The Application of ARGO Data to the Global Ocean Data Assimilation Operational System of NCC

LIU Yimin, ZHANG Renhe, YIN Yonghong, and NIU Tao

National Climate Center of China, Beijing 100081
Chinese Academy of Meteorological Sciences, Beijing 100081

(Received November 2, 2003; revised September 16, 2004)

ABSTRACT

In this paper, we have preliminarily studied the application of ARGO (Array for Real-time Geostrophic Oceanography) data to the Global Ocean Data Assimilation System of National Climate Center of China (NCC-GODAS), which mainly contains 4 sub-systems such as data preprocessing, real-time wind stress calculating, variational analysis and interpolating, and ocean dynamic model. For the sake of using ARGO data, the relevant adjustment and improvement have been made at the corresponding aspects in the sub-systems.

Using the observation data from 1981 to 2003 including the ARGO data of 2001 to July 2003, we have performed a series of numerical experiments on this system. Comparing with the corresponding results of NCEP, it is illustrated that using ARGO data can improve the results of NCC-GODAS in the region of the Middle Pacific, for instance SST, SSTA (SST anomalies), Nino index, sea sub-surface temperature, etc. Furthermore, it is obtained that NCC-GODAS benefits from ARGO data in the other regions such as Atlantic Ocean, Indian Ocean, and extratropical Pacific Ocean much more than in the tropical Pacific.

Key words: ARGO (Array for Real-time Geostrophic Oceanography) data, ocean data assimilation, dynamical ocean model, 3-dimensional variation, SST (sea surface temperature), Nino index

1. Introduction

It is well-known that the state of ocean plays very important role in the climate change. But there is a paucity of the ocean observation data. The data distribution in the space, time and different components is very inhomogeneous, even in some areas, there are no any observation data. Hence, it brings some difficulties to the scientists to study many problems relevant to ocean.

This situation has been being changed since ARGO (Array for Real-time Geostrophic Oceanography) plan was implemented. With ARGO plan’s further implementation and accumulation, it will play more and more important role in the relevant research fields. Though the time span of ARGO is short and the quantity of the data is small in some areas up to now, it is no doubt that it will be significant to study how to use these precious data in the operational system (Xu, 2002; Zhang et al., 2004).

Data assimilation is very powerful in data analyses. When we analyze and diagnose some climate problems, the multi-year averaged field can be used, if its variability is weak. However, for the strong variability, the real signal may be missed by this method. Data assimilation is based on data statistic analyses and model dynamical evolution. Hence it can describe the variability to a certain extent. Meanwhile, there are unavoidable observation errors, we need data assimilation to correct them.

Data assimilation can be used to supply reasonable initial fields to dynamical model. Because of model’s uncertainties and observation errors, they will result in the inaccuracy of performance of modeling. Data assimilation can estimate these errors and balance the errors of observation and model to some degree. Meanwhile, data assimilation can extrapolate information to some areas and some model components even without observation. In addition, the reasonable initial fields offered by data assimilation will make the different components of model in harmony. Therefore, a dynamical model system should have its own data

*Supported by the National Natural Science Foundation of China under Grant No. 40231014.
assimilation system.

There are a variety of algorithms to approach the data assimilation problem (e.g., Daley, 1991; Bennett, 1992; Ide et al., 1997; Kalnay, 2003). The examples of simple methods include nudging method (Kistler, 1974; Kass et al., 1999); optimal interpolation (OI) (Bengtsson et al., 1981); 3-dimensional variation (3DV) (Lorenc, 1986; Parrish and Derber, 1992; Ji and Leetmaa, 1997; Bell et al., 2000; Liu et al., 2000, 2005). The examples of advanced methods include 4-dimensional variation (4DV) (Lewis and Derber, 1985; Coutier and Talagrand, 1990; Thacker and Long, 1988; Derber, 1989; Bouttier and Rabier, 1997; Bennett, 1992; Bennett et al., 1997; Uboldi and Kamachi, 2000); extended Kalman Filter and reduced-order KF (Cane et al., 1996; Pham et al., 1998); and ensemble Kalman Filter (Houtekamer et al., 1998, 2001; Evensen, 1994; Hamill et al., 2001; Anderson, 2001).

The advanced methods are quite expensive for computation as using high resolution model systems, and not easy for system to upgrade due to their algorithm depending on model closely. Considering the situation of our computer resource and operational requirements, we choose the 3DV method to establish our Ocean Data Assimilation System.

The 3DV method was first applied to the assimilation of atmosphere observation data (Bengtsson et al., 1981). Since being first used in oceanology, it has been employed to build their own ocean data assimilation systems by a lot of scientist groups and operational systems (Clancy et al., 1988; Deber and Rosati, 1989; Ji et al., 1995; Bell et al., 2000; etc.). Their work supports that this scheme is efficient and effective. Up to now, their systems have played important role in the studies and operations on weather and climate.

In this paper, we will discuss ARGO data assimilation based on the Global Ocean Data Assimilation System of National Climate Center of China (NCC-GODAS) (Liu et al., 2000, 2005). The outline is as follows: given in Section 2 is a brief description of NCC-GODAS, followed by Section 3 is the adjustments of the system in order to use ARGO data. In Section 4, the analyses of assimilation experiment results are present, finally conclusions are drawn in Section 5.

2. Brief description of NCC-GODAS

The NCC-GODAS is established mainly based on “the Pacific and the Indian Ocean Data Assimilation System” (PIODAS) (Liu et al., 2000), and has been employed by “short-term climate prediction model operational system” since October 2001. It contains 4 parts as follows: data preprocessing, real-time wind stress calculating, variational analysis and interpolating, and ocean dynamic model (Liu et al., 2002, 2005)

2.1 Dynamical model

We choose the ocean climate model L30T63 OGCM Version 1.0 of LASG, which is the ocean component model in the present “short-term climate prediction model operational system of NCC”, as our dynamic model. This model has the same horizontal resolution as T63 atmosphere model (1.875° × 1.875°) and 30 layers in the vertical direction, in which the first 10 layers are to 250 m, the second 10 layers distribute in the depth from 250 m to 1000 m, and the last 10 layers are located deeper till 5600 m. Detailed information about this model can be found in the relevant papers in the book “The Study on Short-term Climate Prediction Model” (Jin et al., 2000).

2.2 Data preprocessing

The data we use in GODAS include sea surface atmosphere temperature (SSAT), sea surface atmosphere pressure (SSAP), sea surface wind filed (SSWF), the profile of ocean temperature, and salinity (Expendable Bathythermograph(XBT)). A decode module is designed for the original encoded data from GTS.

In addition, data preprocessing includes Quality Control of Data (QCD). QCD is to remove some ineligible data and the information of feedback observation errors to the part of variational analysis and interpolating (details given in Liu et al., 2002).

2.3 Real-time wind stress calculating

A simple algorithm to calculate sea surface wind...
stress (SSWD) is designed in GODAS. This module is based on the formula \( \tau = c_d V|V| \rho \).

A smooth interpolation scheme is adopted in this module. Firstly, we calculate the difference between the observation data and the mean of climate data offered by original dynamic model at the 4 nearest points. And then, we add this difference to the climate data at the 16 nearest points by a weight that is determined by the distance from the observation position to the model grid. If a model grid gets the information of more than one observation point, an adjusting factor will be introduced. This module is designed to be called once a model day.

### 2.4 Variational analysis and interpolating

Considering the situation of ocean observation data and the characteristics of ocean state changing slowly, we open a 4-week time-window to employ ocean observation data as much as possible (Ji et al., 1995). All the data in this window will contribute their information to GODAS with a time weight, which depends on the interval from observation time to model time.

In the space, we take 3-dimensional variation scheme. The main task is to find the minimum of object functional \( I \) on \( H \):

\[
I = \frac{1}{2} H^T E^{-1} H + \frac{1}{2} (DH - H_0)^T F^{-1} (DH - H_0),
\]

where \( H \) is the difference vector between model and guest field with dimension, \( N = 94 \times 194 \), \( H_0 \) is the difference vector between model filed and observation data with unfixed dimension, \( E \) is the model error matrix with dimension \( N \times N \), \( F \) is the observation error matrix with unfixed dimension, \( D \) is the interpolating matrix that transfers the variables on the model grid to the observation position. In fact, object function \( I \) is the sum of model errors and observation errors.

It is obvious that the effect of an assimilation system will rely on how to choose \( E \) and \( F \). Due to the limit of computer resources, \( E \) is designed to only to be correlated in horizontal direction and to be the same at each layer. The definition of \( E \) is

\[
E = a \exp(-r^2/b^2 \cos \phi),
\]

where \( a \) is the amplitude of \( E \) and is a spatial function, \( b \) is the correlation length of \( E \) and is also a spatial function, \( r \) is the distance between the two model grids, and \( \phi \) is the latitude between the two model grids. These parameters should be determined by a series of numerical experiments.

It is difficult to give \( F \) exactly. In our system, \( F \) is defined as a diagonal matrix with the form \( F = f \cdot \text{cnt} \cdot \text{tm} \). It indicates that there is no correlation between observation data. Otherwise, it will cause a lot of trouble to GODAS. \( f \) is the data spatial distribution factor, which varies with the density of observation data. \( \text{tm} \) is a time factor and its value is 1 to 0 according to the interval between assimilation time and observation time (from 0 day to 15 days). \( \text{cnt} \) is a quality control factor which depends on \(|H_0|\).

### 3. The adjustments of NCC-GODAS for using ARGO

#### 3.1 Dynamical model

There are two vertical mixing parameterization schemes in this model. One is based on the Richardson number applied in the tropic region from 30°S to 30°N. The other is the isopycnal mixing scheme for the remaining area. While using these models, we have made some improvements on the vertical mixing parameterization scheme so that a transition zone to connect two areas mentioned above is designed, and the criterion of the stability depending on the gradient of density in the isopycnal mixing scheme is redefined as a spatial function rather than a constant based on the state equation of sea water and ARGO data.

Figure 1 shows the comparisons of the results between the before and after parameterization adjustment. Figure 1a is the monthly mean sea surface temperature in August by the original dynamical model. Figure 1b is the result with the adjusted parameterization scheme. Figure 1c is the result of Levitus94 (Levitus, 1982). It is very clear that the temperature discontinuity in the original model disappears, and the result of modified scheme coincides with that of Levitus94 very well. By the way, these improvements have been adopted by the corresponding ocean model in the “short term climate prediction operation system”, too.
3.2 Data preprocessing

In addition to the original QCD scheme, some items are added in order to use ARGO data, such as vertical gradient check and T-S relationship check (if temperature and salinity profiles both are available). The classification of data source is made to give different quality criterion for different data, for instance, the eligible temperature range for ARGO and XBT are -2.5°C to 33°C and -3°C to 35°C, respectively. Because the salinity data were not put in assimilation system before ARGO plan, the QCD did not involve the salinity. In this paper, we add the salinity data to the assimilation system. Therefore, QCD for salinity is needed. The QCD criteria for salinity are instituted similar to temperature when ARGO data are used in NCC-GODAS.

3.3 Analysis and interpolation scheme

The background covariance and observation error covariance play important roles in the assimilation system. As ARGO data are put in the assimilation system, the covariance matrixes are modified. Modified assimilation system can assimilate temperature and salinity together, and a weak correlation between temperature and salinity is specified. In addition, a weak vertical correlation is introduced.

The following factors are added to the observation covariance: (1) The observation data are classified into two kinds as above subsection mentioned, and a factor is given to distinguish them. Through statistic analyzing the data from 2001 to 2003, the factors for ARGO and other data are specified as 1 and 1.435, respectively. In addition, the factors have a certain space distribution according to the observation spatial partition density. (2) When there are both ARGO and XBT data near to a model grid, these data all make contribution to their near model grids. But the ARGO data have a priority to XBT data, and it is described by an observation system error factor 1:1.8. (3) A factor is added to receive the information returned from T-S relationship QCD.

The background covariance has the form as

$$ E = \begin{bmatrix} E_T & E_{TS}^T \\ E_{TS} & E_S \end{bmatrix} $$

(3)

where $E_{TS}$ describes the $T$ and $S$ correlation. It is supposed that $E_{TS}$ is smaller than $E_T$ and $E_S$. Thus, the object functional $I$ can be approximately divided into two parts: one is zero order approximation involving $E_T$ and $E_S$; the other is higher order approximation involving $E_{TS}$. Firstly, we deal with the zero order approximation. And then we deal with the higher order approximation. The same algorithm is for vertical correlation.
4. Assimilation experiment and results

In the numeric experiment, we have used the SSAT, SSWF, SSAP, and XBT from 1982 to 2003, in which XBT contains the observation data downloaded from NCEP web site and the data stored in the data bank of NMC and from other sources. ARGO data have been used since 2001. Figure 2 demonstrates the spatial distribution of ARGO and other profile data, and their observation frequency. The colorbar is for the accumulated number of observation (every day only one time is accumulated if available). Figure 2 shows that ARGO location covers ocean fairly well except for most areas in the Southern Hemisphere. The XBT data of over 20 years are located mainly in the Northern Hemisphere, especial in the North Pacific and Atlantic and tropic Pacific TOGA/TAO array. In the Indian Ocean, the observations are mainly from the fixed routine of some VOS (voluntary observer ship). Up to August 2003, even in the focus region of the tropic Pacific, there are still some areas with the data sparse. It is no doubt that this situation will get better with the ARGO plan further implements.

Figure 3 depicts the evolution of Nino index (Nino3 and Nino34 indexes, from 1993 to July 2003): black line is OISST-V 2, red one is NCC-GODAS with ARGO, and green one NCC-GODAS without ARGO. It can be seen that the Nino indexes of NCC-GODAS coincide with that of OISST-V 2. Using ARGO data makes NCC-GODAS’s result tend to that of OISST-V 2 more closely, for instance, the improvement of Nino3 arrives at over $0.5^\circ$C in September 2001, and for Nino34 in October 2002 up to $0.5^\circ$C.

In this paper, we choose the international famous data as the standard to analyze our assimilation results, which are the monthly mean climate temperature data of Levitues94 (Levitus, 1982), the OISST-V2 reanalysis data (Reynolds et al., 2002) and the tropic Pacific regional ocean assimilation data of Environmental Monitoring Center (EMC; Ji, and Deber, 1995). Because of different method and data using in their analysis systems, it can be seen, from Figs.4-6, that there are some differences among these data sets. Hence, these data sets are the relative standards. The former two are used to the comparison of surface temperature, and the last one (EMC) for the subsurface.

NCC-GODAS is run in the following way. Firstly, we run the ocean dynamic model, forced by climate fields offered by the original model, and from ocean static state to 1800 yr, to arrive at the ocean equilibrium state and produce a restart file. And then, running GODAS, with this restart file and all of the available observation data mentioned above, and from 1982 to present, will output monthly mean assimilation fields on every month. Finally, climate monthly mean fields can be made by multi-year averaging the monthly mean assimilation fields from 1982 to 2002.

Figure 4 shows annual mean climate SST, in which the graphs are for the Levitus94, OISST-V 2, dynamic model without assimilation and NCC-GODAS with ARGO, respectively. For the case of dynamic model without assimilation, the simulation results in the east and west tropical Pacific are much stronger than the observations. After the assimilation of NCC-GODAS, this situation changes much better. However there is only a small difference between assimilating results with both ARGO and XBT and with only XBT. The reason for this is that the time span of ARGO (up to 2 yr) is rather shorter than that of XBT (over 20 yr).

Figure 5 gives the difference fields of annual mean climate SST. Figures 5a-5c are the difference of Levitus94 minus model without assimilation, NCC-GODAS without ARGO, and NCC-GODAS with ARGO, respectively, and Fig.5d is NCC-GODAS with ARGO minus NCC-GODAS without ARGO. Note that Fig.5d is multiplied by 20. Figures 5a-5c demonstrate that climate SST in the ENSO area of model is smaller than that of observations by about $1^\circ$C; in the west equatorial Pacific, model results are larger than that of observations over $1^\circ$C; in the middle and high altitude areas, there are even larger differences. NCC-GODAS reduces these distortions successfully. Figure 5d shows that application of ARGO can improve globally the climate state, because the difference of Levitus94 minus model has almost the same tendency as the difference of NCC-GODAS with ARGO minus without ARGO.

Figures 6 and 7 illustrate the time-longitude sector of SSTA and the subsurface (the depth of 62.5 m)
temperature anomaly averaged over 5°N-5°S from April 2001 to August 2003, longitude from 120°E to 80°W, respectively. The recent El Niño event started on May 2002, and reached its peak on December 2002, and was obviously fading on March 2003. The results of GODAS describe this process successfully. The total tendency of warm water evolution from west to east basically coincides with that of OISST and EMC. Having used ARGO data, NCC-GODAS performance is better than that without ARGO. In addition, we can notice the improvement on the two cool processes located in the east equatorial Pacific in December 2001 and May 2003.

ARGO plan is for the global observation. Therefore, the improvement resulting from it will be global. This can be proven by the results shown in Fig. 8. The several main high and low temperature areas, such as the North Atlantic, North Indian, and extratropical Pacific, are well described by the NCC-GOADS with ARGO. The case without ARGO reflects these phenomena very weak.

Figures 9 and 10 represent the root mean square (RMS) values of SSTA between NCC-GODAS and OISST from 2001 to August 2003. The RMS is globally reduced as using ARGO data, except for a few areas such as near 45°E in Fig.9 and 35°S in Fig. 10. The maximum decrease of RMS appears in the North Atlantic, and then at the north high latitude Pacific, the following is the north tropics of the Indian and Pacific. It coincides with the spatial distribution of data we use at the present (refer to Fig.2).

5. Conclusions

The following conclusions may be drawn through the above analyses and discusses.

(1) ARGO can give model climate state a correct adjustment tendency. Since the time span of ARGO is quite short, the modifying amplitude is very weak.

(2) Using ARGO data can improve the results of NCC-GODAS in the region of the Middle Pacific, for instance, SST, SSTA, Nino index, sub-surface temperature, etc. Furthermore, it is obtained that NCC-GODAS benefits from ARGO data in the other regions such as the Atlantic Ocean, Indian Ocean, and extratropical Pacific Ocean much more than in the Middle Pacific. The reason for this is that there have already been certain observed data in the tropic Pacific besides ARGO such as TOGA/TAO array. But in other areas, the observation was quite sparse before.

(3) Meanwhile, the above results indicate that the adjustment on NCC-GODAS for application of ARGO is effective. As the situation of observation is improved, assimilation system makes the correct response and promotes the skill of NCC-GODAS.

In this paper, though we use ARGO data only 2a, the results are very encouraging. It is no doubt that with ARGO plan further implementation and ARGO data accumulation, we will get more and more information about the ocean and it will benefit the ocean data assimilation to offer more and more reasonable initial fields to short-term climate prediction model operational system and more and more reliable analysis data set to relevant fields.

REFERENCES

Bouttier, F., and F. Rabier, 1997: The operational implementation of 4-D-Var. ECMWF Newsletter, 78, 2-5.
NO.3 LIU Yimin, ZHANG Renhe, YIN Yonghong and NIU Tao 365


