction. However, the wind fields seem to be too important to neglect. This is alarming because we know that the quality of the NCEP reanalysis winds used here is questionable in the polar regions. Yet this is also encouraging because we know that the surface winds from satellite scatterometers have shown great potential for Antarctic application (Yuan et al. 1999). Although the satellite-derived wind products are still too short for model training and they are not available over ice, they could be useful for our model initialization when combined with the reanalysis wind field.

As mentioned earlier, this self-evolving Markov model emphasizes the regional processes of air–sea–ice interaction, and as such it does not explicitly address the issue of the Antarctic–low-latitude teleconnection. However, the success of the model does not compromise in any way the importance of the teleconnection. Since the Antarctic interannual disturbances are likely to be excited in the first place by the influence from low latitudes, as evident in their lagged response to ENSO, any models that simulate the evolution of these disturbances implicitly take into account the teleconnection. All the model does is detect significant signals from observed
initial conditions and predict a statistically meaningful path for the movement, growth, or decay of these signals. In principle, such a multivariate model can be used to predict other variables as well as sea ice, but the present version of the model was not really designed for that because a relatively heavy weight was assigned to the sea ice when building the MEOF bases of the model.

**Acknowledgments.** This research is supported by the National Aeronautics and Space Administration through Grant NAG5-12587. We thank Dr. T. Vihma and an anonymous reviewer for their helpful comments on an earlier version of the manuscript.

**REFERENCES**


