Blending TAC and BUFR Marine in Situ Data for ICOADS Near-Real-Time Release 3.0.2

Chunying Liu[*,1,2], Eric Freeman[2,9], Elizabeth C. Kent[3], David I. Berry[3], Steven J. Worley[4], Shawn R. Smith[5], Boyin Huang[2], Huai-min Zhang[2], Thomas Cram[4], Zaihua Ji[4],

Mathieu Ouellet[10], Isabelle Gaboury[6], Frank Oliva[6], Axel Andersson[7]

William E. Angel[2], Angela R. Sallis[2,8], Adedoja Adeyeye[2]

[*] *Corresponding author address: chunying.liu@noaa.gov*

[1] *Riverside Technology, inc., Asheville, NC, USA*

[2] *NOAA National Centers for Environmental Information, Asheville, NC, USA*

[3] *National Oceanography Centre, Southampton, UK*

[4] *National Center for Atmospheric Research, Boulder, CO, USA*

[5] *Center for Ocean-Atmospheric Prediction Studies, Florida State University, Tallahassee, FL, USA*

[6] *Fisheries and Oceans Canada, Ottawa, ON, Canada*

[7] *Deutscher Wetterdienst, Hamburg Germany*

[8] *General Dynamics Information Technology, Falls Church, VA, USA*

[9] *University of Maryland, College Park, MD USA*

[10] *Natural Resources Canada, Ottawa, ON, Canada*

(Submitted to J. Atmos. Oceanic Tech.)

Abstract

This paper describes the new International Comprehensive Ocean-Atmosphere Data Set (ICOADS) near-real-time (NRT) Release, (R3.0.2), with greatly enhanced completeness over the previous version (R3.0.1). R3.0.1 had been operationally produced monthly from January 2015 onward, with input data from the World Meteorological Organization (WMO) Global Telecommunication Systems (GTS) transmissions in the Traditional Alphanumeric Codes (TAC) format. Since the release of R3.0.1, however, many observing platforms have changed, or are in the process of transitioning, to the Binary Universal Form for the Representation (BUFR) of meteorological data format. R3.0.2 combines input data from both BUFR and TAC formats.

In this paper, we describe: input data sources; the BUFR decoding process for observations from drifting buoys, moored buoys, and ships; and the data quality control of the TAC and BUFR data streams. We also describe how the TAC and BUFR streams were merged to upgrade R3.0.1 into R3.0.2 with duplicates removed. Finally, we compare the number of reports and spatial coverage of essential climate variables (ECVs) between R3.0.1 and R3.0.2.

ICOADS NRT R3.0.2 shows both quantitative and qualitative gains from the inclusion of BUFR reports. The number of observations in R3.0.2 increased by nearly one million reports per month, and the coverage of buoy and ship sea surface temperatures (SSTs) on monthly $2° \times 2°$ degree grids increased by 20%. The number of reported ECVs also increased in R3.0.2. For example, observations of SST and sea level pressure (SLP) increased by around 30% and 20%, respectively, as compared to R3.0.1, and salinity is a new addition to the ICOADS NRT product in R3.0.2.

Significance statement

The International Comprehensive Ocean-Atmosphere Data Set (ICOADS) is the largest collection of surface marine observations spanning from 1662 to present. A new version, ICOADS near-real-time 3.0.2, includes data transmitted in the Binary Universal Form for the Representation of Meteorological Data (BUFR) format, in combination with Traditional Alphanumeric Codes (TAC) data. Many of the organizations that report observations in near-real-time have moved to BUFR, so this update brings ICOADS into alignment with collections and archives of these international data distributions. By including the BUFR reports, the number of observations in the upgraded version of ICOADS increased by nearly one million reports per month and spatial coverage of buoy and ship SSTs increased by 20% over the previous version.

1. Introduction:

The International Comprehensive Ocean-Atmosphere Data Set (ICOADS) (originally COADS) was first released in 1985 (Slutz *et al.* 1985) and covered the period 1854-1979 (Woodruff *et al.* 1987). The most recent release, R3.0.0, spans from 1662 to 2014 (Freeman et al. 2017) with near-real-time (NRT) monthly updates from 2015-present, provided as ICOADS R3.0.1. The ICOADS is an integrated source of environmental observations such as sea surface temperature (SST), air temperature (AT), sea level pressure (SLP), and wind speed and direction. These observations are gathered from a variety of observing platforms including ships, moored and drifting buoys, and fixed platforms like oil rigs and coastal offshore structures. ICOADS provides access to a range of observation types, globally, and through the entire marine instrumental record.

As a foundational dataset, ICOADS has been widely used in various climate-related data and assessment products, such as the US National Oceanic and Atmospheric Administration's (NOAA's) NOAAGlobalTemp (Zhang et al. 2019, Vose et al. 2012); the NOAA Extended Reconstructed Sea Surface Temperature (ERSST), (Huang et al. 2015, 2017); the NOAA Optimum Interpolation Sea Surface Temperature (OISST) (Reynolds et al. 2007, Huang et al. 2021); the United Kingdom (UK) Meteorological Office (UKMO) Hadley Centre Sea Ice/Sea Surface

3

Temperature/global surface temperature data sets HadISST (Rayner et al. 2003), HadSST4 (Kennedy 2019), and HadCRUT5 (Morice et al. 2021); Nighttime Marine Air Temperature data sets HadNMAT2 (Kent et al. 2013), CLASSnmat (Cornes et al. 2020), and UAHNMATv1 (Junod 2019); the Intergovernmental Panel on Climate Change Climate Assessment Report (IPCC 2021); the Japan Meteorological Agency centennial-scale SST analysis (Hirahara et al. 2014); the NOAA-CIRES-DOE Twentieth Century Reanalysis (Slivinski et al. 2019); and the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA 5 reanalysis (Hersbach et. al. 2020).

In 2003, the World Meteorological Organization (WMO) approved a policy that all observational data exchanged internationally on the WMO Global Telecommunication System (GTS) be migrated from the Traditional Alphanumeric Codes (TAC) to the table driven code forms (TDCF), such as the Binary Universal Form for the Representation (BUFR) of meteorological data format (WMO 2019). To align with the WMO policy, many countries and programs have moved, or are currently in the process of moving, to the BUFR format in transmitting NRT marine observational data.

BUFR is designed to represent a wide range of new ocean observing platforms, e.g saildrones and other unmanned autonomous vehicles. BUFR's table-driven structure provides greater flexibility compared to the TAC format; BUFR can easily be extended and is self-describing. These table-driven and self-descriptive features in BUFR allow for greater reporting precision and more metadata (e.g., Pelletier 2008). In TAC, limited by the coding method, the observational values have limited precision; in BUFR, value precisions are prescribed in separate WMO tables thus the data can have their native precisions. For example, SST precision is $0.1°C$ in TAC but $0.01°K$ in BUFR; wind direction is in tens of degrees in TAC versus one degree in BUFR. Similarly, the platform metadata, instrument information, and data meanings can be fully documented in BUFR, while TAC only provides observation datetime, location, and maybe limited instrument information. Furthermore, the TDCF allows the modification of tables, providing a route for changes and additions in precision, metadata, and data type.

The marine in situ reports distributed over the GTS in the older TAC format, such as the FM-13 SHIP code (WMO 2019), have been available for decades and have been a major input source for ICOADS monthly operational updates. With more GTS marine data transmissions being made in

4

BUFR starting in 2016, there were fewer observations in TAC available for inclusion in R3.0.1. This was particularly true for the observations from drifting buoys, as they were among the first marine observation systems fully transitioned from TAC to BUFR. Even in those cases, however, the TAC streams still serve as backup sources in cases of BUFR transmission failures or other unexpected BUFR data distribution issues, so combining different data streams from various GTS data collection centers further enhances the reliability of the merged data.

To ensure that the ICOADS NRT dataset meets the needs of users, including product developers (Freeman et al. 2019, Kent et al. 2019, Huang et al. 2020), the NOAA National Centers for the Environmental Information (NCEI), in collaboration with ICOADS partners, began the process of recovering and decoding the NRT BUFR reports from ships, moored buoys, and drifting buoys. Realizing the impacts of the international format transitions to the operational R3.0.1 product, NCEI began working with the NOAA National Data Buoy Center (NDBC), the Canadian Department of Fisheries and Oceans (DFO) Global Data Assembly Center for Drifting Buoys (GDAC-CA), and the NOAA National Weather Service Telecommunications Gateway (NWSTG). The objectives of the collaboration are to collect previously transmitted BUFR reports but no longer available on the GTS, to establish operational ingest of real time BUFR transmissions from the NWSTG, and to ensure that the BUFR data are preserved in the NCEI archives. Meanwhile, a BUFR tool (see Section 3.a) was installed to decode the BUFR reports, then convert them to the ICOADS International Maritime Meteorological Archive format (IMMA1) (Smith et al. 2016). The decoded drifting buoy data was immediately made publicly available as an auxiliary data set starting in January 2016 until February 2019. Concurrently further development to decode other platform types transitioning to BUFR transmissions, such as moored buoys and ships, as well as needed adjustment to the quality control processes was undertaken to produce the new R3.0.2 with a goal to blend the NRT TAC and BUFR for a more complete marine in-situ dataset.

In this paper, we describe the ICOADS input data sources in Section 2. In Section 3, we describe the methodologies of BUFR decoding; data processing for the data collected by drifting buoys, moored buoys and ships; and the data merging and duplicate record elimination to produce ICOADS R3.0.2 "Total" and "Final" files (Freeman 2017). Section 4 compares the number of reports and coverage for essential climate variables (ECVs) between R3.0.1 and the newly-merged

TAC + BUFR data set, R3.0.2. Section 5 describes both quantitative and qualitative gains from this much-needed improvement for the ICOADS NRT stream. Section 6 provides insight into the future of ICOADS.

2. Input data sources

The ICOADS R3.0.2 covers the period from 1 January 2015 to the present, and is updated every month. Input data are collected in TAC and BUFR formats from the various sources of the WMO GTS data and archived at NCEI. The BUFR input data (Figure 1) includes messages provided by NWSTG, NDBC, and GDAC-CA. The BUFR data from NWSTG were less complete from January 2015 to September 2017; thus, in order to make a more complete data set, data from NDBC and GDAC-CA BUFR streams were collected and combined to help fill in missing data. Duplicated records were subsequently identified in the Total dataset and removed from the Final dataset (see Section 3c). Starting from September 2017, operational collection of the NWSTG stream was set up and this was used as the operational source for R3.0.2.
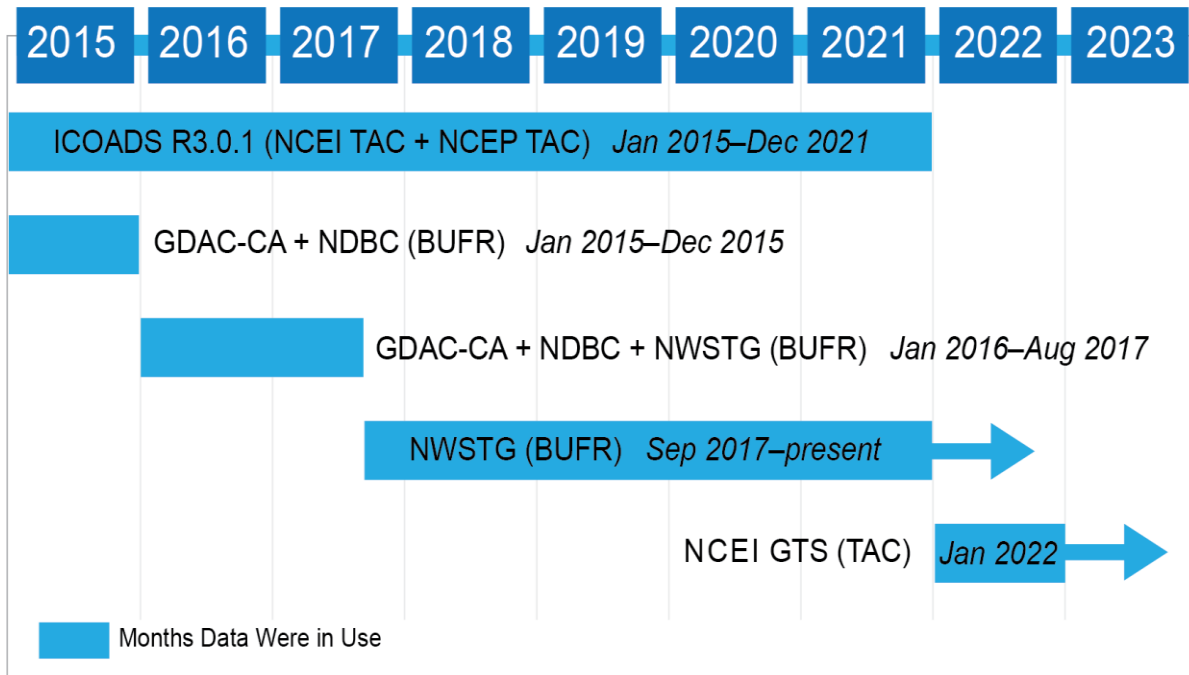

Figure 1. TAC and BUFR input data sources and time periods

6

Historically, the TAC data were acquired from both NCEI and NOAA's National Center for Environmental Prediction (NCEP). There was a period in which R3.0.1 and R3.0.2 overlapped (January 2015 to December 2021). For this period, the R3.0.1 blend of NCEI and NCEP TAC streams serve as the TAC input source for R3.0.2, and were merged with BUFR data records. Starting in January 2022 and onward, R3.0.2 became operational and R3.0.1 production as a stand-alone version ceased.

3. Methodologies

a. BUFR decoding method

Figure 2 shows the process flow of the BUFR decoding and TAC+BUFR merging. The ECMWF ecCodes software (https://confluence.ecmwf.int/display/ECC/ecCodes+Home) was used for decoding the BUFR data. The first step was to identify which observed variables were transmitted in the BUFR messages. Next, the corresponding key for each ICOADS variable was located in the BUFR tables and mapped to the associated IMMA1 format field. The appropriate WMO tables provide the correct unit and scale factor for each variable. Appendix A shows the ICOADS IMMA1 variables and their designated BUFR keys, where direct mappings were available. More BUFR keys can be found in the WMO Manual on Codes (WMO 2016).
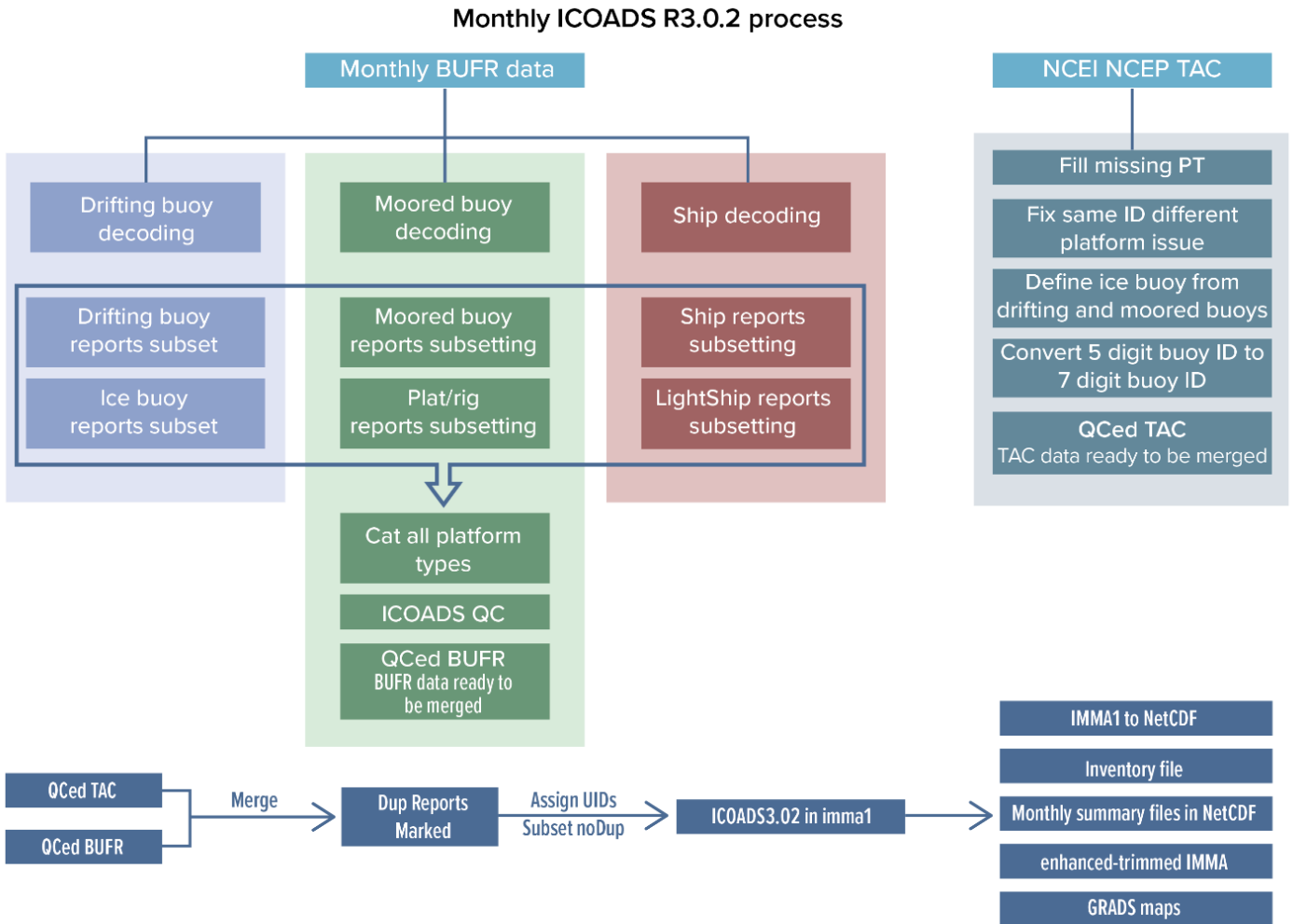
7

Monthly ICOADS R3.0.2 process

Figure 2. Schematic diagram describing the BUFR + TAC merging procedure. Acronyms are: QCed - Quality Controlled; Dup - Duplication; UID - Unique Identifier; C98 - an ICOADS attachment which provides the UIDs; PT - Platform Type.

b. Converting decoded BUFR data to the ICOADS common format

Following the mapping of the BUFR data to the ICOADS IMMA1 data and metadata fields, the BUFR values were converted to IMMA1 format. Due to the different category schemes between BUFR and IMMA1, new mapping tables were required (Appendix A).

8

Additional checks were conducted to quality control the BUFR platform type assignments. For example, through detailed examination and testing of the buoy ID rules from the WMO (https://community.wmo.int/rules-allocating-wmo-numbers) and visual inspection of reports, we were able to determine that some buoys (e.g., 4802008 and 4802009), distributed as moored buoys based on the WMO-assigned ID structure in the BUFR messages, were not stationary. Quality control allowed these buoys to be reassigned the correct platform type of drifting buoy (see https://rda.ucar.edu/datasets/ds539.5/docs/drift_buoy_bufr2imma1_v5.pdf for more information). There were reports which missed platform types; these platform types were inferred and assigned from their station IDs.

c. Unique identifier (UID) assignments

The ICOADS UID is a unique six-digit, base-36 value assigned to each individual ICOADS report. The UID provides a record tracking mechanism since the UID assignments are permanently assigned to the report (Freeman et al. 2017). This makes locating specific records possible, as well as being able to associate error adjustments or other information (i.e., reanalysis feedback, or observational metadata) back to a specific ICOADS report. Additional background information on UIDs can be found in Smith et al. (2016). UIDs are assigned to all reports prior to duplicate removal and were first assigned in Release 2.5.1, and UIDs are also assigned in the R3.0.0 and R3.0.1 Total files. From January 2015 to December 2021, UIDs from ICOADS R3.0.1 are carried forward into ICOADS R3.0.2 in order to preserve existing UIDs. New UIDs were assigned to the reports from the BUFR stream. From January 2022, after cessation of R3.0.1, each source will be processed separately and a new UID assigned to every report, i.e., the NCEI TAC and NCEI BUFR streams.

d. Data quality control (QC)

ICOADS data originate from many different types of observing platforms, such as ships, drifting buoys, moored buoys, coastal stations, and fixed ocean platforms. Random errors and systematic biases may be introduced during the observation and transmission processes (Kennedy et al. 2011). This has been more common in ship observations e.g., due to poor data management of historical

9

data and poor siting of instrumentation (Kent et al. 2017). Quality controls (QCs) have been used starting from COADS Release 1 (Slutz 1985) onward. Updated trimming limits were used for R3.0 as described at https://icoads.noaa.gov/e-doc/R3.0-stat_trim.pdf. Data that exceed climatological thresholds are also flagged during the QC process (Wolter 1997, Smith 2021).

Using the ICOADS QC and trimming software, we assigned QC flags as shown in Appendix B. QC flags were assigned to 14 selected weather and ocean variables, and the sum of the weighted flags (see quality code in Appendix C) determined which report to retain among the duplicates. The QC procedures were slightly different for the TAC and BUFR streams, as described below.

Quality control of the BUFR data used the standard ICOADS QC software, with an additional step which compared the platform types between the TAC and BUFR streams to ensure that the platform types assigned to observation with the same ID were consistent between the TAC and BUFR streams.

To make the older TAC data compatible with the modern BUFR reports, the following checks and modifications were made to the TAC data: (a) dates were validated, and if an erroneous date was detected, they were kept in the Total Files but removed from the Final files; (b) platform types were evaluated and corrected for observations in which TAC and BUFR reported the same platform ID but different platform types; approximately 300 IDs were found with this issue; (c) missing platform types were filled in wherever possible; (d) 5-digit buoy IDs in the TAC format were converted to 7-digit buoy IDs; (e) attachment C98 was added, which provides the UID attachment for each R3.0.2 IMMA1 report; (f) ice platform type reports from TAC drifting and moored buoy reports were identified and marked with the correct platform type. The ice buoy platform type was previously unused in R3.0.1 and was defined simply as drifting or moored buoys, but defined as ice buoys in BUFR.; (g) fixing incorrect IDs, e.g., converting ID 9077070 to ID C6OM7 due to the International Maritime Organization IMO number (9077070) erroneously being encoded as the ID rather than the true call sign (C6OM7); and (h) fixing incorrect ID indicators (II) (e.g., "II" should be "11" if it is a 7-digit buoy ID. "II" was assigned to zero in ICOADS R3.0.1 when it was a 7-digit buoy ID).

e. Duplicate elimination (dupelim)

10

An important step in the production of R3.0.2 is the identification of duplicate observational records for flagging and removal from the Final files (duplicates are retained in the Total files). The QC is important in the duplicate elimination process as QC flags (see Appendix C) are assigned to weather variables (see Appendix B) automatically by the ICOADS QC software suite. The sum of the weighted flags helps determine which report to retain as unique or as the "best" duplicate in cases where multiple exact or near-duplicate matches occur. Ideally, duplicate reports can be identified by exact comparison (exact match) of location, time, and observed elements within the reports. Due to the different input data sources and transmission formats, however, slight differences in TAC and BUFR precisions (retained digits) and encoding/decoding techniques have made duplicates more challenging to identify. Reports that previously were exact duplicates now have precision-related differences in one or more of their fields. For example, the latitude and longitude values from some reports in the TAC stream have a precision of 0.1 degrees, but the same reports from the BUFR stream have a precision of 0.01 degrees. This apparent discrepancy would distinguish and identify them as separate and unique reports in the R3.0.1 algorithm. To account for potential near-duplicates due to these changes in precision, we explicitly allow a 0.05-degree tolerance level of latitude or longitude to define a duplicate in the R3.0.2 algorithm. By relaxing the location comparison criteria in the BUFR data, most of the near-duplicate matches can now be detected.

Similar checks are applied to other variables during the duplicate flagging process: First, potential duplicates (dups) are defined within the same $1°\times1°$ box and within $\pm$ one hour ("hour cross") or one day ("day cross"). A check is then performed for seven observed elements (wind speed, visibility, present weather, past weather, SLP, AT, and SST) to identify duplicates. The checks involving present and past weather fields uses "allowances", which consider weather elements that match under some environmental circumstances even though they were not identical as discussed above (see the ICOADS dupelim documentation at https://icoads.noaa.gov/e-doc/R3.0-dupelim.pdf).

Next, we determine which duplicate report should be retained in the Final dataset. The quality code (Release 1, supp. K. Slutz *et al.* 1985), as computed by the NCDC-QC, is the basis for the selection of one duplicate report over another. If quality codes are identical, a priority list by deck is used to select one of the duplicates. If priorities are identical, the report with the highest deck number

(DCK) is preferred. The deck number for BUFR data is 798. SID for BUFR is 172. The priority for BUFR is 1. If the decks are identical, the report with the highest source ID (SID) is preferred. If the SIDs are identical, the second report in sort order is selected. The duplicate status is assigned from 0 to 14, from unique to worst duplicate (Table 1), respectively. No reports, however, are deleted at this stage and all duplicates are retained in a "Total" file in case further examination is needed at a later time. BUFR data were given higher priority over TAC, as BUFR retains higher precisions, contains more metadata, and provides additional variables sometimes not available in the TAC stream.

DUPS	Description
0	unique
1	most unique
2	OK unique with substitution
3	*worse duplicate: uncertain weather element match with hour cross*
4	worse duplicate: uncertain weather element match with no cross
5	*worse duplicate: uncertain weather element match with day cross2*
6	*worse duplicate: time/space match with ID mismatch (unused until 1950)*
7	*worse duplicate: certain weather element match with hour cross*
8	worse duplicate: certain weather element match with no cross
9	worse duplicate: combined DUPS 4 and 6
10	worse duplicate: combined DUPS 6 and 8
11	worse duplicate: time/space/ID match
12	worse duplicate: combined DUPS 4 and 11
13	worse duplicate: combined DUPS 8 and 11
14	automatic data rejection

Table 1. Duplicate status (DUPS) assignments as noted in Table D8 from Smith et al. 2016. For more information, see the ICOADS R3.0 Duplicate Elimination and Related Processing documentation at

12

https://icoads.noaa.gov/e-doc/R3.0-dupelim.pdf. Italic items are dup status not found in this TAC + BUFR merge practice.

Finally, reports that are not flagged as landlocked and have a duplicate status of ≤ 2, where 0 is unique, 1 is most unique, and 2 is best duplicate with substitute (DUPS=2 not assigned in R3.0.2), are retained for the Final file without duplicates.

4. Comparisons between ICOADS R3.0.2 and R3.0.1

a. Number of reports in R3.0.2 versus R3.0.1

Figure 3 shows the monthly number of reports identified as unique from ship, surface drifting buoy and moored buoy observations in the ICOADS R3.0.2 (TAC + BUFR) and ICOADS R3.0.1 (TAC - only). An important feature of ICOADS NRT R3.0.2 is that it has much higher numbers of drifting buoy observations (Figure 3b) because the transition from TAC to BUFR formats started earlier for surface drifters. The number of surface drifting buoy observations in the R3.0.1 Final file decreased from about 1.1 Million (M) in 2015 to less than 0.1M by the end of 2020 (Figure 3b) as TAC transmission was phased out quickly for these data (Figure 4). In contrast, the observation numbers of surface drifting buoys in the R3.0.2 Final file were steadily above about 1.2M from 2015 onward. The number of ship observations was slightly higher in the R3.0.2 Final file than in R3.0.1 after January 2016 (Figure 3a), with a slow but steady increase, in particular in 2020.
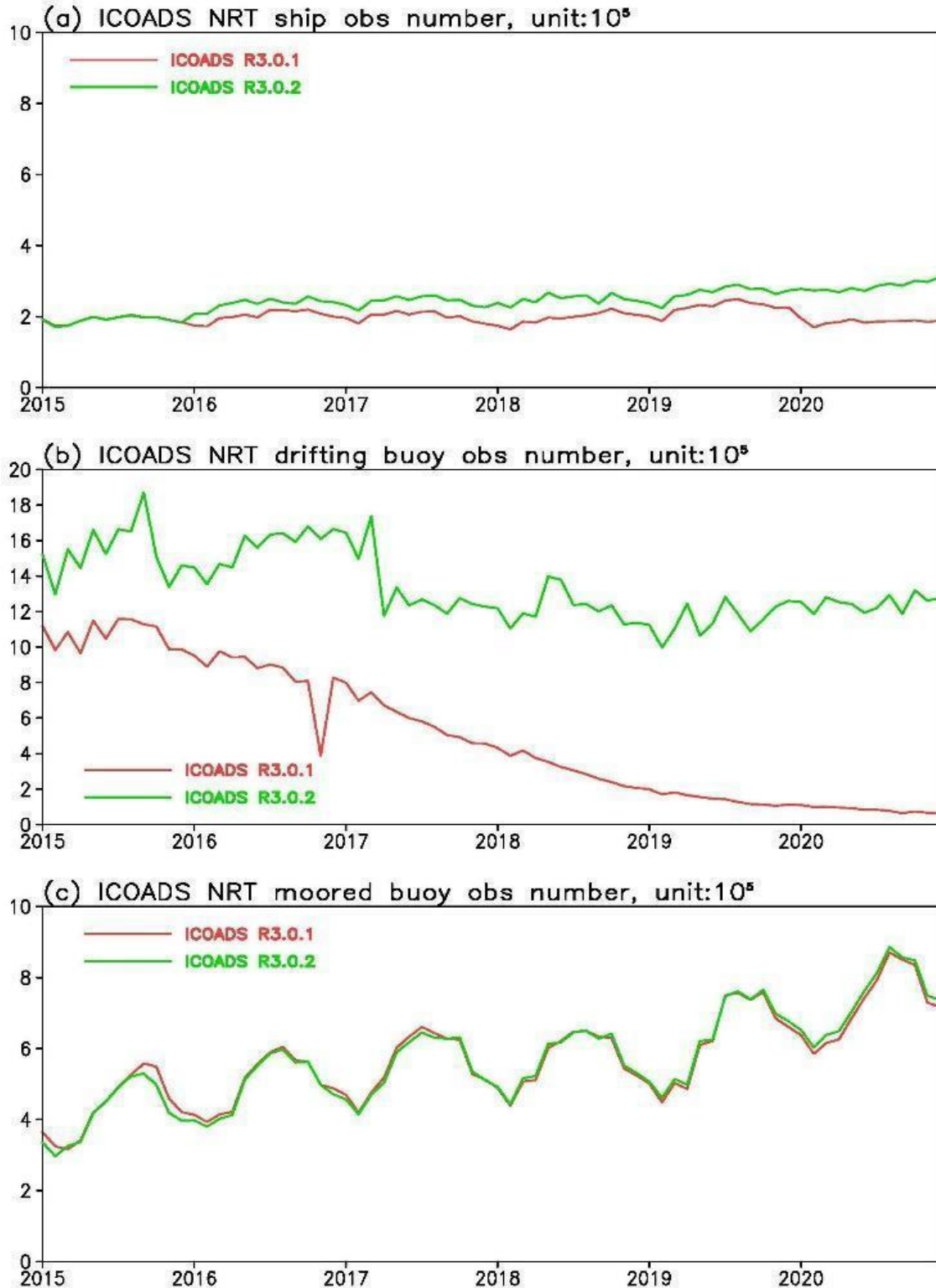
13

Figure 3. The monthly number of in situ reports (x 10^5) in ICOADS R3.0.2 from TAC+BUFR (green) and ICOADS R3.0.1 from TAC (red) for the period Jan. 2015 – Dec. 2020. Panels (a), (b), and (c) show the number of unique reports from ship, surface drifting buoy and moored buoy observations, respectively.

14

The TAC ship reports showed an obvious drop at the end of 2019. The moored buoy report numbers showed relative consistency between R3.0.2 and R3.0.1, largely due to the transition from TAC to BUFR formats started later for surface moored buoys than drifters. Many moored buoys are still transmitting in TAC format. We also noticed that the number of moored buoy observations was slightly lower in R3.0.2 than in R3.0.1 between January 2015 and September 2017. Three factors explain the slightly reduced report numbers of moored buoy in the R3.0.2 Final file: Firstly, some reports with invalid dates were included in R3.0.1 but were removed in R3.0.2 Final files. Secondly, duplicate identification criteria in R3.0.2 were slightly different from R3.0.1 (Section 3). Thirdly, note that there was no specific ice buoy category in R3.0.1 and ice buoys were classified as moored or drifting buoys, while in R3.0.2 ice buoys are their own classification, causing the number of moored and drifting buoys to be lower in R3.0.2.

b. Detailed contribution to R3.0.2: BUFR vs TAC

Figures 4a, 4b and 4c show the report numbers for the raw BUFR and R3.0.1 Total files as input to the ICOADS NRT R3.0.2. Figure 4a shows that there were little BUFR reports for ship observations before January 2016. After January 2016, the number of BUFR reports began to appear (2×10^5) and increased gradually to approximately 3×10^5 by the end of 2020 except for a sudden drop in the mid-2016. In contrast, the raw TAC ship reports maintained a steady level until near the end of 2019, decreased rapidly into early 2020, and then remained steady again.

Figure 4d shows the relative contributions from the TAC and BUFR reports for ship observations after duplicate elimination for the merged and processed ICOADS NRT R3.0.2. Recall that when there are duplicated reports between TAC and BUFR, BUFR reports were given a higher priority selection for R3.0.2 (Section 3). This procedure is shown in Figure 4d: When BUFR report numbers were low (most of 2015 and a dip in mid-2016), contributions from TAC reports were high, and vice versa. Combined with duplicate elimination, the TAC-BUFR merged reports remained quite steady with an overall slow increase during 2015-2020.

Similar results were noted for the surface drifting buoys and moored buoys. For the surface drifting buoys, the TAC remained the primary distribution format through most of 2015, followed by a

15

steady decrease (Figure 4b). The raw BUFR surface drifting buoy reports showed a rapid increase in mid-2015; by mid-2016, the BUFR reports exceeded the TAC reports. Once BUFR was the predominant distribution format for drifting buoys, BUFR reports were then given a higher priority in R3.0.2. Figure 4e shows that the BUFR report contribution surpassed the TAC report contribution by mid-2015 and since late 2016 the TAC report contribution for the R3.0.2 Final file has been small.

The moored buoy reports (Figures 4c and 4f) showed a persistent and dominant contribution from TAC reports until late 2019 when the number of BUFR reports exceeded TAC reports and remained higher thereafter. The TAC report contribution has a clear seasonal cycle; this was due to the seasonal summer mooring deployments in the U.S. Great Lakes and coastal (including lake) moored buoys being primarily serviced in the summer months. In contrast, the number of moored buoy observations from BUFR had been slowly increasing from 2015 to near the end of 2019, at which time there was a noticeable shift from TAC observations to a majority of BUFR observations.
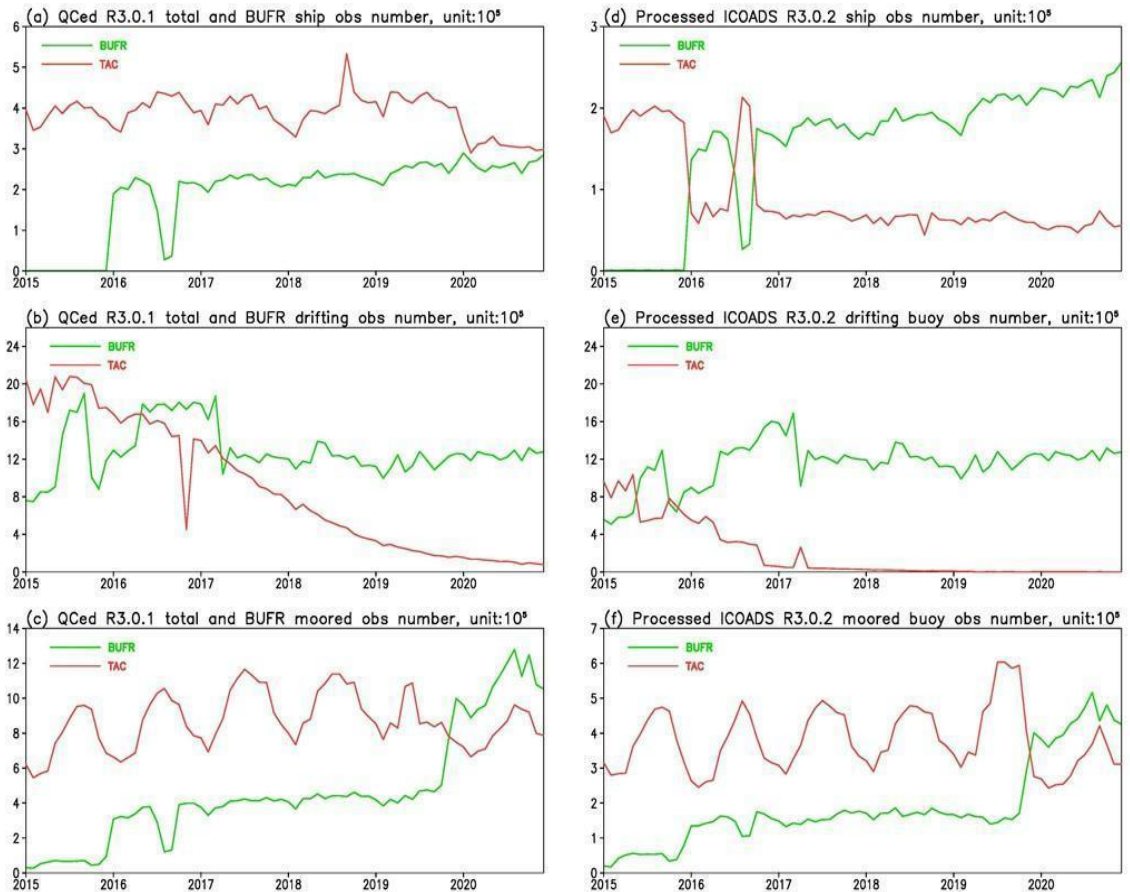
Figure 4. Number of in situ reports (x 10^5) from the raw BUFR (green) and the Total R3.0.1 (red) streams (as input to ICOADS R3.0.2; left panels (a)-(c)) and ICOADS processed data (output from the R3.0.2 Final files with duplicates removed; right panels (d)-(f)), for the period January 2015 – December 2020. Panels (a) and (d) show the ship observations, panels (b) and (e) show the surface drifting buoy observations, and panels (c) and (f) show the moored buoy observations. Note the different y-axis scales used for ship and moored-buoy plots to fully utilize the plot space to clearly show the time series patterns.

c. ECV reports in R3.0.2 versus R3.0.1

The numbers of ECV observations are generally higher in the ICOADS NRT R3.0.2 Final file than in R3.0.1, as quantified below. The numbers of wind speed and direction (Figure 5a), AT (Figure 5c) and relative humidity (RH; Figure 5e) observations were slightly higher in R3.0.2 than in R3.0.1. The number of SLP observations (Figure 5b) was about 2.8M in R3.0.1, but increased in R3.0.2 to 3.7M by 2019 and has remained relatively steady since then. The number of SST observations (Figure 5d) decreased in R3.0.1 from about 2.7M in 2015 to 2.2M in 2019, but increased in R3.0.2 from about 2.7M in 2015 to 3.5M since mid-2019. In addition, there were no salinity observations (Figure 5f) in R3.0.1, but 0.05M are now available in R3.0.2 from the BUFR stream.
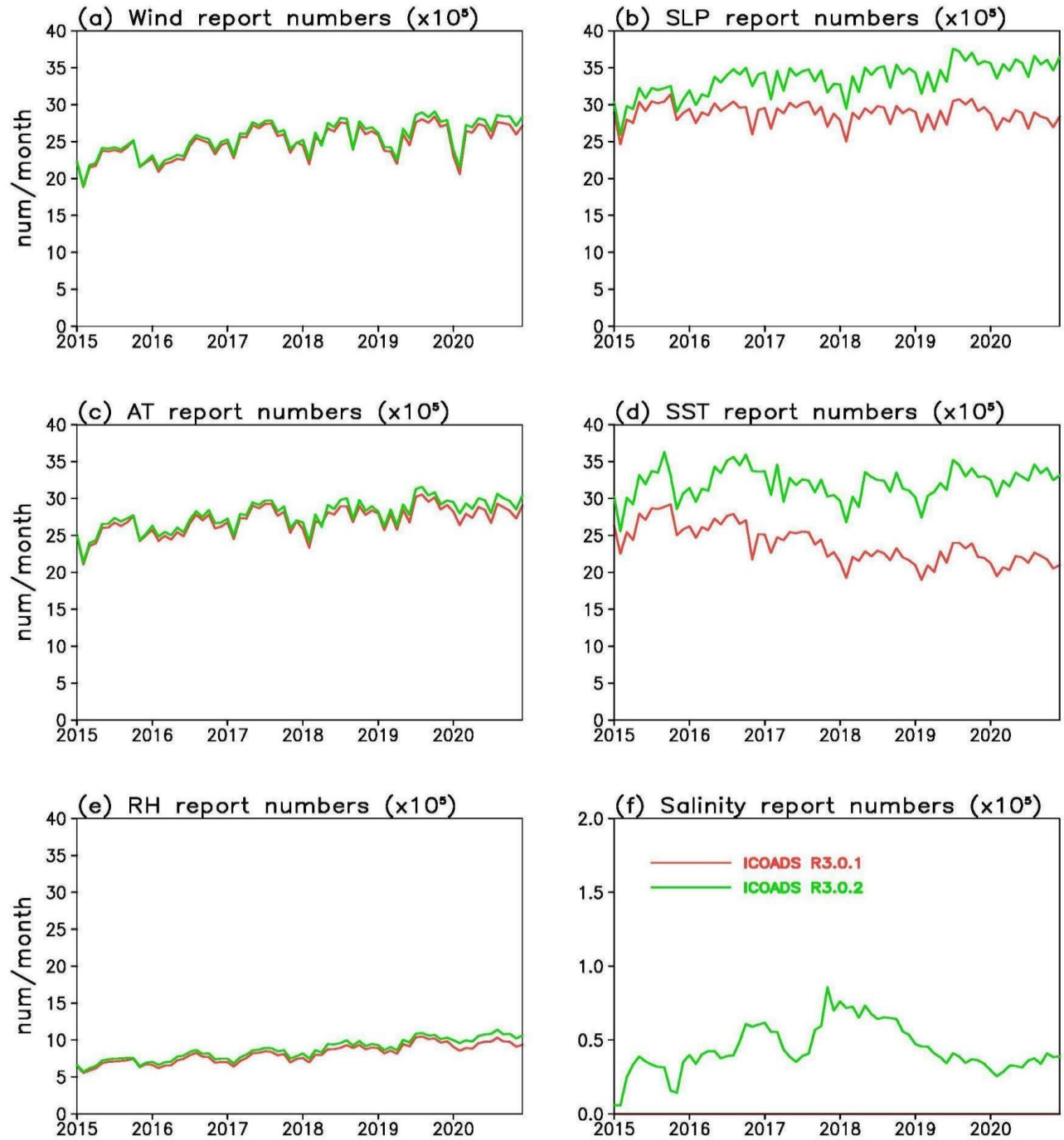
17

Figure 5. Number of essential climate variables (ECVs) reports (x 10^5) in ICOADS R3.0.2 (green) and ICOADS R3.0.1 (red) for the period Jan. 2015 – Dec. 2019. The figure shows the number of reports for wind direction (a), sea level pressure (b), air temperature (c), sea surface temperature (d), relative humidity (e), and sea surface salinity (SSS) (f), respectively. Note the y-axis for SSS is different from those for other variables.

18

e. Spatial coverage: R3.0.2 versus R3.0.1

Figure 6 shows the spatial coverage of the ECVs in May 2019, which exhibited an increase in SST and SLP coverage and a smaller increase for AT.
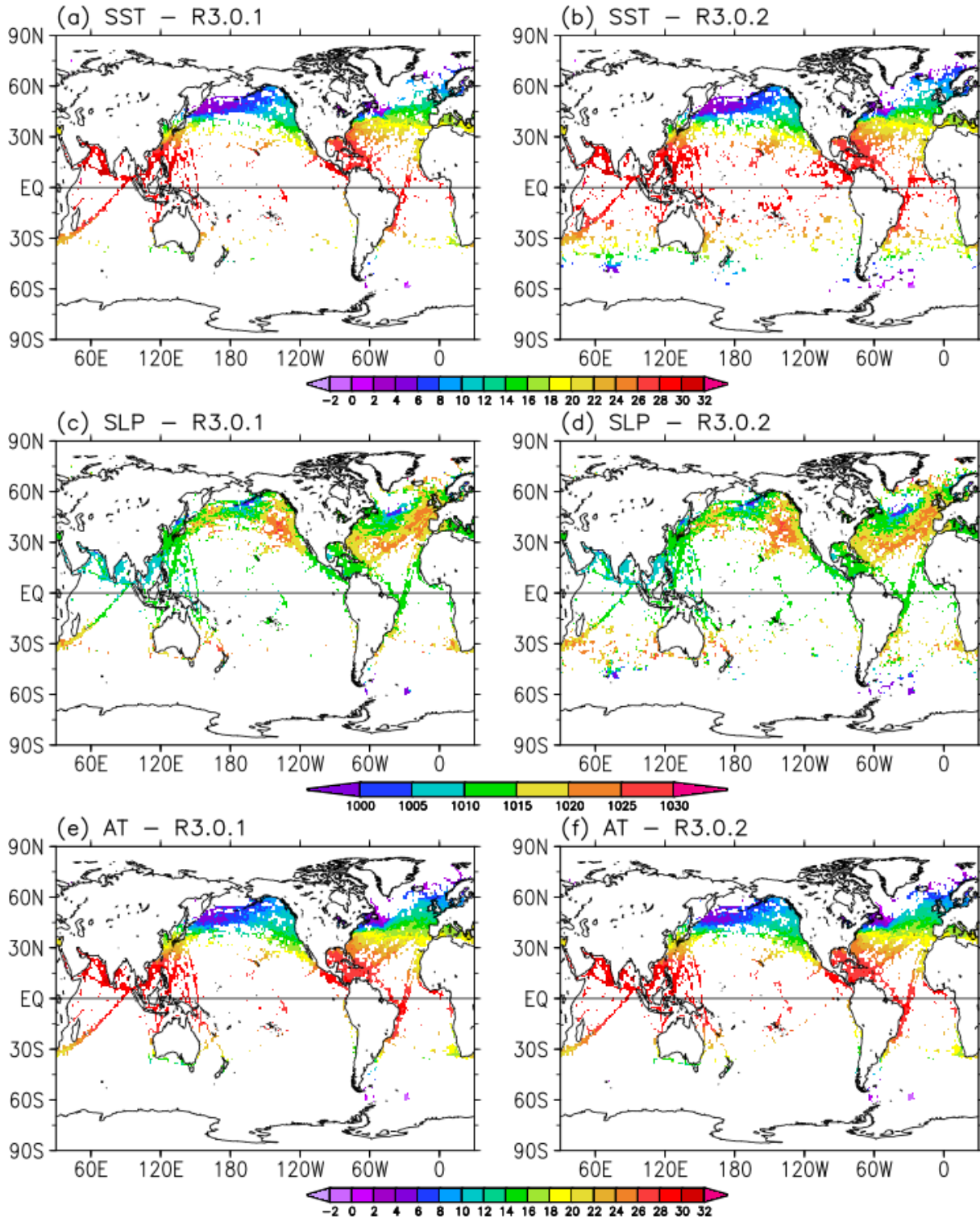
Figure 6. Distribution of ECVs in May 2019. Panels (a) and (b) show the SST distribution of R3.0.1 and R3.0.2, respectively. Panels (c) and (d) show the SLP distribution of R3.0.1 and R3.0.2, respectively. Panels (e) and (f) show the AT distribution of R3.0.1 and R3.0.2, respectively.

Figure 7 presents the difference between the number of months with observations (SST, AT, SLP, and wind) in R3.0.2 and R3.0.1 in 2021. The metric is expressed as the difference of the number of months with data relative to the 12 months (in %). For example, 50% means that R3.0.2 has six additional months with observations compared to R3.0.1. Both the SST and SLP show significant increases (>20%) in most of the areas, particularly in the South Pacific Ocean, Indian Ocean, the North Atlantic Ocean, and the Arctic Ocean. AT and wind speed have the same pattern but smaller changes with much less spatial coverage. This is because the observation number increase mainly comes from drifting buoys in the BUFR stream. Drifting buoys usually report SST and SLP but not AT and wind speed. Since AT and wind are mostly from the ship reports, their spatial patterns are similar to the ship report availability shown in Figure 10.
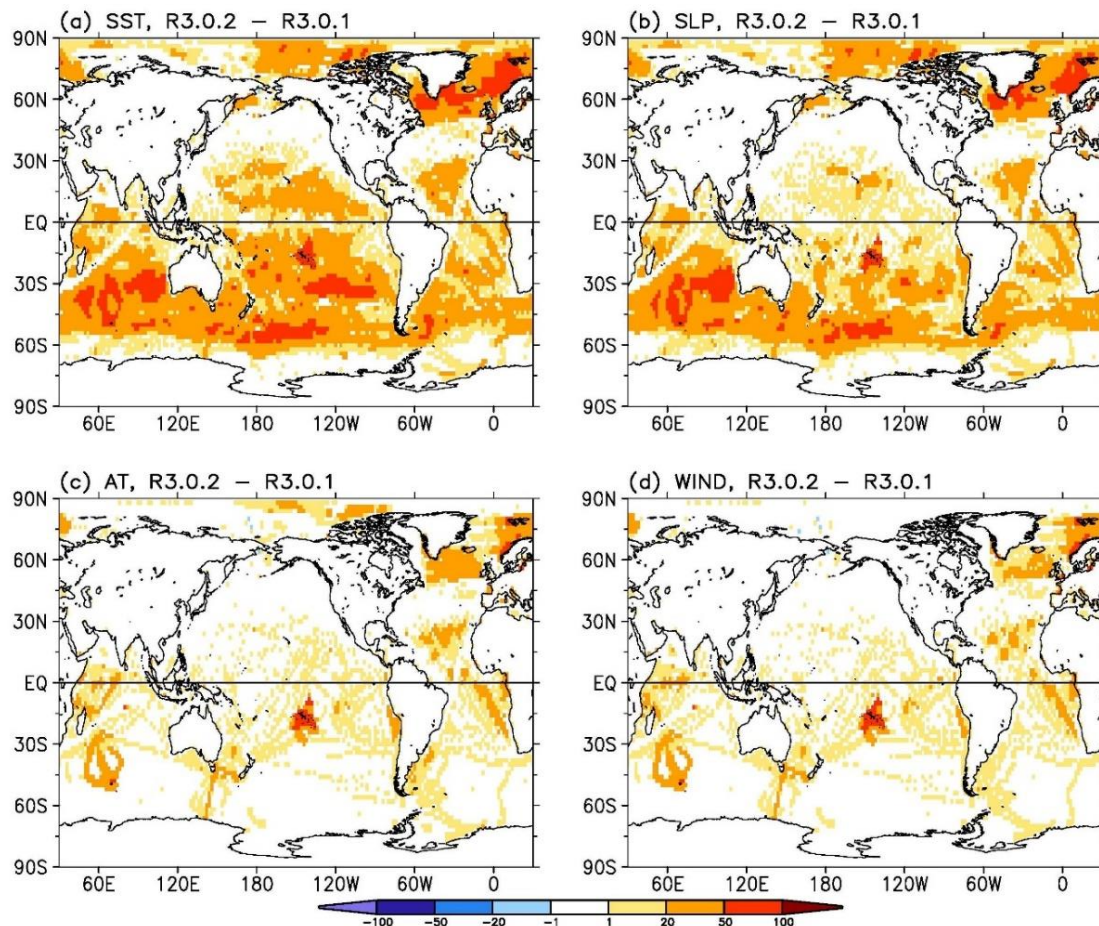
Figure 7. Differences of observation coverage (%) between R3.0.2 and R3.0.1 in 2021 for ECVs of (a) sea surface temperature, (b) sea level pressure, (c) air temperature, and (d) wind speed. The observation coverage is calculated as the ratio of months with observations over the total 12 months in a year in a 2-degree grid box.

Figure 8 shows the percentage of ocean two-degree boxes covered by ECVs. SST coverage increased about 20% (panel a), SLP coverage increased about 15% (panel b), and AT and wind coverage increased about 3% (panels c and d). The small dip in R3.0.2 in 2018 is likely due to the relaxed duplicate removal criteria, but more investigation is required.
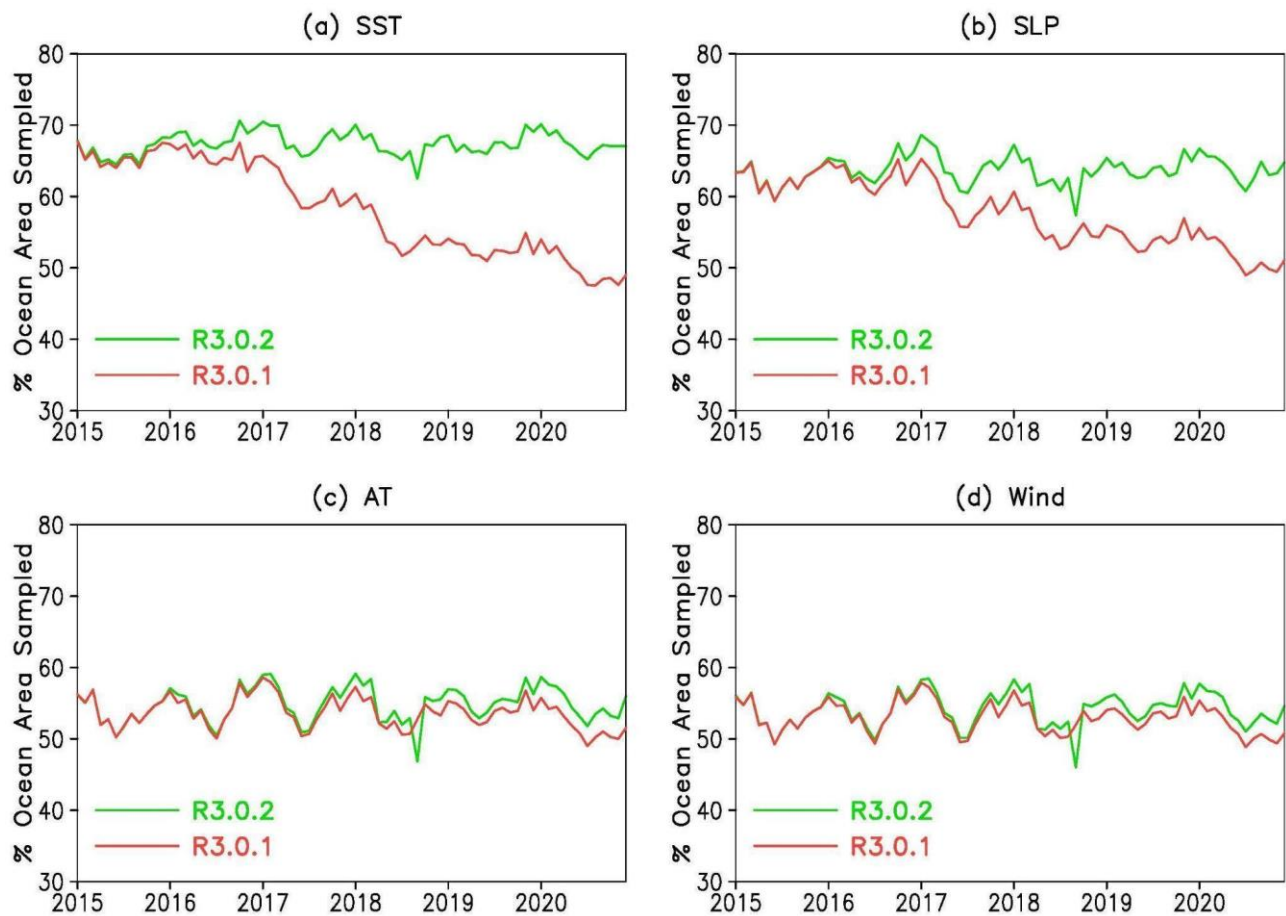

Figure 8. Monthly percentage of the ECV coverage from January 2015 to December 2020 for two-degree ocean boxes. Panel (a) is for sea surface temperature; Panel (b) is for sea level pressure; Panel (c) is for air temperature; and Panel (d) is for wind speed. R3.0.1 is the red line, and R3.0.2 is the green line, respectively.

21

In addition to the ECV coverage, Figure 9 shows the overall increase in the number of observations in the R3.0.2 Final file (Figure 9b) over R3.0.1 (Figure 9a) for May 2019. The increase in spatial coverage is mostly attributed to the increase in surface drifting buoys as discussed previously. Studies indicate that spatial data coverage is critical to downstream higher-level SST analysis and other data products, such as in the widely used Optimum Interpolation Sea Surface Temperature (OISST) product from combined satellite and in situ observations as the increased coverage of in situ observations in R3.0.2 is critical to correct satellite biases (e.g., Huang et. al. 2019, 2021). The increased observation numbers and spatial coverages also increase the quality of uncertainty qualifications in surface marine observations and many products with ICOADS as an input dataset, such as SST (Kent 2019), air-sea fluxes (Berry and Kent 2011), the World Ocean Database (Boyer et al. 2018), nighttime marine air temperature (e.g., Kent et al. 2013, Hirahara et al. 2014, Kennedy et al. 2019, Huang et al 2017), and climate change assessments (e.g., Hausfather et al. 2016, IPCC 2021). The increase in Sea Level Pressure is also critical to the quality of weather and climate forecast and reanalysis (e.g., Slivinski et al. 2019).

In situ observations ICOADS R3.0.1

drifting buoy · moored buoy · ship

In situ observations ICOADS R3.0.2

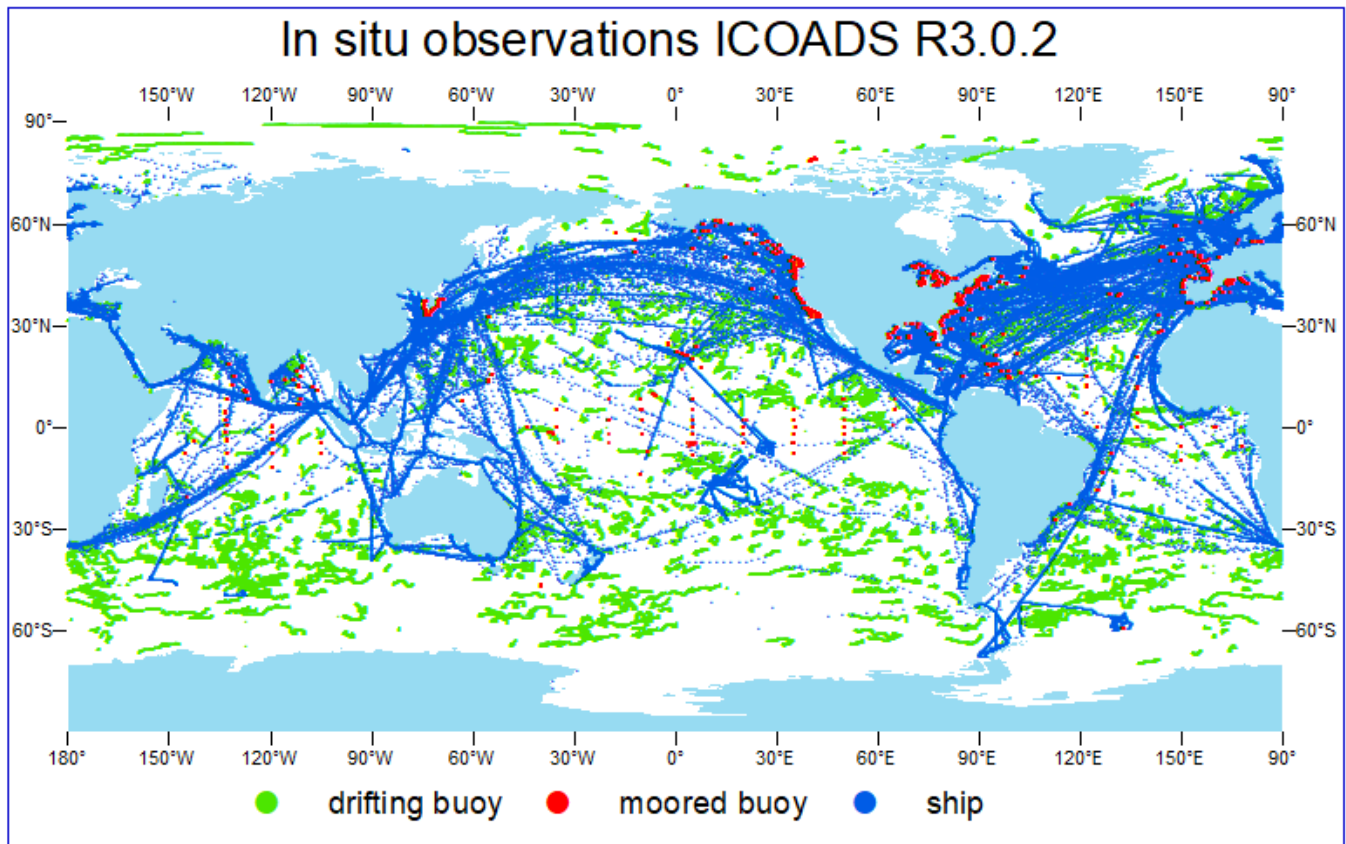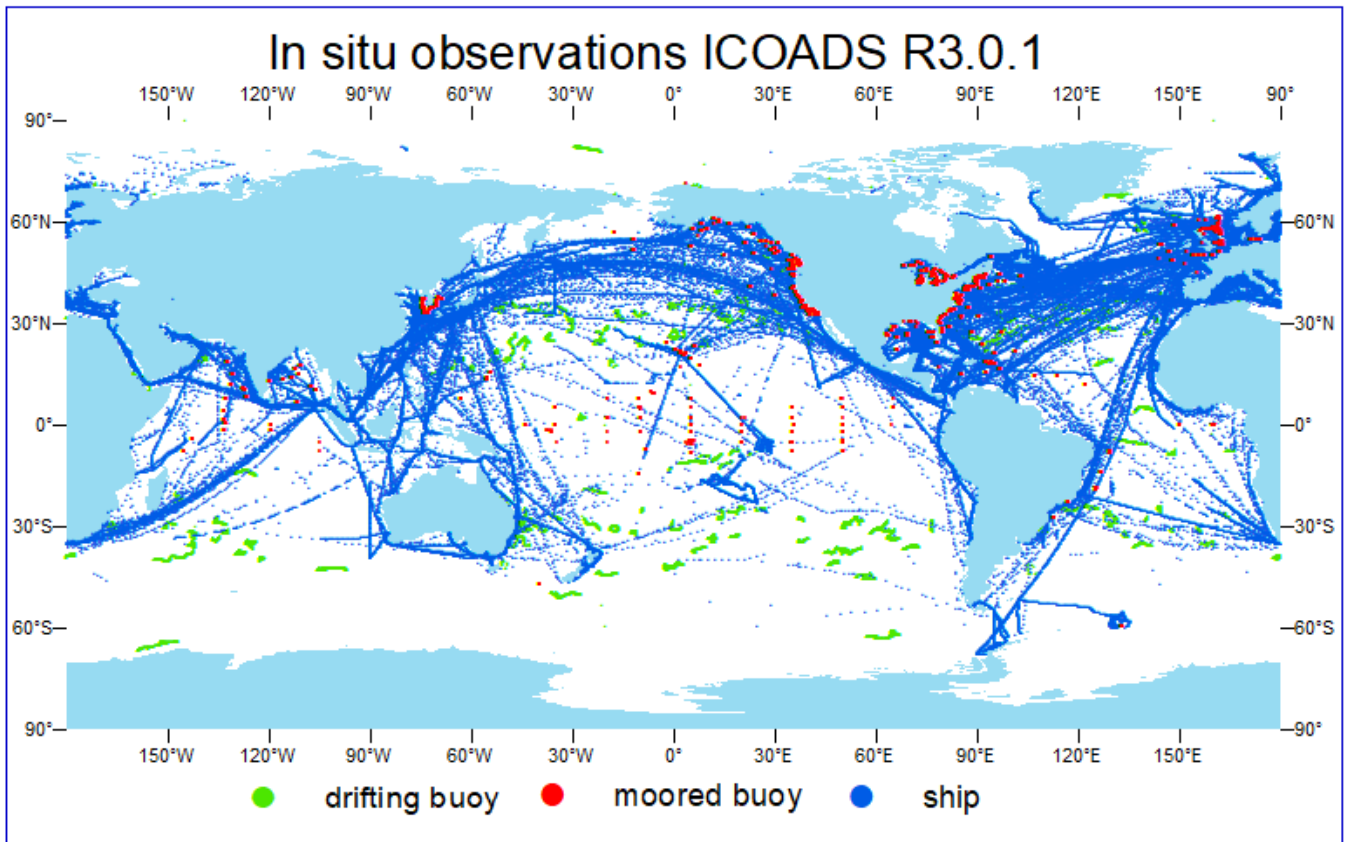drifting buoy · moored buoy · ship

23

Figure 9. Spatial distribution of in situ observations for May 2019 in ICOADS R3.0.1 (top panel) and ICOADS R3.0.2 (bottom panel). Locations for ship, surface drifting buoy, and moored buoy observations are shown in blue, green, and red, respectively.

Figure 10 shows the relative difference between the number of months with observations (drifting buoys, moored buoys, and ships) in R3.0.2 and R3.0.1 in 2021. For drifting buoys, there is an obvious increase of 20% in nearly all the oceans, particularly in most of the mid-latitude oceans and almost all the North Atlantic Ocean, the increase is over 50%. Spatial coverage of moored buoy is patchy but the data availability has increased.

It is worth mentioning that in a small region of the North Sea, there seems a 100% decrease in moored buoy coverage. This is a spurious signal due to incorrect platform type assignment in some TAC data and corrected in R3.0.2. For example, ID 6200114 was incorrectly assigned as moored buoy in TAC but in R3.0.2 it was correctly assigned as a fixed ocean platform.

The coverage increase of ship data is overall less than 20% but over 20% along some tracks (e.g., in the Atlantic Ocean between Europe and North America, in South Pacific between Australia and New Zealand, and Antarctica).

24

(a) Drifting buoy, R3.0.2 − R3.0.1

(b) Moored buoy, R3.0.2 − R3.0.1

(c) Ship, R3.0.2 − R3.0.1

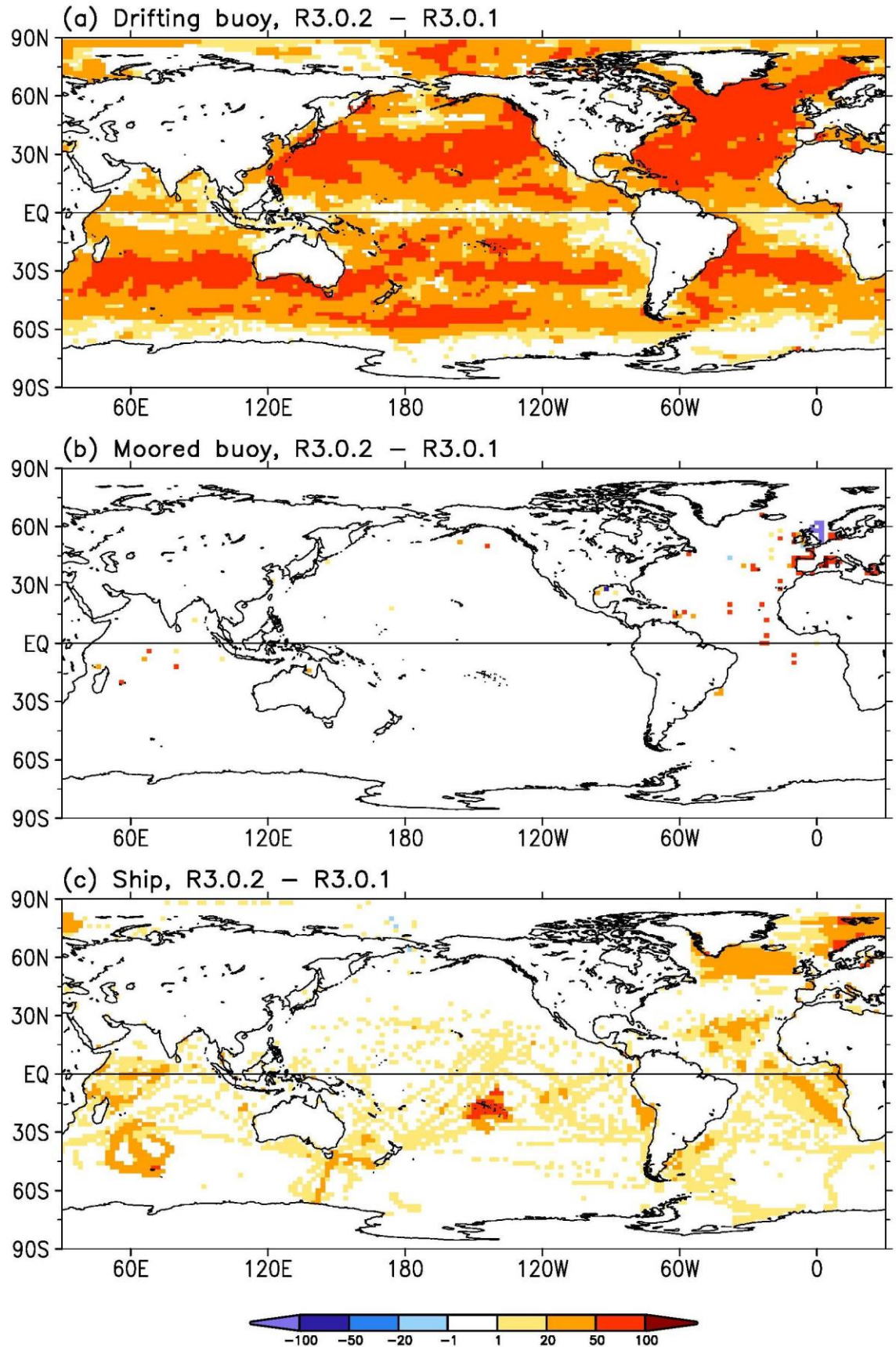−100 −50 −20 −1   1   20   50 100

25

Figure 10. Differences of observation coverage (%) between R3.0.2 and R3.0.1 in 2021 in platforms of (a) Drifting buoy, (b) Moored buoy, and (c) Ship. The observation coverage is calculated as the ratio of months with observations over the total 12 months in a year in a 2-degree grid box.

5. Data Access

ICOADS NRT R3.0.2 data includes the following:

• R3.0.2 Total files: Raw TAC+BUFR blended, duplicate status marked.

• R3.0.2 Final files: Processed TAC+BUFR merged data with duplicate reports removed. Land mask checked. R3.0.2 Final files are also available in NetCDF format.

• Monthly Summary Groups (MSG) statistic files, as described below.

Both the standard and enhanced trimming level data are available on demand from the National Center for Atmospheric Research (NCAR) ICOADS RDA dataset ds548.0 (Research Data Archive/Computational and Information Systems Laboratory/National Center for Atmospheric Research/University Corporation for Atmospheric Research, et al. 2016). Users can request data subsets and customize their own data QC trimming and filtering on ICOADS records.

The enhanced and standard trimming level data are used in turn to calculate the statistics for the monthly summaries on 2°×2° and 1°×1° grids. Ten statistical measures (such as the mean and standard deviation) are calculated for 22 observed and derived variables (Slutz *et al.* 1985). These variables and their statistics computed using the available data are shown in Table 2 and Table 3. These non-interpolated summary data provide a quick look of the key environmental variables and their statistical values.

ICOADS NRT R3.0.2 data access options have been updated on the ICOADS Data and Products page at https://icoads.noaa.gov/products.html, where several file format and data access options are available.

Observed Variables		
Variable	**Abbr.**	**Precision and Units**
Sea surface temperature	S	0.01 deg C
Air temperature	A	0.01 deg C
Scalar wind	W	0.01 m s^{-1}
Vector wind eastward component	U	0.01 m s^{-1}
Vector wind northward component	V	0.01 m s^{-1}
Sea level pressure *(SLP)*	P	0.01 hPa
Total cloudiness	C	0.1 okta
Specific humidity	Q	0.01 g kg^{-1}
Derived Variables		
Variable	**Abbr.**	**Units**
Relative humidity	R	0.1 percent
Sea-air temperature difference = *S - A*	D	0.01 deg C
Sensible heat parameter = *(S – A)W*	E	0.1 deg C m s^{-1}
(Saturation Q at S) minus Q = *(Q_s - Q)*	F	0.01 g kg^{-1}
Latent heat parameter = *(Q_s – Q)W*	G	0.1 g kg^{-1} m s^{-1}
U-wind stress = *UW*	X	0.1 m^2 s^2
V-wind Stress = *VW*	Y	0.1 m^2 s^2
Sensible heat transport, eastward **=** *UA*	I	0.1 deg C m s^{-1}
Sensible heat transport, northward = *VA*	J	0.1 deg C m s^{-1}
Latent heat transport, eastward = *UQ*	K	0.1 g kg^{-1} m s^{-1}
Latent heat transport, northward = *VQ*	L	0.1 g kg^{-1} m s^{-1}

Zonal latent heat parameter = $U(Q_s - Q)$	M	$0.1\ g\ kg^{-1}\ m\ s^{-1}$
Meridional Latent Heat Parameter = $V(Q_s - Q)$	N	$0.1\ g\ kg^{-1}\ m\ s^{-1}$
Scalar wind cubed (W^3), high-resolution	B1	$0.5\ m^3\ s^{-3}$
Scalar wind cubed (W^3), low-resolution	B2	$5\ m^3\ s^{-3}$

Table 2. Variables in the ICOADS NRT R3.0.2 Monthly Summary Group (MSG). For more information, see the ICOADS R2.5 General Information about Statistics documentation at https://icoads.noaa.gov/e-doc/R2.5-stat_doc.pdf.

Statistic	Abbr.	Units
First sextile (estimate of $m - 1s$)	s1	Table 2
Third sextile (median)	s3	Table 2
Fifth sextile (estimate of $m + 1s$)	s5	Table 2
Mean	m	Table 2
Number of observations	n	1
Standard deviation	s	Table 2
Mean day-of-month of observations	d	2 dy
Fraction of observations in daylight	h_t	0.1
Mean longitude (from lower-left [southwest] corner of box) of observations	x	0.2 deg (2 deg box size) 0.1 deg (1 deg box size)
Mean latitude (from lower-left [southwest] corner of box) of observations	y	0.2 deg (2 deg box size) 0.1 deg (1 deg box size)

Table 3. Statistics in Monthly Summary Group (MSG). Units refer to Table 2 if the units depend on a variable. For more information, see the ICOADS R2.5 General Information about Statistics documentation at https://icoads.noaa.gov/e-doc/R2.5-stat_doc.pdf.

6. Summary and Future Work

By combining new BUFR data with existing TAC data, an improved ICOADS NRT product has been developed and released as ICOADS NRT R3.0.2. The number of reports in ICOADS NRT R3.0.2 increased by about 30% as compared to R3.0.1. The spatial ocean coverage for essential climate variables (ECVs) increased by about 20% on monthly 2°×2° grids, and added a new variable for salinity. The new release will improve downstream products significantly, as demonstrated in OISST v2.1 (Huang et al. 2021). With enhanced ingest monitoring, monthly-generated plots to assess the health of the observing system to identify any abnormal platform type distributions, and daily use/monitoring by products such as OISST, NCEI is better equipped to handle future NRT format changes. NCEI will also be able to more speedily include new platform types from the BUFR format as they become available.

In the future, NRT GTS data streams from other GTS data collection centers (such as the Navy FNMOC and European Centre for Medium-Range Weather Forecasts [ECMWF]) will be utilized as well. Although all are collected from the same WMO GTS transmission system, collections from different centers are slightly different; merging them will increase the dataset completeness, and it is a good practice to monitor the input data streams for dropped or missing data and respond appropriately to the data providers about such issues.

We will include other observing platforms such as Argo floats, and the NDBC's Coastal Marine Automated Network (C-MAN) stations in NRT dataset at a future date. Inclusion of Argo floats will increase ocean coverage in the open oceans, outside of main shipping lanes. CMAN stations will enhance coastal data availability for both ocean and land data users.

The QC process can also be improved, by using new climatology data for the trimming, and including additional checks such as internal consistency check (e.g., track checks and spike checks), and mutual consistency check or buddy check (e.g., Lorenc and Hammon 1988). The merging and duplicate identification processes need to be modernized. In the current TAC and BUFR merge process, whole reports were either excluded or kept in the R3.0.2 Final files, resulting in the possible loss of any additional parameters contained in the excluded report. We will examine the possibility of blending near-duplicate reports in order to composite reports for improved quality and completeness of the data.

Acknowledgments

Data Availability Statement

ICOADS R3.0.2 data are available at

https://www.ncei.noaa.gov/data/marine/nrt/

https://www.ncei.noaa.gov/data/marine/msg/

https://rda.ucar.edu/datasets/ds548.0/

Appendix A. ICOADS IMMA1 variables and their keys in the BUFR reports.

Table A1. ICOADS IMMA1 variables and their keys in the BUFR reports. Note: For more detailed information on these parameters, see the IMMA1 format documentation at https://icoads.noaa.gov/e-doc/imma/R3.0-imma1.pdf.

Element Abbreviation	Element description	Key in BUFR
YR	year UTC	year

MO	month UTC	month
DY	day UTC	day
HR	hour UTC	time
LAT	latitude	latitude
LON	longitude	longitude
DS	ship course	directionOfMovingObservingPlatform
VS	ship speed	movingObservingPlatformSpeed
ID	identification/callsign	marineObservingPlatformIdentifier
D	wind direction (true)	windDirection
WI	wind speed indicator	instrumentationForWindMeasurement
W	wind speed	windSpeed
VV	visibility	horizontalVisibility
WW	present weather	presentWeather
W1	past weather	pastWeather
SLP	sea level pressure	pressureReducedToMeanSeaLevel
A	characteristic of PPP	characteristicOfPressureTendency
PPP	amt. pressure tend.	3HourPressureChange
AT	air temperature	airTemperature
WBTI	WBT indic.	methodOfWetBulbTemperatureMeasurement
WBT	wet-bulb temperature	wetBulbTemperature
DPT	dew-point temperature	dewPointTemperature
SI	SST meas. method	methodOfWaterTemperatureAndOrOrSalinityMeasurement
SST	sea surface temp.	oceanographicWaterTemperature(drifters and ships) seaSurfaceTemperature(moorings and float)
N	total cloud amount	cloudCoverTotal
NH	lower cloud amount	cloudAmount
CL	low cloud type	cloudType
H	cloud height	heightOfBaseOfCloud
CM	middle cloud type	cloudType

31

CH	high cloud type	cloudType	
WD	wave direction	wavesDirection	
WP	wave period	periodOfWaves	
WH	wave height	heightOfWaves	
SD	swell direction	swellWavesDirection	
SP	swell period	periodOfSwellWaves	
SH	swell height	heightOfSwellWaves	
PT	platform type	dataBuoyType Platform types for BUFR Key = dataBuoyType and their mapping to IMMA1 Platform Types. Blank means no direct mapping. <table><tr><td>Code Figure</td><td>Code description</td><td>IMMA1 PT Mapping</td></tr><tr><td>0</td><td>Unspecified drifting buoy</td><td>PT = 7</td></tr><tr><td>1</td><td>Standard Lagrangian drifter (Global Drifter Programme)</td><td>PT = 7</td></tr><tr><td>2</td><td>Standard FGGE type drifting buoy (non-Lagrangian meteorological drifting buoy)</td><td>PT = 7</td></tr><tr><td>3</td><td>Wind Measuring FGGE type drifting buoy (non-Lagrangian meteorological drifting buoy)</td><td>PT = 7</td></tr></table>	

32

		4	Ice drifter	PT = 8	
		5	SVPG Standard Lagrangian drifter with GPS	PT = 7	
		6	SVP-HR drifter with high-resolution temperature or thermistor string	PT = 7	
		7	Reserved		
		8	Unspecified subsurface float	PT = 18	
		9	SOFAR	PT = 18	
		10	ALACE	PT = 18	
		11	MARVOR	PT = 18	
		12	RAFOS	PT = 18	
		13	PROVOR	PT = 18	
		14	SOLO	PT = 18	
		15	APEX	PT = 18	

33

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0182.1.

Unauthenticated | Downloaded 11/10/22 04:35 PM UTC

		4	Ice drifter	PT = 8	
		5	SVPG Standard Lagrangian drifter with GPS	PT = 7	
		6	SVP-HR drifter with high-resolution temperature or thermistor string	PT = 7	
		7	Reserved		
		8	Unspecified subsurface float	PT = 18	
		9	SOFAR	PT = 18	
		10	ALACE	PT = 18	
		11	MARVOR	PT = 18	
		12	RAFOS	PT = 18	
		13	PROVOR	PT = 18	
		14	SOLO	PT = 18	
		15	APEX	PT = 18	

33

		16	Unspecified moored buoy	PT = 6
		17	Nomad	PT = 6
		18	3-metre discus	PT = 6
		19	10-12-metre discus	PT = 6
		20	ODAS 30 series	PT = 6
		21	ATLAS (e.g., TAO area)	PT = 6
		22	TRITON buoy	PT = 6
		23	FLEX mooring (e.g., TIP area)	PT = 6
		24	Omnidirectional waverider	PT = 6
		25	Directional waverider	PT = 6
		26	Subsurface ARGO float	PT = 18
		27	PALACE	PT =18

34

		28	NEMO	PT =18
		29	NINJA	PT =18
		30	Ice buoy/float (POPS or ITP)	PT=8
		31–33	Reserved	
		34	Mooring oceanographic	PT = 6
		35	Mooring meteorological	PT = 6
		36	Mooring multidisciplinary (OceanSITES)	PT = 6
		37	Mooring tide gauge or tsunami buoy	PT = 6
		38	Ice beacon	PT = 8
		39	Ice mass balance buoy	PT = 8
		40–62	Reserved	
		63	Missing value	
DOS	depth of SST meas.	depthBelowWaterSurface		

35

HOP	height of visual observation platform	heightOfStationGroundAboveMeanSeaLevel
HOT	height of AT sensor	heightOfSensorAboveWaterSurface
HOB	height of barometer	heightOfBarometerAboveMeanSeaLevel
HOA	height of anemometer	heightOfSensorAboveLocalGroundOrDeckOfMarinePlatform
OTV	temperature value	oceanographicWaterTemperature
OTZ	temperature depth	depthBelowWaterSurface
OSV	salinity value	seaSurfaceSalinity
OSZ	salinity depth	depthBelowWaterSurface
W2	2nd past weather	pastWeather2
SD2	dir. of second. swell	swellWavesDirection
SP2	per. of second. swell	periodOfSwellWaves
SH2	ht. of second. swell	heightOfSwellWaves
IS	ice accretion on ship	causeOfIceAccretion
ES	thickness of IS	iceDepositThickness
RS	rate of IS	rateOfIceAccretionEstimated
IC1	concentration of sea ice	seaIceConcentration
IC2	stage of development	iceDevelopment
IC3	ice of land origin	amountAndTypeOfIce
IC4	true bearing ice edge	iceEdgeBearing
IC5	ice situation/trend	iceSituation
RRR	amount of precip.	totalPrecipitationOrTotalWaterEquivalent
TR	duration of per. RRR	timePeriod
RH	relative humidity	relativeHumidity
DOS	depth of SST meas.	depthBelowWaterSurface
HOP	height of visual observation platform	heightOfStationGroundAboveMeanSeaLevel
HOT	height of AT sensor	heightOfSensorAboveWaterSurface
HOB	height of barometer	heightOfBarometerAboveMeanSeaLevel

HOA	height of anemometer	heightOfSensorAboveLocalGroundOrDeckOfMarinePlatform
OTV	temperature value	oceanographicWaterTemperature
OTZ	temperature depth	depthBelowWaterSurface
OSV	salinity value	seaSurfaceSalinity
OSZ	salinity depth	depthBelowWaterSurface

APPENDIX B. Selected variables and their possible flag values.

Table A2. Selected variables and their possible flag values. Modified from ICOADS Release 1, Supplement J. See Appendix C for definitions of the QC flag values. For more information, see the ICOADS Release 3.0 Quality Control and Related Processing document at https://icoads.noaa.gov/e-doc/R3.0-stat_trim.pdf.

	possible flag values (x)									
	R	A	B	J	K	L	M	N	Q	S
ship position	X						X			
wind	X	X					X		X	X
visibility	X						X			X
present weather	X		X	X		X	X			X
past weather	X			X			X			X
pressure	X					X	X		X	X
air temp.	X			X		X	X	X	X	X
wet bulb temp.	X		X			X	X	X	X	X
dew point temp.	X		X			X	X	X	X	X
sea surface temp.	X					X	X		X	X
cloud	X		X	X				X		X
wave	X	X	X	X			X	X	X	X
swell	X		X	X			X	X	X	X
pressure tendency	X				X		X			X

APPENDIX C. QC Flag Meaning, value, and weight.

Table A3. QC Flag Meaning, value, and weight. Modified from ICOADS Release 1, Supplement J. For more information, see the ICOADS Release 3.0 Quality Control and Related Processing document at https://icoads.noaa.gov/e-doc/R3.0-stat_trim.pdf.

Value	Coded	Weight	Meaning	Reason
R	1	0	correct	
A	2	1	correctable	legality
B	3	1	correctable	internal consistency
J	4	2	suspect	internal consistency
K	5	2	suspect	time
L	6	2	suspect	extreme (outside $\bar{x} \pm 4.8\,\sigma$)
M	7	3	erroneous	legality
N	8	3	erroneous	internal consistency
Q	9	3	erroneous	extreme (outside $\bar{x} \pm 5.8\,\sigma$)
S	10	3	missing	

References

Banzon, V., T. M. Smith, T. M. Chin, C. Liu, C., and W. Hankins, 2016: A long-term record of blended satellite and in situ sea-surface temperature for climate monitoring, modeling and environmental studies. *Earth Syst. Sci. Data*, **8**, 165–176, doi:10.5194/essd-8-165-2016.

Berry, D. I., and E. C. Kent, 2011: Air-Sea fluxes from ICOADS: the construction of a new gridded dataset with uncertainty estimates. *Inter. J. Climate.* **31**, 987–1001. doi: 10.1002/joc.2059

Boyer, T. P., O. K. Baranova, C. Coleman, H. E. Garcia, R. A. Grodsky, A. V. Mishonov, et al., 2018: *World Ocean Database 2018*. Technical report, NOAA Atlas NESDIS **87**

Cornes, R. C., E. C. Kent, D. I. Berry, and J. J. Kennedy, 2020: CLASSnmat: A global night marine air temperature data set, 1880–2019. *Geoscience Data Journal*, **7**, 2,170-184. https://doi.org/10.1002/gdj3.100.

Freeman, E., E. C. Kent, P. Brohan, T. Cram, L. Gates, B. Huang, C. Liu, S. R. Smith, S. J. Worley, and H.-M. Zhang, 2019: The International Comprehensive Ocean-Atmosphere Data Set – Meeting Users Needs and Future Priorities. Front. Mar. Sci. **6**:435. doi: 10.3389/fmars.2019.00435.

Freeman, E., S. D. Woodruff, S. J. Worley, S. J. Lubker, E. C. Kent, W. E. Angel, et al., 2017: ICOADS Release 3.0: a major update to the historical marine climate record. *Int. J. Climatol.* 37, 2211–2232. doi: 10.1002/joc.4775

Hausfather, Z., K. Cowtan, D. C. Clarke, P. Jacobs, M. Richardson, and R. Rohde, 2017: Assessing recent warming using instrumentally homogeneous sea surface temperature records. *Science Adv.* **3**: e1601207. doi: 10.1126/sciadv.1601207

Hirahara, S., Ishii, M., and Y. Fukuda, 2014: Centennial-scale sea surface temperature analysis and its uncertainty. *J. Climate*, **27**, 57-75.

Hersbach, H, and Coauthors, 2020: The ERA5 global reanalysis, Q J R Meteorol Soc., **146**, 730, 1999-2049 https://doi.org/10.1002/qj.3803

Huang, B., V. Banzon, P. W. Thorne, T. Boyer, G. Chepurin, J. H. Lawrimore, M. J. Menne, T. M. Smith, R. S. Vose, and H. Zhang, 2017: Extended Reconstructed Sea Surface Temperature, Version 5 (ERSSTv5): Upgrades, Validations, and Intercomparisons. *J. Climate*, **30**, 8179–8205, https://doi.org/10.1175/JCLI-D-16-0836.1.

Huang, B., V. F. Banzon, E. Freeman, J. Lawrimore, W. Liu, T. C. Peterson, T. M. Smith, P. W.

Thorne, S. D. Woodruff, and H.-M. Zhang, 2015: Extended Reconstructed Sea Surface Temperature version 4 (ERSST.v4), Part I. Upgrades and intercomparisons. J. Climate, **28**, 911-1092 930, doi:10.1175/JCLI-D-14-00006.1.

Huang, B., C. Liu, G. Ren, H.-M. Zhang, and L. Zhang, 2019: The role of buoy and Argo observations in two SST analyses in the global and tropical Pacific oceans. *J. Climate,* **32**, 2517-2535, doi:10.1175/JCLI-D-18-0368.1.

Huang, B., C. Liu, V. Banzon, E. Freeman, G. Graham, B. Hankins, T. Smith, and H.-M. Zhang, 2020: Improvements of the Daily Optimum Interpolation Sea Surface Temperature (DOISST) Version 2.1. *J. Climate*, **34**, 2923-2939, DOI 10.1175/JCLI-D-20-0166.1.

International Comprehensive Ocean-Atmosphere (ICOADS) Global Partners, and Physical Sciences Laboratory/Earth System Research Laboratory/OAR/NOAA/U.S. Department of Commerce. 2002. *ICOADS Input Data Sources*. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. https://doi.org/10.5065/549Q-3G69.

IPCC, 2021: Annex I: Observational Products [Trewin, B. (ed.)]. In: Climate Change 2021: The Physical 21 Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental 22 Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. 23 Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T. K. 24 Maycock, T. Waterfield, O. Yelekçi, R. Yu and B. Zhou (eds.)]. Cambridge University Press. In Press

Jundon, R. A., J. R. Christy, 2019: A new compilation of globally gridded night-time marine air temperatures: The UAHNMATv1 dataset. Int. J. Climatol. **40**, 2609-2623. https://doi.org/10.1002/joc.6354

Kennedy, J. J., N. A. Rayner, C. P. Atkinson, and E. E. Killick, 2019: An ensemble data set of sea- surface temperature change from 1850: the Met Office Hadley Centre HadSST.4.0.0.0 data set. Journal of Geophysical Research: Atmospheres, 124. https://doi.org/10.1029/2018JD029867.

Kent, E. C., N. A. Rayner, D. I. Berry, R. Eastman, V. G. Grigorieva, B. Huang, J. J. Kennedy, S. R. Smith, K. M. Willett, 2019: Observing Requirements for Long-Term Climate Records at the Ocean Surface. Front. Mar. Sci. **6**:441. https://doi.org/10.3389/fmars.2019.00441.

Kent, E. C., and Coauthors, 2017: A Call for New Approaches to Quantifying Biases in

Observations of Sea Surface Temperature. *Bull. Amer. Meteor. Soc.*, **98**, 1601–1616, https://doi.org/10.1175/BAMS-D-15-00251.1.

Kent, E.C., N. A. Rayner, D. I. Berry, M. Saunby, B. I. Moat, J. J. Kennedy, et al., 2013: Global analysis of night marine air temperature and its uncertainty since 1880: The HadNMAT2 data set. *J. Geophys. Res. - Atmospheres*, **118**(3), 1281–1298. https://doi.org/10.1002/jgrd.50152

Lorenc, A. C., and O. Hammon, 1988: Objective quality control of observations using Bayesian methods: Theory, and a practical implementation. Quart. J. Roy. Meteor. Soc., 114, 515–543.

Morice, C. P., J. J. Kennedy, N. A. Rayner, J. P. Winn, E. Hogan, R. E. Killick, R. J. H. Dunn, T. J. Osborn, P. D. Jones, and I. R. Simpson, 2021: An Updated Assessment of Near-Surface Temperature Change From 1850: The HadCRUT5 Data Set. Journal of Geophysical Research (Atmospheres), **126 (3)**. https://doi.org/10.1029/2019JD032361

Pelletier, Y., 2008: A Primer on Writing BUFR templates, Version 1.4, 2008, 25pp., National Prediction Operations Division, Meteorological Service of Canada, https://www.eumetnet.eu/wp-content/ uploads/2017/04/OPERA_BUFR_template_primer_V1.4.pdf

Reynolds, R. W., T. M. Smith, C. Liu, D. B. Chelton, K. S. Casey, and M. G. Schlax, 2007: Daily high-resolution-blended analyses for sea surface temperature. *Journal of Climate*, **20**, 5473–5496, doi:10.1175/JCLI-D-14-00293.1.

Rayner, N. A.; D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century J. Geophys. Res., **108**, No. D14, 4407 10.1029/2002JD002670 (pdf ~9Mb).

Research Data Archive/Computational and Information Systems Laboratory/National Center for Atmospheric Research/University Corporation for Atmospheric Research, Physical Sciences Laboratory/Earth System Research Laboratory/OAR/NOAA/U.S. Department of Commerce, Cooperative Institute for Research in Environmental Sciences/University of Colorado, National Oceanography Centre/University of Southampton, Met Office/Ministry of Defence/United Kingdom, Deutscher Wetterdienst (German Meteorological Service)/Germany, Department of Atmospheric Science/University of Washington, Center for Ocean-Atmospheric Prediction Studies/Florida State University, and National Centers for Environmental Information/NESDIS/NOAA/U.S. Department of Commerce. 2017, updated

monthly. *International Comprehensive Ocean-Atmosphere Data Set (ICOADS) Release 3, Individual Observations*. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. https://doi.org/10.5065/D6ZS2TR3.

Slivinski, L. C., and Coauthors, 2019, Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system. Q J R Meteorol Soc., **145**, 724, doi:10.1002/qj.3598.

Slutz, R. J., S. J. Lubker, J. D. Hiscox., S. D. Woodruff, R. L. Jenne, R. H. Joseph, P. M. Steurer, J. D. Elms, 1985: Comprehensive Ocean-Atmosphere Data Set: Release 1. NTIS PB86-105723, NOAA Environmental Research Laboratories, Climate Research Program, Boulder, CO, 268 pp.

Smith, S. R., R. Richardson, J. P. Stow, M. A.Bourassa, H. McMillan, 2021: A New Technique for Century-Scale Wind Component Indices. Front. Mar. Sci. **9**:661473. https://doi.org/10.3389/feart.2021.661473

Smith SR, E. Freeman, S. J. Lubker, S. D. Woodruff, S. J. Worley, W. E. Angel, D. I. Berry, Brohan P, Z. Ji, and E. C. Kent, 2016. The International Maritime Meteorological Archive (IMMA) Format. http://icoads.noaa.gov/e-doc/imma/R3.0-imma1.pdf (accessed 31 May 2016).

Vose, R.S., D. Arndt, V. F. Banzon, D. R. Easterling, B. Gleason, B. Huang, E. Kearns, J. H. Lawrimore, M. J. Menne, T.C. Peterson, R.W. Reynolds, T. M. Smith, C. N. Williams, Jr., and D. L. Wuertz, 2012: NOAA's merged land-ocean surface temperature analysis. Bulletin of the American Meteorological Society, **93**, 1677–1685, doi:10.1175/BAMS-D-11-00241.1

WMO, 2019: Manual on codes. WMO-No. 306, World Meteorological Organization, Geneva, Switzerland. https://community.wmo.int/activity-areas/wmo-codes/manual-codes. (Accessed 1 May 2020).

Wolter, K., 1997: Trimming problems and remedies in COADS. J. Climate, **10**, 1980-1997.

Woodruff, S. D., R. J. Slutz, R. L. Jenne, P. M. Steurer, 1987: A comprehensive ocean-atmosphere dataset. *Bull. Am. Meteorol. Soc.* **68**: 1239–1250, doi: 10.1175/1520-0477(1987)068<1239: ACOADS>2.0.CO;2.

Woodruff, S. D., S. J. Lubker, K. Wolter, S. J. Worley, J. D. Elms, 1993: Comprehensive

ocean–atmosphere data set (COADS) Release 1a: 1980–92. *Earth Syst. Monit.* **4**(1): 1–8.

Woodruff, S. D., H. F. Diaz, S.J. Worley, R. W. Reynolds, S. J. Lubker, 2005: Early ship observational data and ICOADS. *Clim. Change* **73**, 169–194, doi:10.1007/s10584-005-34563.

Woodruff, S., S. Worley, S. Lubker, Z. Ji, E. Freeman, D. Berry, P. Brohan, E. Kent, D. Reynolds, S. Smith, C. Wilkinson, 2011: ICOADS Release 2.5: extensions and enhancements to the surface marine meteorological archive. *Int. J. Climatol.* **31**: 951–967, doi: 10.1002/ joc.2103.

Zhang, H. -M., Jay H. Lawrimore, Boyin Huang, Matthew J. Menne, Xungang Yin, Ahira Sánchez-Lugo, Byron E. Gleason, Russell Vose, Derek Arndt, J. Jared Rennie, Claude N. Williams, 2019: Updated temperature data give a sharper view of climate trends, Eos, **100**, doi.org/10.1029/2019EO128229. Published on 19 July 2019.