Introduction

Following translation into the IMMA format, Figures 1-2 illustrate the overall processing flow. Sections A-E below discuss each of the major steps in this processing.
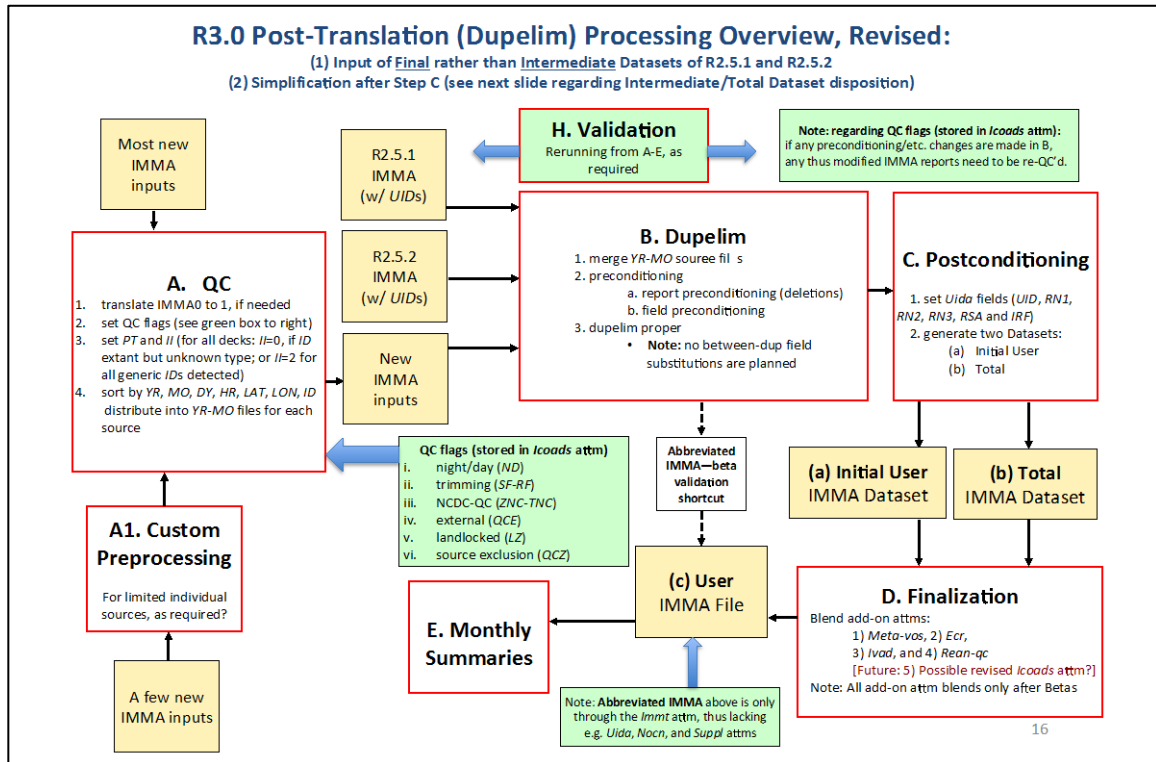

Figure 1. Overview of R3.0 post-translation (i.e. to IMMA format) processing data and product flow. Details of how Total (b), and any Intermediate (not shown, see Fig. 2), datasets will be managed remain to be decided (note: slide excerpted from R3.0-Processing-Logistics-v10.pptx).

Tentatively, the Total (b) dataset will be combined "virtually" with the previously constructed R2.5 intermediate files etc., as shown in Figure 2. But exactly how all these non-final data will be handled and served to users remains to be decided.
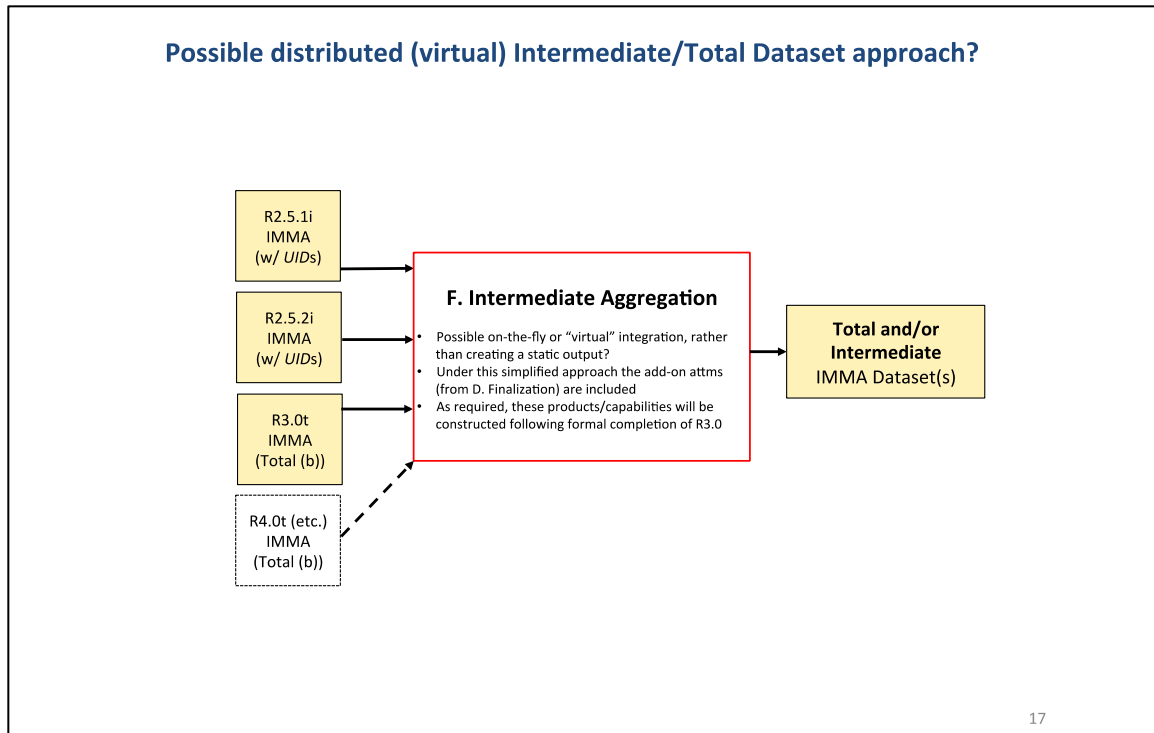
Figure 2. Potential virtual aggregation approach to managing and serving the intermediate vs. total files output from R2.5 vs. R3.0, respectively (note: slide excerpted from R3.0-Processing-Logistics-v10.pptx).

A. Quality control (QC)

These actions were applied to all new/revised decks during the initial quality control (QC) processing step:

a) Translate from IMMA0 to IMMA1 format, if needed.

b) All data inputs were checked, and corrected as needed, to ensure that they adhere to the ICOADS convention for representation of longitude (0.00°-359.99°E).

c) Left-justify *ID* as needed.

d) Compute *DPT,* if missing, from *WBT* and *AT* (and *SLP* if available; details below).

e) Calculate *RH*, if missing, from *AT* and *DPT* (details below).

f) Set all applicable ICOADS QC flags (Table 1; software available at: http://icoads.noaa.gov/software/qctrf1.f)

g) Set *PT* and *II*. For all decks: *II*=0, if *ID* is extant but unknown type; or *II*=2 for all generic *ID*s detected (software available at: http://icoads.noaa.gov/software/other/other/ptii.f)

h) Each input was sorted into ascending order by these respective keys: *YR*, *MO*, *DY*, *HR*, *LAT*, *LON*, *ID*.

i) Distribute into *YR-MO* files for each discrete source input (ref. Fig. 1, R2.5.1/R2.5.2, or new IMMA sources).

Table 1. Overview of QC flag groups set (or carried forward when available for *SQZ-SQA**) during this processing step, and stored in the *Icoads* attm.

	adaptive QC flags (*SQZ-DQA*) (<u>not</u> set during this processing step)*
i.	night/day (*ND*)
ii.	trimming (*SF-RF*)
iii.	NCDC-QC (*ZNC-TNC*)
iv.	external (*QCE*)
v.	landlocked (*LZ*)
vi.	source exclusion (*QCZ*)

* Only *SQZ-SQA* in use presently, these flags were set experimentally in 2002 just on R2.0 data (see also http://icoads.noaa.gov/aqc.html). Thus any data added after R2.0 will not have the flags, and any flags on R2.0 reports that were subtracted during subsequent processing are no longer available.

_____

Following is additional information on c) and d)-e), which are applied to all decks:
Rules:
c) Left-justify ID, with missing right-fill.
d) Compute missing *DPT* if *WBT* and *AT* are extant; if *SLP* is missing 1015.0 is used as *SLP*. This rule is not applied if *WBT* is greater than *AT*. Constants ACON and BCON are set for computation of *DPT* relative to water: ACON=7.5 and BCON=237.3. The following Fortran code is then used to attempt computation of *DPT*:

```
ESW = 6.1078*10.**(WBT*ACON/(WBT+BCON))
E = ESW-(.00066*SLP)*(((.00115*WBT)+1)*(AT-WBT))
IF(E.LT.0.) RETURN
CCON = ALOG10(E/6.1078)
DPT = BCON*CCON/(ACON-CCON)
```

where the resulting *DPT* is rounded to the nearest 0.1°C. To indicate that this calculation has taken place during this step, we set the dew point temperature indicator *DPTI*=1 for all data (for more information, see http://icoads.noaa.gov/software/other/other/qcimma1).
e) Compute missing *RH*, from *AT* and *DPT*, using PROFS function HUM (for more information, see http://icoads.noaa.gov/software/other/profs_short). To indicate that this calculation has taken place during this step, we set the relative humidity indicator *RHI*=3 (see http://icoads.noaa.gov/software/other/other/qcimma1) and add an *Immt* attm if not already extant. [Note: *RH* was not computed in the conversion of R2.5 (IMMA0) to R2.5.1 (IMMA1), and R2.5.1 was not re-QC'd in the R3.0 processing.]

Background:
c) This harmonizes the *ID* field for any inputs that were not left-justified during translation into IMMA format. Similar rules have been applied during earlier ICOADS updates instead at the field preconditioning stage, see e.g. *dupelim_1980*.
d-e) These steps prepare for monthly summary statistics by calculating *DPT* and/or *RH* where they would otherwise be unnecessarily missing. From 1980 forward original *DPTI* information may be extant depending on data source (but is missing for example in GTS data). In earlier ICOADS R2.0 processing, which utilized the obsolete LMR format, a separate second temperature indicator (*T2*) (no longer available in IMMA) was set during preconditioning, which for that earlier processing was the stage when missing *DPT* was calculated (from *dupelim_1980*):

"To indicate that this calculation has taken place during preconditioning, *T2* is set to 3, 4, 5, or 6, simply depending on whether the previous value of T2 was missing, 0, 1, or 2."

Further background information follows about *DPTI* and the associated *WBT*

indicator (*WBTI*), and about contrasting usage of the old *T2* indicator from: http://icoads.noaa.gov/e-doc/imma/R2.5-imma_short.pdf:

> "*WBTI* and *DPTI* are derived from sign positions s_w and s_t in IMMT-4. [Note: For data originally translated into LMR from IMMT formats, the predecessor LMR field *T2* preserved only a subset of information derived from s_w and s_t, coupled with whether *DPT* was computed during ICOADS processing. Future work should seek to recover more complete information for data that were translated to IMMA from LMR, and consider new configurations to separately document ICOADS processing…]"

B. Duplicate elimination (dupelim)

This section documents a sequence of processing steps referred to collectively as duplicate elimination (dupelim). First, following merger of the *YR-MO* source files output from QC (step A), "preconditioning" was used to delete individual marine reports (sec. 1), or to correct or modify individual data fields within a given report (sec. 2; Note: no such rules defined at present). The next step was dupelim proper (sec. 3), during which duplicate reports were flagged (rather than eliminated).

1. Report preconditioning

Deck is the field that initially determined the rule (or rules) to be used; decks that are not specified were not subject to preconditioning.[1] As part of this processing, for each applicable rule as specified below, the intermediate reject flag (*IRF*) was set (or reset) in postconditioning to one of the values listed in Table 2, which determines downstream retention in the intermediate and/or final datasets.

In the case of the final output data from R2.5 used as the largest input for R3.0 (i.e. R2.5.1 data), rules had already been applied during R2.5 preconditioning, which in turn were cumulative to rules applied even earlier (e.g. for R2.1, as detailed in "dupelim" documents available at: http://icoads.noaa.gov/e-doc/other/[2]). For the R2.5 and earlier preconditioning, it's important to note that such report deletions were not preserved in the R2.5 intermediate dataset (in contrast to the new approach for R3.0 to preserve selected deletions in the total file), nor were they preserved through any of the earlier processing, and the *IRF* flag was only recently defined for the latest IMMA1 format version.

Table 2. Intermediate reject flag (*IRF*) values (stored in the *Icoads* attm). *IRF* values 0-1 were already in use (for R2.5.1 and R2.5.2), thus new value *IRF*=2 was added to flag Rejects.

0	Intermediate (i.e. Retain in Intermediate Dataset, Reject from User Datasets)
1	Final (i.e. Retain in both Intermediate and Final User Datasets)
2	Reject (i.e. Reject from both Intermediate and Final Datasets)

---

[1] Any dates indicated as part of the preconditioning and other rules are inclusive, e.g., "October 1991-March 1992" refers to the beginning of October 1991 through the end of March 1992.

[2] Owing to funding problems that emerged in late 2011 after completion of R2.5, and associated staffing changes, the R2.5 documentation situation is still unsatisfactory in some respects, particularly for dupelim and also for the QC flags/screening rules. However, this document and *R3.0-stat_trim* bridge some of these documentation gaps.

Deck 145: PMEL (Daily) Equatorial Moorings and Island Stations
    Rule:
    Delete all deck 145 reports (*IRF*=2 Reject).

    Background:
    From Worley (1992): "…quality controlled records from eight EPOCS moored buoys and four island stations with very low topographic elevation. These reports are all daily averages of measured wind components, air temperature, and *SST* (at the buoys)." Since daily averaged data are very different than near instantaneous observations taken from ships, it was decided for R3.0 to remove this entire data source from the R3.0 final output.

Deck 700: UK Met. Office GTS BUFR Data
    Rule:
    Delete any drifting buoy (*PT*=7) report from deck 700 (*IRF*=2 Reject).

    Background: See deck 794 background.

Deck 714: Canadian Oceanography and Scientific Data (OSD; formerly ISDM/MEDS) Buoys
    Rules:
    a) Delete any report from deck 714 with position flagged doubtful or bad (i.e., lat/lon flag=3 or 4) (Note: *IRF* is not set since this deletion occurred during initial data translation prior to preconditioning). [Note: previously applied through 2007 during R2.5 processing.]
    b) Delete any deck 714 report if the third position of the *ID* falls in the range 0-4 (moored buoys) (*IRF*=2 Reject). [Note: previously applied through 2007 during R2.5 processing; applicable only from Jan. 1986 forward.]
    c) Delete any report from deck 714 containing an *ID* whose first two characters are "91" (*IRF*=2 Reject) [Note: previously applied through 2007 during R2.5 processing.]

    Background: General background on OSD (previously MEDS) data from *dupelim_1980*:
        "The COADS project provided MEDS with NCEP's Office Note 124 (ON124) GTS decode for 1980-January 1986 (ON124 report-type codes 561 and 562, corresponding to moored and drifting buoys, respectively). MEDS quality controlled both the moored and drifting buoy data for 1980-85, as well as MEDS-gathered drifting buoy data for 1986 forward (NOTE: some MEDS-gathered data also were added for the 1980-85 period). For the period starting 1986, MEDS also QC'd data transmitted from moored buoys in the DRIBU or BUOY code, such as PMEL TAO buoys."
    a) Earlier versions of the OSD data had an undesirable transition in 1993 in the frequency of flagging of doubtful reports, which required a time-dependent rule to terminate, starting in 1993, use of the flags (see *dupelim_1980*). However, the latest archive version appears to have resolved the transition problem such that we are able to uniformly use the lat./lon. flag over the entire period (also, additional flag value 4 for "bad" was not previously available).
    b) For major moored buoy arrays we instead use other delayed mode sources, including deck 146 for GTMBA (reporting over GTS in the BUOY FM 18 code), and deck 883 for NDBC coastal moorings (reporting in the SHIP FM 13 code). For simplicity, GTS, rather than OSD, data are used after 1985 for any other (i.e. non-

GTMBA and non-NDBC) moorings.
c) Beginning around November 1989, IDs for some NDBC Western Pacific C-MAN (Westpac) stations took a numeric form resembling a 5-digit WMO buoy number except beginning with "91" (not legitimate starting digits for a buoy number; see *dupelim_1980* for additional background). Reports from these stations have been sporadically misassigned into moored and drifting buoy datastreams (e.g., decks 714, 793, 794). [Note: This rule is less restrictive than the one in *dupelim_1980*, which required that the "91" be followed by three numeric characters.]

Deck 715: German Deep Drifter Data (via OSD; originally from IfM/Univ. Kiel)
Rule:
Delete all deck 715 reports (*IRF*=2 Reject).

Background:
Tests (including "blacklisting" by the Met Office, ref. Atkinson et al. 2013) have shown the *SST* data to be of poor quality relative to other drifting buoy data (see also Appendix D in the R3.0 Workplan).

Deck 732: Russian Marine Met. Data Set (MORMET) (rec'd at NCAR)
Rule:
Delete reports from deck 732 (*IRF*=2 Reject) during 1958-74 and for selected regions according to the method prescribed by Kennedy et al. (2011) in their Tables 1-2.

Background:
From Kennedy et al., 2011): "[17] A manual scan of the data was performed after gridding the observations at monthly 1 degree latitude by 1 degree longitude resolution. Some observations from deck 732 between 1958 and 1974 were identified as being incorrectly located. A number of these areas were identified in R06. Seventeen 5- degree areas or blocks of 5- degree areas were obviously artificially warm or cold relative to neighboring areas and relative to other observations within the areas. Data from deck 732 were not used from the areas specified in Table 1 at the times specified in Table 2." [Note: tables reproduced below]:

Deck 780: NODC/OCL World Ocean Database (WOD) (and formerly Atlas, WOA)
    Rules:
    a) Delete any report from deck 780 with all of *SST*, *AT*, *SLP*, and wind speed and direction missing (*IRF*=2 Reject).
    b) Delete any report from deck 780 except those containing *SID*=149 (*IRF*=2 Reject).
    c) Delete any moored or drifting buoy report from deck 780 (*PT*=6 and 7) (*IRF*=2 Reject).

    Background:
    a) (See *dupelim_1980*.)
    b) The 2013 Edition (WOD05; *SID*=149) provides a complete replacement for earlier versions of the WOD (and WOA) data.
    c) Past inventories have indicated that reports from many of the same moored buoys appear in other sources, and OSD (deck 714) is used as our sole source of drifting buoy data. [Note: These data mixture issues merit additional study in the future.]

Deck 793: NCEP BUFR GTS: Buoy Data (transmitted in FM 13 "SHIP")
    Rule:
    Delete any drifting buoy (*PT*=7) report from deck 793 (*IRF*=2 Reject).

    Background:
    See deck 794 background, adding only that we have found in the past, based on *ID* (buoy number) characteristics for unknown reasons small numbers of drifting buoy reports can also appear in deck 793.

Deck 794: NCEP BUFR GTS: Buoy Data (transmitted in FM 18 "BUOY" code)
    Rule:
    Delete any drifting buoy (*PT*=7) report from deck 794 prior to 2015 (*IRF*=2 Reject).

    Background:

7

Canada OSD provides a quality controlled set of global drifting buoy data (deck 714), based on GTS receipts. It should be noted that these delayed-mode data are significantly down-sampled temporally (through consolidation of fragmentary reports etc.) in comparison to raw GTS receipts, with unanticipated homogeneity impacts therefore possible on analyzed products etc., applicable for users of ICOADS data/products crossing the boundary between R3.0 and its near-real-time GTS-based extension.

Deck 795: NCEP BUFR GTS: Coastal-Marine Automated Network (C-MAN code) Data
Rules:
a) Delete any C-MAN (*PT*=13) report from deck 795 prior to October 2012 (*IRF*=2 Reject).
b) Delete all sub-hourly reports from deck 795 [note: handled separately in advance of preconditioning, thus not assigned *IRF*=2, and not in Total dataset.]

Background:
a) NDBC provides a higher quality and more complete set of official C-MAN data (part of deck 883); this preconditioning step assumes that no other useful data will appear in deck 795. [Note: Related information from *dupelim_1980* (R2.0 processing): NDBC's C-MAN program is documented to begin in March 1983, and inventories confirm that the TD-1171 C-MAN data indeed start at that date. Earlier data appearing in deck 895 have been determined not to be from lightships as was previously thought and the quality of such data was found to be extremely poor. Therefore, as part of the 1980-92 NCEP reconversion, all data prior to March 1983 that would ordinarily go into deck 895 were written out for later study to a separate file, not to be used for Release 1a."]
b) Data volumes and redundancies between C-MAN/tide gauge/coastal decks (i.e. 795, 797, and 995, with the latter containing particularly high data volumes) require further investigation as to the reasons and what are the most useful data to preserve for most users in the final IMMA output.

Deck 797: NCEP BUFR GTS: CREX code
Rule:
Delete all sub-hourly reports from deck 797 [note: handled separately in advance of preconditioning, thus not assigned *IRF*=2, and not in Total dataset.]

Background:
See deck 795 background.

Deck 874: Shipboard Environmental (Data) Acquisition System (SEAS)
Rule:
Delete any report from deck 874 with *SID*=111 (*IRF*=2 Reject).

Background:
This rule eliminates a highly erroneous (ref. http://icoads.noaa.gov/deck874.html) early version of the SEAS data (1991-2007; 438K reports output from R2.5). A new version (*SID*=150) of these data (SEAS9.1; IMMT-5 format) is expected to provide higher quality data going forward (2013-forward only however, the data for 1991-2007 appear likely to be uncorrectable).

Deck 993: NCEI GTS: Buoy Data (transmitted in FM 13 "SHIP" code)

Rule:
Delete any drifting buoy (*PT*=7) report from deck 993 (*IRF*=2 Reject).

Background:
See deck 793 background.

Deck 994: NCEI GTS: Buoy Data (transmitted in FM 18 "BUOY" code)
Rule:
Delete any drifting buoy (*PT*=7) report from deck 994 prior to 2015 (*IRF*=2 Reject).

Background:
See deck 794 background.

Deck 995: NCEI GTS: Coastal-Marine Automated Network (C-MAN code) Data
Rules:
a) Delete any C-MAN (*PT*=13) report from deck 995 prior to October 2012 (*IRF*=2 Reject).
b) Delete all sub-hourly reports from deck 995 [note: handled separately in advance of preconditioning, thus not assigned *IRF*=2, and not in Total dataset.]

Background:
a) See deck 795 background. Noting in addition that NCEI apparently receives some unofficial C-MAN/coastal reports that may be eliminated by this rule, but not replaced via deck 883, as are GTS reports from the official NDBC C-MAN network.
b) See deck 795 background, adding that deck 995 contains particularly high coastal data volumes.

2. Field preconditioning

Deck 780: NODC/OCL World Ocean Database (WOD) (and formerly Atlas, WOA)
Rules:
a) Delete any *AT* from deck 780 SURF reports [note: performed at customized advance QC step].

Background:
a) Tests have shown that most or all of these air temperatures are erroneous.

3. Dupelim overview

We use an upgraded version of the *Release 1* dupelim procedure[3], including a number of modifications and enhancements as described here. Quality code, as computed by the NCDC-QC (*Release 1*, supp. J), is the basis for the selection of one duplicate report over another; or, if quality codes are identical, a priority list by deck is used; or, if priorities are also identical, the report with the highest deck is preferred; or, if the decks are identical, the report with the highest *SID* is preferred; or, if the *SID*s are identical, the

---

[3] The procedure considers reports within the same 1°×1° box and within plus or minus one hour ("hour cross") or day ("day cross") as possible duplicates (dups). It performs a check for seven weather elements (wind speed, visibility, present weather, past weather, *SLP*, *AT*, and *SST*) to identify dups. These checks for weather elements include "allowances," which consider weather elements (or hour) to match under some circumstances even though they were not equal. In the current procedure, no reports are eliminated at this stage, but dup status (*DUPS*) is set to indicate dup certainty depending on a variety of factors, for later elimination of dups and other selected reports during postprocessing.

second report in sort order is selected.

The starting point for this code was the dupelim version used for the near-real-time GTS blend (http://icoads.noaa.gov/merge.html), but modified:
- to eliminate the rule "(i) *ID*s are required to match except if one or both are generic,"
- retaining the rule: "(ii) deck 797 stations are included with a two tenths difference allowance for wind speed and *SST*,"
- data had already had this rule applied "(iii) activestations.xml is used to set (NCEI) or reset (NCEP) locations of C-MAN-format and CREX-format receipts,"
- retaining the rule: "(iv) DUPS 6 (time/space) is disallowed."
- and applying the rule: "Worse duplicate is determined by the following in decreasing significance: (a) larger quality code, (b) larger priority code, (c) generic *ID*, (d) smaller deck number, (e) smaller source id, or (f) first in sort order."

Special deck rules and revised priority structure
Table 3 lists special rules that are applied to some decks, and priority codes for each deck. The special rules, acting independently from the priority codes and other selection criteria, may force a given deck to be selected or not selected, and allow some decks to "pass through" dupelim unchanged.

Table 3. Dupelim special deck rules and priority codes, for R3.0 and as previously used for R2.5 (priorities for R2.5 processing sometimes varied by time period, as indicated; Appendix A provides further details). Unshaded decks are entirely new for R3.0; whereas shaded (i.e. in the deck column) decks were already part of R2.5. Within the latter category, grey-shaded decks are unchanged for R3.0 (no additional/revised input data), whereas tan-shaded decks have been augmented with new/revised data, or entirely replaced. For the purpose of setting *DUPS* and *DUPC*, all decks are considered logbook (or delayed-mode) data unless specifically indicated by a footnote below as GTS. Priorities listed as "(n/a)" refer to decks for which no data are available at the present time, or those listed additionally as "del." refer to decks input, but deleted during report preconditioning, sec. 1.

| R3.0 | | R2.5 | | | |
Rule	*Priority*	*Rule*	*Priority*	*Deck*	*Description*
	2		2	110	US Navy Marine
	2		2	116	US Merchant Marine
	3		3	117	US Navy Hourlies
	2		2	118	Japanese Ships No. 1 (Kobe Collection Data keyed in 1961)
	2		2	119	Japanese Ships No. 2 (Kobe Collection Data keyed in 1961)
	2		2	128*	International Marine (US- or foreign-keyed ship data)
[b]	2	[b]	2	143	Pacific Marine Environmental Laboratory (PMEL) Buoys
[b]	1	[b]	1	144	TAO/TRITON and PIRATA Buoys (from PMEL and JAMSTEC)
(n/a del.)	(n/a del.)	[a]	1	145	PMEL (Daily) Equatorial Moorings and Island Stations
[b+]	1			146	Global Tropical Moored Buoy Array (GTMBA) from PMEL via NOC
	4		4	150	Pacific (US Responsibility) HSST Netherlands Receipts

	4		4	151	Pacific (US Responsibility) HSST German Receipts
	4		4	152	Pacific (US Responsibility) HSST UK Receipts
	4		4	155	Indian (Netherlands Responsibility) HSST
	4		4	156	Atlantic (German Responsibility) HSST
	2		2	184	Great Britain Marine (194 extension)
	2		2	185	USSR Marine IGY
	3		3	186	USSR Ice Stations
	2		2	187	Japanese Whaling Fleet
	2		2	188	Norwegian Antarctic Whaling Factory Ships
	2		2	189	Netherlands Marine
	2		2	192	Deutsche Seewarte Marine
	2		2	193	Netherlands Marine
	2		2	194	Great Britain Marine
	2		2	195	US Navy Ships Logs
	2		2	196	Deutsche Seewarte Marine (192 extension)
	2		2	197	Danish (and Other) Marine (Polar)
	4		varying	201-255	UK Met. Office (MetO) Main Marine Data Bank (MDB) (except with different rules and priorities applicable to the following MDB decks)
[z]	6	[z]	6	215	UK MDB German Marine
[b]	2	[b]	2	245	Royal Navy Ship's Logs (keyed by 2007) (1938-47)
	1		1	246	Atmospheric Circ. Reconstructions over the Earth (ACRE) Digitized Data: Print./Published Expeditions (held at Met. Office) [new: ACRE Expeditions transl. by UKMO (*SID*=167) or by NCEI (*SID*=168)]
	1		1	247	ACRE Digitized Data: *Challenger* Expedition [new: version of *Challenger* Expedition (*SID*=167) vs. old *SID*=138)]
[a]	1			248	English East India Co. (EEIC) Ship Logs
	1			249	Extended WW1 UK Royal Navy Ship's Logs (OldWeather) (1914-23)
	3		3	254	Int. Maritime Met. (IMM) Data (foreign or unknown origin)
	5		5	255	Undocumented TDF-11 Decks or MDB Series [note: previously handed under general 201-255 rules]
	5		5	281	US Navy Monthly Aerological Record (MAR)
[z]	5	[z]	5	555**	US Navy Fleet Num. Met. and Oceano. Center (FNMOC; Monterey) Telecom.
	2		2	666**	Tuna Boats
[a]	1	[a]	1	667	Inter-American Tropical Tuna Commission (IATTC)
	3		2	700**	UK Met. Office GTS BUFR Data [new: additional VOSClim receipts under existing *SID*=110, plus FM13/FM18 (*SID*s=146/147) data]
[b]	1	[a]	1	701	US Maury Collection
	1		1	702	Norwegian Logbook Collection
	2			703	US Lightship Collections
[b]	1	[b]	1	704	US Marine Meteorological Journals Collection (1878-94)

	1		1	705-707	US Merchant Marine Collection (1912-46) (500/600/700 series)
	3			708	US Navy Marine (US-keyed ship data; hourly METAR format)
	2			709	US Navy Marine (IMMA formatted by US Navy)
	1			710	US Arctic Logbooks (OldWeather)
	1			711	Weather Detective Crowdsourcing (Australian Wragge Collection)
[b]	2	[b]	2	714**	Canadian Oceanography and Scientific Data (OSD; formerly ISDM/MEDS) Buoys [new: 2008-14 data]
(n/a; del.)	(n/a del.)	[a]	1	715	German Deep Drifter Data (via ISDM; originally from IfM/Univ. Kiel)
	2		3	720	Deutscher Wetterdienst (DWD) Marine Meteorological Archive [new: German Light Vessels (*SID*=161) plus "HISTOR" project (*SID*=160)] [Beta2 results: Unless *ID* is required to match, deck 720 (including earlier-provided R2.5 DWD data also falling within this deck) is matching various other sources. Generally, however, deck 720 was selected more frequently than the other sources.]
[b]	1			721	German Maury Collection
[a]	1	[a]	1	730	Climatological Database for the World's Oceans (CLIWOC)***
[a]	1	[a]	1	731	Russian S.O. Makarov Collection
[z]	6	[z]	6/5	732	Russian Marine Meteorological Data Set (MORMET)****
	2		2	733	Russian AARI North Pole (NP) Stations (from Polar Science Center)
	1		1	734	Arctic Drift Stations [new: North Pole Expedition of the *Fram*]
	1		1	735	Russian Research Vessel (R/V) Digitization
	1		1	736	Byrd Antarctic Expedition (keyed by Hollings Scholars)
	2		2	740	Research Vessel (R/V) Data Quality-Evaluated by FSU/COAPS [new: SAMOS data (*SID*=131)]
[z]	5	[z]	5	749	First GARP Global Experiment (FGGE) Level IIb (*SID*=53)
[a]	1			750	Australian Navy Vessels: SST Data (1972-77)
	1		1	761	Japanese Whaling Ship Data (CDMP/MIT digitization)
	1		1	762	Japanese Kobe Collection Data (keyed after decks 118-119)
[a]	1	[a]	1	780	NOAA/NCEI World Ocean Database (WOD) (and formerly WOA) [new: WOD13 (*SID*=149)]
	1			781	Chinese/Global Ocean Data Archeology and Rescue (GODAR) Ships
	3			782	Global Ocean Surface Underway Data (GOSUD)*****
	3		3	792**	US Natl. Cntrs. for Environ. Pred. (NCEP) BUFR GTS: Ship Data [new: post-2007 "Dumped" (*SID*=103) data, same for 793-795]
	3		3	793**	NCEP BUFR GTS: Buoy Data (transmitted in FM 13 "SHIP")

	3		3	794**	NCEP BUFR GTS: Buoy Data (transmitted in FM 18 "BUOY")
	3		(n/a; del.)	795**	NCEP BUFR GTS: Coastal Marine Automated Network (C-MAN)
	(n/a)		(n/a)	796**	NCEP BUFR GTS: Miscellaneous (OSV, plat, and rig) Data
	3			797**	NCEP BUFR GTS: CREX code
	4		4	849**	First GARP Global Experiment (FGGE)
	4		4	850**	German FGGE
	2		2	874**	Shipboard Environ. (Data) Acquisition System (SEAS) [new: SEAS IMMT-5 (*SID*=150)]
	2			875	US TurboWin (e-Logbook) VOS Receipts
	1		1	876-882	US National Data Buoy Center (NDBC) Data
[b+]	1	[b]	6/1	883	US National Data Buoy Center (NDBC) Data [new: data from Jan. 2008-Sept. 2012 (when availability of TD1171/F291 formats ended).] [During R2.5 processing, priority 6 was applied only for 1970-79 data.]
	5		5	888**	US Air Force Global Weather Central (GWC)
	4		4	889**	Autodin (US Dept. of Defense Automated Digital Network)
	3		3	892**	US Natl. Cntrs. for Environ. Pred. (NCEP) Ship Data
	3		3	893**	NCEP Moored Buoy Data
(n/a; del.)	(n/a; del.)		(n/a; del.)	894**	NCEP Drifting Buoy Data
(n/a; del.)	(n/a; del.)		(n/a; del.)	895**	NCEP Coastal-Marine Automated Network (C-MAN) Data
	3		3	896**	NCEP Miscellaneous (OSV, plat, and rig) Data
	2		2	897	*Eltanin*
	2		2	898	Japanese
	2		2	899	South African Whaling
	2		2	900	Australian
	3		3	901	FOSDIC Reconstructions (card images from 16mm film)
	2		2	902	Great Britain Marine (184 extension)
[c]	2	[c]	3/2	926*	International Maritime Meteorological (IMM) Data [new: IMMT receipts since R2.5 from the GCCs]
	1		3/1	927*	International Marine (US- or foreign-keyed ship data) [newly keyed data made available since R2.5]
	3		3	928	Same as 927 including Ocean Station Vessels (OSV)
	3			992**	NCEI GTS: Ship Data
	3			993**	NCEI GTS: Buoy Data (transmitted in FM 13 "SHIP" code) [For discussion (as with 793): imperfect, but I think we will need to let dupelim sort out the moored buoy situation.]
	3			994**	NCEI GTS: Buoy Data (transmitted in FM 18 "BUOY" code)
	3			995**	NCEI GTS: Coastal-Marine Automated Network (C-MAN code) Data

	(n/a)		(n/a)	996**	NCEI GTS: Miscellaneous (OSV, plat, and rig) Data
	(n/a)		(n/a)	997**	NCEI GTS: CREX code
[z]	6	[z]	6	999**	US Air Force Environ. Tactical Appl. Center (ETAC)

* Decks 128 and 927 should be entirely ship logbook data. Deck 926 (the successor to deck 128) should be mostly logbook data, but with other platform types possible depending on the setting of *OP*.

** A deck that has been classified as from the Global Telecommunication System (GTS) for the purpose of setting *DUPS* and *DUPC*. Decks 849-850 are considered GTS although they may have been mixed. Also, while deck 714 is quality controlled etc. by Canada OSD in delayed-mode, it is based exclusively on GTS receipts.

*** Deck 730 (CLIWOC) contains IMMA reports exceeding the maximum record length (2048 characters) supported by the default r*wimma1* program. In the current dupelim program the long supplemental attms are not truncated. Users reading CLIWOC reports via the default *rwimma1* should be aware of the likelihood of truncation.

**** Deck 732 (MORMET) is a mixture of GTS and logbook data but is classified as delayed-mode data for setting *DUPS* and *DUPC*.

***** For the voluminous and highly redundant GOSUD sources, separate priorities were suggested by *SID*, but implementation did not prove practical in the tight R3.0 time schedule. Nevertheless, we feel dupelim rules as they stand are working reasonably well. This additional source was received later and did not figure into the earlier *SID*-based priority proposal:

159: Global Ocean Surface Underway Data (GOSUD v2) from NCEI in WOD format received 19 May 2015

which applied only to the following *SID*s:

(a) 162: GOSUD v3 real-time data from NCEI in WOD format received 30 April 2015

(b): 163: GOSUD v3 near real-time data from NCEI in WOD format received 30 April 2015

(c)-(d): 164: GOSUD delayed-mode French research vessels and sailing ship data from NCEI in WOD format received 30 April 2015

and specifically was (5/3/15 10:12 AM e-mail from Steven Worley): "There is need for duplicate elimination on these. I suggest the selection priority is b) first then c), d), and finally a)."

_____

More detailed information follows about the special deck rules and priority codes listed in Table 3. Priority codes are used in the event of a match of two reports with equal quality codes, in which case the deck with the lowest priority code is considered preferable and is selected in that match. (If two reports with equal priority codes match, the report with the highest deck is selected; if two reports with equal decks match, the report with the highest *SID* is selected; if two reports with equal *SID*s match, the second report in sort order is selected.) The special rules may override other duplicate selection criteria, including the quality and priority codes).

The special deck rules listed in Tables 3 are described below in order of precedence i.e. the [a] rules take precedence over the [b] rules in the event of an [a]/[b] match (see *dupelim_1970* for additional background on the rules):

[a] Absolute pass through: These data should be duplicate-free, and not available from any other source. Matches within this deck are ignored (all data are passed through). In the event of a match of a report from this deck with a report from any other deck (including a different [a] deck), the match is ignored (a small amount of matching may be expected, e.g., for ships servicing buoys).

[b+] Limited pass through plus: Matches within this deck are ignored (all data are passed through). When a report from this deck matches a report from a different deck (including any [b] decks), this deck (i.e. applicable only to deck 146 or 883) is automatically selected. We don't expect any matches between decks 146 and 883, but in that event both are preserved.

[b] Limited pass through: Matches within this deck are ignored (all data are passed through). When a report from this deck matches a report from a different deck, dup selection is resolved according to the default rules (i.e., quality code, priority code, deck, *SID*, and sort order). [Note: This is a simplified and modified version of the [b] rule as described in *dupelim _1970*.]

[c] Special rules for IMM data: For matches between a report from deck 926 and any other deck, the default rules hold. For matches within deck 926, dup selection is determined solely by *SID*, such that the report with the largest *SID* is selected (or the default rules apply within a *SID*). If, however, a report that has been selected solely because of its higher *SID* also has a quality code that is higher (i.e., inferior quality) by 6 or more quality points, then the report with the lower *SID* is selected instead.

[k] Automatic data rejection (note: not applicable for R3.0). Some data known to be available from other sources, and some incorrectly located reports, were deliberately introduced into the datastream. In the event of a match of a [k] report with any other report (including another [k] report), the match is ignored. Following such testing for all possible duplicates with [k] reports, all [k] reports are automatically deleted from the dupelim output. [Note: not applicable for R3.0.]

[z] Non-selection except for unique reports: Matches within this deck or another [z] deck are resolved by selection according to the default rules. Matches of a report from this deck with a report from any other deck automatically result in the non-[z] report being considered the best duplicate; thus under these circumstances [z] reports are deleted unless they are unique (or uncertain duplicates).

C. Postconditioning

1. Set the *Uida* attm fields (*UID*, *RN1*, *RN2*, *RN3*, *RSA*, and *IRF*).
2. Compute missing *RH*, applicable only to R2.5 sources.
3. Left-justify deck 730 Climatological Database for the World's Oceans (CLIWOC) *ID*s.
4. Output two Datasets:
   - Initial User
   - Total (i.e. Initial User + Rejects)

D. Finalization

So as to allow more detailed analysis of the dupelim input/output results, both by the ICOADS program and any interested users, and possible output adjustments, no duplicate reports were actually eliminated during dupelim (step B) or during postconditioning (step C). However, the following reports are eliminated during this final step:
   - Certain and uncertain dups: *DUPS* > 2 (see Table 4)
   - Landlocked: *LZ* = 1
   - Intermediate and reject reports (*IRF*=0 and *IRF*=2)

Similar to the situation with preconditioning as discussed in sec. 1, in the case of the Intermediate data from R2.5 being used as the largest input for R3.0, rules and priorities

had already been applied during R2.5 dupelim (listed for reference in Table 3), which in turn were cumulative rules applied during past ICOADS updates extending back to *Release 1*. At R2.5, however, we began preserving the duplicates (and other suspect reports including landlocked) in the intermediate dataset, and for R3.0 the *IRF* flag was defined for the latest IMMA1 format version with the additional value *IRF*=2 to also indicate that some data should be rejected from any intermediate and/or final datasets (see Fig. 2).

Table 4 (note: Table D8 from *R3.0-imma1*). Duplicate status (*DUPS*) assignments. In previous Releases, "certain" (C) duplicates were eliminated from the LMR output, and then "uncertain" (U) duplicates were eliminated from LMRF. Prior to R3.0 processing, settings marked by footnotes apply only to pre-1980 data. For R3.0 processing, *DUPS*=3, 5, 6, and 7 were no longer used (indicated by grey shading).

DUPS	U/C	Description
0		unique
1		best duplicate
2		best duplicate with substitution
3	U	worse duplicate: uncertain weather element match with hour cross[1]
4	U	worse duplicate: uncertain weather element match with no cross
5	U	worse duplicate: uncertain weather element match with day cross[2]
6	U	worse duplicate: time/space match with *ID* mismatch (unused until 1950)
7	U	worse duplicate: certain weather element match with hour cross[1]
8	C	worse duplicate: certain weather element match with no cross
9	C	worse duplicate: combined *DUPS* 4 and 6
10	C	worse duplicate: combined *DUPS* 6 and 8
11	C	worse duplicate: time/space/ID match
12	C	worse duplicate: combined *DUPS* 4 and 11
13	C	worse duplicate: combined *DUPS* 8 and 11
14	C	automatic data rejection

1. For *Release 1*, applied to 1854-1979 data; for R2.0, applied to 1784-1979 data; for R2.5, applied to 1662-1979 data.

2. For *Release 1*, applied to 1854-1969 data; for R2.0, applied to 1784-1969 data; for R2.5, applied to 1662-1969 data.

_____

E. Monthly Summaries

Summaries are calculated for each year and month over the period 1800-recent using 2° latitude × 2° longitude boxes. Owing to sparse coverage statistics are not calculated prior to 1800. In contrast, starting in 1960 when spatial coverage improves, they are also calculated using 1°×1° boxes. Twenty-two observed and derived variables are summarized for each year-month; for each variable, 10 statistics are calculated (for more information, see: http://icoads.noaa.gov/e-doc/R2.5-stat_doc.pdf).

Additionally, the 1°×1° and 2°×2° summaries are produced in two different forms:
- Enhanced statistics were derived using 4.5 standard deviation (σ) trimming limits, so as to accommodate more extreme climate events, and using a broad collection of marine observations including ships, buoys, and near-surface oceanographic profile temperatures.
- Standard statistics were derived using 3.5σ trimming limits and were limited as nearly as possible to observations taken from ships (where identifiable).

For R3.0, we are also considering production (at least for Beta testing) of untrimmed monthly summaries.

References

Atkinson, C.P., N.A. Rayner, J. Roberts-Jones, and R.O. Smith, 2013: Assessing the quality of sea surface temperature observations from drifting buoys and ships on a platform-by-platform basis, *J. Geophys. Res. Oceans*, **118**, 3507–3529, doi:10.1002/jgrc.20257.

Kennedy, J.J., N.A. Rayner, R.O. Smith, D.E. Parker, and M. Saunby, 2011: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization, *J. Geophys. Res.*, **116**, D14104 (doi:10.1029/2010JD015220).

Worley, S.J., 1992: Status of other new data sets for COADS. *In:* Diaz, H.F., Wolter, K., and Woodruff, S.D., (eds.) *Proceedings of the International COADS Workshop, Boulder, Colorado, 13-15 January 1992*, NOAA Environmental Research Laboratories, Boulder, Colorado [paper available at: http://icoads.noaa.gov/Boulder/Boulder.Worley.pdf].

Appendix A.

Table A1. Dupelim priorities applied during R2.5 processing by time period (indicated in table by starting year): 1662-1949, 1950-69, 1970-79, 1980-2004, and 2005-07. [Note: some rules may also have varied by time period, to be checked later (ref. 2 Dec. e-mail spreadsheet).]

DCK	dupelim: 1662	1950	1970	1980	2005
110	2	2	2		
116	2	2	2		
117		3	3		
118	2	2	2		
119		2	2		
128		2	2		
143			2		
144				1	
145			1	1	
150	4	4	4		
151	4	4	4		
152	4	4	4		
155	4	4	4		
156	4	4	4		
184		2	2		
185		2	2		
186		3	3		
187	2	2	2		
188	2	2	2		
189	2	2	2		
192	2	2	2		

193	2	2	2		
194	2	2	2		
195	2	2	2		
196	2	2	2		
197	2	2	2		
201	1	1	1		
202	1	1	1		
203	1	1	1		
204	1	1	1		
205	1	1	1		
206	1	1	1		
207	1	1	1		
209		1	1		
210		1	1		
211		1	1		
213		1	1		
214		1	1		
215	6	6	6		
216	1	1	1		
218		1	1		
221		1	4	3	
223		1	4	3	
224			4	3	
226	1	1	1		
227	1	1	1		
229		1	4	3	
230		1	4		
233				3	
234				3	
239		1	1	3	
245	2	2	2		
246	1	1	1		
247	1				
254	3	3	5	3	
255	1	1	4	3	
281	5	5	5		
555		5	5		
666			2		
667			1	1	
700				2	2
701	1	1	1		

702	1	1	1		
704	1	1	1		
705	1	1	1		
706	1	1	1		
707	1	1	1		
714			2	2	2
715				1	
720	3	3	3		
730	1	1	1		
731	1	1	1		
732	6	6	6	5	
733	2	2	2	2	
734	1	1	1		
735	1	1	1	1	
736	1	1	1		
740				2	
749			5		
761	1	1	1	1	
762	1	1	1		
780	1	1	1	1	1
792				3	3
793				3	3
794				3	3
795					3*
849			4		
850			4		
874				2	2
876			1		
877			1		
878			1		
879			1		
880			1		
881			1		
882			1		
883			6	1	
888			5	5	
889			4	4	
891**		1	1		
892				3	
893				3	
896				3	

897		2	2		
898		2	2		
899	2	2	2		
900	2	2	2		
901	3	3	3		
902		2	2		
926		3	3	2	2
927			3	1	1
928			3		
999		6	6		

* Difference with Table 3.
** Deck not in Table 3 (deleted from R2.5 via preconditioning?).

_____

<div>

Document Revision Information

Previous document version: none.

</div>