Assimilation of sub-surface temperature profiles from Argo floats in the Indian Ocean in an Ocean General Circulation Model

Neeraj Agarwal*, Rashmi Sharma, Sujit Basu and Vijay K. Agarwal

Meteorology and Oceanography Group, Space Applications Centre, Ahmedabad 380 015, India

Effects of assimilation of Argo temperature profiles in an Ocean General Circulation Model (OGCM) are studied in the tropical Indian Ocean. Two assimilation experiments have been performed during the first five months of 2004. In the first one, monthly averaged Argo profiles have been assimilated using nudging technique and in the second, Cressman technique has been used for assimilating daily Argo profiles. The differences in the isothermal layer depth (ILD), depth of the 20°C isotherm (D20) and sea surface temperature (SST) have been examined. The impact of assimilation has also been assessed using independent observations from one of the Indian Ocean Triton buoys. Errors in SST at the Triton buoy location are reduced by 37 and 16% respectively, in the two experiments. ILD and D20 show significant improvement in terms of error reduction in both the experiments, implying improvement in the mixed layer and the thermocline region. The impact of assimilation using Cressman technique is more on ILD and D20 as is evident from the coefficient of determination. This is because in the first experiment monthly averaged data have been assimilated leading to a comparatively poor representation of the high frequency variability of the parameters. Temperatures at the surface, 50 and 100 m depths during the forecast period (June–August 2004) have also been compared with Triton buoy observations. Also, SST in the forecast phase has been compared with available National Institute of Ocean Technology buoy measurements.

Keywords: Argo floats, data assimilation, Ocean General Circulation Model, temperature profiles.

DATA assimilation is one of the frontier areas of research in atmospheric and oceanic sciences. Incomplete model physics and inevitable errors in the initial and boundary conditions generally lead to divergence of model trajectories from reality, which results in substantial forecast error. This forecast error can be corrected to a certain extent using various techniques of data assimilation[1]. There are three broad objectives of data assimilation: (i) Improvement of model parameterization of sub-grid scale processes; (ii) preparation of self-consistent ocean state, and (iii) generation of reasonably accurate initial conditions for the forecast models. Comprehensive reviews of data assimilation are readily available[1–3]. Studies related to this topic can be broadly divided into two classes. The first class is known as sequential data assimilation. Starting from some initial condition, the model solution is sequentially updated whenever measurements are available, using a combination of model and observations according to some prescribed mathematical recipes known as blending, nudging, optimal interpolation and the more sophisticated approach of Kalman filter, etc. The second class of non-sequential methods is known by the general name of variational methods, which can be time-independent like 3D Var as well as time-dependent like 4-D Var using adjoint equations. However, these sophisticated approaches of 4-D Var and Kalman filter suffer from the drawback of high computational cost as well as other drawbacks like the use of model equations as hard constraints in the case of 4-D Var with adjoint equations (of course, the weak constraint formulation allows for weakening this constraint) and use of Kalman filter for nonlinear atmospheric and oceanic models, although it is known that the Kalman filter is a strictly optimal filter only for the class of linear models. In fact, this is the primary reason why researchers in recent times are inclined towards the use of ensemble Kalman filter[4–6], which is applicable for nonlinear models. This method exploits the use of ensembles to estimate the model error covariance matrix. The ensemble Kalman filter is also applicable to (i.e. is optimal for) linear models, but is much easier to apply to nonlinear models than the previous extension of Kalman filter, like the extended Kalman filter.

Historically, ocean modelling has suffered from lack of surface and sub-surface observations, although in recent times there has been enormous number of surface observations by satellites, like the observations of sea surface temperature (SST), sea surface winds (SSW) and sea surface heights (SSH). The large spatial and temporal coverage of satellites provides a unique dataset. The huge dataset of SST and SSH has been successfully assimilated by many researchers in the past in ocean models of various levels of complexity, while the dataset of SSW has been used for driving these models. In more recent times, a large number of sub-surface temperature and salinity profiles have become available with the launch of Argo floats[7]. It is needless to mention here that this dataset provides a unique opportunity to ocean modellers for generating realistic sub-surface analysis using data assimilation. There have been attempts to use these data in ocean data assimilation systems[8,9], where it has been shown that Argo data can significantly improve the ocean model simulations in the region of the middle Pacific. There are more recent studies[10,11] related to this topic.

The present study provides an example of Argo temperature profile assimilation using two simple techniques in an Ocean General Circulation Model (OGCM). The main emphasis of the study is the generation of accurate four-dimensional ocean states.

*For correspondence. (e-mail: neeraj@sac.isro.gov.in)

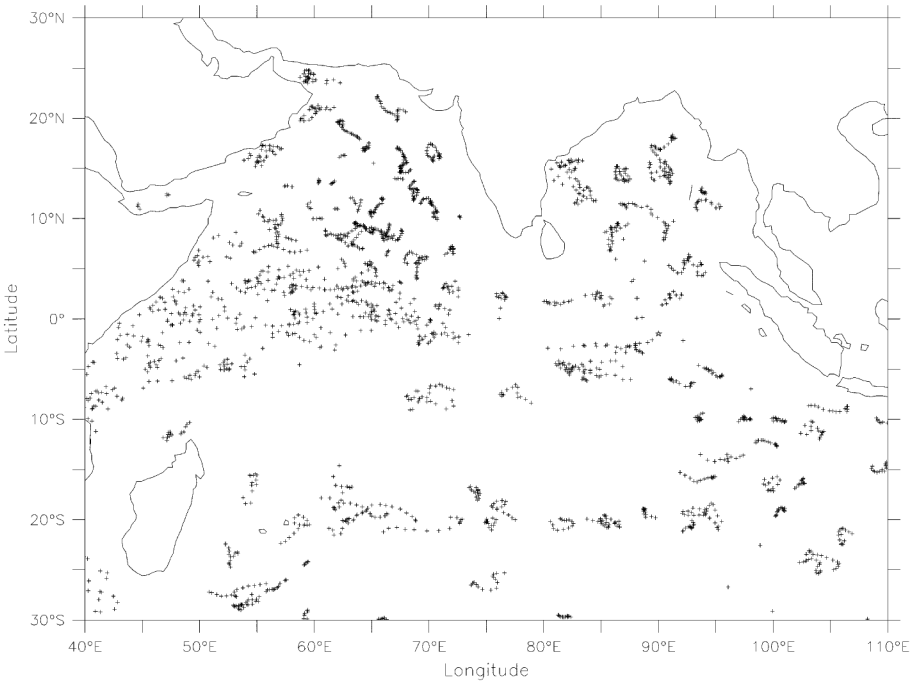

Figure 1. Distribution of Argo profiles assimilated in the OGCM during January–May 2004.

In this study Argo profiles for the period January–May 2004 have been used. Only those Argo profiles have been retained which are available from the surface to depths beyond 500 m. Several standard quality checks were applied to filter the spurious profiles from the database. After these checks, there were ~10,000 temperature profiles available for assimilation for the period January–May 2004. Figure 1 shows a map of all the Argo observations available for the study period. Assimilation of these profiles in the OGCM was carried out using two different techniques, which are discussed below.

In the nudging technique, the model is slowly nudged (relaxed) towards observations at each time step via a Newtonian damping term in the prognostic equation of the relevant variable (in this case, temperature):

$$\frac{\partial T}{\partial t} + f(T, u, v, s) = -\lambda(T - T_{\text{obs}}), \tag{1}$$

where the model variable T (temperature) is nudged towards a reference value or observation T_{obs} and $f(T, u, v, s)$ represents the remaining terms like advection, diffusion and source. λ is a depth-dependent nudging coefficient given by

$$\lambda = \frac{\mathrm{d}z}{86,400 \times \text{days (s)}}. \tag{2}$$

In this method correction to the model temperature is made at every time-step and hence the observations need to be interpolated in time. Moreover, nudging has to be carried out at all the model grid points. Since the Argo data are not dense and not available on daily timescale, monthly averaged Argo profiles have been used with Levitus climatology filled in the data gaps. The spatial fields have been smoothened with nine-point averaging, so that there are no sharp discontinuities. The lowest model depth is 5500 m. Argo observations are not available at this depth. Therefore, once again Levitus climatology has been used to extend the observed Argo profile up to 5500 m depth and the discontinuities are removed by simple three-point smoothing. Thus a database of T_{obs} was made on a monthly basis and then the model was slowly nudged towards this database using eq. (1).

In the Cressman technique[12], the model state is set equal to the observed values in the vicinity of available observations and to an arbitrary state (say, climatology or a previous forecast) otherwise. If we represent a background state (which can be model forecast, climatology, etc.) by x_{b} and a set of n observations of the same parameter by $y(i)$, where $i = 1, 2, \ldots, n$, then the analysed field of the parameter x_{a} using Cressman technique at each model grid point j can be represented as:

$$x_{\text{a}}(j) = x_{\text{b}}(j) + \frac{\sum_{i=1}^{n} w(i, j)\{y(i) - x_{\text{b}}(i)\}}{\sum_{i=1}^{n} w(i, j)}, \tag{3}$$

where

$$w(i, j) = \max\left\{0, \frac{R^2 - d_{i,j}^2}{R^2 + d_{i,j}^2}\right\}. \tag{4}$$

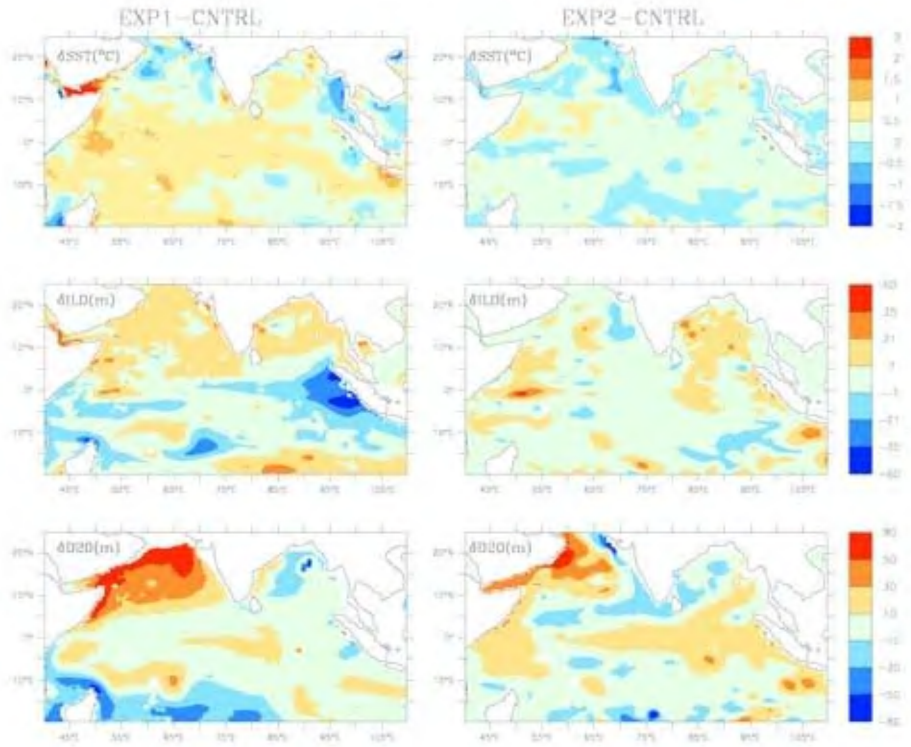

Figure 2. Spatial differences of SST (°C), ILD (m) and D20 (m) between EXP1 and control run (left), as well as EXP2 and control run (right).

Here $d_{i,j}$ is the distance between the model grid point j and observation point i and $x_b(i)$ is the background value interpolated to the observation point. R is the influence radius which is user-defined. $w(i, j)$ is a measure of weight which decreases with increasing $d_{i,j}$, becoming zero if $d_{i,j} = R$.

In our case the temperature field of the OGCM has been used as the background field and keeping R as 100 km the analysed field of temperature was generated using the above technique with Argo observations. This analysis technique was applied to each level of the model. Once the analysed field for a particular day (say, 1 January 2004, in our case) is generated, then using that as the initial condition the model is integrated for a day. The output temperature fields are then again considered as background fields and the next day's Argo observations (wherever available) are assimilated into it to generate analysed fields. This procedure is carried out for January–May 2004.

The model used in the present study is a modified version of the Modular Ocean Model version 3.1 (MOM-3) from GFDL/NOAA, set up for the global domain, excluding the polar regions (80°S–80°N), with variable horizontal resolution varying from 0.5° in the Indian Ocean to 2° in the rest of the oceans. There are 38 levels in the vertical with 16 in the upper 110 m. The bottom topography is based on 1/12° by 1/12° resolution data from the US National Geophysical Data Centre. A wind-dependent drag coefficient was used in the model. The model was spun up from the rest for about 20 years using climatological winds and restoring boundary conditions for SST and sea surface salinity (SSS). After this, the model was further run with the actual daily winds and flux data for 1996–2003 without any restoring. The freshwater flux used was from daily climatological dataset. A suite of three experiments have been performed for 2004. Using the initial conditions generated by model integration from 1996 to 2003, the model is integrated for 2004 using QuikSCAT scatterometer winds and heat flux parameters (net shortwave, net longwave, air temperature and specific humidity) from NCEP/NCAR Reanalysis. Monthly climatology of river discharge data have been used. In addition, daily data of diffuse attenuation coefficients derived from SeaWIFS have also been used. Latent and sensible heat terms were computed using model SST. Hence SST and SSS were not restored. This run was designated as control run (CNTRL). In the second run, monthly averaged Argo data were assimilated using nudging technique (EXP1) and in the third run (EXP2), Cressman technique was used to assimilate the Argo data on daily timescale. Forcing parameters were kept the same in all the three runs.

Figure 2 shows the differences in SST (δSST), ILD (δILD) and D20 (δD20) between CNTRL and the two assimilation experiments, EXP1 and EXP2 for April. ILD is defined as the depth at which the temperature is lower by 1°C from the surface value. The left panel shows differences between EXP1 and CNTRL, while the right panel shows differences between EXP2 and CNTRL. One obvious inference that can be made is the large values of

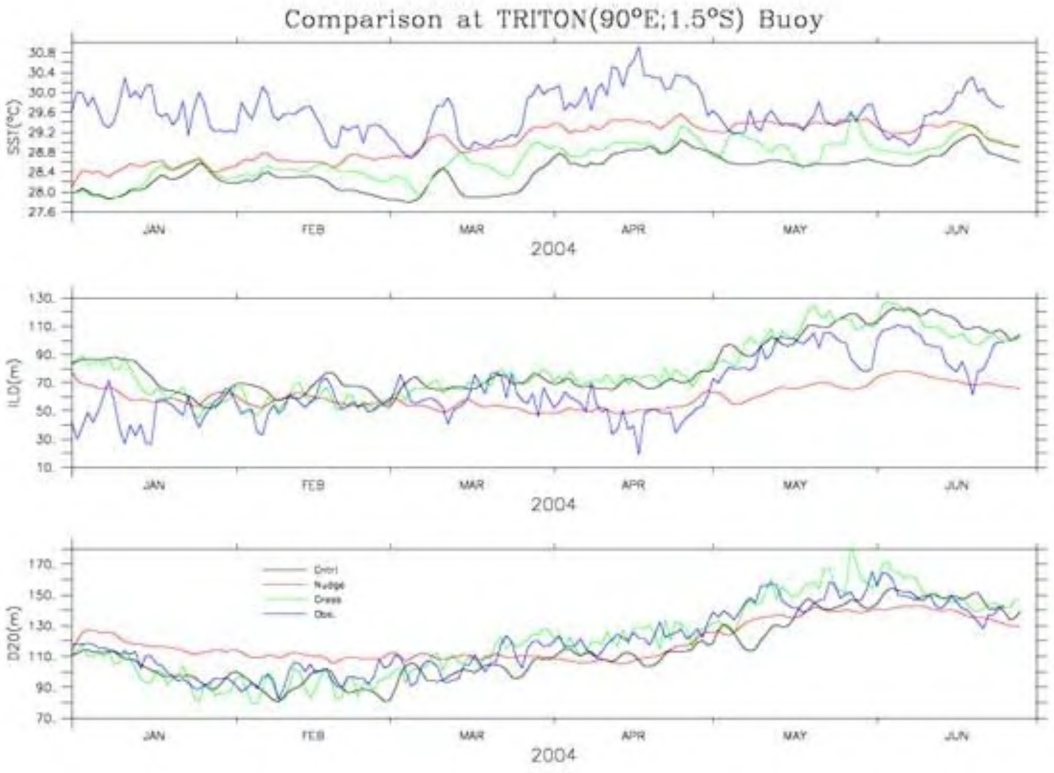

Figure 3. Time series of SST (°C; upper panel), ILD (m; middle panel) and depth of 20° isotherm (m; lower panel) for control run, EXP1, EXP2 and observations at TRITON buoy location (during assimilation phase).

δSST in EXP1 compared to EXP2, possibly because EXP1 runs are constrained by Levitus profiles. Similar behaviour can be observed in δILD and δD20 also. The impact of assimilation is more in the deeper layers, which is reflected in the difference maps of ILD and D20. In EXP1, ILD has become shallower than the control run in the eastern Indian Ocean. In the western Arabian Sea, both EXP1 and EXP2 have shown a deepening in the D20 with respect to the control run. However, these are qualitative differences. The exact impact of the assimilation can be inferred by comparing the simulations with measurements.

An independent dataset consisting of continuous temperature profiles from Triton buoy located in the Eastern Equatorial Indian Ocean region (90°E and 1.5°S) has been used to assess the relative performance of the two experiments. These observations have been linearly interpolated to the model depths for comparison. Although the assimilation of temperature will affect other model variables (like currents and salinity) too, the analysis has been carried out only for temperature and some derived parameters.

Time evolution of SST, ILD and D20 in the control run, EXP1, EXP2 and observations is shown in Figure 3. Model-simulated surface temperature in all the three runs shows an underestimation of ~1°. This could be due to the thermodynamic imbalance caused between scatterometer winds and other air–sea exchange parameters from NCEP. However, the model picks up the patterns well. This underestimation in surface temperature is reduced with time in the two experimental runs. In EXP1 (nudging scheme) the surface temperature becomes closer to the observation as time progresses, whereas in EXP2 the effect is seen as and when the Argo observations closer to the TRITON buoy are getting assimilated. This is so because in EXP1 assimilation is performed at each time-step of the model, and in case the Argo data are absent, it is the Levitus climatology which affects the model. However, in EXP2 it is only when Argo data are available that the impact is seen.

Significant impact of assimilation on the isothermal layer depth simulation can be seen in EXP1, whereas EXP2 is close to the control run, except for the time when Argo observations are available. In EXP1, nudging of monthly data has smoothened the high frequency variability of the model, which otherwise was in better agreement with the observations, except for events that resulted in ILD shoaling (April–May). After April, ILD from EXP1 remains consistently underestimated. In EXP2, the ILD patterns have slightly improved with even high fluctuations being captured. However, EXP2 also failed to simulate the shallow ILD during April due to non-availability of Argo observations during that time.

D20, which was well simulated in the control run, was further improved using the Cressman technique for assimilation. High period oscillations, which were missing in the control run were induced and further maintained even

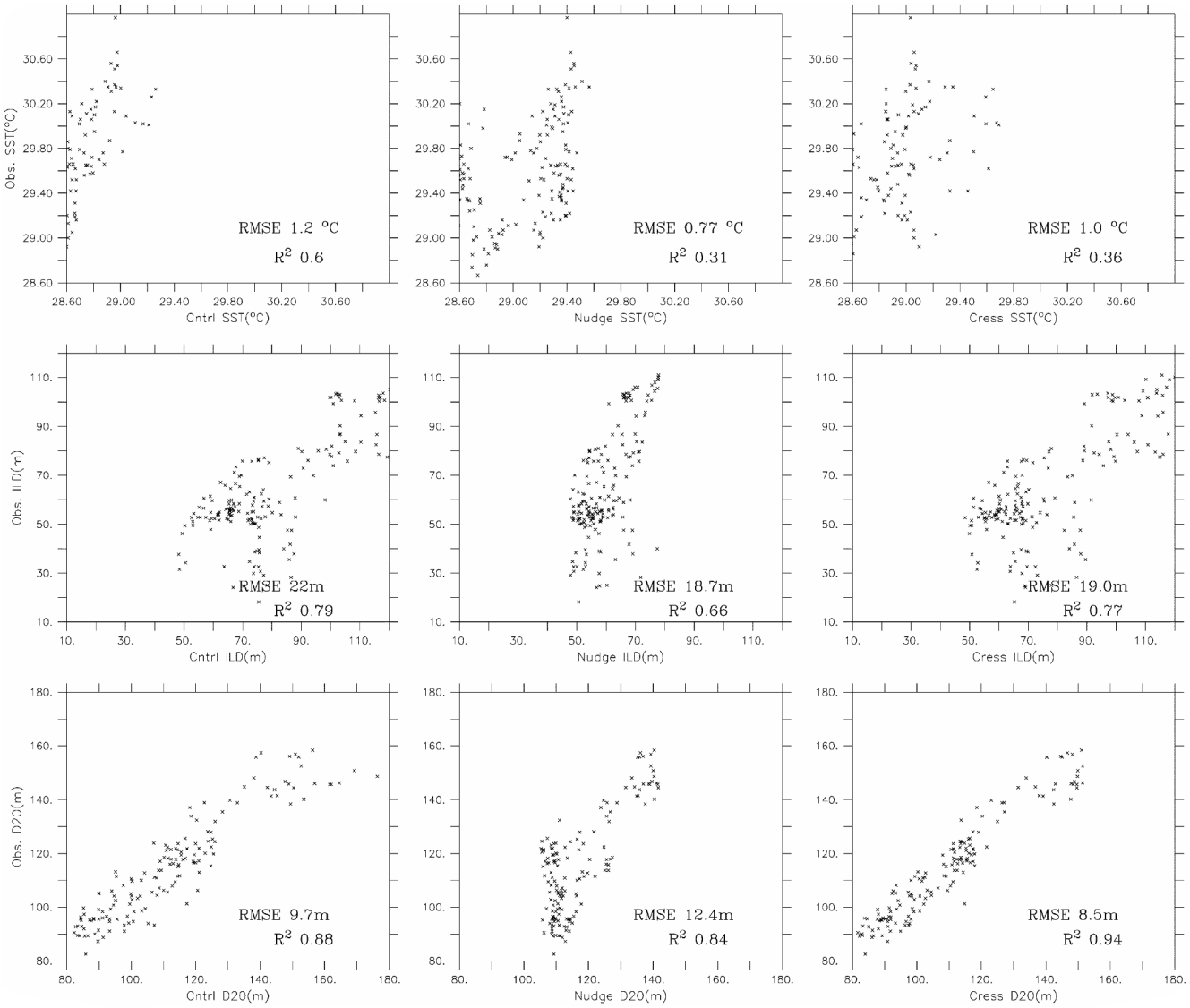

Figure 4. Scatter plots of observed and model-simulated SST (°C), ILD (m) and D20 (m) for control run, EXP1 and EXP2 at the TRITON buoy location (during assimilation phase).

when Argo data for assimilation were not available. Nudging technique of assimilation had completely failed in simulating the correct variability of D20. In fact, it had made D20 simulations worse than the control run. This is again because of the use of monthly average Argo data in EXP1, which lacked daily variability of the observations. The depth around D20 experiences maximum temperature variability (figure not shown), and therefore the performance of the nudging scheme is worst in that region. It was also observed that the standard deviation of the temperature profile from EXP2 is in close agreement with the observations.

Figure 4 shows the scatter plots of SST, ILD and D20 between the simulations and observations. It can be seen that after assimilation the root mean square error (RMSE) reduces by 37 and 16% in EXP1 and EXP2 respectively. However, the coefficient of determination (in both the experimental runs) is reduced compared with the control run, with EXP2 (34%) performing marginally better than EXP1 (31%). Reduction in RMSE of similar order can also be seen in the ILD simulations. In EXP2, the coefficient of determination is 0.77, compared to 0.79 and 0.66 in the control run and EXP1 respectively. As in the case of time-series plot, the maximum impact of assimilation is seen in D20 simulations. It can be seen that in EXP2 the scatter has become organized and the coefficient of determination of D20 with respect to observations is 0.94. RMSE for D20 increased in EXP1 and the coefficient of determination has decreased to 0.84 compared to 0.88 in the control run, which implies that assimilation of monthly data restricts the model from simulating low frequency variability.

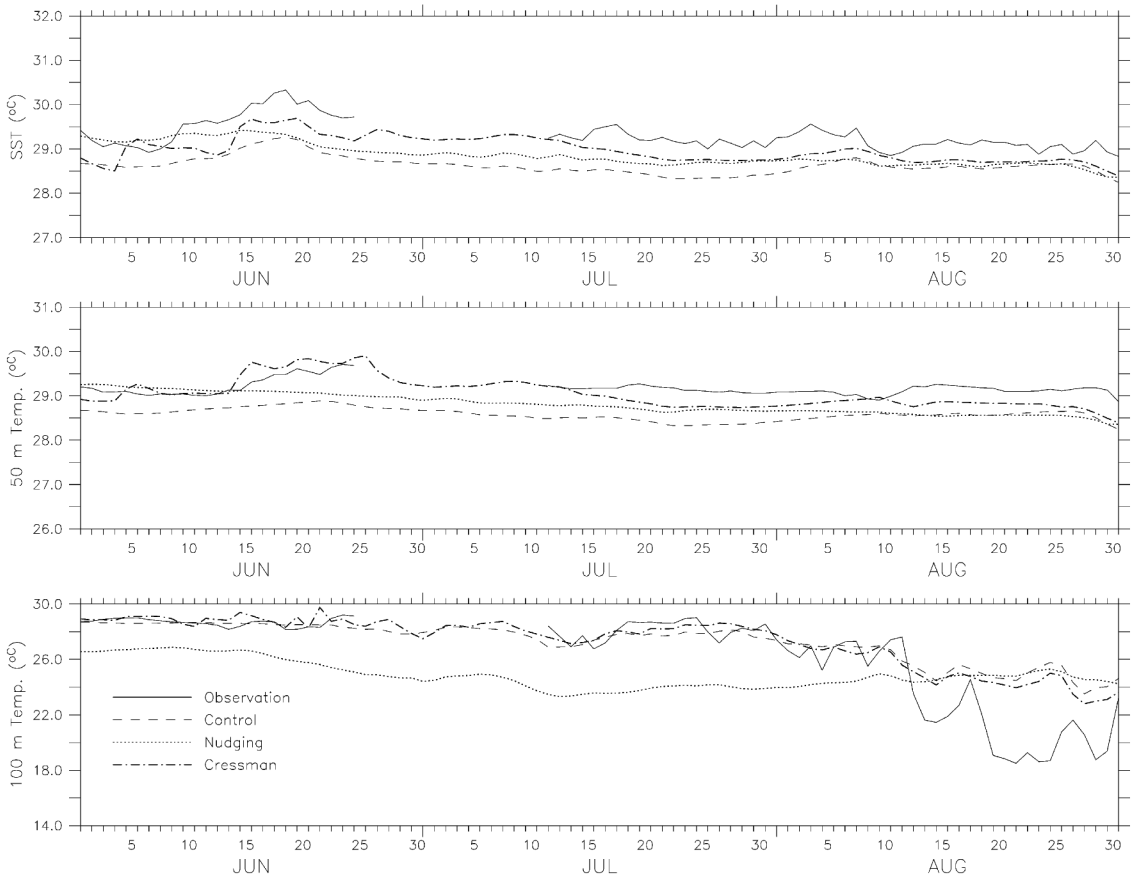

Figure 5. Time series of observed and model-simulated SST (°C), and temperatures at 50 and 100 m depth in the forecast phase during control and assimilation runs.

Having observed the effect of assimilation on SST, ILD and D20 during the assimilation phase, it would be instructive to assess the model performance during the forecast period (June–August 2004). In Figure 5, we show the temperatures at the surface, 50 and 100 m depths at the Triton buoy location. Observations from the buoy are also shown in Figure 5. At all depths, the correlation coefficients in the forecast phase after nudging (0.7 at the surface, 0.3 at 50 m and 0.22 at 100 m as against 0.87 at the surface, 0.6 at 50 m and 0.96 at 100 m for control run) are much less than the control run simulations. On the other hand, the Cressman technique has improved the simulations at 50 m ($R = 0.73$) and at 100 m (0.98). Maximum impact of assimilation in the forecast phase (after Cressman assimilation) is observed at 100 m. RMSE of 2.2°C in the control run decreased to 1.9°C in this case.

We had examined the impact of assimilation in the forecast run at a single point in the eastern equatorial Indian Ocean (Triton buoy location). Here we evaluate the impact of assimilation on SST at one of the National Institute of Technology buoy locations (14°N, 83°E) in the forecast phase (June–August 2004). The correlation increased from 0.41 in the control run to 0.92 in the forecast after nudging, and to 0.95 in the forecast after Cressman assimilation. Similarly, RMSE reduced from 0.34°C to 0.28°C in the case of nudging and 0.16°C in the case of Cressman assimilation. These facts suggest that in the forecast phase the impact of Cressman technique of assimilation is more than the nudging technique. The analysis was restricted to only one buoy, as there was no other buoy with continuous SST observations in this period.

Sub-surface temperature profiles available from Argo floats have been assimilated in an OGCM. Two experiments differing in the nature of their assimilation schemes were performed. The first experiment (EXP1) involves the use of nudging technique, while in the other experiment (EXP2) Cressman technique has been used. In EXP1 monthly averaged Argo profiles have been assimilated, whereas in EXP2 daily data have been assimilated. Impact of assimilation in general and the relative performance of the two schemes in particular, have been assessed using an Eastern Equatorial Indian Ocean buoy data (Triton). While the RMSE in SST has been reduced after assimilation in both the experiments, the variability in SST became less than the explained variance in the control

run. Subsurface features like ILD and D20 show significant improvement in terms of error reduction, implying improvement in the mixed layer and thermocline region. EXP2 scores over EXP1 in terms of the coefficient of determination of ILD and D20. This is because EXP1 is handicapped by the use of monthly averaged data. In future it will be interesting to investigate the impact of Argo profile assimilation on model-simulated currents and sea-level fields. Also advanced techniques of data assimilation (like ensemble Kalman filter) could be applied for better performance.

1. Malanotte-Rizzoli, P. (ed.), *Modern Approaches to Data Assimilation in Ocean Modeling*, Elsevier Oceanography Series, 1996, p. 455.
2. Wunsch, C., *The Ocean Circulation Inverse Problem*, Cambridge University Press, Cambridge, 1996, p. 456.
3. Kalnay, E., *Atmospheric Modeling, Data Assimilation, and Predictability*, Cambridge University Press, Cambridge, 2002, p. 342.
4. Evensen, G., Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, 1994, **99**, 10143–10162.
5. Fukumori, I., A partitioned Kalman filter and smoother. *Mon. Weather Rev.*, 2002, **130**, 1370–1383.
6. Hamill, T. M., Ensemble-based atmospheric data assimilation: A tutorial, University of Colorado and NOAA–CIRES Climate Diagnostics Center, Boulder, CO, USA, 2003, p. 46.
7. Argo Science Team, Argo: The global array of profiling floats. In *Observing the Oceans in the 21st Century* (eds Koblinsky, C. J. and Smith, N. R.), GODAE Project Office, 2001, pp. 248–258.
8. Liu, Y., Zhang, R., Yin, Y. and Niu, T., The application of ARGO data to the global ocean data assimilation operational system of NCC. *Acta Met. Sin.*, 2005, **19**, 355–365.
9. Yin, Y., Zhang, R., Shi, L. and Niu, T., In The Scientific and Technical Workshop of the Data Buoy Cooperation Panel, 18–19 October 2004, Chennai.
10. Balmaseda, M., Anderson, D. and Vidard, D., Impact of ARGO on analysis of the global ocean. *Geophys. Res. Lett.*, 2007, **34**, L16605.
11. Oke, P. R. and Schiller, A., Impact of ARGO, SST and altimeter data on an eddy resolving ocean reanalysis. *Geophys. Res. Lett.*, 2007, **34**, L19601.
12. Cressman, G., An operational objective analysis system. *Mon. Weather Rev.*, 1959, **87**, 367–374.

ACKNOWLEDGEMENTS. We are grateful to the Director, Space Applications Centre (SAC), Ahmedabad; the Deputy Director, Remote Sensing Area, SAC, and the Head, Oceanic Sciences Division, SAC for their encouragement. We also thank the International Argo team for providing the dataset from the floats. The ARGO datasets and QuikSCAT wind vectors were downloaded from ttp.ifremer.fr. Ocean model (MOM-3) was from Geophysical Fluid Dynamics Laboratory (NOAA).

Received 28 September 2007; revised accepted 24 July 2008

Attenuating the seismic interference noise on three-dimensional seismic data by frequency–receiver–shot (*f–x–y*) prediction filters

Sanjeev Rajput[1], N. K. Thakur[2,*], P. Prasada Rao[2] and Anand Joshi[3]

[1]Schlumberger Reservoir Seismic Services, Mumbai, India
[2]National Geophysical Research Institute, Hyderabad 500 007, India
[3]Department of Geophysics, Indian Institute of Technology, Roorkee 247 667, India

Encouraged by the potential prospects of oil/gas and gas hydrates, exploration activities in the marine environment have increased extensively. These seismic experiments however, result in contaminating the signal by interference noise which is generated due to neighbouring seismic vessels working in the same area. This noise in seismic reflection data may be substantially minimized by seismic interference noise attenuation technique. This technique utilizes frequency–receiver–shot (*f–x–y*) prediction filters to estimate the noise. The predicted filters can be applied in the *f–x* domain to remove noisy frequencies in seismic records. We have performed this exercise on a multi-channel seismic reflection dataset and the results indicate that noise from different sources has been successfully eliminated from the recorded datasets. Our study suggests that the *f–x–y* interference attenuation mechanism preserves the primary signal and eliminates the seismic interference in an efficient manner.

Keywords: Noise attenuation techniques, prediction filters, seismic interference.

WHEN more than one seismic vessel acquires data in a common area, the energy generated by sources of neighbouring ships causes interference in the signal recorded by other ships, which is known as seismic interference noise (SI). This is recorded along with the desired reflection data. This interference can be easily identified in the seismic section. Within a single trace, the interference consists of isolated, short-duration events that can hardly be recognized. However, across numerous traces, a characteristic trend develops, which is different from the trends of the real reflection characteristic, representing real geologic configuration. For reflection times greater than the water bottom time, this high-energy noise overrides weak reflections and is destructive to many pre-stack processes such as, surface multiple prediction, pre-stack migration and AVO trends. For seismic reservoir monitoring, interference noise from other seismic operations can cause either delay in data acquisition due to time sharing, or

*For correspondence. (e-mail: nkthakur46@yahoo.com)