Australian Argo National Data Management Report

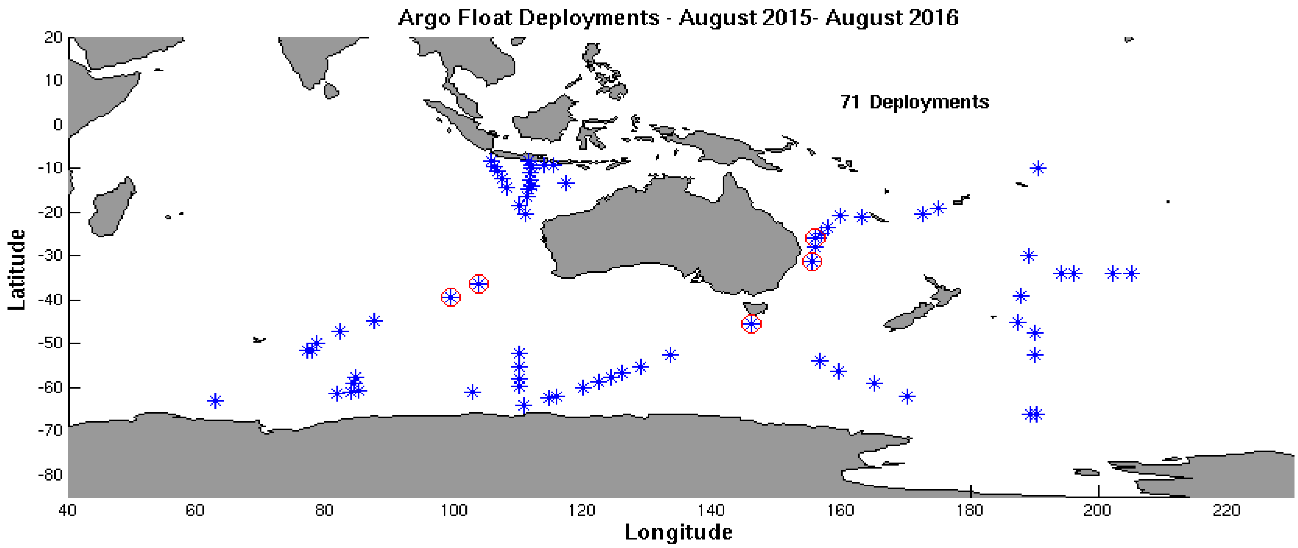
ADMT17

Tianjin China – 28-29 September 2016

Ann Gronell Thresher for the Argo Australia Team (CSIRO, Bureau of Meteorology)

Status of Array

Australian deployments in 2015-16:



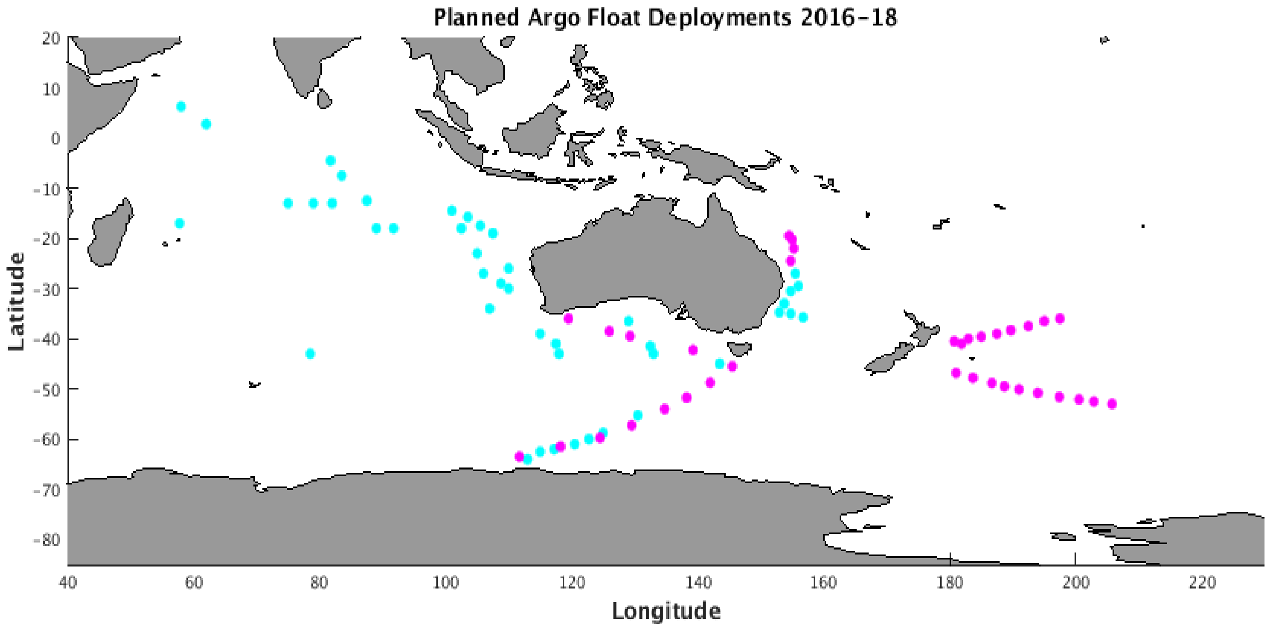
Australian Argo deployments between August 2015 and August 2016.
Red circles indicate bio-Argo deployments

Australia has deployed 71 floats since the last meeting, including 5 complex Bio-Argo models which follows the deployment of 4 such floats the previous year. This has again been a very busy and productive year for us. A significant feature of our deployments this year were those within the Indonesian EEZ, made possible only through the assistance of our Japanese colleagues.

We currently have 431 floats listed as 'live' though this includes some that are under ice or have been missing for over a year, from a total of 750 deployments since 1999. We also have 49 floats in the lab, most with deployment opportunities already identified. We hope to order a further 30 – 40 floats depending on funding outcomes, from multiple funding sources including the Australian Bureau of Meteorology (ABOM), the Royal Australian Navy (RAN), the Integrated Marine Observing System (IMOS) and the Antarctic Climate and Ecosystems Cooperative Research Centre (ACE-CRC). These purchases will help us to maintain float density in the South Indian and South Pacific Oceans.

Known deployment locations for the floats over the next year are shown below. We will continue to re-seed the Indian Ocean and South Pacific Ocean but some deployment locations are still to be decided. Focus this year will be on the central Indian (Kaharoa – Argo19); focus next year will be on the S. Pacific (Kaharoa – Argo20).

Australian Deployment plans 2016-17:



Locations identified for new float deployments. Cyan circles indicate floats to be deployed within the next 6 months, magenta indicates floats on order for deployment later in the year or next year.

Significant improvements:
With the help of Yulong Liu, a colleague from NMDIS who visited CSIRO in July, we have now coded our processing to use the latest versions of Matlab, utilizing the in-built netcdf routines instead of the older, CSIRO built netcdf routines. This makes processing faster and more reliable. We have also added trajectory file production to our real-time processing. All files are generated in format version 3.1.

Issues impacting data delivery in 2014-5:
Aside from further coding to create version 3.1 files, accommodating Bio-Argo data remains our biggest challenge. This has included the version 3.1 coding for the Bio data but also included deployment of floats that report profiles in new data formats.These have proved particularly challenging to process and deliver within the B-files.

The cumulative effect of the coding and getting things through the GDACs, as well as just gaining an understanding of this very complex data, has resulted in a long lag between data arrival and delivery to the GDACs for some of our floats. Other floats, with ‘known’ formats have had much shorter delays.

We also moved our servers and reorganized our data delivery ports. This caused some delays before the servers were correctly set up.

Software development:
We continue to evolve our code to deliver V3.1 for all of the 4 file types, which has been a huge undertaking and impacted the timeliness of both the RT and DM data streams. Trajectory files have only recently been completed and we continue to assess and adjust these files as needed. Profile, Metadata and Tech files are all being delivered in real-time in version 3.1, including B-Profiles. Almost all of our Bio data is now being delivered both in raw and derived form in the new BR data files.

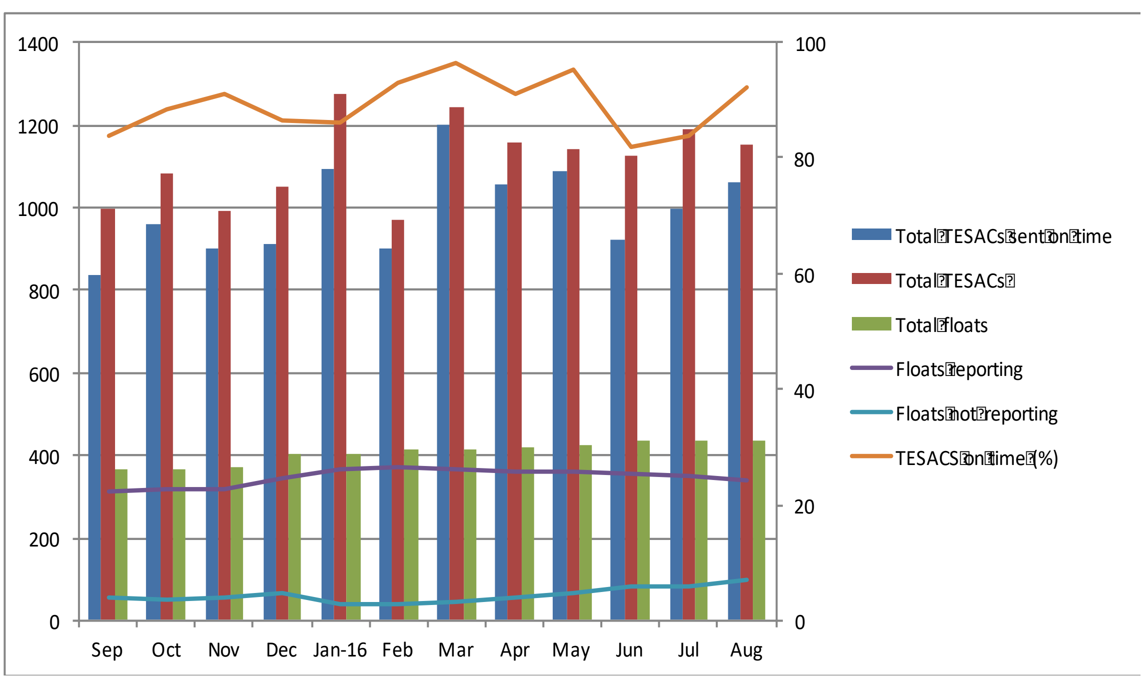
We have also finished redevelopment of our DM software and are now delivering DMQC'd oxygen data. We are working towards final software revisions and other data types will be added as manpower permits.

We continue to work with INCOIS and NMDIS on code for new float formats and new data formats.

Data Acquisition and delivery to the GDACs and GTS:
Our aim is that raw data is processed within a maximum of 18 hours of delivery from either Argos or via Iridium. We are achieving this for most of our floats. Some floats with new formats or which have been newly deployed take longer to enter our processing.

The data is issued to the GTS in both BUFR bulletins and TESAC messages by the Bureau of Meteorology (AMMC). These messages are generated 8 times daily.

Delays in data delivery appear to have improved but we will always have some floats that are under ice or have just been deployed and need additional processing before the data is sent out. Because many of the floats we are deploying tend to have the same formats as previous deployments, these delays are now minimal. Delivery problems this year appear to have been minor.



Summary of the timeliness of the Argo Australia GTS delivery for 2016.

Data is available for delayed mode QC as soon as the real-time data is processed but only considered eligible for DMQC after 12 months. The Delayed Mode report is appended below.

Additional Data Distribution:
As noted in previous years, the National Collaborative Research Infrastructure Strategy (NCRIS) funds the Integrated Marine Observing System (IMOS) which is a major source of Argo funding for Australia. As part of this initiative, it is required that we have a local data delivery pathway. IMOS is now serving Argo data as a mirror to the US GDAC through its data portal which can be accessed at:

http://imos.aodn.org.au/webportal/

All IMOS data, from all platform Facilities, can be accessed through this web site.

Web Pages:
The Australian Argo Real Time web pages are updated with the most recent data during the processing of the reports from the floats. They are therefore up to date as soon as float data is received. We have added web pages that contain details of the technical data from our floats, aiding in the diagnosis of problems. This is done as a float is processed making them up-to-date and easy to find.

Home page for Argo Australia (IMOS)
http://imos.org.au/argo.html

The Australian data portal can be found at:
http://www.imos.org.au/facilities/argo-australia.html

Information on individual floats can be found at:
http://www.marine.csiro.au/~gronell/ArgoRT/

There are links to the technical pages for a float from each profile page.

Statistics of Argo data usage:
Argo data is downloaded to a local mirror daily using the *rsync* capability. We then generate a Matlab file of the data with an index table to make data access easier.

Argo usage is a difficult list to compile, as Argo data are now being used routinely by many researchers nationally and globally. Not much has changed in the past year. In addition to the information below, there are numerous publications from Australian researchers which have used Argo data and have appeared in the last year.

The data is being used with other data on the GTS to inform the Bureau of Meteorology's Seasonal Climate Outlook and is used in a dynamical climate forecast system (POAMA). As part of this the data are ingested into the Bureau's Ocean Analysis (http://www.bom.gov.au/oceanography/analysis.shtml)

- Argo data is also being used in the BLUElink ocean forecasting system. http://www.bom.gov.au/oceanography/forecasts/index.shtml

- We are also incorporating it as a high quality background data field for our upper ocean temperature QC programs (QuOTA archives, SOOP XBT QC).

We report usage to our funders IMOS – the Argo report can be found at:

http://imos.org.au/imospublications.html

A large number of Australian PhD students are using Argo data and it is an integral part of many collaborative research projects which rely on our outputs. Please see the IMOS web site for more details.

Please also see the AST-17 report for a list of research projects using Argo data in Australia.

Delayed Mode QC (DMQC) Report:

Australian DM Statistics (to 12 Sep 2016)

	Core Argo	**Core Argo and EM APEX, Argo Eq., BGC and Bio Argo**
D files submitted to GDAC	79528	97631
Total R files	23393	44758
R files eligible for DMQC	12325	29730
Total eligible files for DMQC	91853	127361
Total files at GDAC	102921	142389
DMQC % eligible files	**87**	77

Table 1. Delayed Mode processing statistics for the Australian array.

We have made good progress towards our DMQC targets this year. Currently, the DMQC percentage stands at 87% of eligible core Argo profiles. Core Argo is defined here as floats with the standard P, T and S sensors including floats in the seasonal ice zone. The statistics for the entire Argo array, expanding core Argo to include Bio, BGC, Argo equivalent and EM Apex Argo we are currently at 77% of eligible profiles completed.

We aim to assess each float once per year and profiles are considered eligible when they are 12 months old to ensure there is an adequate time series to assess for sensor drift or offset. If a float is dead, then we process the entire record as long as profiles are more than 6 months old.

A challenge for our program is the significant increase in data volumes not only of the standard P, T and S floats but those with Bio or BGC sensors. We have spent significant time this year developing new DMQC processes and software to deal with the QC of oxygen and trajectory data. Floats that sample more than one profile per cycle also require manual inspection of both profiles for spikes, inversions etc. which significantly increases the time taken to DMQC these floats and has required a rewrite of all our DM software to deal with the multi-profile data formats.

A major achievement over the past year has been the DMQC of almost all of our Argo floats with DOXY sensors (91% submitted to the GDAC). We have assessed 67 floats with DOXY sensors through DMQC where the DOXY data is either on the primary, the secondary or on both profiles. In our fleet we have assessed 18 floats with Aanderaa Optodes (Stern Volmer calibration), 20 floats with Aanderaa Optodes (polynomial calibration), 9 floats with Aanderaa Optodes (factory calibration), 10 floats with SBE43 (factory calibration) and 10 floats with SBE63 (factory calibration). Of the 67 floats, we have submitted the P, T and S data for 61 floats, 56 of which also have DOXY data (5 floats had complete failure of the DOXY sensor). A further 2 floats were dead on deployment and therefore have no data. Another 4 floats have data issues that need to be resolved before the DOXY data can be submitted.

We have developed a DMQC approach that is modified from the approach used by Takeshita et al. 2013. We first assess the DOXY data on deep potential temperature or density surfaces to examine drift of the DOXY sensor with time. Consistent with other studies we find that once floats are deployed the oxygen sensors are very stable while the float is in the water and we did not find any evidence of sensor drift with time.

We do however find that there is 'storage' drift of the sensor, drift of the sensor that occurs in the lab, between the time of the calibration of the oxygen sensor and deployment (this can be many months for some floats). This means that when there is a co-located CTD on deployment, often the float data does not agree with the co-located CTD data and is generally lower than the climatology. Therefore, we correct the float DOXY data to climatology (CARS 2009) using a model II linear regression of percent oxygen saturation of the float data compared to climatology. We consider a gain only correction as most appropriate (as opposed to a gain and offset correction that Takeshita use) consistent with discussions with S Riser. In addition we exclude bad data and data sampled by the float in strong vertical gradients (exclude data where ascent rate > 0.2 uM/m). We also restricted data points used in the regression to those in the mixed layer and below 1500 db (where floats sample deep enough, otherwise deepest depth data available). Mixed layer data is defined using the density criterion of 0.03 kg/m^3 (Montegut et al., 2004) and a seasonal filter is applied, i.e. only data collected from May through September as the percent oxygen saturation remains very close to solubility equilibrium over these months. We estimate that once the data has been corrected to climatology the uncertainty of the corrected DOXY data is better than 3%.

We are also working on DMQC of other bio-sensors. In particular, we have started with CHL-A and are developing visualization products to help with this QC. This is complicated by non-zero black counts at depth which, when used to auto-correct the data, cause negative chlorophyll calculations. The "*Bio-Argo quality control manual for Cholorphyll-A concentration*" does not properly address this. Along with suggested light quenching and spike testing we have implemented an alternative profile Dark Count adjustment which uses the profile minimum as the Dark Count.

DMQC for optical/backscatter sensors is under development with manuscripts in preparation. For radiometric profiles a statistical method for profile quality classification has been used. The results show that about 50% of radiometric profiles are of good quality. DMQC for backscattering data is more complicated as most of sensors display an initial drift during the first month of operation that needs to be addressed. As recommended, a good-quality pre-deployment measurements are crucial for obtaining correct magnitude of BBP. We also suggest that using dual-channel backscatter sensors can be useful for controlling the quality of BBP profiles.

Results and preliminary visualization tools can be found at

http://www.per.marine.csiro.au/staff/Dirk.Slawinski/bioargo/index.html

And

http://wa-shiny.imos.csiro.au:3838/bioargo/