GEOPHYSICAL RESEARCH LETTERS, VOL. 31, L09212, doi:10.1029/2003GL019209, 2004

A statistical approach to Indian Ocean sea surface temperature prediction using a dynamical ENSO prediction

Jong-Seong Kug and In-Sik Kang
Climate Environment System Research Center, Seoul National University, Seoul, South Korea

June-Yi Lee
Laboratory for Atmosphere, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA

Jong-Ghap Jhun
School of Earth and Environmental Sciences, Seoul National University, Seoul, South Korea

Received 3 December 2003; accepted 19 April 2004; published 15 May 2004.

[1] In this study, a statistical prediction model has been developed to forecast monthly Sea Surface Temperature (SST) in the Indian Ocean. It is a linear regression model based on a lagged relationship between the Indian Ocean SST and the NINO3 SST. A new approach to the statistical modeling has been tried out, in which the model predictors are obtained from not only observed NINO3 SST but also predicted results produced by a dynamical El Niño model. The forecast skill of the present model is better than that of persistence prediction. In particular, the present model has a significantly improved predictive skill during the spring and summer seasons when the boreal summer Indian monsoon is affected by the Indian Ocean SST. *INDEX TERMS:* 4263 Oceanography: General: Ocean prediction; 4219 Oceanography: General: Continental shelf processes; 3339 Meteorology and Atmospheric Dynamics: Ocean/atmosphere interactions (0312, 4504); 3210 Mathematical Geophysics: Modeling; 1620 Global Change: Climate dynamics (3309). **Citation:** Kug, J.-S., I.-S. Kang, J.-Y. Lee, and J.-G. Jhun (2004), A statistical approach to Indian Ocean sea surface temperature prediction using a dynamical ENSO prediction, *Geophys. Res. Lett.*, *31*, L09212, doi:10.1029/2003GL019209.

1. Introduction

[2] As understanding of ENSO dynamics has accumulated over the last two decades, a number of studies have been devoted to a better prediction for the tropical Pacific SST. The tropical Pacific SST has been operationally forecasted by several dynamical and statistical models, which have a predictive skill of up to a 12-month lead time for the equatorial eastern Pacific SST. However, little research effort has been devoted to predicting ocean areas other than the tropical Pacific.

[3] Recently, there has begun to emerge a need for global SST forecasts for boundary conditions of atmospheric GCMs in long-range forecasting. In addition, the climate variability of many local areas is significantly affected by the regional ocean SST rather than the tropical Pacific SST. For example, the Indian Ocean SST variation influences the regional climate over the Asian continent [e.g., *Meehl and Arblaster*, 2002; *Krishnan et al.*, 2003] and Africa [e.g., *Goddard and Graham*, 1999]. The Indian Ocean SST affects the regional climate mostly by a modulation of the monsoon system. Therefore, there is a need for an a priori indication of the expected seasonal variability over the Indian Ocean.

[4] In this study, we have focused on Indian Ocean SST prediction in spite of the request for global SST prediction. It is widely recognized that the warm and cold events of the Indian Ocean SST are strongly linked to those of the equatorial eastern Pacific [*Klein et al.*, 1999; *Xie et al.*, 2002]. *Klein et al.* [1999] suggested that over the eastern Indian Ocean, enhanced subsidence during El Niño reduces cloud cover and increases the solar radiation absorbed by the ocean, thereby enhancing SST warming. Therefore, the eastern Indian Ocean SST lags the ENSO SST forcing since the ocean has a large heat capacity, although atmospheric teleconnection gives a simultaneous response. Recently, *Xie et al.* [2002] showed that the anomalous equatorial wind associated with the ENSO in the equatorial Indian Ocean can explain the warming of the western Indian Ocean. The equatorial easterly wind produces an anticyclonic wind stress curl, and forces downwelling Rossby waves off the equator. The downwelling Rossby waves, which result in SST warming, propagate slowly westward. Due to the propagation time of the Rossby waves, there is still a lag time with the eastern Pacific warming associated with ENSO. The lag relation between the Indian Ocean SST and the eastern Pacific SST is of great importance. This indicates that the Indian Ocean SST can be empirically predicted using the eastern Pacific SST information. In this study, we have constructed a statistical model for Indian Ocean SST prediction based on this lag relation.

[5] A few modeling efforts have been devoted to the equatorial Indian Ocean. Recently, *Landman and Mason* [2001] developed a statistical model for predicting near-global SST using Canonical Correlation Analysis (CCA). They showed that forecast skill for the Indian Ocean outscores persistence after a 3-month lead time. They mentioned that the temporal correlations for the Indian Ocean Index, representing basin-wide warming and cooling, are slightly less than 0.5 for 0-, 3-, 6- and 9-month lead times.

[6] In this study, a statistical model has been developed for the Indian Ocean SST forecasting. The model is based on the lagged linear regression of the Indian Ocean SST

0094-8276/04/2003GL019209$05.00

with respect to the NINO3 SST. A distinct difference from other statistical models is that the predictors of the model are not only observed NINO3 SST but also SST forecast by a dynamical El Niño prediction model. The forecast skill of the statistical model will be investigated and compared to that of persistence prediction.

2. Statistical Model

[7] As mentioned in the introduction, the Indian Ocean SSTs have a significant lag relation with the tropical Pacific SST. This indicates that the Indian Ocean SST can be empirically predicted using the Pacific SST information. In this study, the statistical model is developed to predict the monthly Indian Ocean SSTs. The model is based on a lagged linear regression, hereafter referred as "LLR". The predictor of the model is monthly NINO3 SST averaged over the region 150°–90°W and 5°S–5°N and the predictand is the SST at each grid point in the Indian Ocean. In order to construct the statistical model, we introduce a fitting period and training period. The fitting period is recent 12 years from time of the forecast (or hindcast). This period is used to calculate linear regression coefficient. We checked that general performance of the model is not sensitive to the selection of the fitting period. The training period is for determining optimal lag between predictor and predictand. Although the Indian Ocean SSTs have a significant lag relation with the eastern Pacific SST, the linear relationship and the lag time can be variable in grid points, seasons, and decades. Therefore, it is critical to select optimal lag and linear relationship for model performance. In order to consider two variable factors, the linear regression equation is constructed as follows:

$$SST(x, y, t_f) = \alpha(x, y, t_f) \cdot NINO3(t_f - lag)$$

$$\text{where } \alpha(x, y, t_f) = \frac{1}{lyr} \sum_{t=t_f-lyr}^{t_f-1} \frac{\mathrm{SST(x, y, t)NINO3(t - lag)}}{\sigma^2_{\mathrm{NINO3}}}$$

where t_f denotes a target time of the forecast. SST and NINO3 indicate the monthly Indian Ocean SST (the predictand) and monthly NINO3 SST (the predictor), respectively. α represents a linear regression coefficient, which is calculated during the fitting period. *lyr* is the total number of model fitting year, 12. σ_{NINO3} is a standard deviation of the NINO3 SST. *lag* indicates a lag month between the predictor and the predictand. It is usually important to determine the value of the *lag* in order that a statistical model has a high forecast skill. In the present model, the *lag* is automatically selected by a retrospective prediction process. In advance of real prediction, the retrospective predictions (i.e., hindcast) for the past SST are performed at various lag times from 1 month to 18 months and their skills are compared in order to determine the optimal value of the *lag*. Then, the value of the *lag* is determined as the one having the highest predictive skill for the training period.

[8] For the training period, all available historical SST data from NCEP [*Reynolds and Smith*, 1994] were used from the year 1962 to the time of forecasting. For example, when we try to forecast the SST in July 1982, the retrospective predictions for training are performed for 20 cases from July 1962 to July 1981. From the hindcast skill of the 20 cases, optimal lag is selected and real prediction is performed with the lag. Next, when we try to forecast the SST in July 1983, the training is performed for 21 cases. In a similar manner, the training period is gradually increased, as long as the data are available. If the skill of the retrospective predictions for the training period is lower than that of the persisted SST, the model selects the persisted SST rather than that forecast by the LLR. Note that the whole prediction strategy is stratified by calendar month because the Indian Ocean variability is strongly phase-locked to the seasonal cycle [*Xie et al.*, 2002]. Following this strategy, the Indian Ocean SSTs were predicted during the period Jan. 1981–Dec. 2000.

[9] Most statistical models have used observed data as a predictor. In this case, the predictor must be past values at the initial time of the forecast. For example, when we try to predict with a 6-month lead time, we cannot use the predictor for the 1–6 month lag times because observations are not available at that time. Therefore, the forecast skills of most statistical models are unavoidably degraded as the forecast lead time increases.

[10] It is well known that the NINO3 SST, predictor of the present model, is predictable for long lead times. Therefore, it is possible to use the NINO3 SST forecast as a predictor of the present statistical model. Thus, we have constructed two types of model in this study. One is a model which used only the observed NINO3 SST as predictor. Hereafter, it will be referred to "LLRobs". The other is a model which used both the observed NINO3 SST and those forecast by a dynamical prediction model. It will be refer as "LLRfcst". In the LLRfcst, the predictor is a forecast NINO3 SST wherein the forecast lead time is longer than the lag time between predictor and predictand. For example, when the optimal lag is determined 4-month lag by the retrospective skill, the NINO3 forecast is used for 4-month and the longer lead time forecast.

[11] The dynamical El Niño prediction model used in the NINO3 SST prediction is based on the intermediate coupled model developed by *Kang and Kug* [2000]. The model has a predictive skill up to a 12-month lead time judged to have a correlation of more than 0.6 [*Kug et al.*, 2001].

3. Predictions

[12] Using each of the LLRobs and LLRfcst, Indian Ocean SSTs were predicted with a 12-month lead time during the period Jan. 1981–Dec. 2000. Forecast skill for the total of 240 predictions was calculated and compared to that of the persistence prediction. We defined the persistence prediction as a prediction for which the SST anomaly from the monthly annual cycle is persisted in forecasting the target month. In order to readily compare the forecast skill of the present model, an Indian Ocean SST Index is defined as an area-averaged SST over 48°–104°E, 6°S–6°N. The definition is the same as that of *Landman and Mason* [2001]. The index represents a basin-wide warming and cooling over the Indian Ocean basin.

[13] Figure 1 shows forecast skill for the Indian Ocean SST index as a function of calendar month. The forecast

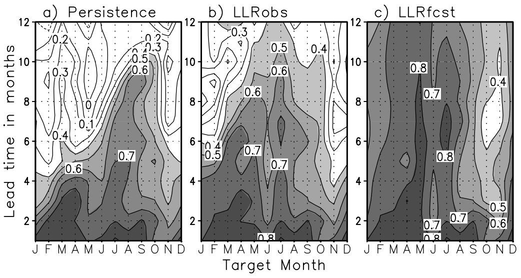

Figure 1. Correlation coefficients between the observed and forecast Indian Ocean SST Index as a function of calendar month. (a) persistence prediction, (b) LLRobs and (c) LLRfcst.

skill is represented as a correlation coefficient between the observed and forecast values for each lead time over the forecast period of 20 years. The x-axis of Figure 1 indicates the target month for SST prediction. The persistence prediction has a low forecast skill, and the correlation coefficients during some seasons degrade less than 0.5 for a 4-month lead time. This implies that it is not correct to use the persisted SST as a boundary condition for the atmospheric GCM for more than one-season forecast. It is worthwhile to note that the forecast skill of the persisted SST has a strong seasonality. While the SST prediction for July to September has a higher forecast skills with a correlation coefficient exceeding 0.5 up to a 9-month lead time, the forecast skill for November to April SST is rapidly degraded. Figure 1b shows the forecast skill of the LLRobs. The seasonality of the correlation is similar to that of the persistence. The forecast skill for the winter SST appears to have a lower correlation compared to that for other seasons. However, the forecast skill for the spring and summer SST is improved compared to that of the persistence.

[14] The forecast skill for the LLRfcst is shown in Figure 1c. The correlations over most seasons are higher than those of the persistence and the LLRobs. However, the correlations for the October and November SST are still significantly lower compared to those for other seasons. One may explain this in terms of the Indian Ocean dipole events contributing to the low predictive skill. *Landman and Mason* [2001] showed using their CCA statistical model that forecast skill for the Indian Ocean SST is poor during boreal autumn when the Indian Ocean dipole mode is active. *Saji et al.* [1999] argued that all Indian dipole events do not concurred with the ENSO events, although there are still some disputes as to their dependence on ENSO [e.g., *Baquero-Bernal et al.*, 2002]. Also, the dipole-like mode tends not to lag but to lead tropical Pacific SST as is apparent from comparing their statistical relations. Therefore, it can be expected that with the present prediction model, which utilizes lag relation with ENSO, it is relatively hard to predict the SST variation associated with the Indian dipole events.

[15] Note that the LLRfcst has a superior predictive skill during the spring and summer. The correlation is more than 0.7 up to a 12-month lead time. The SST variation during these seasons is strongly related to the Indian summer monsoon [*Meehl and Arblaster*, 2002; *Krishnan et al.*, 2003]. Therefore, we expect that the present model can contribute to improving the predictive ability for the Indian summer monsoon.

[16] Retroactive forecast values for each individual month of the period 1981–2000 for the Indian Ocean Index, at 3-,6-,9-, and 12-month lead times are shown in Figure 2. The predictions were made by the LLRfcst. The Indian Ocean Index has a strong ENSO-related variation. Its warm and cold events are concurrent with El Niño and La Niña events with a time-delay. Since the present statistical model uses the NINO3 SST as a predictor, the model predicts well warm and cold events over the Indian Ocean. The temporal correlations between the observed and predicted indices are 0.70, 0.67, 0.65 and 0.65 for the four respective lead times. The skills seem to be better than that of other statistical models in previous studies although direct comparison is impossible. In addition, the model successfully predicts the anomalous warm Indian Ocean SST that has been persisting since the mid-1970s when a large positive shift in SST occurred. Note that the successive prediction of the decadal variation of the Indian Ocean SST was not reflected in the correlation coefficients used to measure forecast skill because the SST anomaly is calculated from the 20-yr climatology during 1981–2000.

[17] A map of the retroactive temporal correlations between the persistence and observed SSTs, and between predicted and observed SSTs are shown in Figure 3. For a 6-month lead forecast, the persistence prediction has a predictive skill over the equatorial central Indian Ocean during the boreal fall season. However, it has a low correlation over the other regions and other seasons. On the other hands, the LLRfcst, in general, has a better predictive skill than persistence prediction. During the boreal winter time, the LLRfcst has a higher forecast skill over the southern Indian Ocean. In this region, SST variability is related to the oceanic Rossby wave propagation associated with the ENSO-related wind stress forcing [*Xie et al.*, 2002]. There-

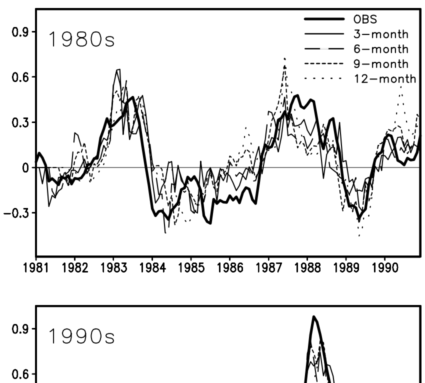

Figure 2. Time series of the observed (thick line) and forecast (thin) Indian Ocean SST index. Solid, long-dashed, short-dashed and dotted lines are 3-, 6-, 9- and 12-month lead forecasts, respectively.

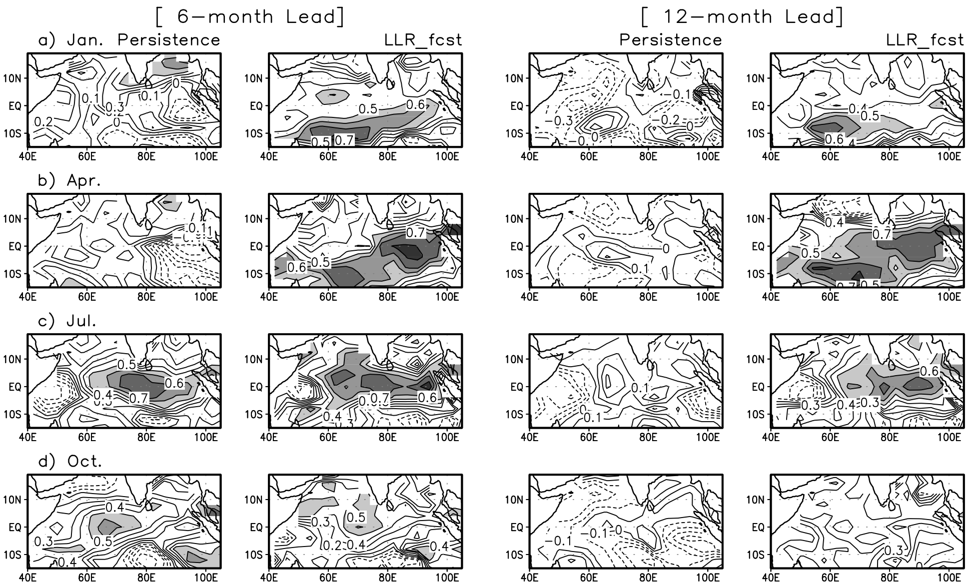

Figure 3. Correlation coefficients between the observed SST and that forecast SST by persistence prediction (Left Panel) and LLRfcst (Right Panel) during Jan. (a), Apr. (b), Jul. (c), and Oct. (d), respectively.

fore, the model can predict well the SST variation in this region. During spring and summer times, the model has a better skill over the equatorial Indian Ocean. This is relevant to basinwide warming and cooling events in the Indian Ocean associated with ENSO as several studies pointed out [*Venzke et al.*, 2000]. For a 12-month lead forecast, the persistence does not have a predictive skill over all the Indian Ocean while the LLRfcst has still predictive skill over the Indian Ocean. Judging with a correlation of more than 0.5, the equatorial Indian Ocean SST can be predicted up to a 12-month lead time during the boreal spring and summer times.

4. Summary and Discussion

[18] A statistical prediction model for the Indian Ocean SST was constructed based on the lag relationship between the NINO3 SST and the Indian Ocean SST. A distinct difference from other statistical models is that the predictors of the present model are obtained from not only observations but also prediction results produced by a dynamical model. Predictive ability of most statistical models degrades as the forecast lead time increases because observational predictors are restrictive, while the present model uses forecast of dynamical model as a predictor, so that the skill of the statistical model depends on the skill of the dynamical model used for the predictor. Because the NINO3 SST, the predictor of the present model, is predictive for long lead times, the LLRfcst could have better skill than LLRobs. If the dynamical model has a high skill for the NINO3 SST, the statistical model can maintain the forecast skill up to a long lead forecast.

[19] The forecast skill of the present model has been examined and compared to that of the persistence prediction and the LLRobs. The model predicts well basin-wide warming and cooling events in the Indian Ocean associated with the El Niño and La Niña. The forecast skills are considerably better than those of the persistence and LLRobs over all the forecast lead time. Especially, the present model has a better predictive skill during the spring and summer season when the summer monsoon is modulated by the Indian Ocean SST. The correlation score for the Indian Ocean SST index is more than 0.5 up to a 12-month lead time.

[20] We applied the present statistical model to other oceans as well as the Indian Ocean (not shown). The model has a predictive skill up to a 12-month lead time over the tropical Pacific Ocean as well as the Indian Ocean. However, over the other ocean regions, the forecast skill of the present model is very poor. This implies that the other ocean SSTs are not linearly affected by the ENSO forcing but are mostly governed by different mechanisms unlike the Indian Ocean SST. Therefore, it may need different approaches in order to predict the other Ocean SST, in order to provide the boundary condition of the atmospheric GCM for long-range global climate prediction.

[21] **Acknowledgments.** This research was performed for the project ''Development of forecast techniques for the East Asian monsoon-Changma circulation system'', one of Meteorological and Earthquake R&D programs funded by KMA. Dr. Kug was supported by the SRC program of the Korean Science and Engineering Foundation. Prof. Kang acknowledges the support of the National Computerization Agency through Application Research Project using KOREN.

References

Baquero-Bernal, A., M. Latif, and L. Stephanie (2002), On dipolelike variability of sea surface temperature in the tropical Indian Ocean, *J. Clim.*, *15*, 1358–1368.

Goddard, L., and N. E. Graham (1999), Importance of the Indian Ocean for simulating rainfall anomalies over eastern and southern Africa, *J. Geophys. Res.*, *104*, 19,099–19,116.

Kang, I.-S., and J.-S. Kug (2000), An El Niño prediction system using an intermediate ocean and a statistical atmosphere, *Geophys. Res. Lett.*, *27*, 1167–1170.

Klein, S. A., B. J. Soden, and N. G. Lau (1999), Remote sea surface temperature variation during ENSO: Evidence for a tropical atmosphere bridge, *J. Clim.*, *12*, 917–932.

Krishnan, R., M. Mujumdar, V. Vaidya, K. V. Ramesh, and V. Satyan (2003), The abnormal Indian summer monsoon of 2000, *J. Clim.*, *16*, 1177–1194.

Kug, J.-S., I.-S. Kang, and S. E. Zebiak (2001), The impacts of the model assimilated wind stress data in the initialization of an intermediate ocean and the ENSO predictability, *Geophys. Res. Lett.*, *28*, 3713–3716.
Landman, W. A., and S. J. Mason (2001), Forecast of near-global sea surface temperature using canonical correlation analysis, *J. Clim.*, *14*, 3819–3833.
Meehl, G. A., and J. M. Arblaster (2002), Indian monsoon GCM sensitivity experiments testing tropospheric biennial oscillation transition conditions, *J. Clim.*, *15*, 923–944.
Reynolds, R. W., and T. M. Smith (1994), Improved global sea surface temperature analysis using optimum interpolation, *J. Clim.*, *8*, 929–948.
Saji, N. H., B. N. Goswami, P. N. Vinayachandran, and T. Yamagata (1999), A dipole mode in the tropical Indian Ocean, *Nature*, *401*, 360–363.
Venzke, S., M. Latif, and A. Villwock (2000), The coupled GCM ECHO-2. Part II: Indian Ocean response to ENSO, *J. Clim.*, *13*, 1371–1383.
Xie, S.-P., H. Annamalai, F. A. Schott, and J. P. McCreary Jr. (2002), Structure and mechanisms of south Indian Ocean climate variability, *J. Clim.*, *15*, 864–878.

J.-G. Jhun, School of Earth and Environmental Sciences, Seoul National University, Seoul, 151-742, Korea.

I.-S. Kang and J.-S. Kug, School of Earth and Environmental Sciences, Seoul National University, Seoul, 151-742, Korea. (kang@climate.snu.ac.kr; jskug@climate.snu.ac.kr)

J.-Y. Lee, Laboratory for Atmosphere, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA.