

ICOADS: A multinational data rescue, digitization, archiving, and access success for the ocean

Steven Worley, NCAR

Scott Woodruff, NOAA/ESRL

Eric Freeman, NOAA/NCDC

International **C**omprehensive **O**cean-**A**tmosphere **D**ata **S**et



**International
COADS**

Topics

- Introduction to ICOADS
- Lessons Learned
 - Data Management
 - International Partnerships
 - Data User and Project Expectations
- ICOADS Value-Added Database



ICOADS – Status Today

Release 2.5

- ICOADS Release 2.5: extensions and enhancements to the surface marine meteorological archive, IJC, First Published Online 2 Mar 2010, DOI: 10.1002/joc.2103
Scott D. Woodruff, Steven J. Worley, Sandra J. Lubker, Zaihua Ji, J. Eric Freeman, David I. Berry, Philip Brohan, Elizabeth C. Kent, Richard W. Reynolds, Shawn R. Smith and Clive Wilkinson
- Recovery of logbooks and international marine data: the RECLAIM project, IJC, First Published Online 2 Mar 2010, DOI: 10.1002/joc.2103
Clive Wilkinson, Scott D. Woodruff, Philip Brohan, Stefan Claesson, Eric Freeman, Frits Koek, Sandra J. Lubker, Catherine Marzin and Dennis Wheeler



ICOADS – Status Today

What it is: A surface marine meteorological and oceanographic dataset.

Date Range: Rel. 2.5 1662-2007, Preliminary Extension 2008-8/2010 (ongoing)

Number of Records: 261 M (not including P. Ext.)

Number of Unique Users: \cong 400 per year

Number of Data Sources: \cong 80

Number of Major Updates/Releases: \cong 10



ICOADS – Status Today

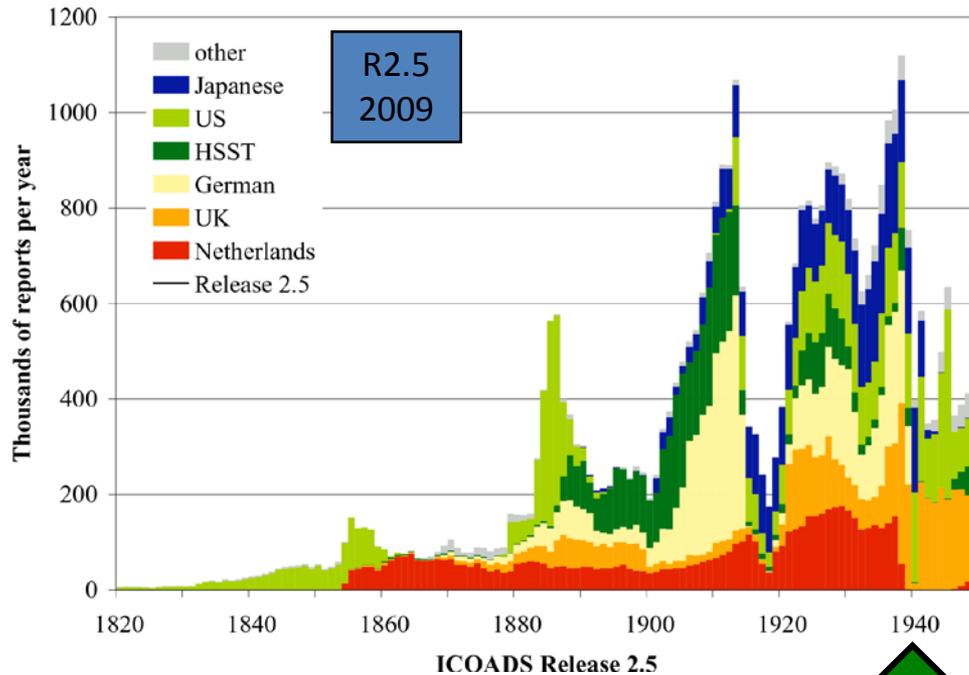
U.S. Primary Partners	Number of People
NOAA ESRL	1.5
NOAA NCDC	0.5 (CDMP ~ 1.0)
NCAR	0.5

International/National Partners and Contributors

- Met Office Hadley Centre
- National Oceanography Centre - Southampton
- Center for Ocean-Atmospheric Prediction Studies, FSU
- Deutscher Wetterdienst, DWD
- Koninklijk Nederlands Meteorologisch Instituut, KNMI
- Climate Research Unit, University of East Anglia
- Many others currently and in the past!



Early Data Mixture Changes: Science Impacts

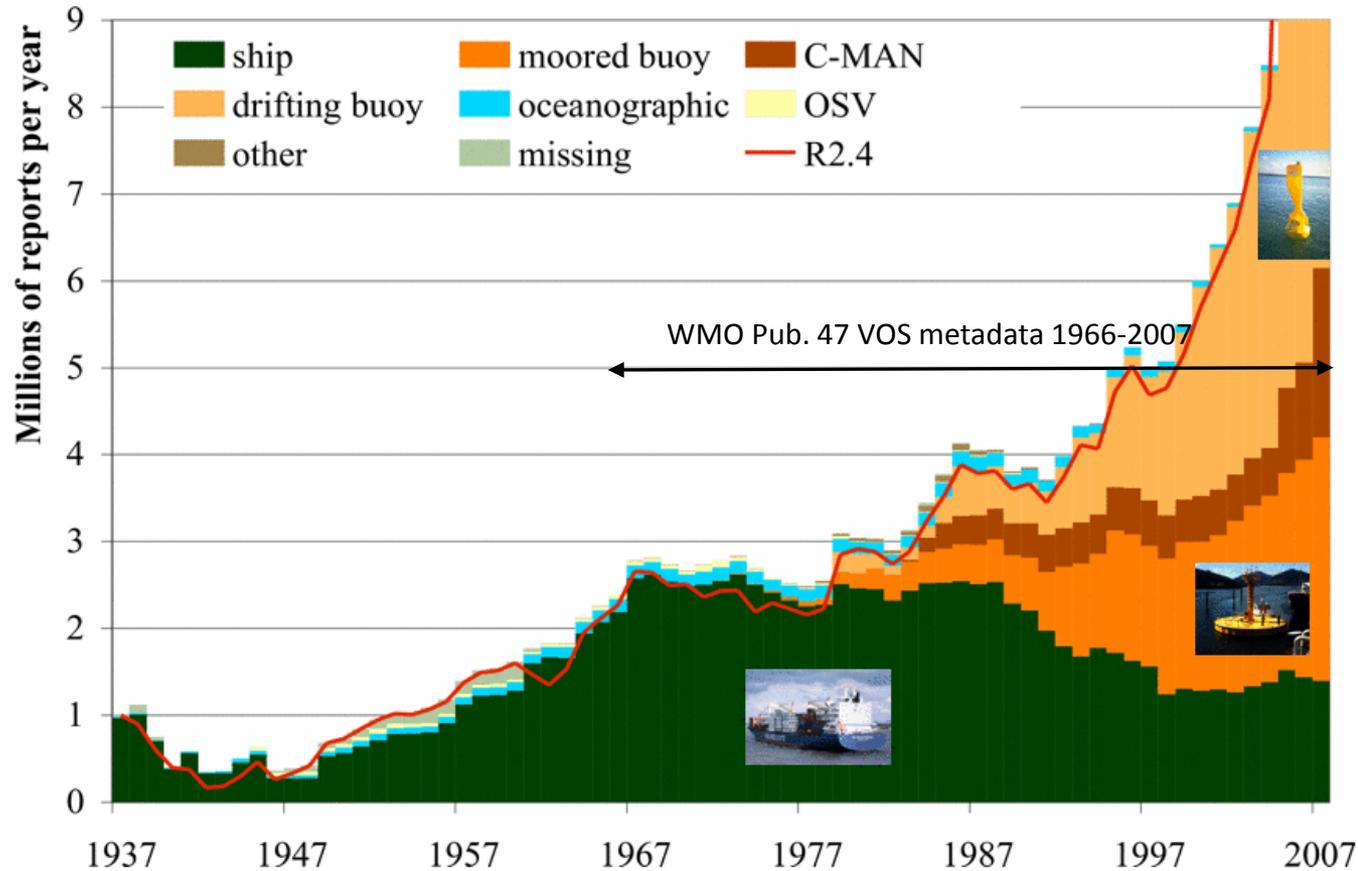


- Relative importance of national collections in early period.
- Changing mixture over time – has science impact
- WWII is drastically different – Thompson et al. 2008, Nature, noted discontinuity in global mean temperature.



**International
COADS**

Platform Mixture



Project Foundation

NOAA special report cover page

Comprehensive Ocean-Atmosphere Data Set Release 1, April 1985

CIRES	University of Colorado/NOAA Cooperative Institute of Research in Environmental Sciences Ralph J. Slutz, Sandra J. Lubker , Jane D. Hiscox
ERL	U.S. Department of Commerce National Oceanic and Atmospheric Administration (NOAA) Environmental Research Laboratories Scott Woodruff
NCAR	National Science Foundation sponsored National Center for Atmospheric Research Roy L. Jenne, Dennis H. Joseph
NCDC	U.S. Department of Commerce National Environmental Satellite, Data, and Information Service National Climatic Data Center Peter M. Steurer, Joe D. Elms

Forward to Release 1, by Joe Fletcher (NOAA)

"The incentive for developing the Comprehensive Ocean-Atmosphere Data Set (COADS) was to **make this record available to the individual investigator in a form that is reliable and easy to use**. The global marine surface data set contains the most detailed record we will ever have of the dynamics of the global climate system over the last century and more. It should trigger rapid progress in understanding by making it possible to delineate the spatial and temporal characteristics of the several sharp adjustments of the global circulation that have occurred, and to glean from them clues to the nature and causes of global climate variability. COADS provides the **material for diagnostic research** to identify and explore the key questions. It also provides the needed boundary conditions for model simulation of the climate system variability.

It has taken four years and much effort by many individuals and several institutions to obtain and process **the hundreds of tapes containing the basic data input**. All of this effort was provided from ongoing activities; there was no appropriation identified for the task. It is a **tribute to the spirit of cooperation among the participating organizations that the task has been successfully completed.**"

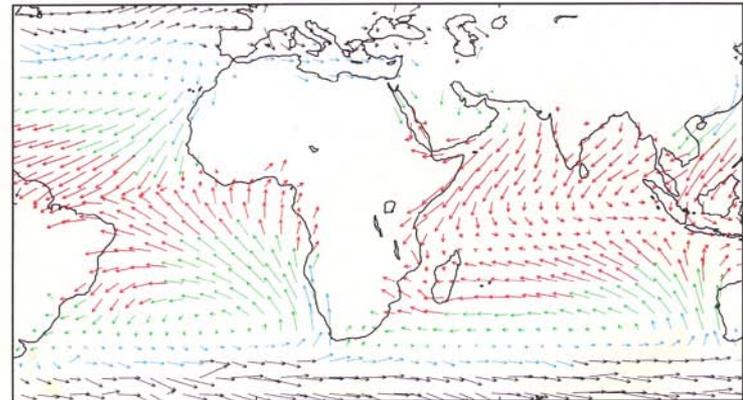
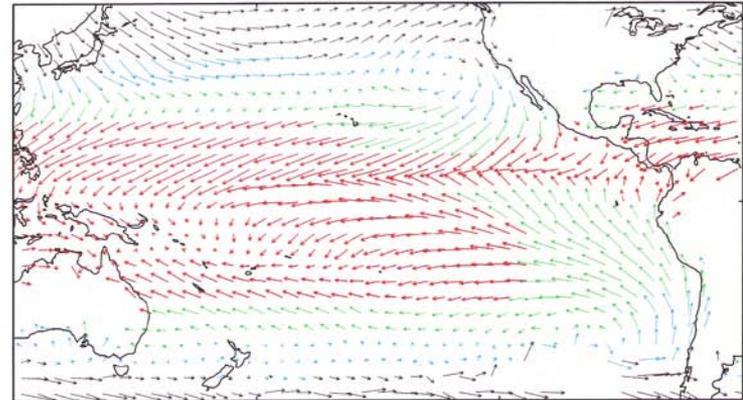
First Big Announcement

Woodruff, Scott D., Ralph J. Slutz,
Roy L. Jenne, Peter M. Steurer,
1987: A Comprehensive Ocean-
Atmosphere Data Set. *Bull. Amer.
Meteor. Soc.*, 68, 1239-1250

Bulletin of the American Meteorological Society

ISSN 0003-0007

Volume 68 Number 10 October 1987



Wind: → = 4 m/s. SST: black < 15°C ≤ blue < 20°C ≤ green < 25°C ≤ red

**International
COADS**



Data Management Lessons Learned

A successful multi-organization team requires balance

- Balance elements
 - Significant contribution of people and facilities
 - Mutually supportive expertise in areas like operations, archival, research
 - Commitment for an extended period
- ICOADS was forged out of ongoing activities
 - Institutional commitment led to sustainability
 - After 4 years, a strong foundation had been established



**International
COADS**

Support (people and \$,£,€) is always a challenge - multiple organizations can help buffer the crises

- Organizations can shift focus and impact progress
 - Climate data archives are research development efforts, but bridge with operational activities at near real time.
 - Too much emphasis on operations will effect resources to improve the historical record.
- When support ebbs at one organization it may increase at another – progress continues possibly in different areas



Long-term involvement of key individuals is essential

- Need key focal points with deep knowledge
- Need individuals to take charge of - or document - the hard questions
- Historical archives have many subtle characteristics
 - Answers come from knowing the history and evolution of the data

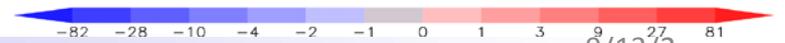
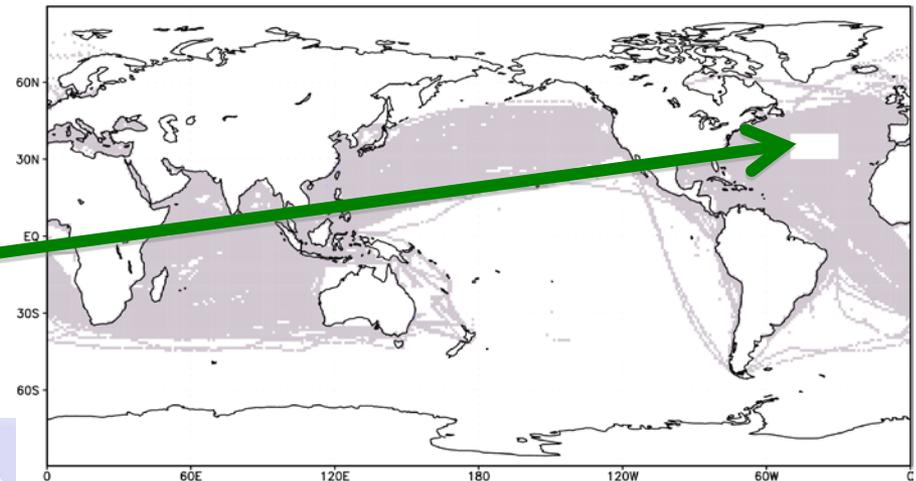


Organizing data sources is very difficult

- Data provenance - not well documented
- Many evaluations are needed
 - Inventories, distributions (time and space), duplicate assessment, date and time / location problems
 - <http://icoads.noaa.gov/r2.5.html>
 - <http://climateaudit.org/2010/08/30/a-first-look-at-icoads/>
- Study formats - QC'd data . Descriptions \neq findings.

Dutch deck 193: empty box
(data never received?)

Release 2.5 minus Release 2.4 193 Nobs 1930's



9/13/2
010



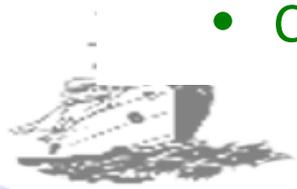
**International
COADS**

Finding duplicates between overlapping sources can be difficult

- Some Factors
 - Truncated values, inaccurate units conversion
 - Proper handling of 'missing' values, 00 ≠ null/blank
 - Early technology influence
 - 80 character punch cards - leading to abbreviated records.
 - Challenge to match GTS and delayed mode archives

Source data and documentation need long-term preservation

- Track the details
 - Data, documentation, event logs
 - Make backup copies of everything
- CDMP has funded small project to image legacy ICOADS documents
 - **Climate Data Modernization Program (CDMP)/NOAA**

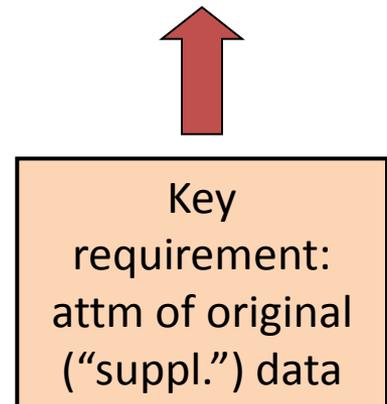


Design and use a robust and extensible data record format.

***ICOADS, IMMA: flexible format
(International Maritime Meteorological Archive)
core + optional attachments***



- ASCII, based on established formats, expanded with metadata for source tracking
- Core is multi-parameter, satisfies +95% of users
- Ship/instrument metadata – thanks to NOCS
- Easy to use and suitable for long-term archiving
 - Does well with these competing objectives



Advantage: exact copy of original permits re-translation and cross-checks at any time.



Data translation into a uniform format is necessary

- Brings differently formatted sources together
- Carefully done translation is the "foundation" for success
 - Undetected errors create negative downstream impacts
 - Keep emphasis on the "data" - system look and feel are secondary
- Mitigate difficulties by standardizing software to write records
- Create, use, share, standardized units-conversion libraries
- Assure source data format understanding/interpretation
 - Use two-person blind translations (expensive)
 - Amazing number of subtle QC issues are found



Evolution of technology will have impact

- Data storage concerns (1985) => compact binary formats
- Evolution factors
 - OSs, compilers, languages, applications, rapidly multiplied
 - Diversity of the user community demanded easier approaches
- Conversion to ASCII (IMMA) (~2000)
- Impact factors
 - Software stack modifications – producers, providers
 - Documentation rewrite
 - Access interfaces upgrade
 - Pushes unwanted changes onto regular data users
 - Large resource drain - seldom adequately funded
 - Risky but necessary

International Partnerships Lessons Learned

International partnerships are a substantial benefit (COADS to ICOADS in 2002)

- Add value through:
 - Content building, data discovery
 - Promoting broad community recognition
 - Sharing the workload
 - Detecting errors and resolving problems

Partnership conundrum - requirements and administration commitments grow without concomitant national resources support

- In the balance ICOADS is very positive on partnerships



Recognition by international bodies aids development

- Easier to get national resources to support the project if it has received 'approval' of international organizations
 - World Meteorological Organization (WMO),
 - Intergovernmental Oceanographic Commission (IOC)
- ICOADS is seeking formal international recognition
 - Through the Joint WMO-IOC Technical Commission on Oceanography and Marine Meteorology
- Goal: Create mirrored/shared development of ICOADS at WMO-IOC Centres



**International
COADS**

Regular international meetings, with a stated data focus, help drive progress and develop shared project ownership

ICOADS, benefited @ CLIMAR and MARCDAT meeting series



**Third International
Workshop on Advances in the Use of Historical Marine
Climate Data (MARCDAT-III)**

2-6 May 2011, Frascati, Italy

<http://icoads.noaa.gov/marcdat3/>
Formal announcement expected 1 October

Data User and Project Expectations Lessons Learned

ICOADS had over-expectations on how well the users would understand the data record format

- ICOADS archive heterogeneity => complex QC flagging
- Early years – many users ignored the flags
 - Results, erroneous extreme values, ship on land
- Solution – offer alternative data access pathways
 - Provide a web interface with standard QC defaults and selectable options
 - Serve access software with bulk file downloads



User expectations will shift over time

- Early years, simple monthly summary statistics
- Now, individual observations are most demanded
- Maybe - simple statistics replaced by climate-quality derived products
- Observational datasets are now small compared personal computers. ICOADS is 90 GB.



Today data access is expected to be immediate in 'my' favorite format

- ICOADS has not fully addressed some areas
 - Interoperable access, web services
 - GIS
- Organizations can share access services - each addressing different user communities
 - Critical point, data must be identical

Open data access can lead to out-dated secondary archives that negatively impact uninformed users

- Release 1a COADS (1993) can still be found on the web!
- Important, have a known authoritative location



**International
COADS**

The scientific reach is surprising

- The ICOADS data users are now more than physical meteorologist and oceanographers.
 - Examples: bird migration, turtle population studies, ship engine exhaust impacts, coral reef ecology, etc...
- This interdisciplinary nature is also expanding because of historical data rescue
 - Program support from:
 - RECLAIM, ACRE, CDMP

REcovery of Logbooks And International Marine data
(RECLAIM) Project
Atmospheric Circulation Reconstructions for the Earth (ACRE)
Climate Data Modernization Program (CDMP)



Need for precise data citation and version control is of hyper new importance

- Growing movement for accurate data referencing in publications, e.g. digital object identifier (DOI) for data
- Version of ICOADS need to be reproducible upon request
- Our approach has been *ad hoc* – we will formalize it in future



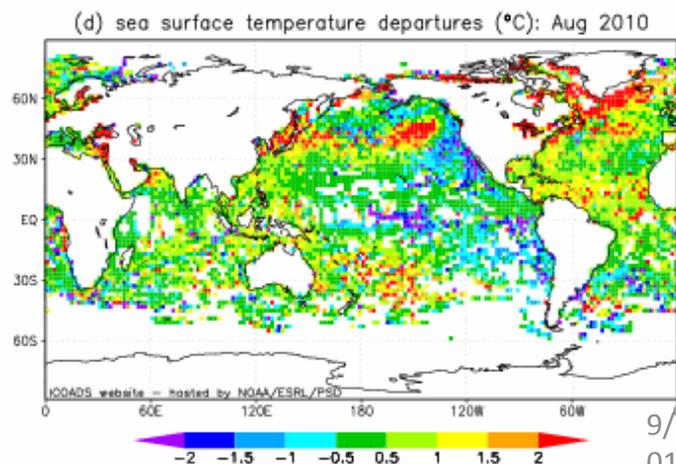
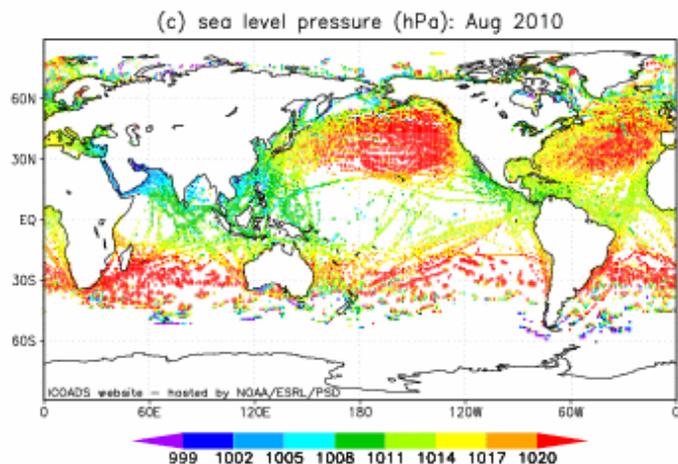
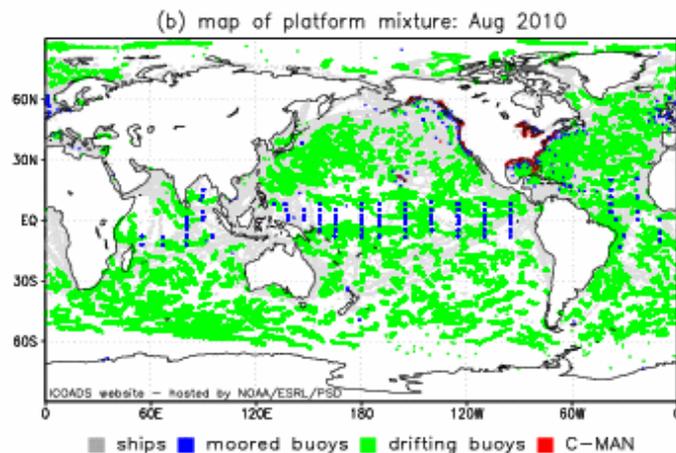
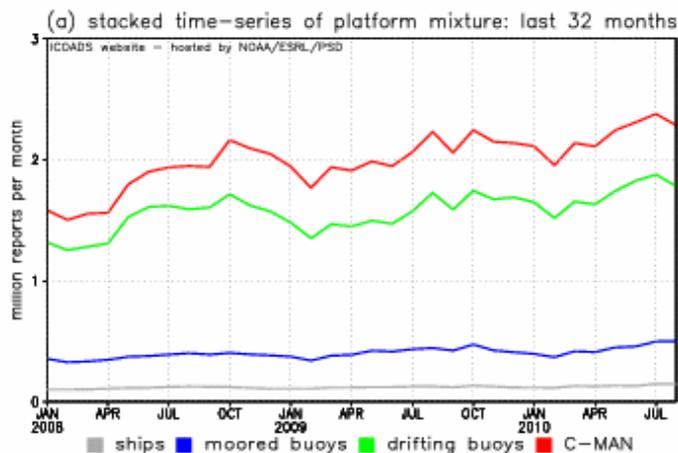
ICOADS Value-Added Database (IVAD)

- Status of a new proposal
 - 3-years, submitted to NOAA Climate Observations and Monitoring Program
- Partners
 - COAPS FSU, NOAA ESRL & NCDC, NCAR
- Main Features
 - Enhancement of IMMMA data format
 - Deploy ICOADS in a database
 - Receive and include some parameter adjustments
 - Provide user access to original and adjusted parameters
 - Demonstrate impact on marine surface flux estimates



<http://icoads.noaa.gov/>

- Last 32 month platform type time series
- Last month geo-located platforms, SLP, and SST departure
- Others – long-term plots of platform counts and parameters



ICOADS: A multinational data rescue, digitization, archiving, and access success for the ocean

International **C**omprehensive **O**cean-**A**tmosphere **D**ata **S**et



**International
COADS**