U. S. DEPARTMENT OF COMMERCE
NAITONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NATIONAL WEATHER SERVICE
NATIONAL CENTERS FOR ENVIRONMENTAL PREDICTION

OFFICE NOTE 424

NEW GLOBAL OROGRAPHY DATA SETS

SONG-YOU HONG
GENERAL SCIENCES CORPORATION

FEBRUARY 1999

Abstract

New orography data sets are constructed based on a United States Geological Survey (USGS) global digital elevation model (DEM) with a horizontal grid spacing of 30 arc seconds (approximately 1 km). Orography statistics including average height, mountain variance, maximum orography, land-sea-lake masks are directly derived from a 30-arc second DEM for a given resolution. Computed global orography data sets are available for three resolutions of 8, 4, and 2 minutes which correspond to approximately 16, 8, 4 km, respectively.

The average height in a 10-min resolution is compared against the NAVY 10 min data that has been used at NCEP. Overall, the new data is more realistic than the NAVY data. The biggest difference between the two data appears in the antarctic region having a dipole patterns of maximum differences as high as 1 km. Near the coastlines along the antarctic and Greenland also are shown differences of 1-km which is mostly due to land-sea mask difference between the two data sets. Over mountainous regions, NAVY data show higher peaks than the new data in some places, but lose mountain peaks in other places, which is due to inconsistent treatment of orography in the NAVY data. This new orography will provide more realistic topographic forcing for NCEP forecast models and may provide improved orography-induced precipitation. In addition, orography statistics based on the new data set are expected to provide realistic treatment of gravity wave drag, for example, in the NCEP Medium-Range Forecast (MRF) model.

1. Computation of orography statistics

a. Extraction of the GTOPO30 data

The GTOPO30 data was constructed by United States Geological Survey (USGS). The Digital Elevation Model (DEM) contains orography with 30-arc second resolution and the data source index. About 50 % of the globe was derived from the orography source with a 3-arc second resolution. The data has "open ocean mask", but "not for lakes". Over the lake regions, a constant height was assigned (for example, 178 m over the Great Lakes). A detailed information of this data is available in the appendix. *[web site : http://www1.gsi-mc.go.jp/gtopo30/gtopo30.html].*

The 8mm tape consists of 33 sub-data files with a special binary format. The files were downloaded in a "sgi" machine, and each file was converted to an ascii format. A care needs to be taken in downloading the data from tape, by declaring "setblksize 512", although the tape file was packed with 10240 bytes. The size of ascii data is 4.6 Giga bytes (GB), and compressed file is 464 Mega bytes (MB).

b. Computation of orography statistics

Orography statistics including average height, variance, maximum values are derived from the GTOPO 30 data at a given resolution. The highest possible resolution is 2 min for orography statistics computation. The 4 and 8 min resolution data sets are also constructed.

- Average : $\overline{h} = \dfrac{\sum_{n=1}^{N} h}{N}$

- Variance : $\sigma = \left[\sum_{n=1}^{N} (h-\overline{h})^2 / N\right]^{1/2} = \dfrac{\sum_{n=1}^{N} h^2}{N} - \overline{h}^2$

- Maximum : $h_{max} = MAX(h, h_{max})$, for n = 1, N

- Ocean mask : m = 0 (ocean), when the number of land points in a grid box are smaller than N/2.

where N is the grid number in a grid box, for example, N = 16 x 16 = 256 for 8 min resolution data sets. Likewise, N = 4 x 4 = 16 for 2 min, and N = 8 x 8 = 64 for 4 min data sets. Since GTOPO30 does not contain lake-mask, a NAVY 10-min resolution lake-mask file is used to fill up the lake regions inland. Table 1 shows the percentages of land, ocean, lake for each resolution data sets.

2. Data structure and format

The compressed data sets are located in

/export/sgi100/data/wd20sh/INT30 for GTOPO30 ascii files,
/export/sgi100/data/wd20sh/topo/ieee for ieee format, and
/export/sgi100/data/wd20sh/topo/ascii for ascii format.

Each resolution data set contains 4-variables as summarized in Tables 2 and 3.

Fortran source files are in

/export/sgi100/data/wd20sh/topo/src and grads control files are in
/export/sgi100/data/wd20sh/topo/ctl directories.

where the ocean-land mask derived from GTOPO data (0 for ocean, and 1 for land and lakes is filled up by a 10-min resolution lake mask data for lake regions (0 for ocean and lakes, 1 for land). Maximum and minimum values, and ocean treatments are given in Table 4.

Reading format and structure are described for the 8-min data sets,

read (1) data(2700,1350) for ieee format, and the ascii format is

read(1,10) data(2700,1350)
10 format(20I4) for orography files and (80I1) for mask files.

Data points are arranged as in the NAVY 10 min orography file,

(1,1) : (0.0666666666, -89.933333333),
(2,1) : (0.2 , -89.933333333),
.
.
(1,2) : (0.0666666666, -89.8 ,
.
.
(2700,1350) : (359.933333333, 89.93333333)

where I,J increase in the longitudinal and latitudinal directions. Latitude and longitude values in degree units will be,

Longitude (I) = $dx/2 + (I-1) * dx$,
Latitude (J) = $- 90 + dx/2 + (J-1) * dx$

where dx = 8/60, 4/60, 2/60 for 8, 4, 2 min resolutions, respectively.

3. Comparison of the new data against the NAVY data.

The 10-min NAVY data (NAVY) is compared to the new data sets (USGS). A 10-min average height is computed for fair comparison (Figs. 1-6). A significant difference is found over the ant-arctic region with a maximum of about 1km (Fig. 2). A similar amount of difference is found along the coasts over the antarctic and Greenland regions (Fig.3). Over the Tibetan plateau (Fig. 4), the USGS shows a smaller height than the NAVY in most regions. This is rooted from the fact that the NAVY data is constructed in 10 min resolution and treated with a special function to recover mountain peaks. A detailed comparison over the US west coast (Fig. 5) shows that the USGS reveals a more realistic orography distribution than the NAVY. For example, a steep orography over the Sierra-Nevada mountains is well reproduced. Mountain peaks over Washington are more realistic in the USGS than the NAVY. Over the East Asia (Fig. 6), the NAVY shows a unrealistically uniform orography shape over northern China and Mongolian regions, while they are represented well in the USGS. Over Japan, mountain complex near the Fuji Mt. is more realistic in USGS than NAVY.

4. Comparison of different resolution data sets.

Figures 7-9 shows average height and variance for the 8, 4, 2 min data sets over the west coast. Averaged orography reveals much details over the mountains as the resolution increases. For example, Mt. Rainier height reduces as resolution decreases while it is only 1 km in the NAVY data (Table 5). Mountain variances show a smaller horizontal scales but absolute magnitudes are similar since all variances are computed directly from the 30-arc second orography data.

5. A concluding remarks

This new orography will provide more realistic topographic forcing for NCEP forecast models and may provide improved orography-induced precipitation. In addition, orography statistics based on the new data set are expected to provide realistic treatment of gravity wave drag, for example, in the NCEP Medium-Range Forecast (MRF) model.

The averaged orography can be interpolated onto the model grid. The variance data for the model (m) can be interpolated from a new data set (g),

$$\sigma_m^2 = \sum_{n=1}^{N} [\sigma_g^2 + (\overline{h}_g)^2] / N - \left(\frac{\sum_{n=N}^{N} \overline{h}_g}{N} \right)^2$$

$$= \frac{\sum_{n=1}^{N} \sigma_g^2}{N} + \frac{1}{N} \sum_{n=1}^{N} (\overline{h}_g)^2 - \left(\frac{\sum_{n=1}^{N} \overline{h}_g}{N} \right)^2$$

where N designates the number of grid points of orography data sets in a model grid box. Computed variances in the model grid still represent orography variance based on the 30-arc second orography.

The convexity and mountain asymmetry in the gravity wave drag (Kim and Arakawa 1995, J. Atmos. Sci.) cannot be computed directly from the 30-arc second data. The computed values are unrealistically large because of their 4-th power nature. For the MRF having a resolution lower than 20 km, convexity can be computed from the averaged height of 8-min data. Mountain asymmetry computed from maximum orography and averaged values in the 8-min data would be reasonable.

Lake masks in the new data sets are based on a 10-min resolution. A compatible resolution data for lake information to the 30-arc second orography needs to be constructed in the future.

Acknowledgements : The author express his sincere gratitude to M. Kanamitsu for his introduction of this work and continuous encouragement. M. Ji kindly provided huge amount of disk space and computer resources that should be acknowledged. Thanks are also due to F. Mesinger for a careful internal review.

Table 1. Percentages of ocean, land, lake points in the new data sets (%).

Resolution	Ocean	Land	Lakes
2 min	66.81	32.55	0.63
4 min	66.82	32.60	0.57
8 min	66.83	32.65	0.51

Table 2. File size

Resolution	Dimension	File number	Size (uncompressed)	Size (compressed)
GTOPO30	43200 X 21600	33	Ascii (4.6 GB)	Ascii (464 MB)
2 min	10800 X 5400	4	Ascii (767MB) or ieee (932MB)	Ascii (85MB) ieee(146MB)
4 min	5400 X 2700	4	Ascii(192MB) or ieee (233MB)	Ascii (23MB) ieee(26MB)
8 min	2700 X 1350	4	Ascii (48MB) or ieee (58MB)	Ascii (6MB) ieee(12MB)

Table 3. File naming for 8-min resolution data sets

Average height	: TOP8M_avg.ieee or TOP8M_avg.20I4.asc
Variance	: TOP8M_var.ieee or TOP8M_var.20I4.asc
Maximum	: TOP8M_max.ieee or TOP8M_max.20I4.asc
Ocean-lake-land mask:	TOP8M_slm.ieee or TOP8M_slm.80I1.asc

Table 4. Maxim and minimum values for each orography statistics, and assigned ocean values.

Resolution	maximum (m)	minimum (m)	ocean	remarks
GTOPO30	8752	-407	9999	No lake mask
2M_avg	8207	-390	-999	No lake mask
4M_avg	7381	-380	-999	No lake mask
8M_avg	6600	-371	-999	No lake mask
2M_var	1150	0	0	0 near lakes
4M_var	1181	0	0	0 near lakes
8M_var	1487	0	0	0 near lakes
2M_max	8752	-405	0	
4M_max	8752	-405	0	
8M_max	8752	-252	0	
2M_slm	1	0	0	0 over lakes
4M_slm	1	0	0	0 over lakes
8M_slm	1	0	0	0 over lakes

Table 5. Average height of Mt. Rainier (m).

Resolution	30 sec	2min	4min	8min	NAVY
	4200	3500	2700	2100	1000

Figure lists

Fig. 1. Difference of the 10-min averaged USGS and NAVY (m). Contour intervals at 500 m without zero lines.

Fig. 2. The USGS 10-min orography (upper) and the differences (bottom) [USGS-NAVY].

Fig. 3. Same as in Fig. 2, but for the region between 150E and 110W.

Fig. 4. Same as in Fig. 2, but for the Tibetan Plateau region.

Fig. 5. Same as in Fig. 2, but for the US west coast region.

Fig. 6. Same as in Fig. 2, but for the East Asian region.

Fig. 7. The 10-min averaged orography (Contour intervals at 500 m and shaded over 1500 m) and mountain variance (Contour intervals at 200 m and shaded over 200 m).

Fig. 8. Same as in Fig. 7, but for the 4-min data sets.

Fig. 9. Same as in Fig. 7, but for the 2-min data sets.

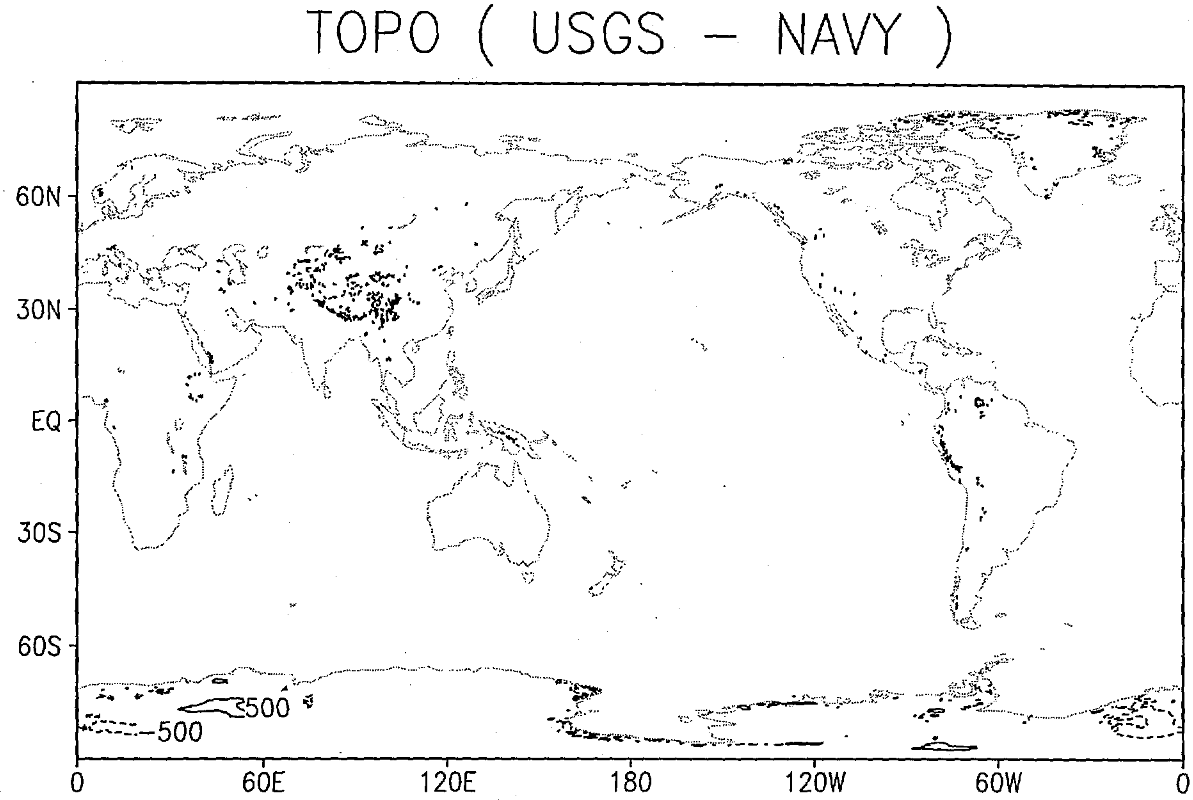

TOPO (USGS - NAVY)
60N
30N
EQ
30S
60S
-500
-500
0
60E
120E
180
120W
60W
0

Fig. 1

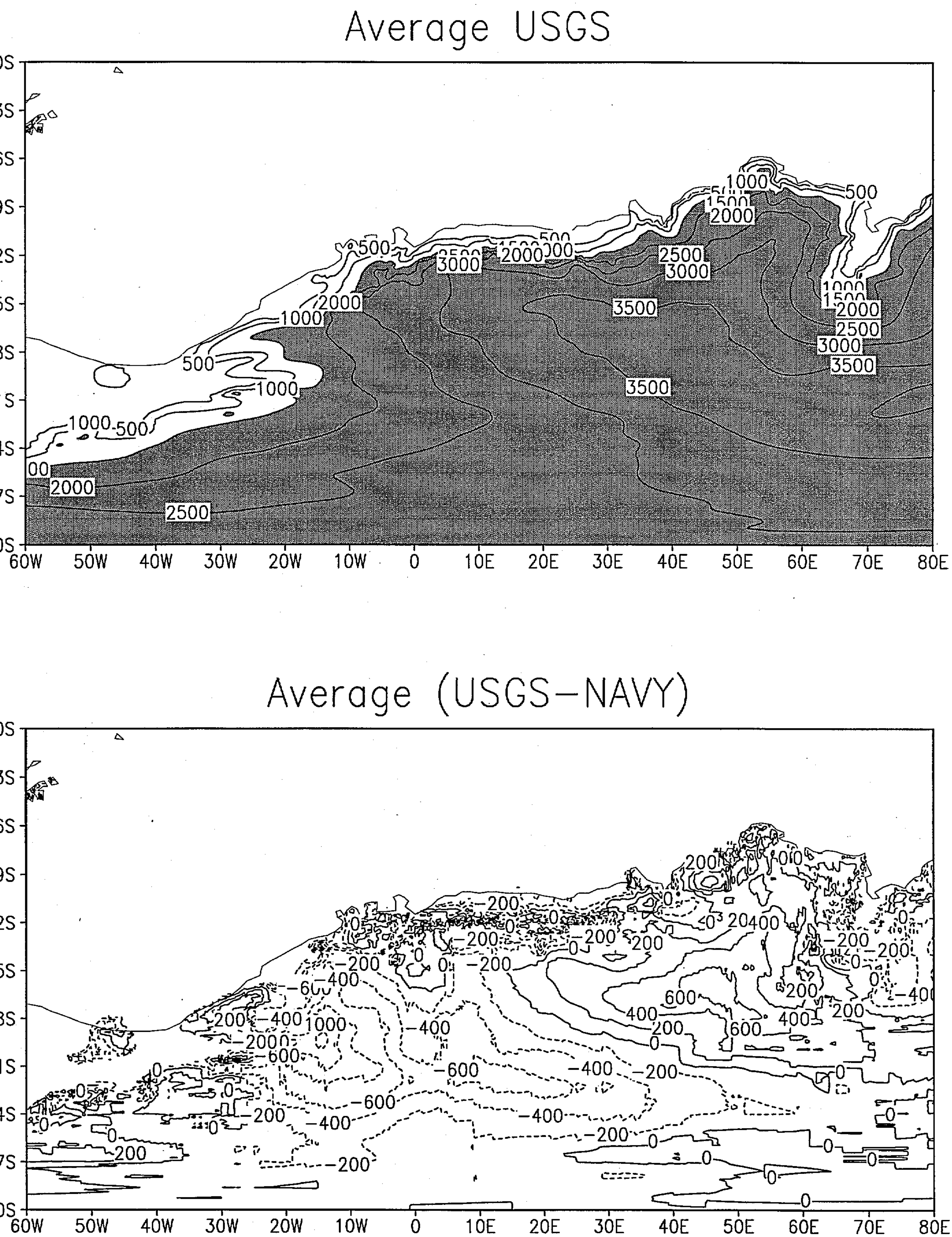
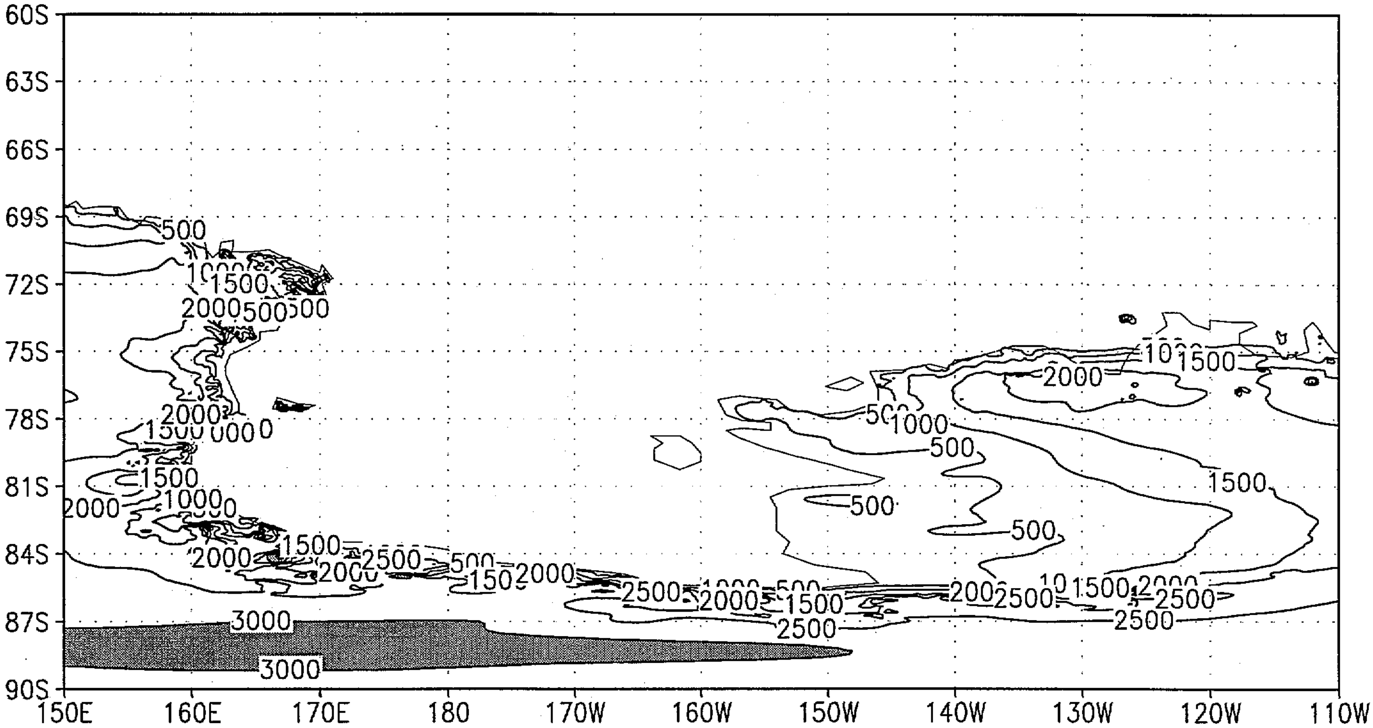

Average USGS
60S
63S
66S
69S
72S
75S
78S
81S
84S
87S
90S
60W 50W 40W 30W 20W 10W 0 10E 20E 30E 40E 50E 60E 70E 80E
Average (USGS−NAVY)
60S
63S
66S
69S
72S
75S
78S
81S
84S
87S
90S
60W 50W 40W 30W 20W 10W 0 10E 20E 30E 40E 50E 60E 70E 80E

Fig. 2

Average USGS



Average (USGS–NAVY)

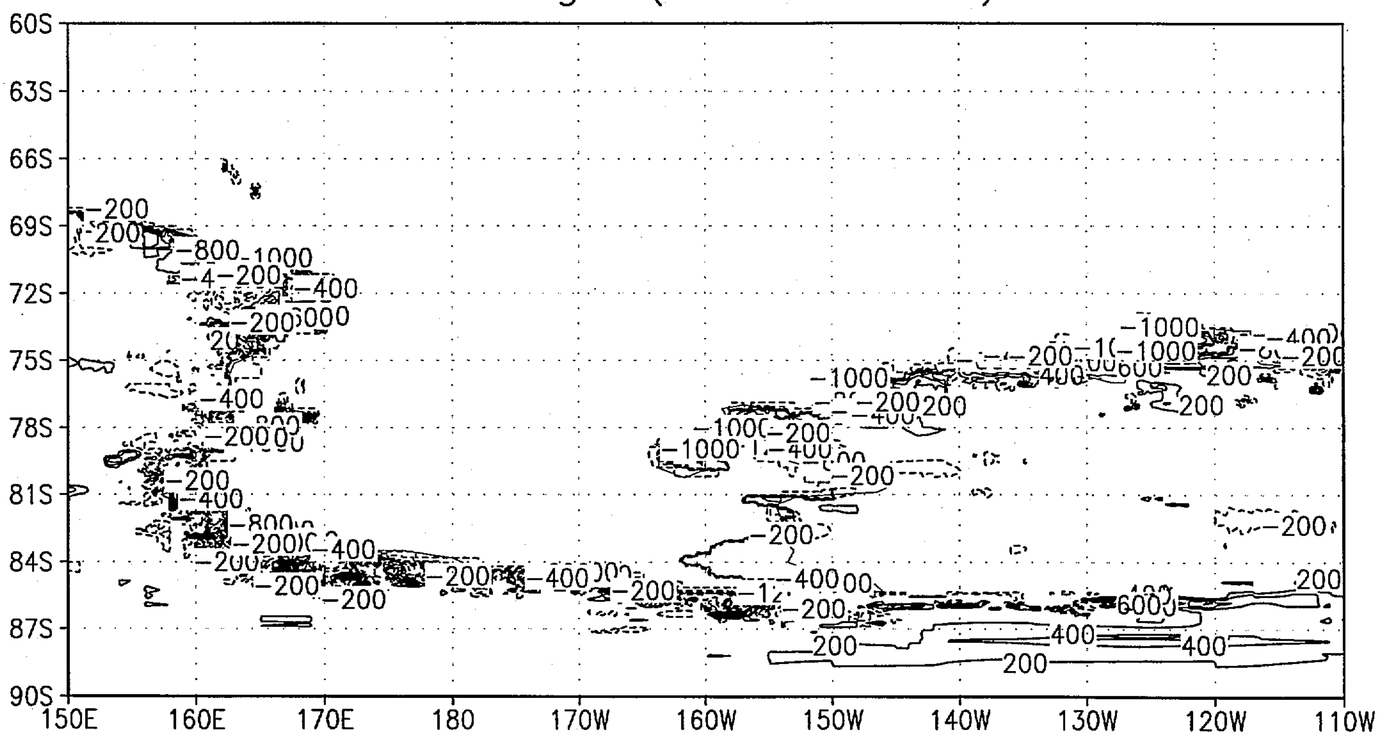


Fig. 3

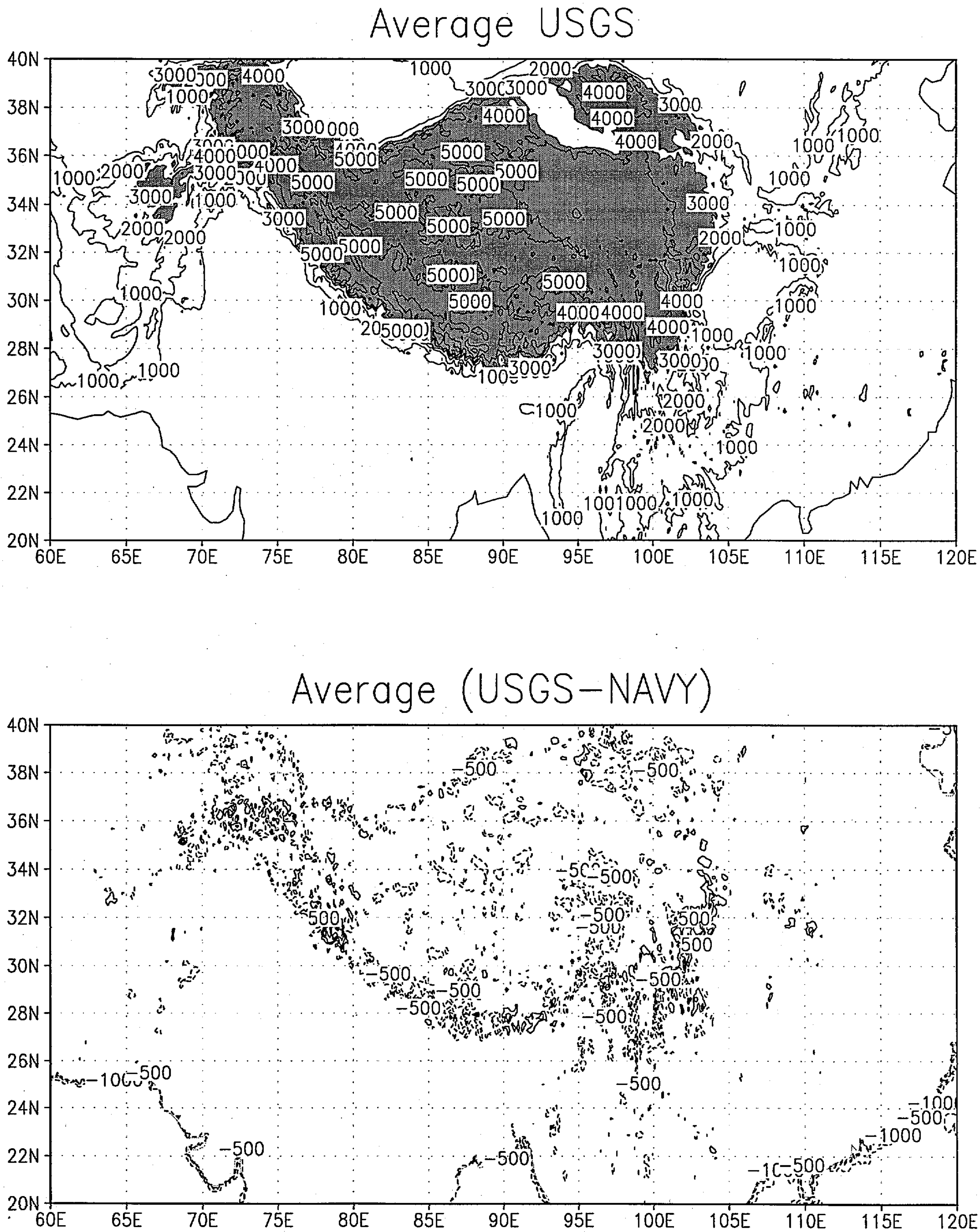

Average USGS
Average (USGS−NAVY)
40N
38N
36N
34N
32N
30N
28N
26N
24N
22N
20N
60E
65E
70E
75E
80E
85E
90E
95E
100E
105E
110E
115E
120E

Fig. 4

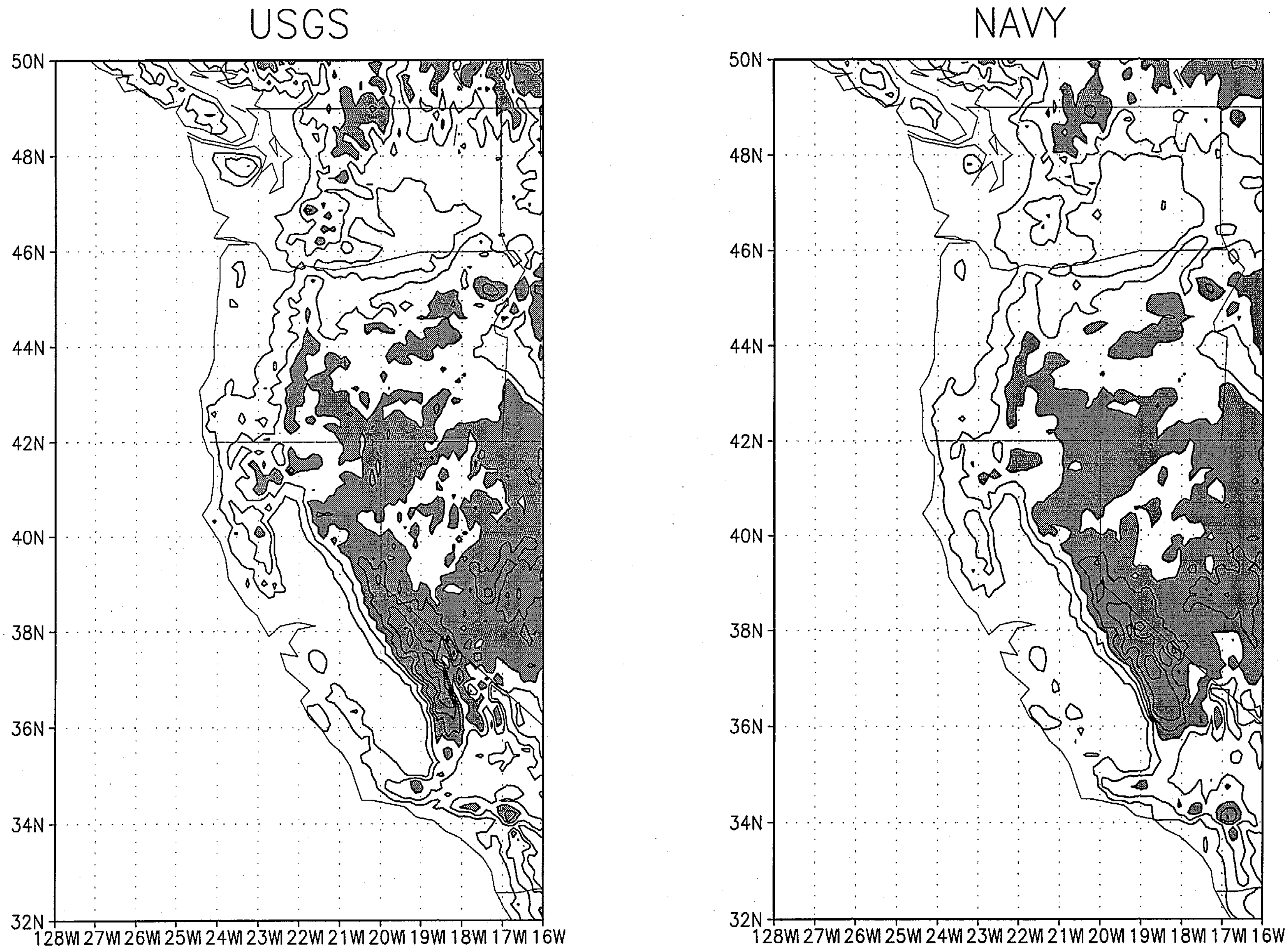
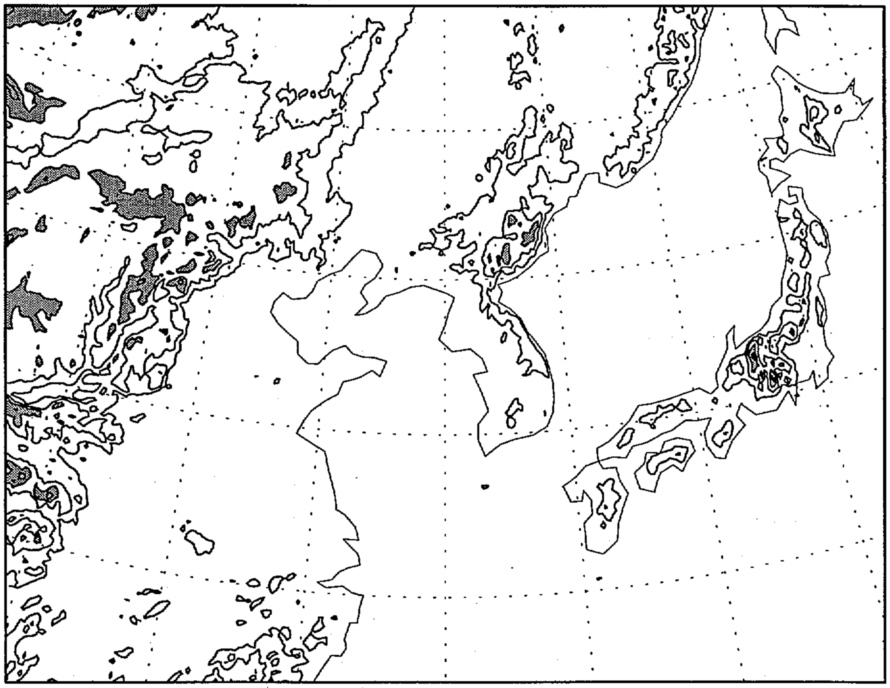

USGS
NAVY
50N
48N
46N
44N
42N
40N
38N
36N
34N
32N
128W 27W 26W 25W 24W 23W 22W 21W 20W 19W 18W 17W 16W

Fig. 5

USGS



NAVY

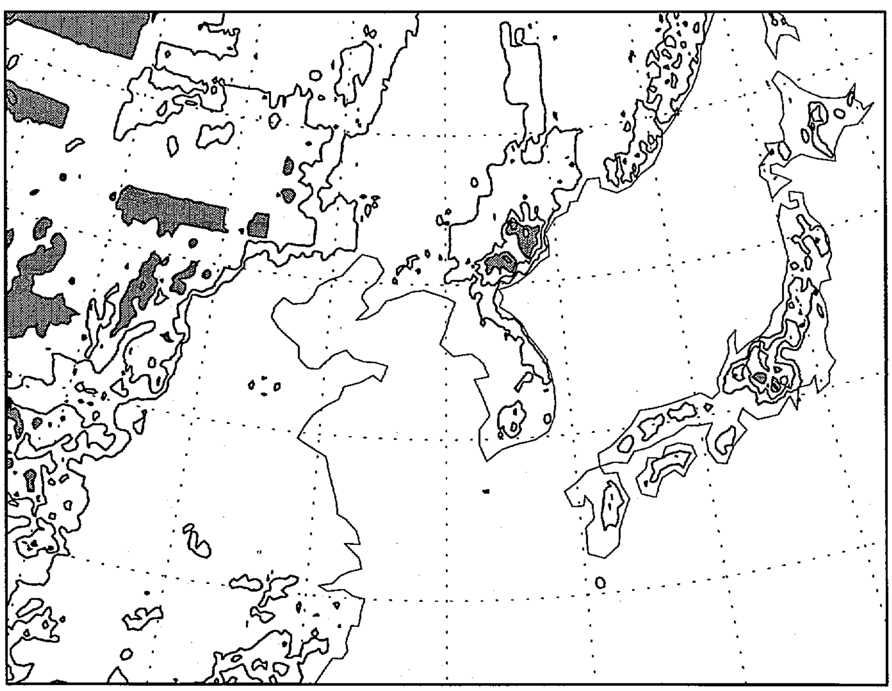


Fig.6

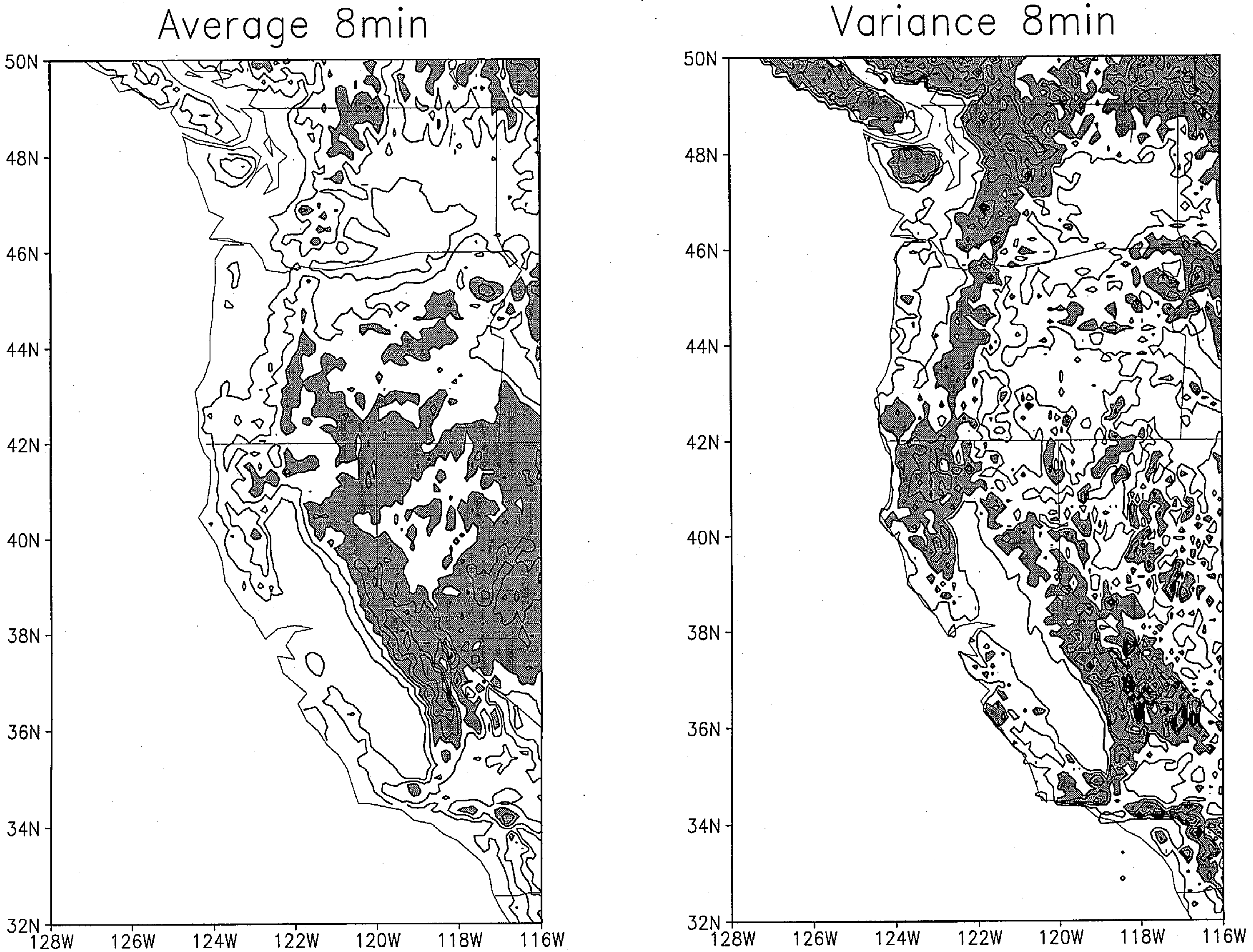

Average 8min
Variance 8min
50N
48N
46N
44N
42N
40N
38N
36N
34N
32N
128W
126W
124W
122W
120W
118W
116W

Fig. 7

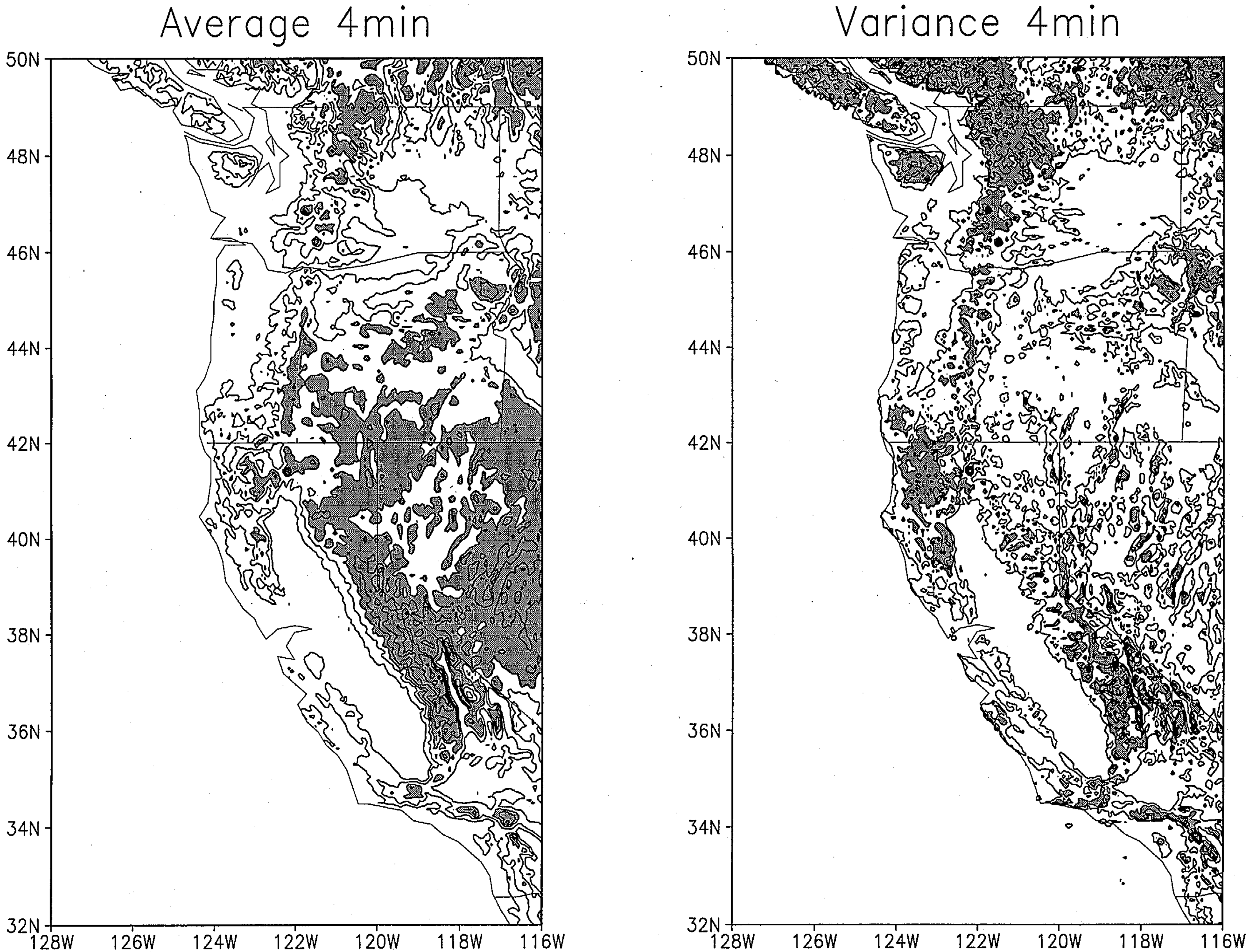

Average 4min
Variance 4min
50N
48N
46N
44N
42N
40N
38N
36N
34N
32N
128W
126W
124W
122W
120W
118W
116W

Fig. 9

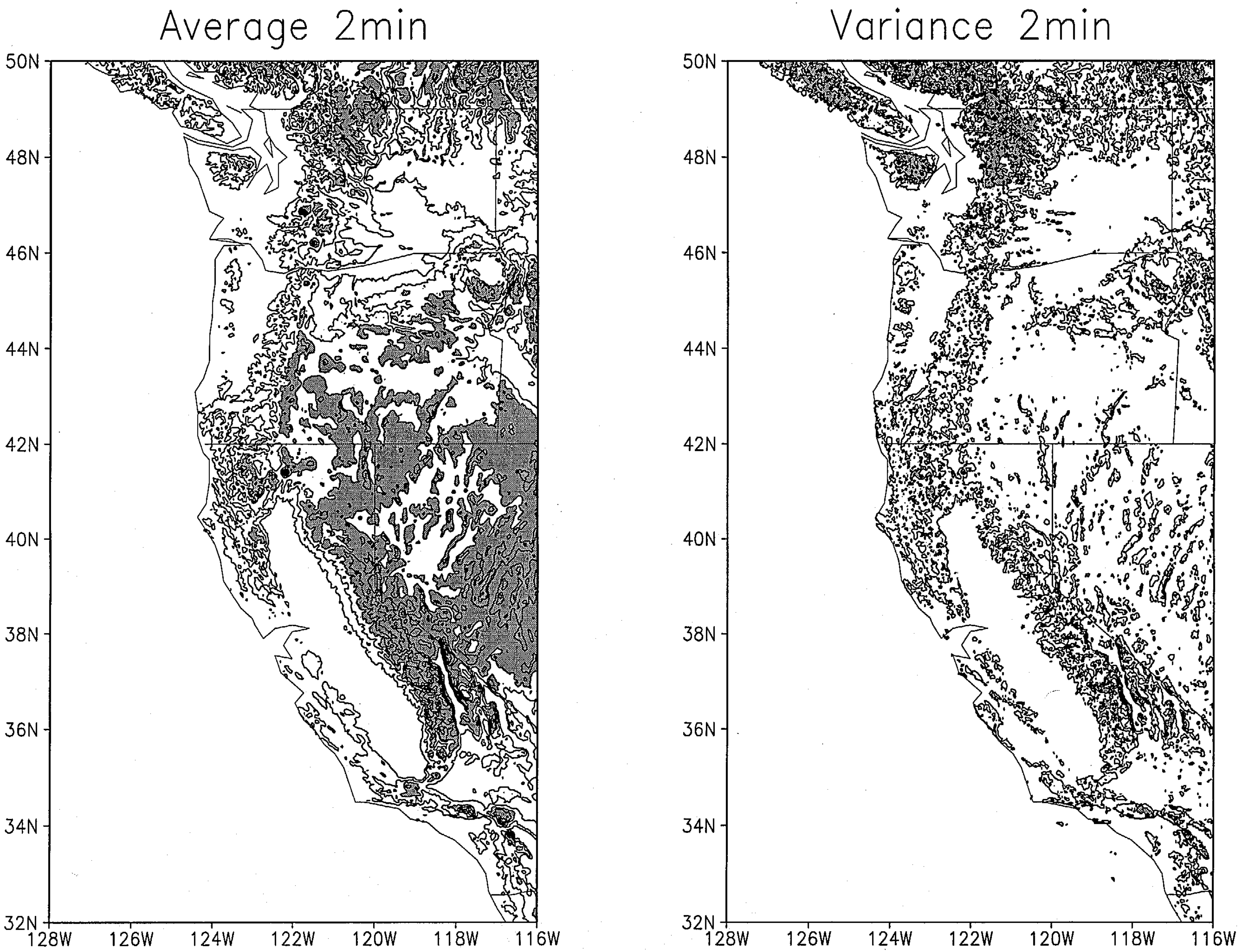

Average 2min
Variance 2min
50N
48N
46N
44N
42N
40N
38N
36N
34N
32N
128W
126W
124W
122W
120W
118W
116W

Fig 9

Appendix : GTOPO30 Documentation

1.0 Introduction

GTOPO30 is a global digital elevation model (DEM) resulting from a collaborative effort led by the staff at the U.S. Geological Survey's EROS Data Center in Sioux Falls, South Dakota. Elevations in GTOPO30 are regularly spaced at 30-arc seconds (approximately 1 kilometer). GTOPO30 was developed to meet the needs of the geospatial data user community for regional and continental scale topographic data. This release represents the completion of global coverage of 30-arc second elevation data that have been available from the EROS Data Center beginning in 1993. Several areas have been updated and the entire global data set has been repackaged, so these data supersede the previously released continental data sets. Comments from users of GTOPO30 are welcomed and encouraged. Please send your comments to Dean Gesch at gesch@edcmail.cr.usgs.gov or to Sue Greenlee at sgreenlee@edcmail.cr.usgs.gov.

2.0 Data Set Characteristics

GTOPO30 is a global data set covering the full extent of latitude from 90 degrees south to 90 degrees north, and the full extent of longitude from 180 degrees west to 180 degrees east. The horizontal grid spacing is 30-arc seconds (0.008333333333333 degrees), resulting in a DEM having dimensions of 21,600 rows and 43,200 columns. The horizontal coordinate system is decimal degrees of latitude and longitude referenced to WGS84. The vertical units represent elevation in meters above mean sea level. The elevation values range from -407 to 8,752 meters. In the DEM, ocean areas have been masked as "no data" and have been assigned a value of -9999. Lowland coastal areas have an elevation of at least 1 meter, so in the event that a user reassigns the ocean value from -9999 to 0 the land boundary portrayal will be maintained. Due to the nature of the raster structure of the DEM, small islands in the ocean less than approximately 1 square kilometer will not be represented.

3.0 Data Format

To facilitate electronic distribution, GTOPO30 has been divided into 33 smaller pieces, or tiles. The area from 60 degrees south latitude to 90 degrees north latitude and from 180 degrees west longitude to 180 degrees east longitude is covered by 27 tiles, with each tile covering 50 degrees of latitude and 40 degrees of longitude. Antarctica (90 degrees south latitude to 60 degrees south latitude and 180 degrees west longitude to 180 degrees east longitude) is covered by 6 tiles, with each tile covering 30 degrees of latitude and 60 degrees of longitude. The tiles names refer to the longitude and latitude of the upper-left (northwest) corner of the tile. For example, the coordinates of the upper-left corner of tile E020N40 are 20 degrees east longitude and 40 degrees north latitude. There is one additional tile that covers all of Antarctica with data in a polar stereographic projection. The following table lists the name, latitude and longitude extent, and elevation statistics for each tile.

	Latitude		Longitude		Elevation			
Tile	Minimum	Maximum	Minimum	Maximum	Minimum	Maximum	Mean	Std.Dev.
W180N90	40	90	-180	-140	1	6098	448	482
W140N90	40	90	-140	-100	1	4635	730	596
W100N90	40	90	-100	-60	1	2416	333	280
W060N90	40	90	-60	-20	1	3940	1624	933
W020N90	40	90	-20	20	-30	4536	399	425
E020N90	40	90	20	60	-137	5483	213	312
E060N90	40	90	60	100	-152	7169	509	698
E100N90	40	90	100	140	1	3877	597	455
E140N90	40	90	140	180	1	4588	414	401
W180N40	-10	40	-180	-140	1	4148	827	862
W140N40	-10	40	-140	-100	-79	4328	1321	744
W100N40	-10	40	-100	-60	1	6710	375	610
W060N40	-10	40	-60	-20	1	2843	212	168
W020N40	-10	40	-20	20	-103	4059	445	298
E020N40	-10	40	20	60	-407	5825	727	561
E060N40	-10	40	60	100	1	8752	1804	1892
E100N40	-10	40	100	140	-40	7213	692	910
E140N40	-10	40	140	180	1	4628	549	715
W180S10	-60	-10	-180	-140	1	2732	188	297
W140S10	-60	-10	-140	-100	1	910	65	124
W100S10	-60	-10	-100	-60	1	6795	1076	1356
W060S10	-60	-10	-60	-20	1	2863	412	292
W020S10	-60	-10	-20	20	1	2590	1085	403
E020S10	-60	-10	20	60	1	3484	893	450
E060S10	-60	-10	60	100	1	2687	246	303
E100S10	-60	-10	100	140	1	1499	313	182
E140S10	-60	-10	140	180	1	3405	282	252
W180S60	-90	-60	-180	-120	1	4009	1616	1043
W120S60	-90	-60	-120	-60	1	4743	1616	774
W060S60	-90	-60	-60	0	1	2916	1866	732
W000S60	-90	-60	0	60	1	3839	2867	689
E060S60	-90	-60	60	120	1	4039	2951	781
E120S60	-90	-60	120	180	1	4363	2450	665
ANTARCPS	-90	-60	-180	180	1	4748	2198	1016

The 27 tiles that individually cover 50 degrees of latitude and 40 degrees of longitude each have 6,000 rows and 4,800 columns. The 6 Antarctica tiles that individually cover 30 degrees of latitude and 60 degrees of longitude each have 3,600 rows and 7,200 columns. There is no overlap among the tiles so the global data set may be assembled by simply abutting the adjacent tiles.

The tile named ANTARCPS includes the same data as the 6 geographic tiles covering Antarctica, but is presented in a polar stereographic projection. The horizontal grid spacing is 1,000 meters, and the tile has 5,400 rows and 5,400 columns. The projection parameters used for the polar stereographic projection are: 0 degrees for the longitude of the central meridian, 71 degrees south for the latitude of true scale, and 0 for the false easting and false northing.

Data for each tile are provided in a set of 8 files. The files are named with the tile name and a file name extension indicating the contents of the file. The following extensions are used:

Extension	Contents
DEM	digital elevation model data
HDR	header file for DEM
DMW	world file
STX	statistics file
PRJ	projection information file
GIF	shaded relief image
SRC	source map
SCH	header file for source map

The source map is a simple 8-bit binary image which has values that indicate the source used to derive the elevation for every cell in the DEM. The source map is the same resolution and has the same dimensions and coordinate system as the DEM. Like the DEM, it has no header or trailer bytes and is stored in row major order. These codes are used in the source map image:

Value	Source
0	Ocean
1	Digital Terrain Elevation Data
2	Digital Chart of the World
3	USGS 1-degree DEM's
4	Army Map Service 1:1,000,000-scale maps
5	International Map of the World 1:1,000,000-scale maps
6	Peru 1:1,000,000-scale map
7	New Zealand DEM
8	Antarctic Digital Database

The absolute vertical accuracy of GTOPO30 varies by location according to the source data. Generally, the areas derived from the raster source data have higher accuracy than those derived from the vector source data. The full resolution 3-arc second DTED and USGS DEM's have a vertical accuracy of + or - 30 meters linear error at the 90 percent confidence level (Defense Mapping Agency, 1986; U.S. Geological Survey, 1993). If the error distribution is assumed to be Gaussian with a mean of zero, the statistical standard deviation of the errors is equivalent to the root mean square error (RMSE). Under those assumptions, vertical accuracy expressed as + or - 30 meters linear error at 90 percent can also be described as a RMSE of 18 meters. The areas of GTOPO30 derived from DTED and USGS DEM's retain that same level of accuracy because through generalization a representative elevation value derived from the full resolution cells is chosen to represent the area of the reduced resolution cell (although the area on the ground represented by that one elevation value is now much larger than the area covered by one full resolution cell).

Source	Vertical accuracy (meters) L.E. at 90%	RMSE	Estimation method
DTED	30	18	product specification
DCW	160	97	calculated vs. DTED
USGS DEM	30	18	product specification
AMS maps	250	152	estimated from 500-meter interval
IMW maps	50	30	estimated from 100-meter interval
Peru map	500	304	estimated from 1,000-meter interval
N.Z. DEM	15	9	estimated from 100-foot interval
ADD	highly variable		wide range of scales and intervals

Latitude (degrees)	Ground distance (meters) E/W	N/S
Equator	928	921
10	914	922
20	872	923
30	804	924
40	712	925
50	598	927
60	465	929
70	318	930
73	272	930
78	193	930
82	130	931