International Comprehensive Ocean-Atmosphere Data Set (ICOADS)	**Release 3.0**
General Information about Statistics	14 July 2016
(DRAFT) Document name:	*R3.0-stat_doc*

This document gives a general description of the sets of monthly summary statistics available from ICOADS Release 3.0 (R3.0) (Freeman et al., 2016), based on R3.0 individual observations in the International Maritime Meteorological Archive (IMMA) format (see *R3.0-imma1*). The summaries were calculated for each year and month over the period 1800-2014 using 2° latitude × 2° longitude boxes. Coverage was extremely sparse, and thus statistics were not calculated, prior to 1800. In contrast, starting in 1960 when spatial coverage improves, they were also calculated using 1°×1° boxes. Twenty-two observed and derived variables (Table 1) were summarized for each year-month. For each variable, 10 statistics (Table 2) were calculated.

Following the official end date of R3.0 (December 2014), the IMMA individual observations and summary products are extended in near-real-time based on Global Telecommunication System (GTS) receipts. These "preliminary" (note: subject to change) data and monthly summary products for 2015-forward are updated monthly (i.e., lagging the data month by up to 5 days).

Table 1. Variables in Monthly Summary Group (MSG) format. Each variable is abbreviated with an upper-case letter (followed "1" or "2" for variable W^3, which is available at two resolutions). "Units" gives the smallest increment of the data that has been encoded.

No.	*Abbr.*	*Variable*	*Units*
		"Observed"	
1	*S*	sea surface temperature	0.01 °C
2	*A*	air temperature	0.01 °C
3	*W*	scalar wind	0.01 m/s
4	*U*	vector wind eastward* component	0.01 m/s
5	*V*	vector wind northward* component	0.01 m/s
6	*P*	sea level pressure (SLP)	0.01 hPa
7	*C*	total cloudiness	0.1 okta
8	*Q*	specific humidity	0.01 g/kg
		Derived	
9	*R*	relative humidity	0.1 %
10	*D*	$S - A$ = sea-air temperature difference	0.01 °C
11	*E*	$(S - A)W$	0.1 °C m/s
12	*F*	$Q_S - Q$ = (saturation Q at S) – Q	0.01 g/kg
13	*G*	$FW = (Q_S - Q)W$ = (evaporation parameter)	0.1 g/kg m/s
14	*X*	*WU* (wind stress	0.1 m^2/s^2
15	*Y*	*WV* parameters)	0.1 m^2/s^2
16	*I*	*UA* (sensible-heat—transport	0.1 °C m/s
17	*J*	*VA* parameters)	0.1 °C m/s
18	*K*	*UQ* (latent-heat—transport	0.1 g/kg m/s
19	*L*	*VQ* parameters)	0.1 g/kg m/s
20	*M*	*FU*	0.1 g/kg m/s
21	*N*	*FV*	0.1 g/kg m/s
22	*B1*	$B = W^3$ (high-resolution)	0.5 m^3/s^3
23	*B2*	$B = W^3$ (low-resolution)	5 m^3/s^3

* Using the oceanographic convention, which is <u>not</u> the westerly and southerly positive (i.e.,

blowing from) convention used in meteorology. E.g., a positive *U* vector wind eastward component is wind blowing toward the east. This also follows the mathematical vector reference framework.

_____

Table 2. Statistics available in Monthly Summary Group (MSG) format. The statistics are abbreviated by lower-case letters, or a lower-case letter followed by a digit. "Units" (with reference to Table 1, if the units depend on the variable) gives the smallest increment of the data that has been encoded.

Abbr.	*Statistic*	*Units*
s1	1/6 sextile (estimate of *m* – 1*s*)	Table 1
s3	3/6 sextile (the median)	Table 1
s5	5/6 sextile (estimate of *m* + 1*s*)	Table 1
m	mean	Table 1
n	number of observations	1
s	standard deviation*	Table 1
d	mean day-of-month of observations	2 dy
h_t	fraction of obs. in daylight	0.1
x	mean longitude of observations: 2° box size	0.2°
	1° box size	0.1°
y	mean latitude of observations: 2° box size	0.2°
	1° box size	0.1°

* Note: a standard deviation estimate is computable from the sextiles, defined as $e = (s5 - s1)/2$.

_____

The statistics were broken up into "groups" each containing four variables, in Monthly Summary Group (MSG) format (version 1). Each group (numbered 3-7, plus 9) contains the 10 statistics for each of the four variables (Table 3).

Table 3. Variables included in each trimmed (3-7, 9) group in the MSG format (group 8 is undefined for MSG and refers to the previously available MSTG format). Each group contains four variables (from Table 1) and ten statistics (Table 2) for each variable. For technical details for the MSG format, see *msg*, and for quality control ("trimming") and data selection rules used to construct the products, please refer to *R3.0-stat_trim*.

3	*4*	*5*	*6*	*7*	*9*
Sea surface temperature (*S*)	Scalar wind (*W*)	Total cloudiness (*C*)	$D=S-A$	$I=UA$	$M=FU$
Air temperature (*A*)	Wind U-component (*U*)	*R*	$E=(S-A)W$	$J=VA$	$N=FV$
Specific humidity (*Q*)	Wind V-component (*V*)	$X=WU$	$F=Q_S-Q$	$K=UQ$	$B1=(W^3)$*
Relative humidity (*R*)	SLP (*P*)	$Y=WV$	$G=FW$	$L=VQ$	$B2=(W^3)$*

* *B1* and *B2* are high- and low-resolution representations of $B = W^3$ (i.e., data increments of 0.5 or 5 m^3/s^3).

_____

Product, domain, and group abbreviations

The following abbreviations are combined (together with the inclusive period of record

covered by a given file) to create Unix tar file names (e.g., *MSG2_R3.0.0_STD_G3*, *MSG1_R3.0.0_ENH_G3*, and *MSG1_R3.0.0_ENH_G3_EQ*):

MSGn: Monthly Summary Groups

Six groups (numbered 3-7, plus 9), each containing four variables × 10 statistics (Table 3), where *n* = 1 for 1°, or *n* = 2 for 2°, latitude/longitude boxes.

R3.0.0: Current ICOADS Release number

ENH: Enhanced statistics

The 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.

STD: Standard statistics

The standard statistics were derived using 3.5σ trimming limits and were limited as nearly as possible to observations taken from ships (where identifiable).

Gn: Group numbering

Gn where *n* = 3, 4, 5, 6, 7 or 9 identifies the MSG Group number (see Table 3).

EQ: Equatorial domain

Domain limits: equatorial latitude band 10.5°N to 10.5°S; global with respect to longitude. These 1° boxes were shifted half a degree in latitude (only) in comparison to the global domain, such that the center-latitude of the central row of boxes is the equator (e.g., 0°–1°E, 0.5°S–0.5°N).

For the global domain, in contrast, the boundaries of 1° boxes fall on units of latitude and longitude (e.g., 0°–1°E, 89°–90°N). All the 2° products also utilize a global domain, with box boundaries falling on even degrees of latitude and longitude (e.g., 0°–2°E, 88°–90°N).

Temporal coverage of MSG products

MSG2_R3.0.0_ENH_Gn [1800----------2014]
MSG2_R3.0.0_STD_Gn [1800----------2014]

MSG1_R3.0.0_ENH_Gn [1960----------2014]
MSG1_R3.0.0_STD_Gn [1960----------2014]
MSG1_R3.0.0_ENH_Gn_EQ [1960----------2014]
MSG1_R3.0.0_STD_Gn_EQ [1960----------2014]

Frequently asked questions about the statistics

1) Are the summaries interpolated or analyzed?
No. No interpolation or extrapolation is done to fill in temporal or spatial data gaps. Only boxes that have observations will have data records. No analysis is done to smooth or modify the statistics.

2) Are the summaries adjusted for changes in observing practices?
No. The available summaries incorporate no adjustments for changes in observing practices, such as changes from bucket to engine room intake measurements of sea surface temperature, or changes from estimated (Beaufort) to anemometer measured wind speeds.

3) How were the statistics (e.g., mean, median, sextiles) calculated?
Refer to *Release 1* (Slutz et al. 1985), supp. A, and to *msg*. If there is a single observation in a box, that value is used to calculate the statistics (e.g., the mean is that value divided by one).

4) For a given variable, is the number of observations as provided in MSG format used to calculate all the statistics?
Yes, for all statistics except the mean day-of-month and the fraction of observations in daylight. For those two statistics the numbers of observations may be fewer (and are not stored in MSG format). This is because day and hour may be missing in IMMA (but latitude, longitude, year, and month may not). However, any differences are generally minor since most data sources contain extant day and hour, with the exception of significant amounts of (e.g., US Maury Collection) data prior to about 1861.

5) Why is the lower-left (SW) corner of the 2° or 1° box included in the MSG format (BLO and BLA), rather than the box center?
The box corner should NOT be interpreted as a grid-point center from an analyzed dataset (ref. question 1). The weighted mean position of all observations within a box can be obtained from the corner coordinates plus the longitude and latitude offset statistics (*x* and *y*) for each variable. In a well and uniformly sampled box the weighted mean position of all observations will be the box center. However, in a sparsely sampled box or box that overlaps land masses the mean sample position may not be the box center.

6) How is the mean sample position within a box obtained?
By: mean longitude = *BLO* + *x*, and mean latitude = *BLA* + *y*. Note that longitude is always measured in East coordinates. Because the *x* values can range up to two degrees (one degree) for 2° (1°) boxes, the resultant range of mean longitude is 0°-360°E.

7) How were the derived variables computed?
The derived variables were computed using individual observations of the constituent variables exactly as shown in Table 1. For example, the mean of *D* results from calculating (*S* – *A*) for each individual (e.g., ship or buoy) report, and then averaging those values. Refer to *Release 1*, supp. A for computational details, including references to the formulae (available in *profs_short* software) used to calculate specific and relative humidities.

8) How were the pseudo-fluxes calculated?

We have chosen not to apply any constants or adjustments to variables when calculating heat and momentum flux parameters. This is because the constants for the bulk formulae and adjustments for atmospheric stability of the boundary layer are not universally agreed upon. Therefore, the heat and momentum flux parameters within ICOADS are actually psuedo-fluxes. Thus we do not define or use factors such as density, specific heat of air, etc.

9) If there are 22 variables, why does Table 1 list 23?

A large numerical range is possible for wind speed cubed ($B = W^3$). Owing to MSG format limitations, it was necessary to separate this variable into high- and low-resolution representations, i.e., data increments of 0.5 or 5 m^3/s^3 for *B1* and *B2*. *B2* can store the full data range, but *B1* offers better resolution. Statistics (*s1*, *s3*, *s5*, *m*, and *s*) for the low-resolution field *B2* are always available, if there were data. But the *B1* statistics provide better data resolution, unless they are missing because a given statistic can only be represented in *B2*.

References

Freeman, E., S.D. Woodruff, S.J. Worley, S.J. Lubker, E.C. Kent, W.E. Angel, D.I . Berry, P. Brohan, R. Eastman, L. Gates, W. Gloeden, Z. Ji, J. Lawrimore, N.A. Rayner, G. Rosenhagen, and S.R. Smith, 2016: ICOADS Release 3.0: A major update to the historical marine climate record. *Int. J. Climatol.* (doi:10.1002/joc.4775).

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

Document Revision Information

Previous document version: Derived from *R2.5-stat_doc* of 13 February 2012, with various updates to describe the R3.0 products, add the Freeman et al. 2016 reference, etc.