Comprehensive Ocean-Atmosphere Data Set; Release 1
Supplement B: 2 ° Monthly Summary Groups
Formats: MSUG.1, MSTG.1

0. Introduction

The seven group files are relatively compact alternatives to the full Monthly Summary Trimmed and Untrimmed (MST or MSU) formats, intended for studies using only a few variables or statistics. Eight important statistics for each of four related variables are grouped together in each file using a packed binary format. Thus five files are needed to represent all 19 MST variables, and two files are needed to represent all eight MSU variables. The statistics were chosen to bring together information that can be used to analyze the variability of the data and inhomogeneities of their distribution in time and space.

Cross reference is made to supp. A for standardized unpacking information, and the same notation for variables and statistics is followed or extended, also using the same type of two-dimensional table presentation.

1. Monthly Summary Trimmed Groups (MSTG)

The five trimmed groups were derived from MST (described in supp. A). Each MST was split into five MSTG records; these were written out onto the five separate group files even if every $\alpha\beta$ (the value of the statistic α for the variable β) was missing. Thus the record structure is identical for all the groups and their parent MST file. The five trimmed groups are numbered 3-7 to distinguish them from the untrimmed groups (numbered 1-2). Groups 3-7 contain four variables each: 3 = (S, A, Q, R), 4 = (W, U, V, P), 5 = (C, R, X, Y), 6 = (D, E, F, G), and 7 = (I, J, K, L). Table B1-1a shows the bit layout in common to any MSTG, and Tables B1-1b through B1-1f show the bit layout of each of the 64-bit or 16-bit sections of groups 3 through 7, respectively, in sequential bit-order reading from top to bottom. An example showing the bit-order is given following Table B1-1f.

Table B1-1a
MSTG.1

#	α	Statistic	Bits
		rptin	16
		year	8
		month	4
		2° box	14
		10° box	10
		identification checksum	12
11	3	3/6 sextile (the median)	64
6	m	mean	64
5	n	number of observations	64
15	e	standard deviation estimate	64
1	d	mean day-of-month of observations	16
2	h_t	fraction of observations in daylight	16
3	x	mean longitude of observations	16
4	y	mean latitude of observations	16
		total	384

Table B1-1b
Group 3
64-bit or 16-bit Sections

#	β	Variable	Bits	Bits
1	S	sea surface temperature	16	4
2	A	air temperature	16	4
8	Q	specific humidity	16	4
9	R	relative humidity	16	4
		total	64	16

Table B1-1c
Group 4
64-bit or 16-bit Sections

#	β	Variable	Bits	Bits
3	W	scalar wind	16	4
4	U	vector wind eastward component	16	4
5	V	vector wind northward component	16	4
6	P	sea level pressure	16	4
		total	64	16

Table B1-1d
Group 5
64-bit or 16-bit Sections

#	β	Variable	Bits	Bits
7	C	total cloudiness	16	4
9	R	relative humidity	16	4
14	X	WU	16	4
15	Y	WV (14-15 are wind stress parameters)	16	4
		total	64	16

Table B1-1e
Group 6
64-bit or 16-bit Sections

#	β	Variable	Bits	Bits
10	D	$S - A$ = sea-air temperature difference	16	4
11	E	$(S - A)W$ = sea-air temperature difference*wind magnitude	16	4
12	F	$Q_s - Q$ = (saturation Q at S) $- Q$	16	4
13	G	$FW = (Q_s - Q)W$ (evaporation parameter)	16	4
		total	64	16

Table B1-1f
Group 7
64-bit or 16-bit Sections

#	β	Variable	Bits	Bits
16	I	UA	16	4
17	J	VA	16	4
18	K	UQ	16	4
19	L	VQ (16-19 are sensible and latent heat transport parameters)	16	4
		total	64	16

For example, group 3 contains, in order: rptin, year, month, 2° box, 10° box, and identification checksum, followed by

$$((\alpha\beta,\ \beta = S,A,Q,R),\ \alpha = 3,m,n,e,d,h,x,y)$$

which defines the following matrix, with 4 rows and 8 columns:

	α	3	m	n	e	d	h	x	y
β	#	11	6	5	15	1	2	3	4
S	1	$3S$	mS	nS	eS	dS	hS	xS	yS
A	2	$3A$	mA	nA	eA	dA	hA	xA	yA
Q	8	$3Q$	mQ	nQ	eQ	dQ	hQ	xQ	yQ
R	9	$3R$	mR	nR	eR	dR	hR	xR	yR

stored in the order:

column 1, row 1,..., row 4; column 2, row 1,..., row 4;...; column 8, row 1,..., row 4.

2. Monthly Summary Untrimmed Groups (MSUG)

The two untrimmed groups were derived from MSU (described in supp. A). Each MSU was split into two MSUG records; these were written out onto the two separate group files even if every $\alpha\beta$ (the value of the statistic α for the variable β) was missing. Thus the record structure is identical for all the groups and their parent MSU file. Groups 1-2 contain four variables each: 1 = (S, A, P, Q), 2 = (W, U, V, C). Table B2-1a shows the bit layout in common to any MSUG, and Tables B2-1b and B2-1c show the bit layout of each of the 64-bit or 16-bit sections of group 1 and group 2, respectively, in sequential bit-order reading from top to bottom.

Table B2-1a
MSUG.1

#	α	Statistic	Bits
		rptin	16
		year	8
		month	4
		2° box	14
		10° box	10
		identification checksum	12
11	3	3/6 sextile (the median)	64
6	m	mean	64
5	n	number of observations	64
15	e	standard deviation estimate	64
1	d	mean day-of-month of observations	16
2	h_u	mean hour of observations	16
3	x	mean longitude of observations	16
4	y	mean latitude of observations	16
		total	384

Table B2-1b
Group 1
64-bit or 16-bit Sections

#	β	Variable	Bits	Bits
1	S	sea surface temperature	16	4
2	A	air temperature	16	4
6	P	sea level pressure	16	4
8	Q	specific humidity	16	4
		total	64	16

Table B2-1c
Group 2
64-bit or 16-bit Sections

#	β	Variable	Bits	Bits
3	W	scalar wind	16	4
4	U	vector wind eastward component	16	4
5	V	vector wind northward component	16	4
7	C	total cloudiness	16	4
		total	64	16

3. Reconstruction of Floating Point Data

The *coded* and *true value* ranges, the *units*, and the *base* for the fields that are unique in representation to the groups are given in Table B3-1.

Table B3-1
Unpacking Groups

#	α	Statistic	True value	Units	Base	Coded
1	d	mean day-of-month of observations	2⩽ 30	2 days	0.0	1⩽ 15
2	h_t	fraction of observations in daylight	0.0⩽ 1.0	0.1	-1	1⩽ 11
2	h_u	mean hour of observations	1⩽ 23	2 hours	-0.5	1⩽ 12
3	x	mean longitude of observations	0.1⩽ 1.9	0.2 °	-0.5	1⩽ 10
4	y	mean latitude of observations	0.1⩽ 1.9	0.2 °	-0.5	1⩽ 10

Further descriptions of the fields in Table B3-1 follow. All other fields are common to the MST or MSU, with characteristics as given in sec. 2.3 of supp. A, except that some fields have different names and other differences as noted.

- identification checksum

The group number, 1 for (S, A, P, Q), 2 for (W, U, V, C), 3 for (S, A, Q, R), 4 for (W, U, V, P), 5 for (C, R, X, Y), 6 for (D, E, F, G), or 7 for (I, J, K, L), must be added into the usual checksum prior to the modulo for proper identification. For example, supposing that the *coded* group 3 data matrix is available in an array MSTG, the checksum CK is computed and verified against the stored checksum CKS in FORTRAN as follows:

```
          INTEGER CK,J,I,MSTG(4,8),YEAR,MONTH,BOX2,BOX10,CKS
          CK = 0
          DO 500 J = 1,8
               DO 400 I = 1,4
                    CK = CK + MSTG(I,J)
400            CONTINUE
500       CONTINUE
          CK = CK + YEAR + MONTH + BOX2 + BOX10 + 3
          CK = MOD(CK,4095)
          IF(CK .NE. CKS) THEN
               PRINT *,'ERROR. CK = ',CK,' .NE. CKS = ',CKS
               STOP
          ENDIF
```

- standard deviation estimate

Instead of the standard deviation about the mean (statistic 7, s), this robust estimate is provided from the fifth and first sextiles: $e = (s_5 - s_1)/2$. (This was computed using integer truncating division on the *coded* quantity $s_5 - s_1$, i.e., rounding down.) For unpacking purposes, e is treated exactly like the corresponding standard deviation of each respective variable.

- mean day-of-month of observations
  fraction of observations in daylight
  mean hour of observations
  mean longitude of observations
  mean latitude of observations

The centroids of observational location in time and space are shortened in length and precision from their representation in MST and MSU. "Nice" ***true values*** are reported for d, h, x, and y using the aforementioned ***units*** and ***base***. Because of successive rounding steps, these values actually represent intervals whose trimmed or untrimmed behavior is shown in Tables B3-2a or B3-2b. The lowest and highest intervals are always exceptions to the behavior of the central intervals, so these extreme intervals are shown explicitly with the deviation (δ) of the actual midpoint of this interval from the reported ***true value***. ***Coded*** values greater than one and less than the maximum ***coded*** value correspond to central intervals that can be obtained by subtracting the minus value and adding the plus value to the reported ***true value***, yielding an inclusive lower and exclusive upper bound. The actual midpoint of each central interval can be obtained in any of three ways: by taking the mean of the upper bound and the lower bound, by plugging the ***basex*** shown into the usual formula in place of ***base***, or by adding the midpoint deviation (δ) to the reported ***true value***. For example, the intervals and actual midpoints corresponding to ***true*** h_u ***values*** 1,3,5,...,23 are [0,2.05), [2.05,4.05), [4.05,6.05),...,[22.05,23] and 1.025,3.05,5.05,...,22.525.

Table B3-2a
MSTG Interval Behavior

#	α	Lowest		Central			Highest	
		Interval	δ	± reported	Basex	δ	Interval	δ
1	d	[1,3.1)	+0.05	[−0.9,+1.1)	0.05	+0.1	[29.1,31]	+0.05
2	h_t	[0,0.055)	+0.0275	[−0.045,+0.055)	−0.95	+0.005	[0.955,1]	−0.0225
3	x	[0,0.205)	+0.0025	[−0.095,+0.105)	−0.475	+0.005	[1.805,2]	+0.0025
4	y	[0,0.205)	+0.0025	[−0.095,+0.105)	−0.475	+0.005	[1.805,2]	+0.0025

Table B3-2b
MSUG Interval Behavior

#	α	Lowest		Central			Highest	
		Interval	δ	± reported	Basex	δ	Interval	δ
1	d	[1,3.05)	+0.025	[−0.95,+1.05)	0.025	+0.05	[29.05,31]	+0.025
2	h_u	[0,2.05)	+0.025	[−0.95,+1.05)	−0.475	+0.05	[22.05,23]	−0.475
3	x	[0,0.205)	+0.0025	[−0.095,+0.105)	−0.475	+0.005	[1.805,2]	+0.0025
4	y	[0,0.205)	+0.0025	[−0.095,+0.105)	−0.475	+0.005	[1.805,2]	+0.0025