Adding Value to ICOADS through Data Rescue and Other Enhancements

Scott Woodruff\(^1\),
Steve Worley\(^2\), Eric Freeman\(^3\),
Clive Wilkinson\(^4\), Shawn Smith\(^5\)

1) NOAA Earth System Research Laboratory, USA
2) National Center for Atmospheric Research (NCAR), USA
3) Sourcecorp & CDMP/NCDC, USA
4) CRU/Univ. of East Anglia, UK & CDMP/NCDC, USA
5) Florida State University/COAPS, USA

3rd ACRE Workshop, Reanalysis and Applications, Baltimore, USA, 3-5 November 2010
Main Topics

• (1) ICOADS – joint project in US between:
  – NOAA (ESRL and NCDC) and NCAR
  – plus national/international contributions
  – plans to improve both recent and historical data

• (2) RECLAIM project
  – RECovery of Logbooks And International Marine data
  – Closely linked with ACRE and NOAA’s CDMP

• (3) IVAD proposal
  – ICOADS Value-Added Database
(1) ICOADS  [http://icoads.noaa.gov](http://icoads.noaa.gov)

- *International Comprehensive Ocean-Atmos. Data Set*
- Widely used for climate research etc. – reanalyses
- VOS plus buoys and other surface marine data types

The official Release 2.5 period (1662-2007) is now extended monthly with “preliminary” real-time data and products based on GTS data.

WMO Pub. 47 VOS metadata 1966-2007

1955-65 problematic
Improvements in ocean area coverage
Release 2.5 (1662-2007): completed 2009

US Marine Met. Journals

Keyed (or electronic) ship logbooks
Early Data Mixture Changes: Science Impacts

- Relative importance of national collections in early period
- WWII is drastically different – Thompson et al. (2008) *Nature*, described discontinuity in global mean temperature
**IMMA: A Robust and Extensible Observational Data Format**

- International Maritime Met. Archive (IMMA) format (ASCII)
- Core + optional “attachments”
- With metadata for source tracking
- Ship instrumental metadata – thanks to UK NOCS
- Suitable for long-term archiving
  - Carefully validated translations form “foundation” for all subsequent work

Key requirement: attm of original data forms: experience demonstrates format translations frequently contain errors or omissions

Advantage: exact copy of original permits re-translation and cross-checks at any time
Other Contemporary Data Challenges (in JCOMM Framework)

- VOS callsign encryption: since ~Dec 2007
  - Impacts NCEP datastream now used for ICOADS
- Need for improved international metadata for rigs and platforms
- UK Real-Time Monitoring Center (RTMC) to provide model feedbacks for GTS data
  - planned to populate ICOADS records
  - existing IMMA attm dedicated to these feedbacks
  - similar storage of e.g. 20CR reanalysis feedbacks?
Adding SAMOS Data to ICOADS

- SAMOS data center has QC’d marine R/V data since 2005
- Data at 1-minute intervals
- Will be sub-sampled to hourly and converted to IMMA
- Method used to provide earlier R/V data to ICOADS
- Est.: ~300K new obs.; 27 RVs
- Once implemented, updates can be sent to ICOADS on monthly basis as part of routine archival process with NODC
(2) RECLAIM Project

- Support from NOAA Climate Database Modernization Program (CDMP)
- Closely linked with ACRE and other international projects
- UK/other archive surveys (e.g. WW2, WW1)
- Published data and literature inventory
- Locating early platform and instrumental metadata


As practical, initiate **concurrent** processing:
- (b) prior to completing (a)
- (c) prior to completing (a-b)

Can be helpful to explore data quality/characteristics in advance (e.g. dups)

Translation into IMMA format

Among IMMA benefits:
- suitability for permanent archival (contrast e.g. WMO’s BUFR)

**Critical US resource bottleneck:** the translation of data in unique formats to IMMA is very expensive and presently not adequately resourced
Untapped marine instrumental data: probably most are ~1854-forward

Brussels Maritime Conference, 1853

But successful efforts starting with EU’s CLIWOC (1750-1850) contributing important pre-instrumental data (wind, weather, etc.)

Variables available from the US Maury Collection
Completion of another ICOADS delayed-mode update by ~mid 2012 is desirable for reanalyses
**US Lightship Collection**

1916-82 component; ~430K daily obs.
(CDMP-funding project initiated by WHOI)

<table>
<thead>
<tr>
<th>Lightship Name</th>
<th>Period(s) of record</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ambrose</td>
<td>1937-74</td>
</tr>
<tr>
<td>Barnegat</td>
<td>1947-70</td>
</tr>
<tr>
<td>Boston</td>
<td>1958-75</td>
</tr>
<tr>
<td>Buzzards Bay</td>
<td>1958-80</td>
</tr>
<tr>
<td>Chesapeake</td>
<td>1947-79</td>
</tr>
<tr>
<td>Delaware</td>
<td>1961-70</td>
</tr>
<tr>
<td>Diamond Shoals</td>
<td>1947-74</td>
</tr>
<tr>
<td>Five Fathoms</td>
<td>1957-72</td>
</tr>
<tr>
<td>Frying Pan Shoals</td>
<td>1936-79</td>
</tr>
<tr>
<td>Georges Shoal AFS</td>
<td>1956-60</td>
</tr>
<tr>
<td>Nantucket</td>
<td>1916-18</td>
</tr>
<tr>
<td>and</td>
<td>1947-82</td>
</tr>
<tr>
<td>Pollock Rip</td>
<td>1947-69</td>
</tr>
<tr>
<td>Portland</td>
<td>1956-66</td>
</tr>
<tr>
<td>Savannah</td>
<td>1954-64</td>
</tr>
</tbody>
</table>

Greenwich Mean Noon (GMN) Obs. 1910-47; Daily and Monthly Forms (plus other 19thC published forms etc.)

- Imaging complete
- Digitization underway  ✔ Goal: late 2011/early 2012
- 2.6M global reports
- Pipelining candidate
- Concern: Globally, this sampling strategy creates a day/night bias

R.M.S. Laconia (1912), courtesy of the Steamship Historical Society of America, Inc.
German Maury (1845-67)

- 850 logbooks loaned by DWD
- Imaged and digitized by CDMP; ~544K reports
- QC and translation into IMMA underway (by ~mid 2011)
  ✓ problems with interpretation of time elements and location
- SLP biases in other data from this era (needs research)
Proposal for Formal WMO-IOC Recognition through JCOMM

- To establish a network of mirrored WMO-IOC Centres for Marine-meteorological and Ocean Climatological Data (CMOC)
- Proposed requirements:
  - Host standardized formats and QC processing
  - Reliably mirror data and products
  - Open data access; **WIS (WMO Information System)** interoperability
- Benefits e.g. historical data exchange
  - Countries can be reluctant to exchange historical data without assurance of formal international repository
(3) ICOADS Value-Added Database (IVAD)

- 3-year research proposal
  - To NOAA/CPO Climate Obs. and Monitoring Program

- Partners
  - COAPS FSU, NOAA ESRL & NCDC, NCAR

- Main Features
  - Enhance IMMA data format
  - Deploy ICOADS in a database
  - Receive and include some community parameter adjustments (e.g. wind, SST, AT)
  - Provide user access to both original and adjusted parameters
  - Demonstrate impact on marine surface flux estimates

IMMA format improvements will include: Unique Record ID (UID) and improved ICOADS version tracking
International Marine Climatology Workshops

- Regular meetings, with a stated data focus, help drive progress and develop shared project ownership
- Next: MARCDAT-III 2-6 May 2011, Frascati, Italy
- Closely linked with satellite community (ESA, NASA)
  ✓ also linked with the International Land Surface Temperature Databank Initiative (initial meeting: Sept. 2010, Exeter UK)
- Overall objective of recommending a 10-year action plan for improved integration & accessibility of climatological observations

http://icoads.noaa.gov/marcdat3/
Abstract submission deadline: 15 Nov
Data Rescue: A Need and Challenge

• Need for augmented (and documented) “Best practices”
  • For imaging/digitizing e.g. 19th and 20thC marine obs.
  • For translations (e.g. standardized software libraries)

• Rationale
  ✓ Past efforts often limited by early technology (e.g. punched cards)
  ✓ Thus omitting key metadata (e.g. ship names: Challenger)
  ✓ But enhancing existing digital collections can be difficult to justify/fund

• Challenge
  – How to prioritize candidates given limited resources?
  – Seeking community agreement, but different research applications may foresee different priorities
  – Content assessment (if feasible) may be useful prior to decision (e.g. est. of parameter availability in time/space, accessibility)
  – Comparative studies needed to gauge cost/benefits against existing digital records?

Overall objective: Get the most climate data record benefit from highly distributed and very limited international resources