WORLD METEOROLOGICAL ORGANIZATION

INTERGOVERNMENTAL OCEANOGRAPHIC COMMISSION (OF UNESCO)

JOINT WMO/IOC TECHNICAL COMMISSION FOR OCEANOGRAPHY AND MARINE METEOROLOGY (JCOMM)
EXPERT TEAM ON MARINE CLIMATOLOGY

FIRST SESSION

GDYNIA, POLAND, 7 TO 10 JULY 2004

ETMC-I/Doc. 3.2
(30.VI.2004)

ITEM 3.2

Original: ENGLISH

REVIEW OF BUFR TEMPLATE FOR SHIP DATA

(Submitted by Mr Scott Woodruff)

Summary and Purpose of Document

This document presents a limited review offering general comments on the BUFR template (WMO, 2003) for ship data, and on the storage of marine data in BUFR, based on assessments of marine BUFR data from NOAA's National Centers for Environmental Prediction (NCEP).

ACTION PROPOSED

The Expert Team on Marine Climatology is invited to:

(a) Review the information provided;

(b) Make a recommendation to relevant WMO bodies including the Commission for Basic Systems (CBS) as appropriate.

Appendix: Comparisons of NCEP Marine GTS Data for ICOADS

DISCUSSION

Introduction

This is a limited review offering general comments on the BUFR template (WMO, 2003) for ship data, and on the storage of marine data in BUFR. The comments are based on assessments of marine BUFR data from NOAA's National Centers for Environmental Prediction (NCEP). An early set of comparisons of the BUFR data with overlapping data, which were recorded in the format that was previously used by NCEP, revealed some initial BUFR data continuity problems (Table 1).

Table 1. Examples of initial data continuity problems in NCEP's version of BUFR marine GTS data, based on comparisons for March 1997 data.*

Temperature biases (0.1°C)	Usage of the standard factor 273.15 for conversion of Celsius temperatures, and rounding to tenths Kelvin precision (which until approximately 17 Feb. 1999 was the maximum precision available), lead to some unrecoverable temperature errors of 0.1°C.
Wind speed indic. (measured/est.)	Indicator omitted until approximately 21 October 1997.
Wind codes	Incomplete conventions to store originally reported FM 13 code combinations for calm and variable winds.
Cloud amounts	Oktas converted to percent, such that BUFR did not preserve the distinction between code figures 9 (sky obscured by fog, snow, or other meteorological phenomena), "/" (cloud cover indiscernible for reasons other than code figure 9, or observation is not made), and a missing code figure.

* Starting in March 1997, data are available processed by NCEP into BUFR. In addition, overlapping data were processed into NCEP's previous ON124 format until 19 April 1997. Limited comparisons were made between the overlapping BUFR and ON124 data, and also against BUFR data encoded by the US Navy (more information about the comparisons is available here: http://www.cdc.noaa.gov/coads/real-time.html). Some of the data continuity problems were later alleviated, as noted.

A more recent set of comparisons (Annex) was used to select and prepare 1998-2002 "real-time" data for the International Comprehensive Ocean-Atmosphere Data Set (ICOADS) (Worley et al., 2005). NCEP's BUFR archive included the raw FM 13 SHIP (or similarly FM 18 BUOY message) attached to the BUFR message. The more recent comparisons translated the BUFR portion into IMMA, and independently decoded the FM 13 (or FM 18) portion into IMMA. The two files should in theory be identical. However, mechanical verification (character-by-character agreement) revealed some continuing problems.

The following are the general findings from the Annex, and limitations of the comparisons, which have a bearing on the BUFR ship (and buoy) template, and on the storage and archival of marine data in BUFR:

- Among the (limited) set of elements compared, we found that data problems or omissions still exist in NCEP's version of BUFR. The source and significance of each field difference has not been systematically determined.
- Generally, however, these problems appear to be confined to secondary elements such as clouds and waves, or special codes mixed with data that otherwise are translated properly to BUFR (e.g., the code 99 for variable wind direction in FM 13, which apparently does not yet have a place in the BUFR template).
- Improvements have been made in the NCEP data since the comparisons summarized by Table 1, but even some of the problems listed in Table 1 still appear to exist.
- In addition to improperly or incompletely converted data, fields may be completely absent in BUFR that are present in the original GTS data. The Annex comparisons were limited in scope, concentrating on the most widely used and often reported elements. Not all FM 13 and FM 18 elements were checked for their presence and accuracy as represented in BUFR.
- Retention by NCEP of the original GTS message strings is critically important. This is recommended for consideration for inclusion in the BUFR template, because it permits

effective, automated verifications of the translations made into BUFR by National Meteorological Centers (NMCs) from the original data.

- For example, differences were known to exist between the BUFR data produced by NCEP, the UK Met Office, and the US Navy (FNMOC). Among these centres it appeared that only NCEP attached (or otherwise archived) the original message strings.
- Even if the BUFR template was able to prescribe a complete set of field-by-field mappings of FM 13 and FM 18 into BUFR (which it does not appear to have achieved yet), differences likely would arise between versions of BUFR produced at different NMCs due to software errors and biases, or different translation approaches, in the event that the handling of a given field is not 100% specified (as for instance in the case of variable wind direction).

References

WMO, 1995: *Manual on Codes*. WMO–No.306, Geneva, Switzerland (Vol. I.1: 1995 Ed. including Suppl. through No. 3 (VIII.2001); Vol. I.2: 2001 Ed.).

WMO, 2003: BUFR/CREX Templates to Transmit in Table Driven Code Forms: Automatic Weather Station Data, Traditional WMO Observations, Tropical Cyclone Tracks Data from EPS, AMDAR Data (March 2003 paper containing the outcome of the work of the ET on DR&C for BUFR/CREX templates).

Worley, S.J, S.D. Woodruff, R.W. Reynolds, S.J. Lubker, and N. Lott, 2005: ICOADS Release 2.1 data and products. (Draft paper for submission to the special issue for CLIMAR-II of the Int. J. Climatol.)

Appendix: 1

Comparisons of NCEP Marine GTS Data for ICOADS:
Conversions of BUFR and of FM 13/FM 18 Messages into IMMA Format

1. Introduction

This Annex describes comparisons of marine GTS data from NOAA's National Centers for Environmental Prediction (NCEP). These comparisons were used to select and prepare data for the "real-time" (RT) archive of the International Comprehensive Ocean-Atmosphere Data Set (ICOADS). GTS data for March 1997 through 2002 were processed into the ICOADS.RT archive (further information is available at: http://www.cdc.noaa.gov/coads/rt.html).

The data were translated by NCEP from the original GTS message strings (i.e., FM 13 SHIP or FM 18 BUOY code), into the Binary Universal Form for the Representation of Meteorological Data (BUFR) (WMO, 1995). NCEP included in their versions of the BUFR format one (or more) of the input original GTS message string(s) used to construct each BUFR report, thus the NCEP format is referred to as "BUFR+string" in this document.

Including the strings was extremely beneficial, because it permitted automated comparisons using the originally reported data, in order to help determine the completeness and accuracy of BUFR in representing the original data. Including the strings also allowed the resolution of some significant data omissions or problems, particularly in the earlier BUFR data, through access to the originally reported data.

NCDC is also translating and archiving GTS data obtained from NOAAPort and other GTS sources. These NCDC data are planned for utilization in the ICOADS.RT archive at a future date. However, for the March 1997-2002 period, only the NCEP data have been used.

2. Comparison method

Since the translated BUFR data and the original string(s) were both available in NCEP's BUFR+string format, this allowed two different translations to be performed, and cross-checked, for March 1997-2002, using as input:

(i) translated BUFR data
(ii) original GTS string

A third translation (iii), which processes original GTS strings similarly to translation (ii), is in operation at NCDC to process its GTS data. All three translations were made into the new International Maritime Meteorological Archive (IMMA) format, with the original GTS message attached as supplementary data (similarly to NCEP's approach in BUFR).

Not all original GTS fields were translated into IMMA fields—Figure 1 illustrates the scope of the translations from FM 13 and FM 18 into IMMA. This also then outlines the extent of the comparisons made against the BUFR data. Additional fields might exist in the original GTS data that are not yet being translated by NCEP into BUFR, or are not yet part of the BUFR template (WMO, 2003).

The availability of the three translations provided opportunities for extensive comparisons:

- Checks between translations (i) and (ii) revealed some problems in earlier versions of the BUFR format at NCEP. Therefore, data of better quality and completeness were obtained in some cases from translation (ii).

- Checks between (ii) and (iii) continue to be helpful to improve the quality of both of those (independently written) translations. The original GTS formats are very complex and can contain errors (e.g., introduced during transmission), therefore translations (ii) and (iii) still produce different results under some circumstances.

- Availability of overlapping delayed-mode (ICOADS.DM) data (incorporating GTS plus additional sources, and more fully processed) for March-December 1997 provided opportunities for additional checks.

3. Data and processing for ICOADS.RT

For the final ICOADS.RT archive, translation (ii) was used for the period March 1997 through September 1999, and translation (i) for the period October 1999 through December 2002. October 1999 also marked a transition in the NCEP archive to "dumped""data, in which some partial duplicate reports had been merged and quality control (QC) applied by NCEP.

To further improve the quality of October 1999-December 2002 data, corrections of two known BUFR problems, which could be made to the BUFR data directly without accessing the original GTS string, also were applied as part of translation (i). These and other details of the ICOADS.RT processing are covered in Figure 2 (and accompanying processing notes). Some residual problems are known to exist in the BUFR data through 2002, which were propagated into ICOADS.RT. Generally, however, these problems appear to be confined to secondary elements such as clouds and waves (see sec. 4).

In addition to the new IMMA format, the widely used ICOADS "LMRF" observational format was produced, as well as full sets of 2° and 1° monthly summaries in MSG format (see Figure 2).

Data for March-December 1997 that overlap ICOADS.DM were processed primarily for validation purposes (using the original message strings). However, those overlapping data also provide a source for some data elements that are currently missing or incorrect in ICOADS.DM, related to problems in the earlier BUFR data.

The dumped data used for ICOADS.RT for October 1999-December 2002 contained "quality marks" that were used at NCEP to correct the BUFR data. However, the quality marks themselves were not retained in the translation into ICOADS.RT. In addition, GTS "bulletin header" information was stored in BUFR, which was not retained in the translation into ICOADS.RT. Some of this information can be important for specialized applications such as tracing the GTS routing of data. In the future we may wish to reassess whether more metadata from BUFR such as NCEP quality marks and bulletin header information should be preserved in ICOADS.RT.

4. Comparison results

As discussed above, translation (ii) of the original GTS messages was used for the period March 1997-September 1999 to create the ICOADS.RT archive, and translation (i) of dumped BUFR data was used for the period October 1999-December 2002. Figure 3 provides field-by-field comparison results between the two translations, separately for the two periods (upper and lower panels), and for all platform types combined (FM 13 and FM 18 data). This clearly shows some improvements in BUFR for the later period compared to the earlier period, which helped guide our decision on which translation to use for ICOADS.RT.

For example, Figure 3 (upper) shows that many wind speed indicator (WI) values are missing from BUFR in the earlier period. This indicator (from FM 13) was not added by NCEP until late in 1997. In contrast, Figure 3 (lower) shows fairly good (but not complete) agreement for WI between the two translations in the later period (these residual differences arise largely from FM 18 data, for which it appears NCEP did not add WI to BUFR until sometime in 2000).

Similarly, there are significant amounts of disagreement between wet bulb, dew point, air, and sea surface temperatures (WBT, DPT, AT, and SST) between the two translations in the earlier period, whereas agreement is improved in the later period. Storage of temperature values in BUFR was extended to hundredths Kelvin at NCEP starting during February 1999. Usage at NCEP of the factor 273.15 for conversion of Celsius temperatures, and rounding to tenths Kelvin precision (until that date

the maximum precision available in BUFR), previously lead to some temperature errors of 0.1°C.

Similar field-by-field (early versus late period) comparisons are provided for data reported in FM 13 only, from ships (Figure 4) and from buoys (Figure 5). Figure 6 provides comparisons for all the FM 18 BUOY data (drifting buoys probably compose the largest amount of these data, plus some moorings). Figure 7 provides comparisons for data from the Coastal-Marine Automated Network (C-MAN), operated by NOAA's National Data Buoy Center (NDBC), which are reported in a variant of FM 13.

These figures indicate that some residual differences exist in the later BUFR data, among the limited set of major elements that has been compared (ref., Figure 1). Generally, however, the surveyed problems appear to be confined to secondary elements such as clouds and waves, plus some residual differences in major weather elements.

A systematic assessment has not been made of the sources and significance of individual field differences, which can sometimes arise from relatively small numerical problems or with the representation of special code combinations within otherwise continuous numeric data. For example, a code for variable wind direction (FM 13: dd=99) may not yet exist in the BUFR template. Since approximately 21 October 1997, NCEP has set wind direction to zero (which signifies calm—not 0°—in FM 13), when GTS code 99 (variable direction) was reported with wind speed greater than zero. The procedure used at NCEP to handle variable wind direction, prior to that date in 1997, is not known.

It should be noted that field differences, between IMMA as derived from BUFR versus the original GTS data, in some cases may have arisen from dup-merge processing by NCEP of multiple messages and associated QC corrections (applicable only to the later period October 1999-December 2002). The dup-merging and QC changes were not applied to the original GTS data, and thus did not become a part of that IMMA translation.

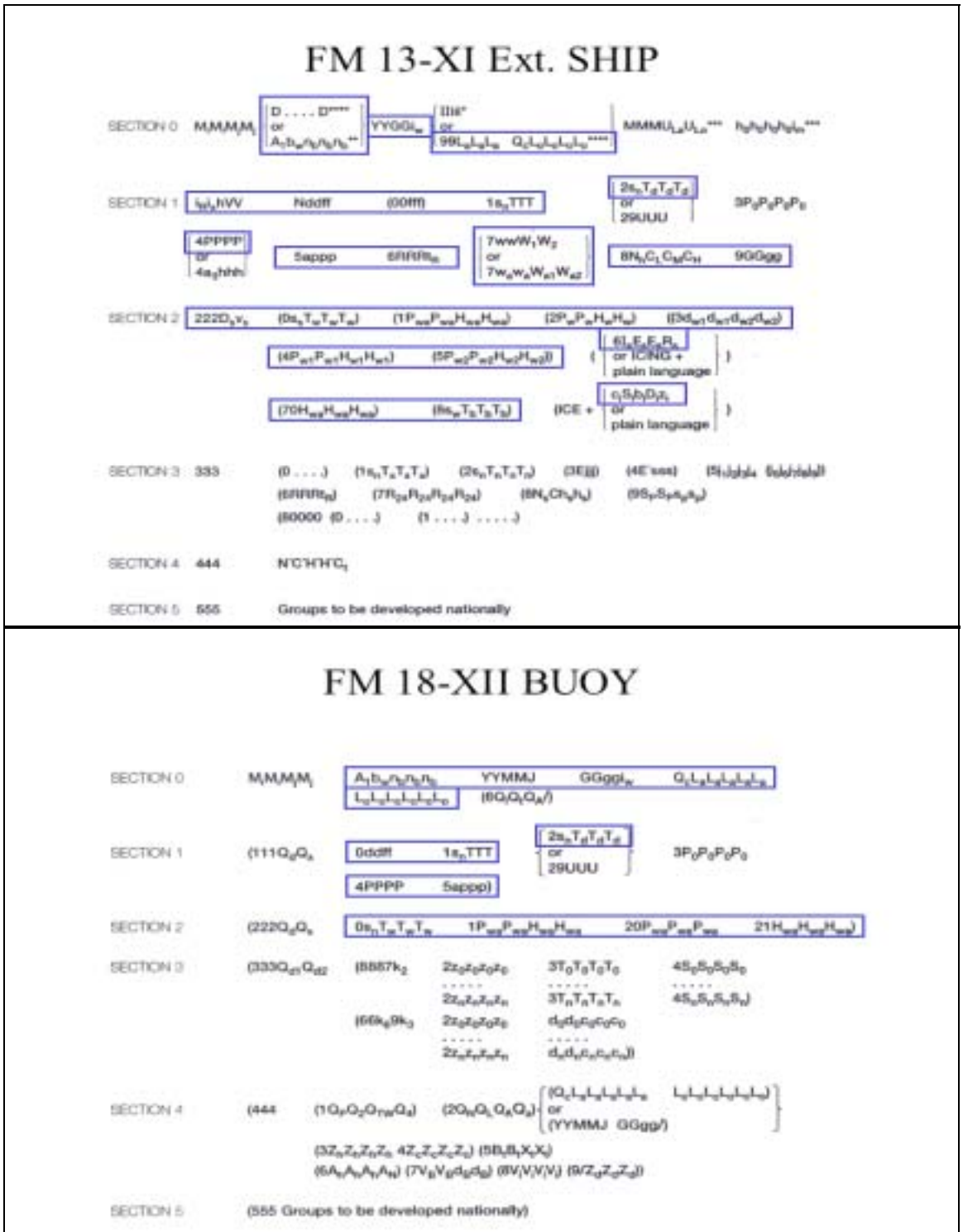


Figure 1. Blue boxes enclose portions of the GTS FM 13 (upper) and FM 18 (lower) code forms (from WMO, 1995) that were translated for ICOADS.RT into the IMMA format (FM 13 groups marked * and *** are not used for SHIP data).

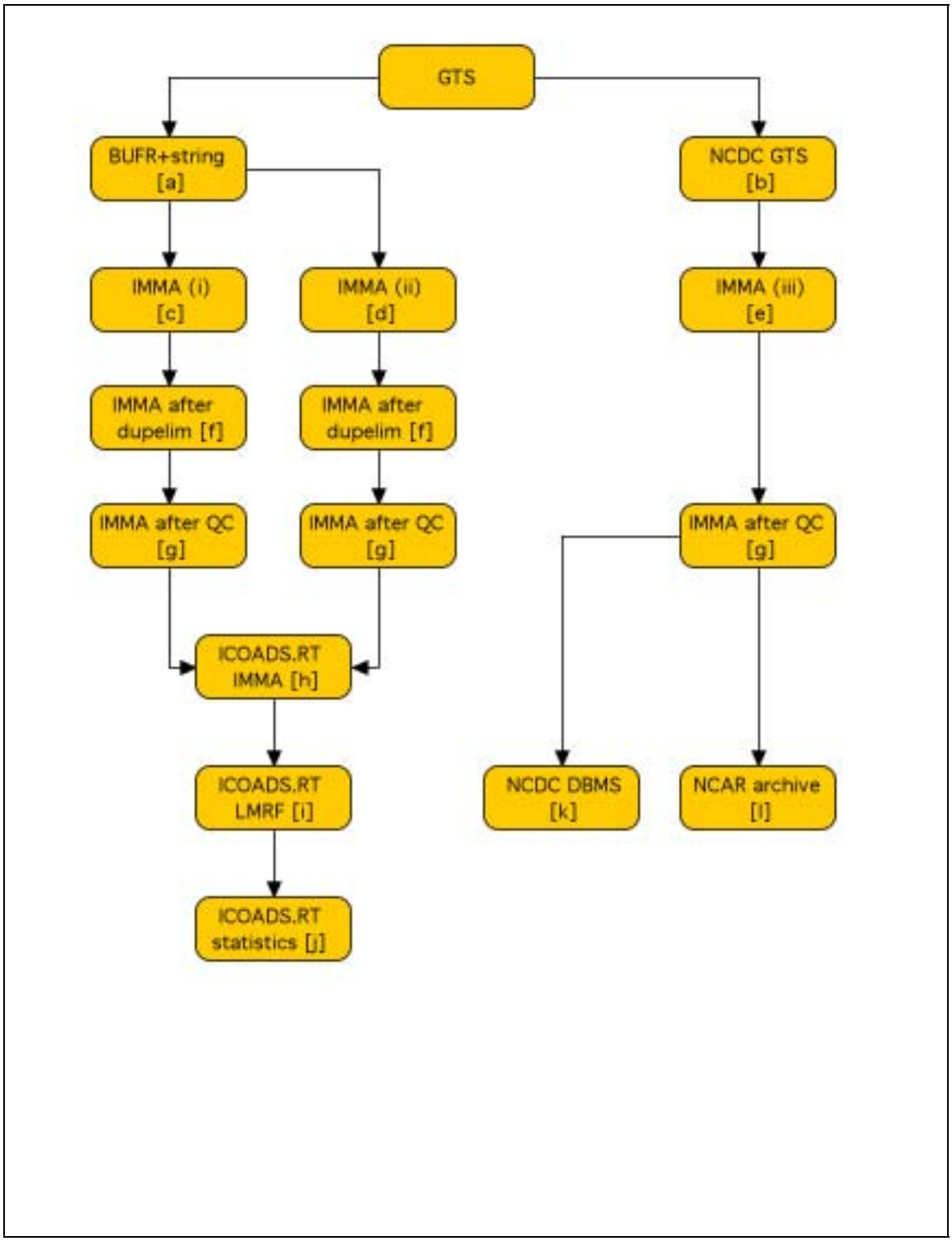


Figure 2. Overview of the ICOADS.RT processing flow (simplified), and of related NCDC processing (data not yet used for ICOADS.RT). Letters in [brackets] refer to processing notes (overleaf).

Processing notes accompanying Figure 2
(see corresponding [letters] on Figure)

[a] The input NCEP BUFR+string data. Regular "operational" data were used for March 1997-September 1999, and "dumped" data for October 1999-December 2002. Dup-merge processing was applied by NCEP to the dumped data. Also, "quality marks" from interactive QC were used at NCEP to correct the dumped BUFR data, and the quality marks were supplied with the BUFR data. However, we did not retain the quality marks (and some GTS "bulletin header" information available in BUFR) in ICOADS.RT.

[b] The input NCDC GTS data. Presently, the main source for these data is NOAAPort.

[c] Translation (i) of the BUFR part of BUFR+string into IMMA. Until 15Z 23 July 2002, NCEP was inadvertently storing in BUFR the original GTS code values for the wet bulb temperature indicator (WBTI) and sea surface temperature method indicator (SI), rather than the defined BUFR values for these fields. As part of translation (i), these BUFR fields were interpreted as GTS codes through 23 July 2002, and thereafter as BUFR codes (thus a small amount of data after 15Z on 23 July was erroneously translated). Tables A1 and A2 define the mappings between the GTS and BUFR codes (WMO, 1995), and the resultant fields WBTI and SI in IMMA (see Table A3).

Table A1. WMO Code 3855 is the "Indicator for the sign and type of wet-bulb temperature reported" in FM 13. BUFR Code 0 02 039 is the "Method of wet-bulb temperature measurement." The mappings between these codes (by NCEP starting 15Z 23 July 2002), and from the BUFR code into IMMA WBTI (if applicable), are also listed.

Code 3855	Map	BUFR Code 0 02 039	WBTI
0 positive or zero measured	0 ⇒ 0	0 measured	0
1 negative measured	1 ⇒ 0	1 iced bulb measured	2
2 iced bulb measured	2 ⇒ 1	2 computed	1
5 positive or zero computed	5 ⇒ 2	3 iced bulb computed	3
6 negative computed	6 ⇒ 2	4-6 reserved	N/A
7 iced bulb computed	7 ⇒ 3	7 missing value	missing

Table A2. WMO Code 3850 is the "Indicator for sign and type of measurement of sea-surface temperature" reported in FM 13. BUFR Code 0 02 038 is the "Method of sea-surface temperature measurement." The mappings between these codes (by NCEP starting 15Z 23 July 2002), and from the BUFR code into IMMA SI (if applicable), are also listed.

Code 3850		Map	BUFR Code 0 02 038	SI
0 positive or 0	intake	0 ⇒ 0	0 ship intake	1
1 negative	intake	1 ⇒ 0	1 bucket	0
2 positive or 0	bucket	2 ⇒ 1	2 hull contact sensor	3
3 negative	bucket	3 ⇒ 1	3 reversing thermometer	N/A
4 positive or 0	hull contact sensor	4 ⇒ 2	4 STD/CDT sensor	N/A
5 negative	hull contact sensor	5 ⇒ 2	5 mechanical BT	N/A
6 positive or 0	other [or missing indic.]	6 ⇒ 15[1]	6 expendable BT	N/A
7 negative	other [or missing indic.]	7 ⇒ 15[1]	7 digital BT	N/A
			8 thermistor chain	N/A
			9 infared scanner	N/A
			10 microwave scanner	N/A
			11-14 reserved	N/A
			15 missing value	missing

1. Note that unless SST is missing it is not possible to report a "missing" indicator in the original GTS code (because the indicator must be extant to provide the sign of SST), and there is no value for "other" in BUFR. Therefore, "other" in GTS was mapped to "missing" in BUFR and SI. There is a value for "others" in SI, but this was derived from WMO's delayed-mode IMMT format, in which there are separate values for "others" and "missing" (plus instrument types in addition to the three listed in this table).

[d] Translation (ii) of the string part of BUFR+string into IMMA. Starting sometime in March 2002 multiple reports may exist in the string portion of the dumped BUFR data. In this case, the ICOADS.RT translation (ii) of the original data used only the first (usually most recent) GTS message, since the messages were generally attached in reverse order of receipt time. A complicating factor also exists for some buoys in that the data were reported as two separate (subsurface and surface) messages. In this case, ICOADS.RT processing still utilized only the first message.

[e] Translation (iii) by NCDC of the NCDC GTS data. The NCDC data can be compared with the NCEP data, to check for possible gaps or variations in receipts at the different GTS centers. In addition, the attachment of the original GTS string to IMMA allows the re-translation of the same messages (e.g., using translation software (ii) versus (iii)) and comparison of those results to identify differences in the translation software.

[f] As part of ICOADS.RT processing, a simple form of duplicate elimination (dupelim) was applied to all the data. This “exact” dupelim rejected all but one report among any set of reports with the same year, month, day, hour, latitude, longitude, and ID. As expected, the exact dupelim had much less impact on the dumped data than the regular operational data.

[g] Common software (qctrf) was used on all the data streams to add the regular ICOADS quality control and “trimming” flags into the IMMA format.

[h] The final ICOADS.RT archive was assembled using the IMMA output from translation (ii) for March 1997-September 1999, and from translation (i) for October 1999-December 2002.

[i] The older “LMRF” observational format is widely used, and interfaces with the current ICOADS statistics program. We plan to continue to offer it as a product until IMMA becomes widely used. (The production “LMR” format also played a role, not shown in Figure 2, in this processing.)

[j] For 1998-2002 we generated the full suite of ICOADS statistics: standard and enhanced, 2° and 1°, global and equatorial.

[k] NCDC-translated GTS data will be offered from NCDC in the future, but are not yet a regular part of ICOADS.RT. NCDC plans to ingest the data into a database management system (DBMS), for user access, but to archive the IMMA+string data.

[l] The NCDC GTS data are also flowing monthly to NCAR as a backup archive.

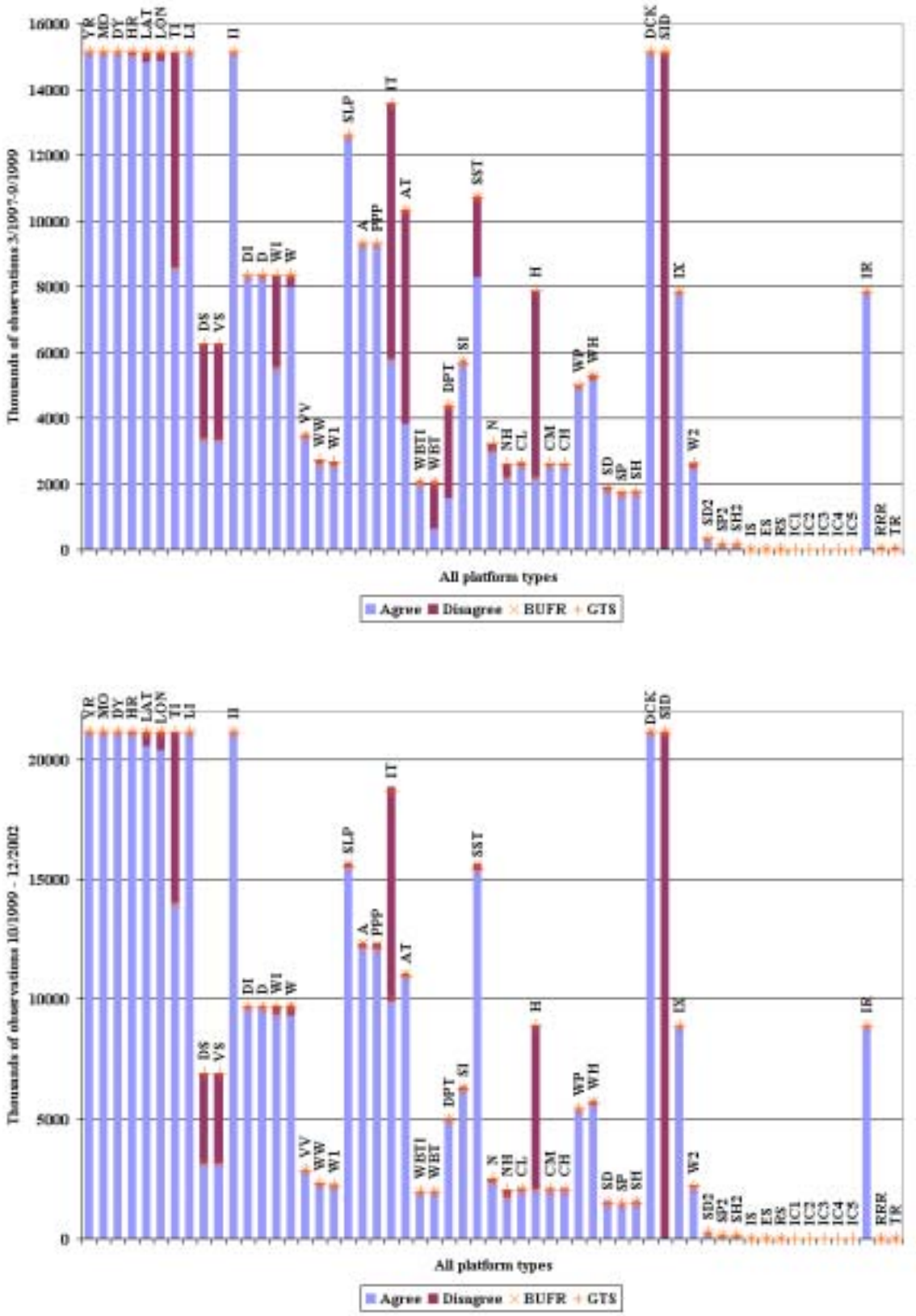


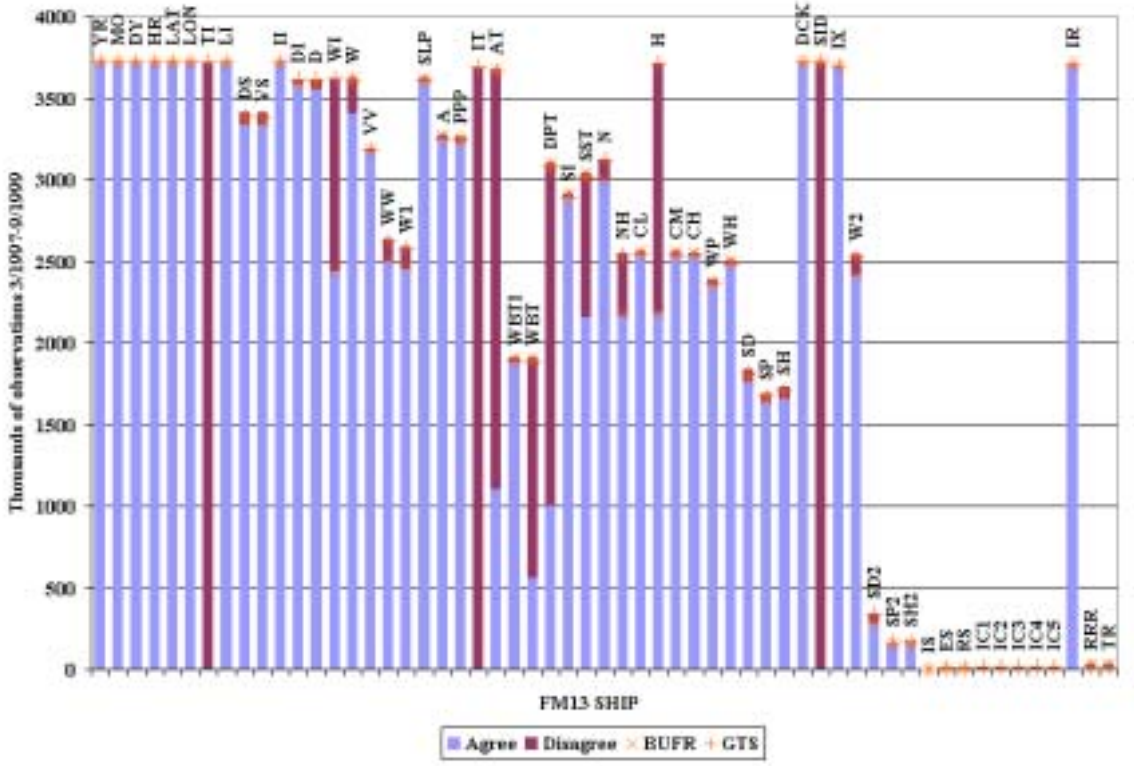
Figure 3 (see caption overleaf).

Figure 3 caption. (Upper) Total number of observations for March 1997-September 1999, shown separately for each IMMA field (Table A3) that had any data (all platform types). Blue (red) bars indicate that the translation of the BUFR data agrees (disagrees) with the translation of the original GTS message, into IMMA format. Disagreement includes the case of missing versus extant data, as well as small numerical differences. The amount of extant data resulting from the BUFR (GTS) translation is indicated by an orange "X" ("+"). If a field is completely missing the "X" or "+" is not shown. (Lower) As for upper, except for October 1999-December 2002.

NOTE: Source ID (SID) and a few other fields including the time and temperatures indicators (TI and IT, which were unknown for some of these code types in the BUFR translation) were expected to have total or large amounts of disagreement.

Table A3. Abbreviations and descriptions for the IMMA fields used in Figures 3-7.

Abbr	Element description	Abbr	Element description
YR	year UTC	*N*	total cloud amount
MO	month UTC	*NH*	lower cloud amount
DY	day UTC	*CL*	low cloud type
HR	hour UTC	*H*	cloud height
LAT	latitude	*CM*	middle cloud type
LON	longitude	*CH*	high cloud type
TI	time indicator	*WP*	wave period
LI	latitude/long. indic.	*WH*	wave height
DS	ship course	*SD*	swell direction
VS	ship speed	*SP*	swell period
II	*ID* indicator	*SH*	swell height
DI	wind direction indic.	*DCK*	deck
D	wind direction (true)	*SID*	source ID
WI	wind speed indicator	*IX*	station/weather indic.
W	wind speed	*W2*	2nd past weather
VV	visibility	*SD2*	dir. of second. swell
WW	present weather	*SP2*	per. of second. swell
W1	past weather	*SH2*	ht. of second. swell
SLP	sea level pressure	*IS*	ice accretion on ship
A	characteristic of *PPP*	*ES*	thickness of I_s
PPP	amt. pressure tend.	*RS*	rate of I_s
IT	indic. for temperatures	*IC1*	concentration of sea ice
AT	air temperature	*IC2*	stage of development
WBTI	indic. for WBT	*IC3*	ice of land origin
WBT	wet-bulb temperature	*IC4*	true bearing ice edge
DPT	dew-point temp.	*IC5*	ice situation/trend
SI	*SST* meas. method	*IR*	indic. for precip. data
SST	sea surface temp.	*RRR*	amount of precip.
		TR	duration of per. *RRR*



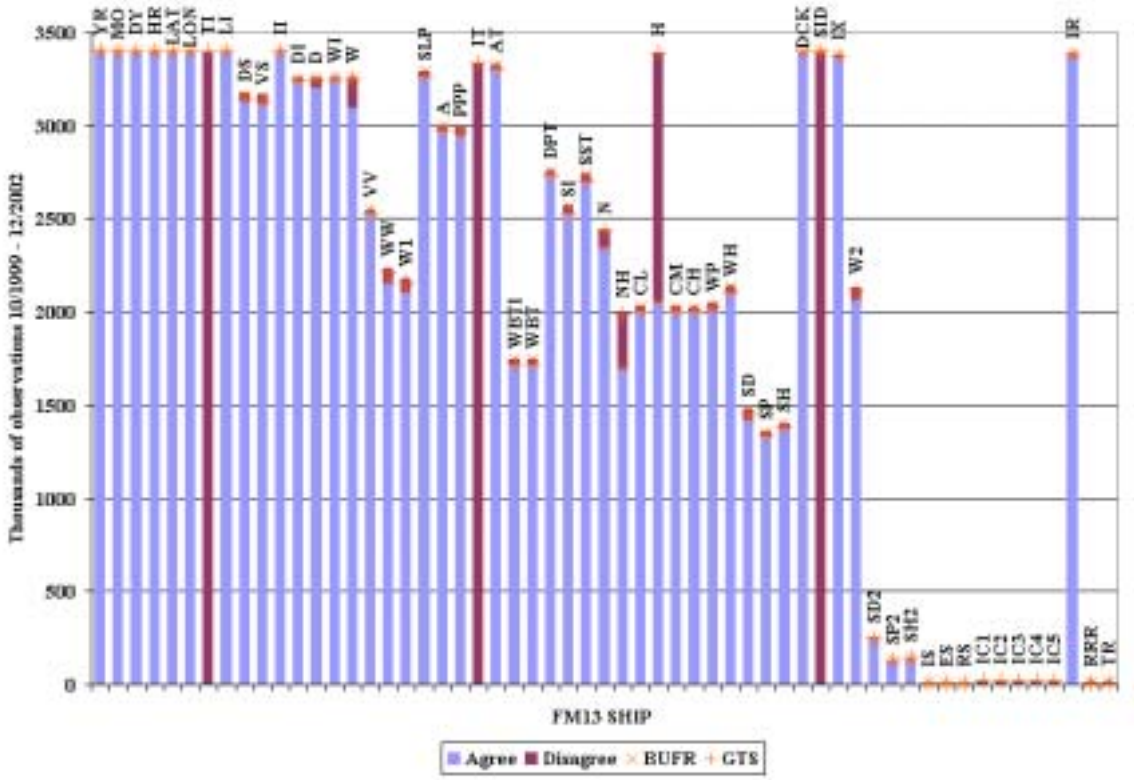


Figure 4. As for Figure 3, except the observations are plotted separately for ships reporting in the FM 13 code.

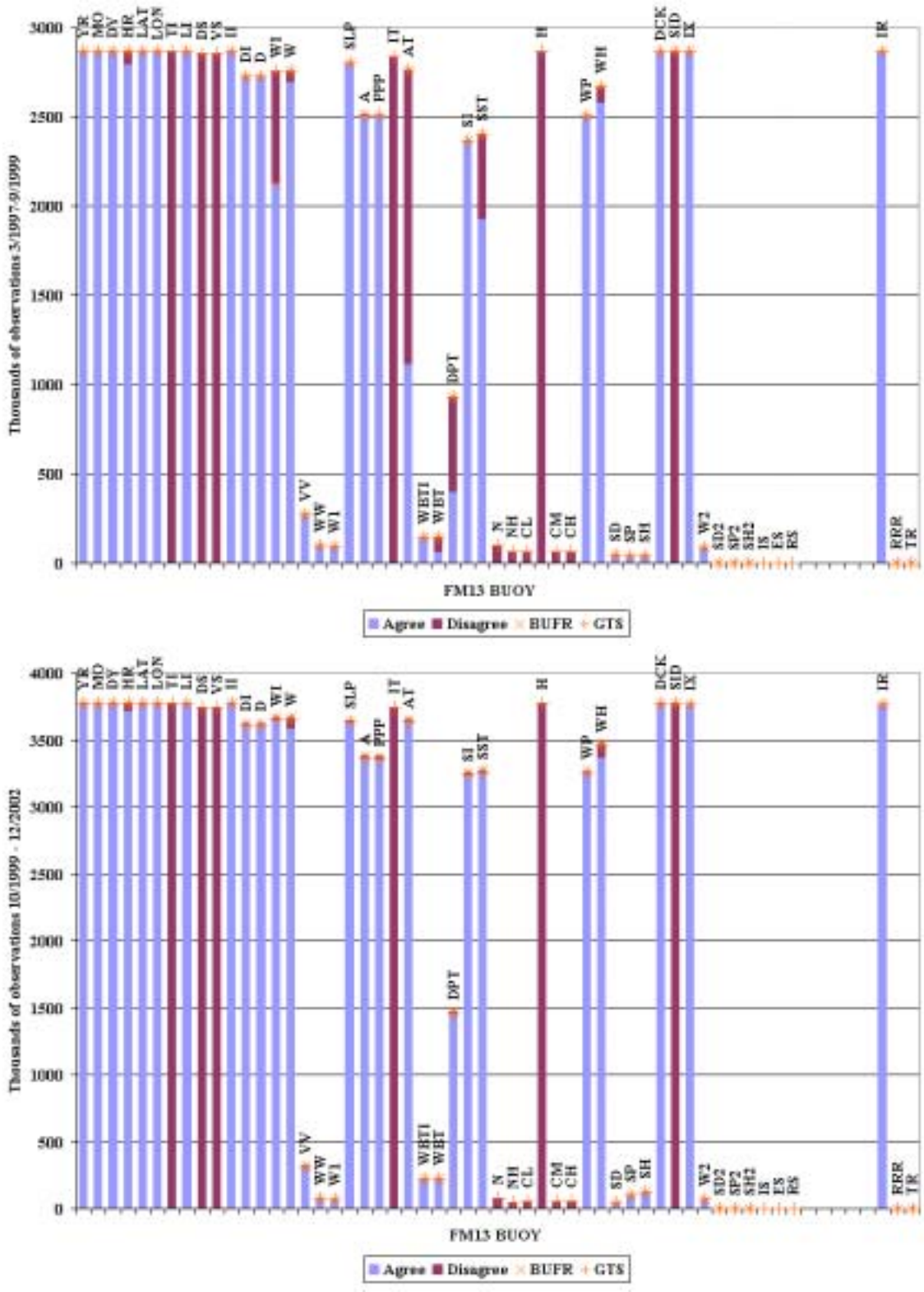
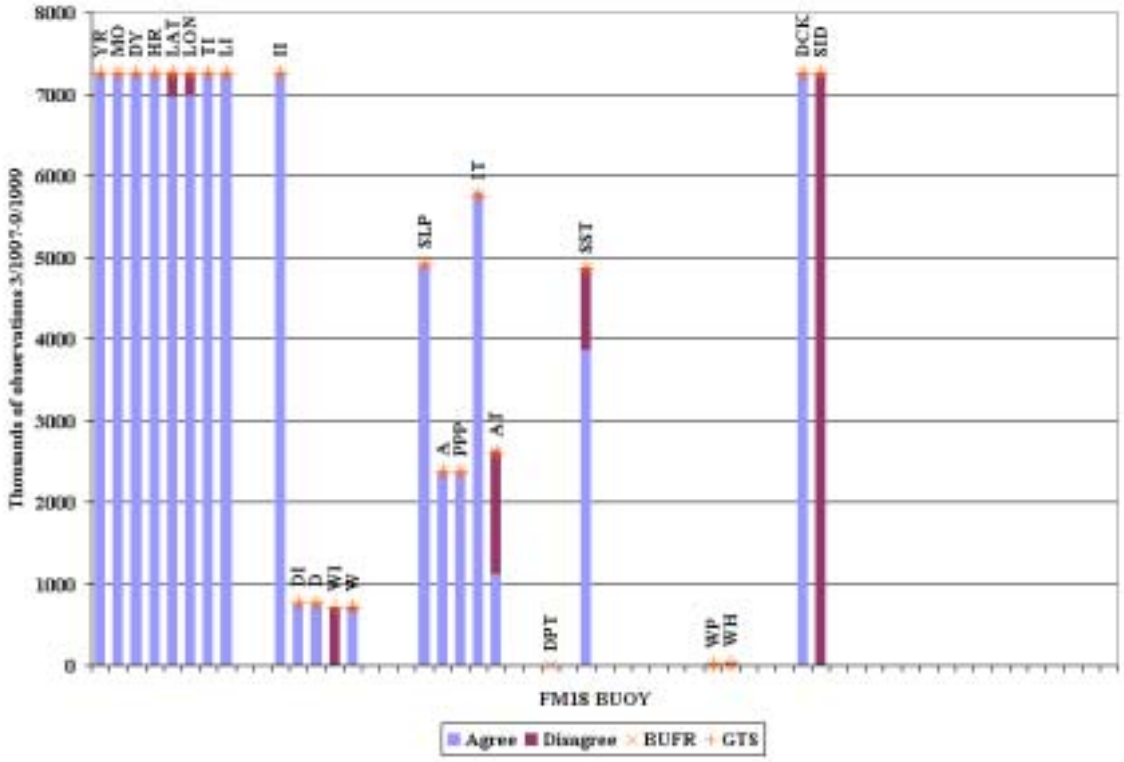


Figure 5. As for Figure 3, except the observations are plotted separately for buoys reporting in the FM 13 code.



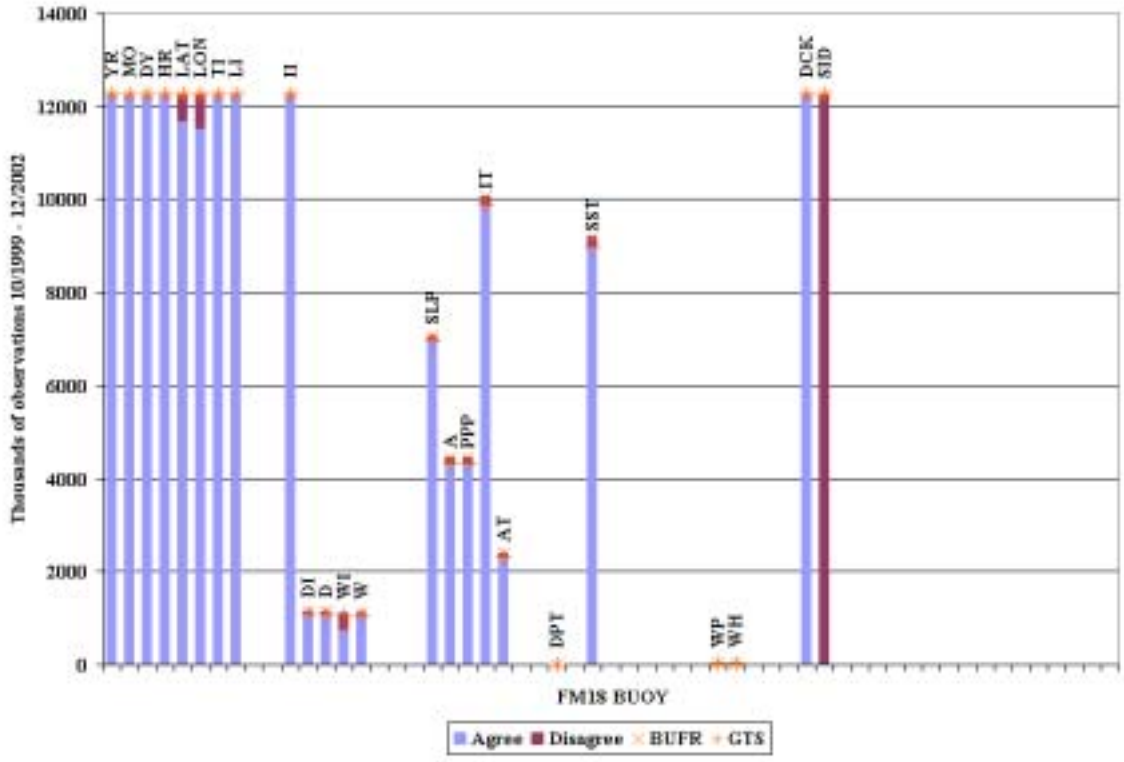


Figure 6. As for Figure 3, except the observations are plotted separately for buoys reporting in the FM 18 code.

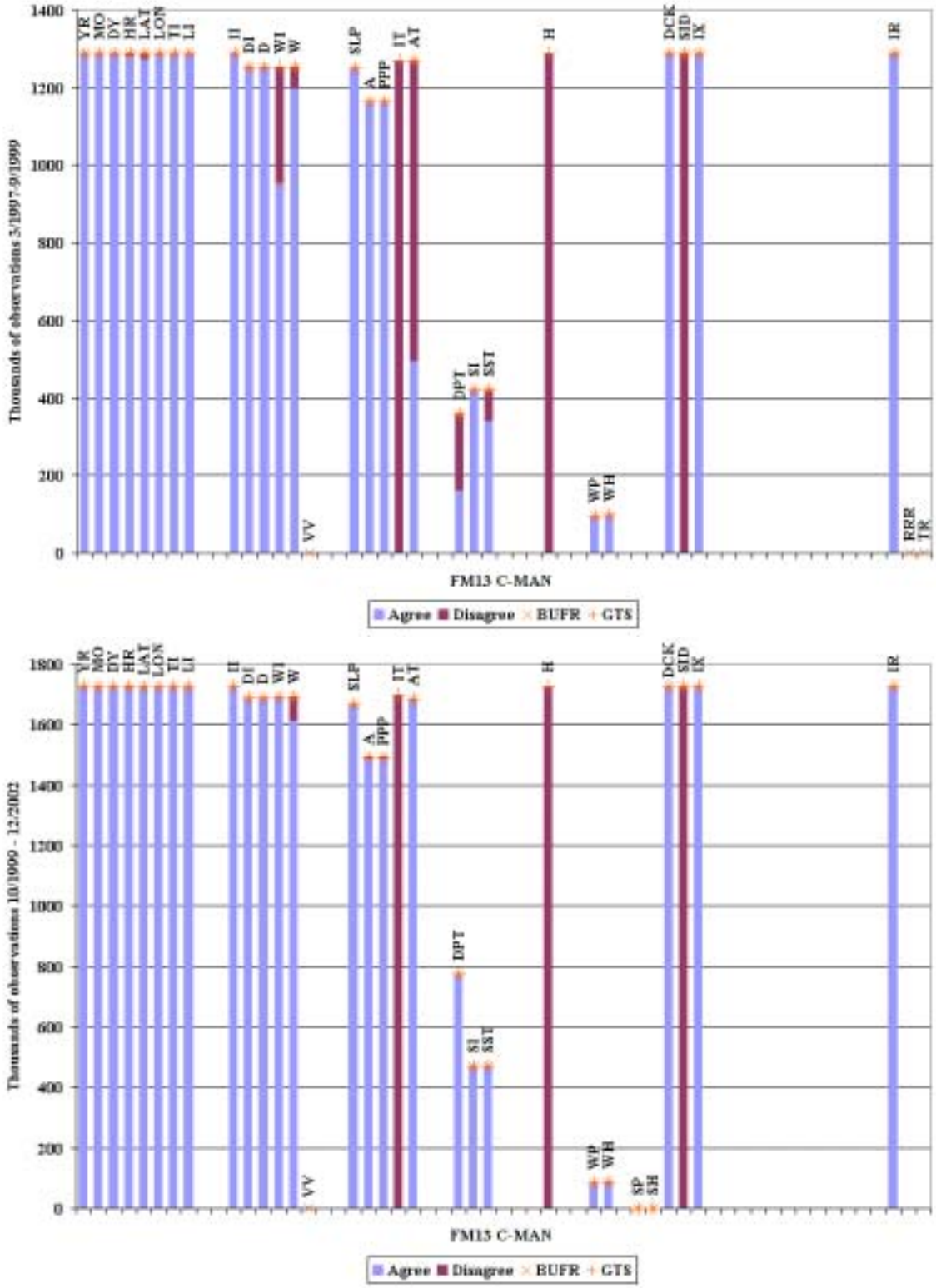


Figure 7. As for Figure 3, except the observations are plotted separately for C-MAN data reported in the FM 13 code.