Atlantic Atmosphere–Ocean Interaction: A Stochastic Climate Model–Based Diagnosis

TIMOTHY J. MOSEDALE AND DAVID B. STEPHENSON
Department of Meteorology, University of Reading, Reading, United Kingdom

MATTHEW COLLINS
Hadley Centre, Met Office, Exeter, United Kingdom

(Manuscript received 17 November 2003, in final form 15 September 2004)

ABSTRACT

A simple linear stochastic climate model of extratropical wintertime ocean–atmosphere coupling is used to diagnose the daily interactions between the ocean and the atmosphere in a fully coupled general circulation model. Monte Carlo simulations with the simple model show that the influence of the ocean on the atmosphere can be difficult to estimate, being biased low even with multiple decades of daily data. Despite this, fitting the simple model to the surface air temperature and sea surface temperature data from the complex general circulation model reveals an ocean-to-atmosphere influence in the northeastern Atlantic. Furthermore, the simple model is used to demonstrate that the ocean in this region greatly enhances the autocorrelation in overlying lower-tropospheric temperatures at lags from a few days to many months.

1. Introduction

The coupling between the extratropical ocean and the atmosphere, and in particular the small effect that the ocean has on the atmosphere, is difficult to understand and quantify in observations and models (Kushnir et al. 2002). Studies suggest that there is a small positive effect of the midlatitude ocean on the large-scale atmospheric circulation (e.g., Robinson 2000). Many of these studies rely on descriptive statistical covariance/correlation techniques such as maximum covariance analysis to identify the main coupled patterns of variation in the ocean and in the atmosphere (e.g., Czaja and Frankignoul 2002; Rodwell and Folland 2002; Deser and Timlin 1997). Lag correlations between oceanic fields and subsequent atmospheric fields are interpreted as causal influences from the ocean to the atmosphere (e.g., Frankignoul et al. 1998; von Storch 2000). Here, a more rigorous time series modeling framework based on a simple coupled stochastic model is used to diagnose the coupling and causality between the atmosphere and the ocean.

Stochastic models have been used in coupled climate modeling since Hasselmann (1976) and Davis (1976). Frankignoul and Hasselmann (1977) developed a simple stochastic atmosphere coupled to a slave ocean and used this to model observed variability in the atmosphere and ocean. Barsugli and Battisti (1998, hereafter BB98), based on the work of Barsugli (1995), proposed a similar two-variable autoregressive model that incorporates two-way interactions between the extratropical atmosphere and the ocean. In BB98 and later in Bretherton and Battisti (2000), this model was used to investigate coupling by comparing output from the model run in coupled and uncoupled modes. In this study, the model is used to diagnose interactions in a more complex coupled general circulation model (GCM).

The following section describes the simple stochastic model and the GCM data. Monte Carlo simulations of the simple model are described in section 3. The simple model is fit to the complex model data, and results of how the ocean affects the autocorrelation in the atmosphere are given in section 4. Conclusions and some future directions are presented in the final section.

2. Models

The aim of this work is to use a very simple model of ocean–atmosphere interaction to understand the interactions in a very complex model. To achieve this, the simple model should be simple enough that the interactions in this model can be easily understood.

Figure 1 shows a schematic of ocean–atmosphere interaction in coupled systems—the state of the atmo-
The simple model used here models each of these sets of processes by a single numerical value that represents local ocean–atmosphere interactions.

*a. The complex model: HadCM3*

The complex model investigated in this study is the Third Hadley Centre Coupled Ocean–Atmosphere GCM (HadCM3), which has been used extensively in previous climate variability, sensitivity, and prediction studies (e.g., Gordon et al. 2000; Collins et al. 2001). Fifty years of daily model output has been analyzed to allow the day-to-day coupling between the ocean and the atmosphere to be examined, as these processes occur on time scales shorter than monthly averages (Wallace et al. 1990; Ciasto and Thompson 2004). The time scale of 1 day allows for passing weather systems to be resolved, while being long enough so that the effects of the diurnal cycle can be neglected. The model ocean is coupled to the atmosphere in HadCM3 once per day, making the daily time step a natural choice. Daily ocean fields were not previously archived from the HadCM3 control simulation so a separate 50-yr simulation had to be made. The thermal variables chosen are surface air temperature (at 1.5 m) and sea surface temperature (SST). These are representative of the two well-mixed layers on either side of the air–sea interface. The air temperature at 1.5 m was chosen rather than at a higher level as heat fluxes between the atmosphere and ocean depend on these surface properties, so the surface air temperature is an obvious choice.

Anomalies were calculated by removing the mean and the annual cycle composed of the mean annual and semiannual harmonics at each grid point (note that in this version of the model a 360-day calendar was adopted). The harmonics were obtained by Fourier analysis of the 50 years of the simulation. The domain was chosen to be the northern and tropical Atlantic (20°S–87.5°N, 101.25°W–22.5°E). Only long winters have been considered (October–March inclusive), and the time series of atmospheric and oceanic anomalies are both standardized by dividing by their respective standard deviations at each grid point to facilitate comparison of results at different geographical locations. The resulting atmospheric temperature anomalies are denoted by $T$, and the SST anomalies by $S$. Figure 2 shows some basic summary statistics: the sample means and standard deviations prior to standardizing and the lag-1 autocorrelation of each time series. By comparing Figs. 2e and 2f, it can be seen that the autocorrelation is much greater in the languardus ocean than in the more volatile atmosphere.

*b. The simple model: Barsugli and Battisti (1998)*

The simple model used here is based on the two-variable linear stochastic model proposed by BB98 and Barsugli (1995). The model, as explained in BB98, is derived from a local energy balance model for a one-layer extratropical atmosphere overlying a slab ocean. The stochastic forcing is assumed to be atmospheric only and represents internal atmospheric wintertime variability. The rate of change of the atmospheric and oceanic variables are linear functions of the current state of the system with model parameters defined by physical values of quantities such as the heat capacities of the atmosphere and ocean and the fluxes between them.

In this study, a discrete time representation of the model is employed, with model equations forming the following first-order vector autoregressive [var(1)] model:

$$T_{n+1} = aT_n + bS_n + e_n^T,$$

$$S_{n+1} = cT_n + dS_n + e_n^S,$$

where $e_n$ is Gaussian white noise forcing. Note that, in contrast to the continuous form given in BB98, this discrete form includes a stochastic term in the ocean component. This enables more flexibility in the model interpretation and represents effects of short time-scale oceanic noise, for example, from advective or mixing processes. The stochastic term for the ocean has a much smaller amplitude than that for the atmosphere ($a$ for $e^T$ compared to $c$ for $e^S$; values follow).

Note that the $a$, $b$, $c$, and $d$ in this matrix correspond to those shown schematically in Fig. 1 (not those used in BB98). Parameter $a$ determines how much of today’s atmosphere affects tomorrow’s atmosphere, and pa-
Parameter $d$ similarly for the ocean. Parameter $b$ is of great interest since it quantifies how much today’s ocean influences tomorrow’s atmosphere—the causal effect of the ocean on the atmosphere. Parameter $c$ quantifies how much today’s atmosphere influences tomorrow’s ocean. Typical midlatitude values derived from BB98 are $(a, b, c, d) = (0.79, 0.092, 0.0046, 0.995)$.

Figure 3 shows time series from a simulation of this model and time series simulated by the complex GCM. Although there is by definition no seasonal dependence in this simple model, half of the year designated “summer” is removed for this model data as was done for the complex GCM data that does have seasonal variations. The simple model ocean and atmosphere time series...
are similar to those from the GCM, with a noisy-looking atmospheric time series and a smoother ocean time series with long-term (up and down) trends. The behavior of the atmosphere in the Tropics is different from the extratropics, as the tropical atmosphere is less prone to daily variations from passing baroclinic systems and has a slower trending behavior coupled to that of the ocean. Trends in the ocean and the atmosphere in Figs. 3e and 3f are clearly collocated, suggesting that interaction between the ocean and the atmosphere is important in this region.

3. Monte Carlo simulations with the simple model

Before fitting the simple model to the complex model data, tests of the statistical techniques were performed using a simulation procedure whereby the simple model was fitted to data generated by that same model. Ordinary least squares regression is used to estimate separately the model parameters in Eqs. (1) and (2). For simulations from the simple model, the estimates should be able to recover the original parameters used to perform the simulation (up to the addition of sampling errors).

a. Fifty-year simulations

One thousand Monte Carlo simulations of 50-yr length were performed using the simple model. Table 1 shows the median and standard deviation of the distributions of the parameter estimates. The estimates for parameters \( a, c, \) and \( d \) are all close to the value used to make the simulations, and the errors are small, as expected for a large Monte Carlo simulation of long daily datasets (the sample size is 9000 for each simulation). However, parameter \( b \) is significantly underestimated, with a large error even for these very long artificial datasets. More complex regression procedures such as the Cochrane-Orcutt (1949) method have been tested and were not found to give any significant improvement. The parameter \( b \) is difficult to estimate due to the covariability of long-term trends in \( S \) and \( T \). Due to oceanic persistence, one of the eigenvalues of the mapping matrix is close to unity, and such unit roots lead to nonstationary random walk behavior that is difficult to estimate (see Mills 1999).

These experiments with the simple model demonstrate that, not only is it difficult to estimate the effect that the ocean has on the atmosphere, but that the effect is likely to be underestimated in a regression estimation even when using long time series from models or observations.

b. The variation of parameter estimates with simulation length

A further investigation was carried out into the effects of changing the length of the Monte Carlo simulations. One thousand simulations with lengths of 10, 20, 50, 100, and 200 years were performed. Figure 4 shows the estimates versus the length of the simulations. For the simple BB98 model the estimates for \( b \) and \( d \) and those for \( a \) and \( c \) are linearly related. Whereas for \( a \) and \( c \), the best estimate is close to the true value even with a short simulation of 10 years, for \( b \) and \( d \) there is a only slow convergence to the true value as simulation length increases. The underestimate in \( d \) at 10 years is less than 1%, but for \( b \) the underes-
The estimate is around 40%. There is a systematic underestimate of the effect that the ocean has on the atmosphere, which improves only slowly with extended simulation length.

There is a disparity in the size of the error compared to the size of the parameter for $a$, $c$, and $d$ compared with $b$. For $a$, $c$, and $d$ the confidence interval is only a few percent of the actual value, in stark contrast to the large relative uncertainty in $b$—for a simulation length of 10 years, this is around 300% of the estimated value. In particular, note that the 90% confidence interval contains zero for simulation lengths of less than 50 years and so at least 50 years would be needed to reject a no-ocean-effect hypothesis at the 10% level of significance. The no-effect hypothesis (see Czaja et al. 2003) of $b/H_{1100}$ reduces the atmospheric component of the BB98 model to a simple autoregressive [AR(1)] uncoupled process, as proposed by Hasselmann (1976). For this value of $b$ (= 0.092), 50 years of daily data are required to be able to be able to state, at 90% confidence, that there is an effect of the ocean on the atmosphere.

4. Diagnosis of the complex model

a. Estimates of model parameters

In this section, the simple BB98 model is fitted to $T$ and $S$ time series from the HadCM3 coupled model. The model parameters are estimated using ordinary least squares, as in the previous section, to give $a$, $b$, $c$, and $d$ at each grid point. Note that there are discontinuities in the time series between each winter. In all of the following analysis, care is taken to not relate two temporally adjacent data points in the time series when these two points are, in fact, 181 days (the whole summer) apart. Instead, subsets of the 50 winters, each 179 days long, are considered, where $\{T_{n}\}$ consists of days 1 to 179 of each winter and $\{T_{n+1}\}$ consists of days 2 to 180.

Figure 5 shows estimates of the parameters. Due to the standardization of the $T$ and $S$ time series before estimation of $a$, $b$, $c$, and $d$, the values of the parameters at different grid points are directly comparable. Tests of how well the simple model fits the complex model data are shown in the following section. It should be empha-
sized at this point that the simple model was designed by BB98 for extratropical conditions, and this is the focus of this study. Results are also shown for the Tropics but the model fit is less reliable here, so results should be interpreted with caution. The analysis has been repeated using upper-ocean heat content as the oceanic variable, with very similar results outside of the Tropics.

Figure 5a shows a large region of small values of parameter $a$ (atmosphere to atmosphere influence) in the vicinity of the North Atlantic storm track, where rapid changes in conditions reduce the day to day persistence of the atmosphere. The maps of $c$ and $d$ are more spatially uniform throughout the domain, although the estimates of $c$ (the atmospheric forcing of the ocean) are slightly increased in the storm-track region where the active atmosphere affects the underlying ocean. The most striking feature in the extratropics, Fig. 5, is the well-defined and spatially coherent region of nonzero $b$ in the Gulf Stream/North Atlantic Current region. There is also a smaller positive patch in the Norwegian Sea. In these regions the simple model diagnosis suggests that the ocean is influencing the atmosphere in the complex model. It is noteworthy that this interaction occurs in regions where there is a strong horizontal gradient in the mean SST. Based on the simple model results of the previous section, it is likely that this $b$ is biased low and so the influence of the ocean may be greater than estimated here.

There is also strong ocean to atmosphere influence in the Tropics (Fig. 5b). However, this area is subject to strong colinearity problems in the estimates. As surface air temperature and SST are so highly correlated here ($r > 0.8$), the regression model is unable to reliably decide which of these variables is causing any predictive effect. Regressing out the part of $S$ that covaries with $T$ to retain the “oceanlike” part of SST removes the tropical regions of $b$ while leaving the extratropical region. This demonstrates the differences in ocean–atmosphere interaction between the Tropics and the extratropics. This colinearity is also seen in this region by comparing maps for $c$ and $d$.

b. Goodness of fit

In fitting any model to data it is important, not only to look at the results, but to analyze the residuals, that is, the part of the data not explained by the model, to see how well the model fits the data (see Draper and Smith 1998).

The most commonly used statistic to test this is the
$R^2$ statistic—the fraction of variance of the data that is explained by the model, ranging from 0 for no fit to 1 for a perfect fit. Maps of $R^2$ for the atmosphere component [Eq. (1)] and for the ocean component [Eq. (2)] are shown in Fig. 6. The ocean BB98 equation provides a good fit to the data, with $R^2$ values in excess of 0.99 almost everywhere (Fig. 6b) mainly due to the very strong day to day persistence of the ocean. Figure 6a shows that the BB98 model is able to capture around 50% of variance in the center of the storm track region, increasing to about 80% poleward and equatorward of this. Again, this follows the pattern of persistence in the atmosphere where the atmosphere persists for a long time; the model captures this well but fails to capture the more complicated atmospheric variability. This is to be expected for a simple autoregressive model such as in BB98.

Implicit in the BB98 model are assumptions about the form of the residuals, so it is important to analyze residuals to test whether these assumptions hold. The three main assumptions of the residuals are

1) **Normality**: The residuals should be Gaussian distributed. Histograms of $T$ and $S$ residuals show that this condition is fairly well met (not shown). However, the complex model $T$ is slightly negatively skewed over much of the North Atlantic due to more extreme and shorter cold events than warm events. None of this skew in the data is accounted for by the simple model, so some slight skewness still remains in the residuals.

2) **Linearity of the fit**: The residuals should not have any dependence on the variables $T$ and $S$. This is best analyzed by making scatterplots of the residuals against these variables. Departures from linearity can be seen by an irregular scatter of points forming, for example, a wedge or curved shape. For HadCM3 over the Atlantic, only uniform bands of scatter exist (not shown), suggesting that the linear model is adequate.

3) **Seriality**: The residuals should be uncorrelated in time. For HadCM3, this was the least well-satisfied criterion, as can be seen in the lag-1 autocorrelation maps in Fig. 6. In the atmosphere, the lag-1 correlation of the residuals is small, around 0.1 to 0.3 in all but isolated patches. The ocean residuals exhibit larger autocorrelation at lag 1, with 0.6 a more typical value. As the residuals are still serially correlated, a higher-order autoregressive model than
BB98 would be required to obtain a white noise structure for the residuals. Tests using a higher-order model helped reduce the problem substantially, but the more parsimonious BB98 model was retained here because of its ease of interpretation.

These residual diagnostics show that the simple model captures many of the major features of the more complex model.

c. Atmospheric persistence and predictability

Where the ocean affects the atmosphere there is the possibility of improved predictability for the atmosphere, given prior knowledge of the ocean. Using lag covariance analysis, Rodwell and Folland (2002) investigated the large-scale predictability of the North Atlantic Oscillation (NAO) based on ocean–atmosphere interactions and found that the ocean can add some modest predictive skill by the use of May SSTs to predict the following winter NAO. In the present study, predictability is explored by examining the persistence of atmospheric temperature anomalies as measured by the autocorrelation function. Due to the short decorrelation time (about 1–2 weeks) in the atmosphere, the expected atmospheric behavior in the absence of oceanic effects would be a rapid decrease in autocorrelation with increasing time lag. Figure 7 shows autocorrelation functions from the three points shown previously (Fig. 3): one at high latitude (60°N, 26.25°W), one at midlatitude (47.5°N, 26.25°W), and one in the Tropics (15°N, 26.25°W). Note that, as lag increases, the amount of data used to compute the autocorrelation decreases, due to the noncontiguous winters. However, even at lag 120 the sample size is 3000 data points. The three points exhibit different behavior. At high and midlatitudes there is a fast decrease over the first few days. In midlatitudes the decrease is much slower at large lags and the subsequent decrease is then very slow over the following months, with this transition occurring at a lag of about 5 days. The Tropics show quite different behavior, with atmosphere–ocean coupling there leading to a much slower and more uniform decay in autocorrelation.

Using the BB98 model, the autocorrelation of the atmospheric variable $T$ can be split into the sum of a direct atmospheric and an indirect oceanic part (Junge and Stephenson 2003; Charlton et al. 2003):

$$r(T_n, T_{n+lag}) = a_{lag} + b_{lag} r(T_n, S_n),$$

where $r(T_n, S_n)$ is the correlation between $T$ and $S$, and $a_{lag}$ and $b_{lag}$ are the $a$ and $b$ estimated using multiple regression of $T$ and $S$ at different time lags. The first term on the right-hand side is the “direct” effect of the atmosphere on the atmosphere and the second term is the “indirect” effect of the ocean on the atmosphere.

The three plots in Fig. 7 show very different roles for the direct and indirect parts. First, at high latitudes the direct part is virtually indistinguishable from the auto-
correlation itself because the oceanic indirect part is negligible; hence at high latitudes there is generally little ocean influence on the persistence of atmospheric boundary layer temperature. In midlatitudes, where the map of $b$ indicates that the ocean does influence the atmospheric boundary layer on a daily time scale (Fig. 5b), the indirect part is a much greater fraction of the total autocorrelation. It explains the “shoulder” in the autocorrelation, caused by the slower decrease after lags of 5 days. The direct part shows a similar behavior to that at high latitudes with no significant autocorrelation after about 10 days. This analysis shows that after 10 days$S$ it is the ocean that gives the atmosphere persistence and therefore potential predictability in midlatitudes. Finally, the Tropics (Fig. 5c) show a strong indirect part over the first 3-month lag, given by the stability of the air–sea interface in this region.
5. Conclusions

The main conclusions of this study are as follows.

1) The simple stochastic model proposed by Barsugli and Battisti (1998) provides a useful tool for diagnosing the daily ocean–atmosphere coupling in a very complex fully coupled GCM.

2) Monte Carlo experiments using the simple model have shown that the parameter $b$, a measure of the effect of the ocean on the atmosphere, is systematically underestimated and has a large sampling error. The estimation is worse for short simulations, with an underestimate of $b$ by 40% for 10-yr simulations, slowly decreasing to 17% even for 50 years of daily data. This suggests that the oceanic effect on the atmosphere is likely to always be underestimated by regression procedures in similar studies with finite samples of data.

3) Using Monte Carlo tests, 50 years of daily data, equivalent to 9000 data points, were required to be able to detect that the “standard” value of $b$ is different from zero and thus that there is an ocean to atmosphere influence, at the 90% confidence level. The estimation problems with the parameter $b$ are intriguing, particularly in the light of the very good estimation of the other parameters. One of the problems involved in estimating the effect of the ocean on the atmosphere is that the two systems have very different time scales. As $d$ approaches unity in the simple model, the ocean component tends toward a nonstationary random walk process. Such a process is much harder to estimate than a simple decaying stationary behavior. Resorting to higher-order VAR models might help to improve the situation, but then the parsimony of the simple model and its ease of interpretation will be lost.

4) The simple model was fitted to the daily data from a control simulation of HadCM3. Parameter estimates revealed spatially coherent regions. In particular, there is a region where the ocean influences the atmospheric boundary layer in the North Atlantic, in the region of the Gulf Stream/North Atlantic Current. The area of greatest influence is in the eastern North Atlantic. Comparison with a second complex coupled GCM, the Bergen climate model (Furevik et al. 2003), gives very similar results for this parameter, both in spatial structure and in strength (not shown). While not confirming that this is also present in the real climate system, the agreement does improve the robustness of this result.

5) Diagnosis of the HadCM3 data in terms of the simple model allows the autocorrelation of the atmospheric variable to be split into a direct part, which comes about solely through atmospheric interaction, and an indirect part, which is the product of ocean–atmosphere coupling. In the midlatitudes, where the ocean affects the atmosphere, the atmospheric autocorrelation is increased by the presence of this interaction on lags from a few days to several months. This confirms that the ocean here leads to increased low-frequency variability of the atmosphere.

The simple model used in this study is, of course, a very limited model of ocean–atmosphere interaction. Perhaps the greatest omission is in not modeling advection of ocean or atmosphere temperatures, which will certainly have an effect on longer time scales. The BB98 model does not allow any nonlocal influences on either the ocean or the atmosphere. Such influences include that of the tropical Atlantic on the extratropical atmosphere in the North Atlantic region (e.g., Czaja et al. 2003). A nonlocal influence would be communicated via the atmosphere on the time scales discussed here, so any external influence on the local atmosphere–ocean system will be part of the residual variance in the system and may manifest itself in the map of $c$ where this external influence affects the extratropical ocean. The map of $b$ will not be influenced by any external effects, as the lag involved in estimating this parameter allows influences from the underlying ocean to the atmospheric boundary layer only, and external influences communicated through the atmosphere will have the reverse direction. The possibility remains that the persistence in the extratropical ocean is caused by a persistent external influence via the extratropical atmosphere. Further assumptions and limitations of the simple model are discussed by BB98.

The stochastic climate model–based approach is currently being applied to fields outside the boundary layer, including pressure fields. In particular, an analysis of the North Atlantic Oscillation using these and more powerful time series techniques for testing causality is being performed.

Acknowledgments. We wish to thank Rowan Sutton and Helge Drange for their thoughts and contributions, and the NERC COAPEC project for their funding and support.

REFERENCES


Cochrane, D., and G. H. Orcutt, 1949: Application of least
squares relationships containing autocorrelated error terms. 


