R. Kahn
April, 1992

N94-23616
171312
p. 28

Experience of the JPL Exploratory Data Analysis Team at Validating HIRS2/MSU Cloud Parameters

Ralph Kahn, Robert Haskins,

Stephanie Granger-Gallegos, and Andy Pursch

Jet Propulsion Laboratory, California Institute of Technology

and

Anthony Del Genio

Goddard Institute for Space Studies

Where We Began:

The Cloud/Climate Feedback Problem

Key Measurements Addressing the Cloud-Climate Feedback Problem

Microphysical parameters:

1. $\beta = d(\ln m) / dT$ — dependence of cloud water (liquid and ice) content on temperature (including liquid to solid transition temperature and small ice particle concentrations)

2. $\gamma = d(\ln \tau) / dT$ — dependence of cloud opacity on temperature (implicitly, dr / dm; dr / dT)

Cloud properties:

3. $n(q, w, T)$ — dependence of cloud amount on relative humidity, vertical velocity, temp., and other environmental parameters

4. cloud top height dependence on temperature, relative humidity, vertical velocity, and and other environmental parameters

5. variability in cloud behavior (diurnal, seasonal, interannual; land & ocean)

Cloud-related processes:

6. distinguish T from dynamical effects on clouds (sign & size of feedbacks)

7. determine large-scale conditions for formation and breakup of marine stratocumulus (Cloud Top Entrainment Instability)

8. determine the relationship between deep convection and upper troposphere water

SENSITIVITY OF DERIVED EFFECTIVE CLOUD AMOUNT TO SURFACE TEMPERATURE

$$\left.\frac{\partial N}{\partial T_s}\right|_{p_c} = \frac{\varepsilon_s e^{-t_0} (1 - N) \dfrac{\partial B_\nu(T_s)}{\partial T_s}}{\varepsilon_s B_\nu(T_s) e^{-t_0} + R_s e^{-t_0} + \displaystyle\int_{t_1}^{t_0} B_\nu(T') e^{-t'} dt' - B_\nu(T_c) e^{-t_1}}$$

On the right side, the terms in the denominator account for:

(1) direct radiation from the surface

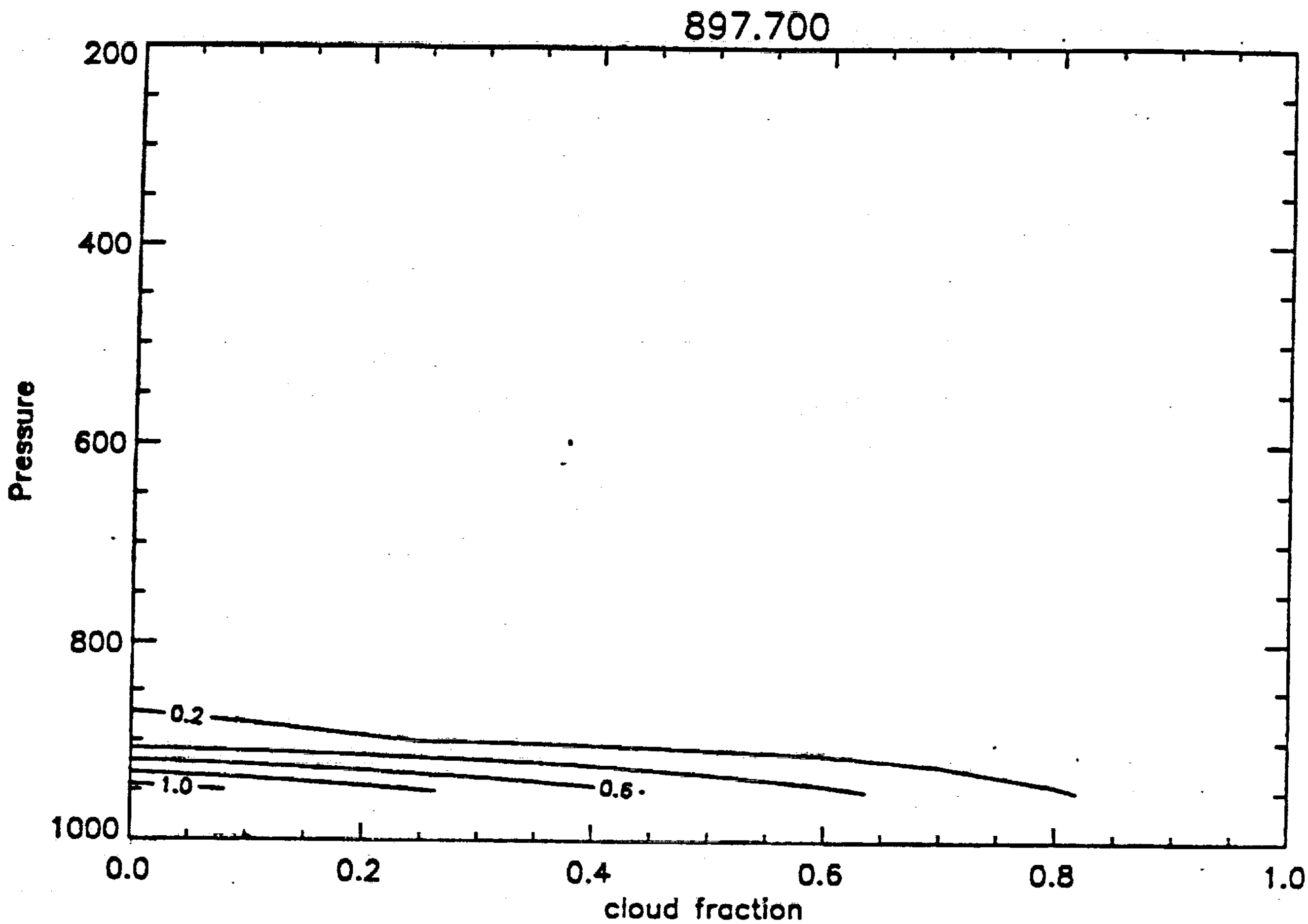
(2) solar radiation reflected by the surface

(3) emission of the atmosphere below the cloud level

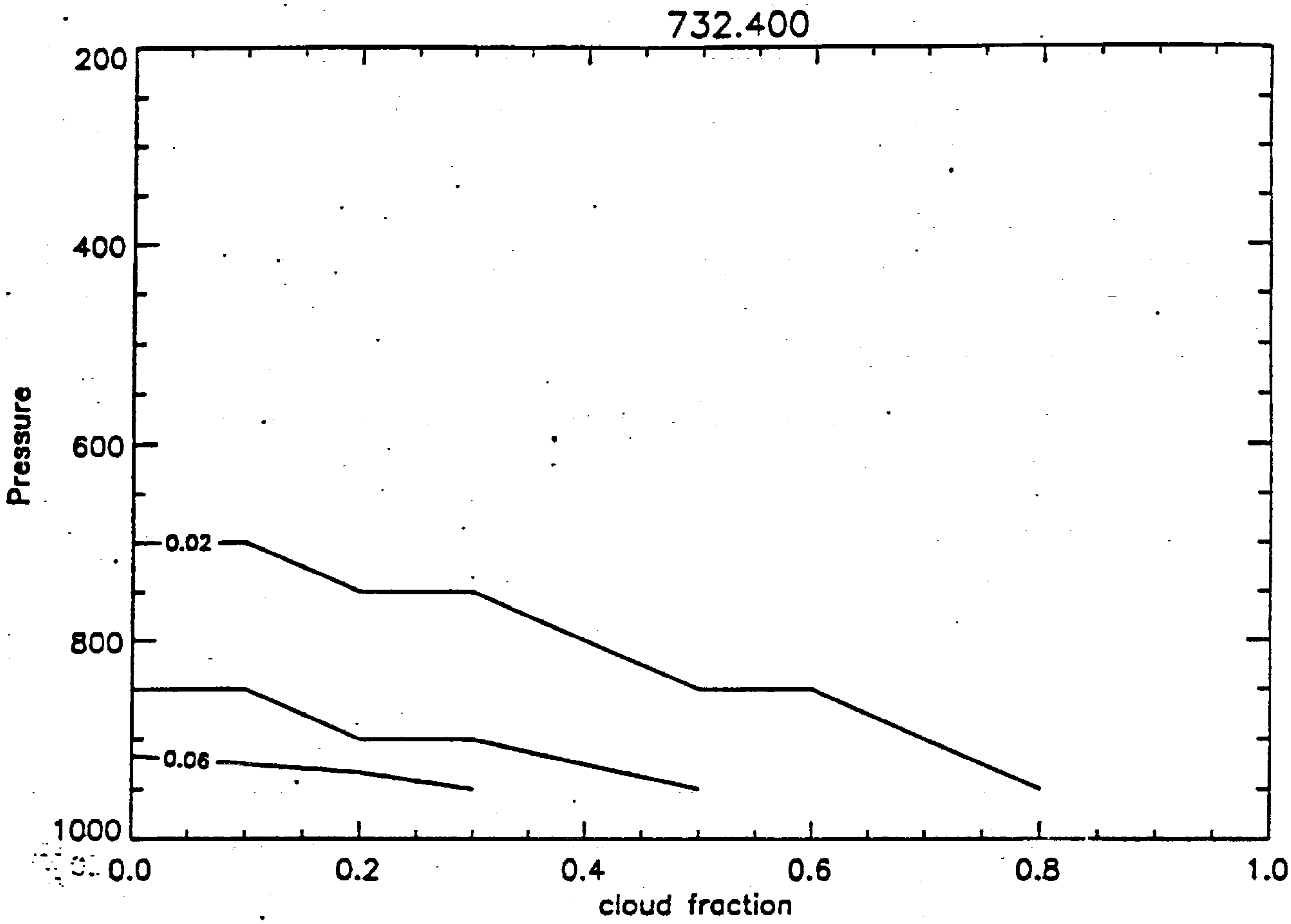
(4) emission from the cloud surface.

The terms are wavelength dependent

The derived cloud amount is less sensitive to surface temperature for higher clouds. This occurs because as the cloud elevation increases, the difference between T_S and T_C increases, so only a small change in cloud amount is needed to effect a large change in radiance at the detector.



$\left.\frac{\partial N}{\partial T_s}\right|_{p_c}$ [window channel]



$\left.\frac{\partial N}{\partial T_s}\right|_{p_c}$ [725mb peak channel]

Definition of Validation

By **'Validation'**, we mean **'developing a quantitative sense for the physical meaning of the measured parameters'**, by:

(1) identifying the **assumptions** involved in deriving parameters from the measured radiances

(2) **testing the input** data and derived parameters for statistical error, sensitivity, and internal consistency

(3) **comparing with similar parameters** obtained from other sources using other techniques

Identifying the Assumptions

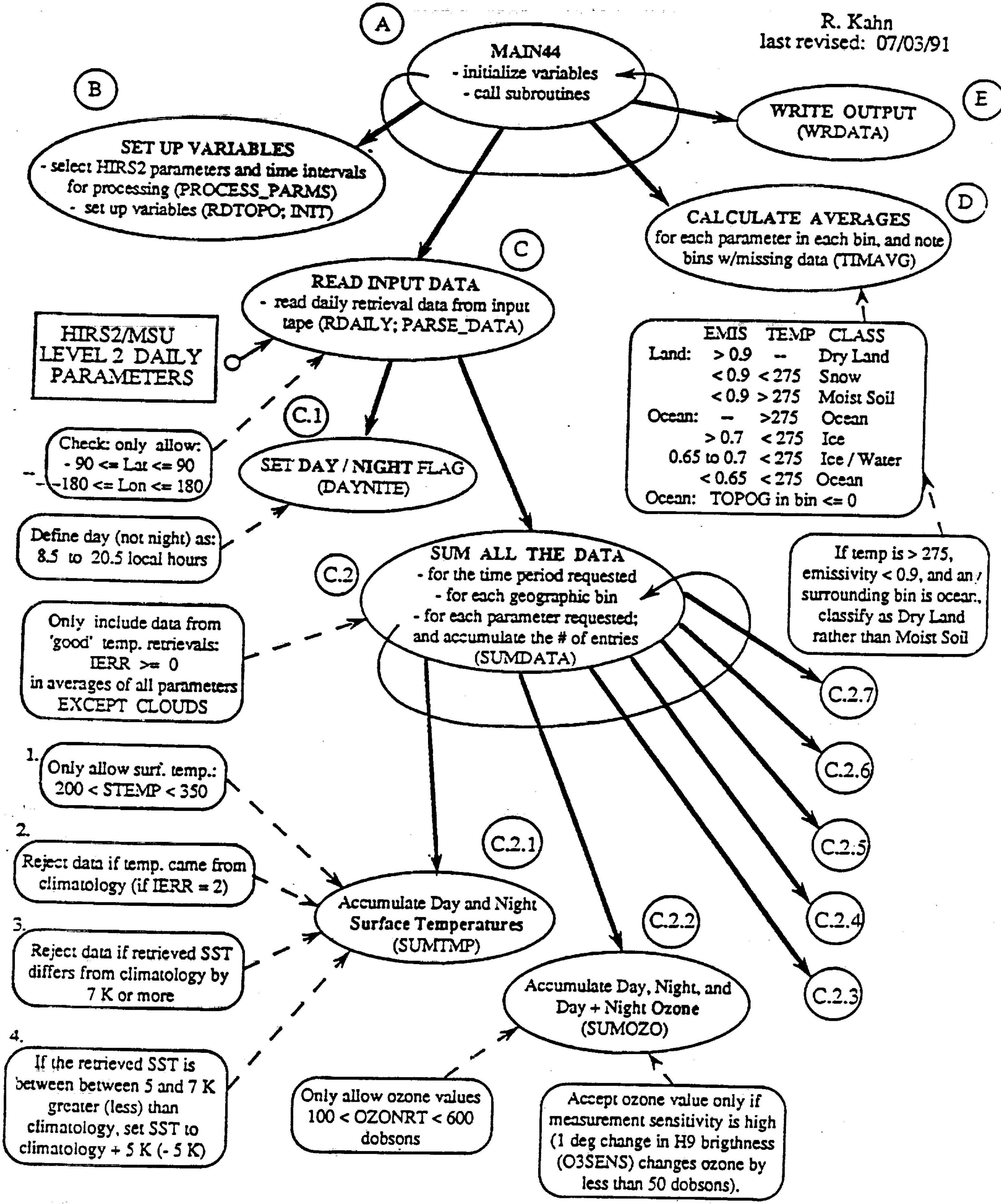
-- in the **Measurements** (instrument, technique)

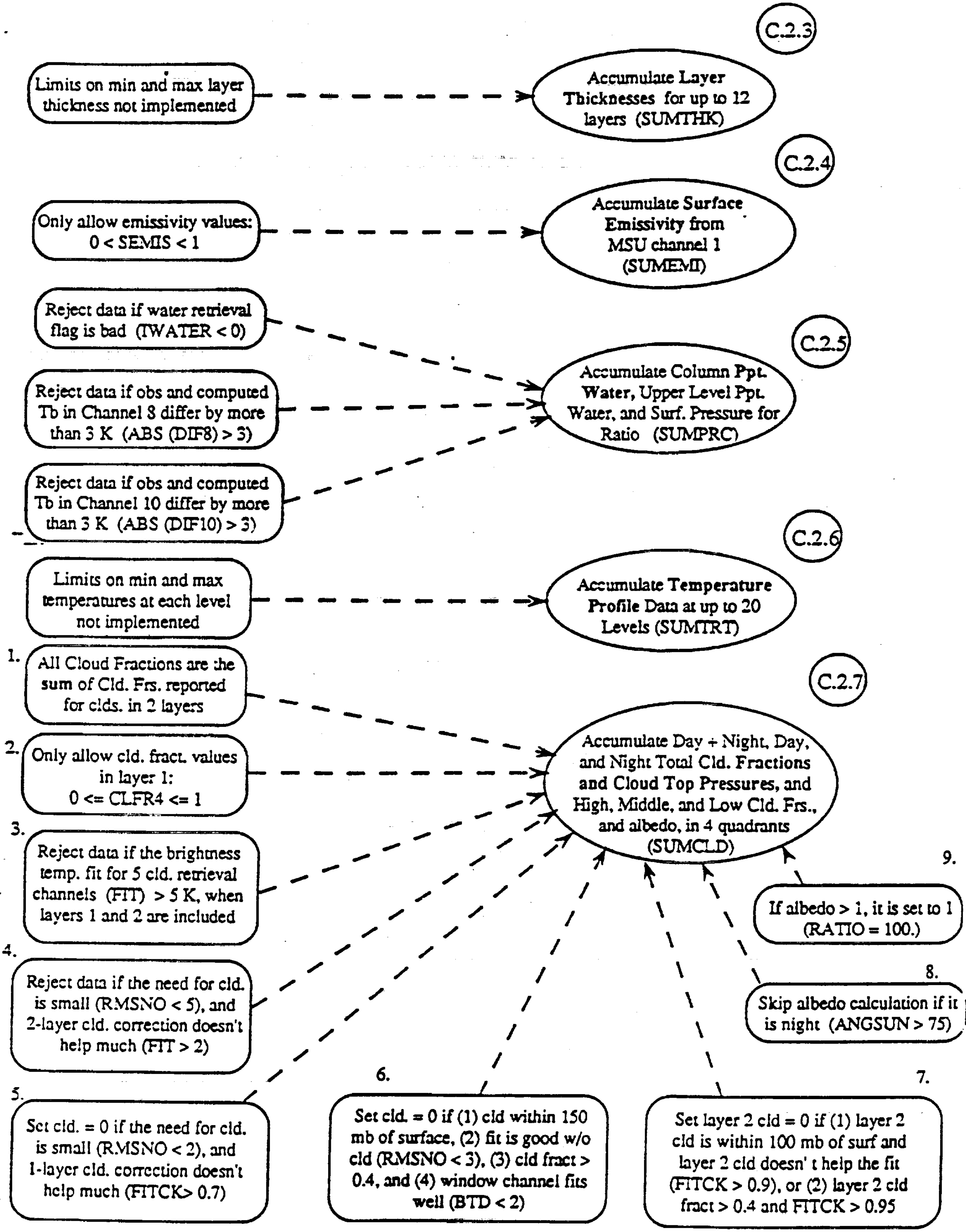
-- in the **Algorithm** (retrieval equations)

-- in the **Code** ('if' statements)

GSFC HIRS2 Level 2 to 3 Software Overview / Assumptions

R. Kahn
last revised: 07/03/91

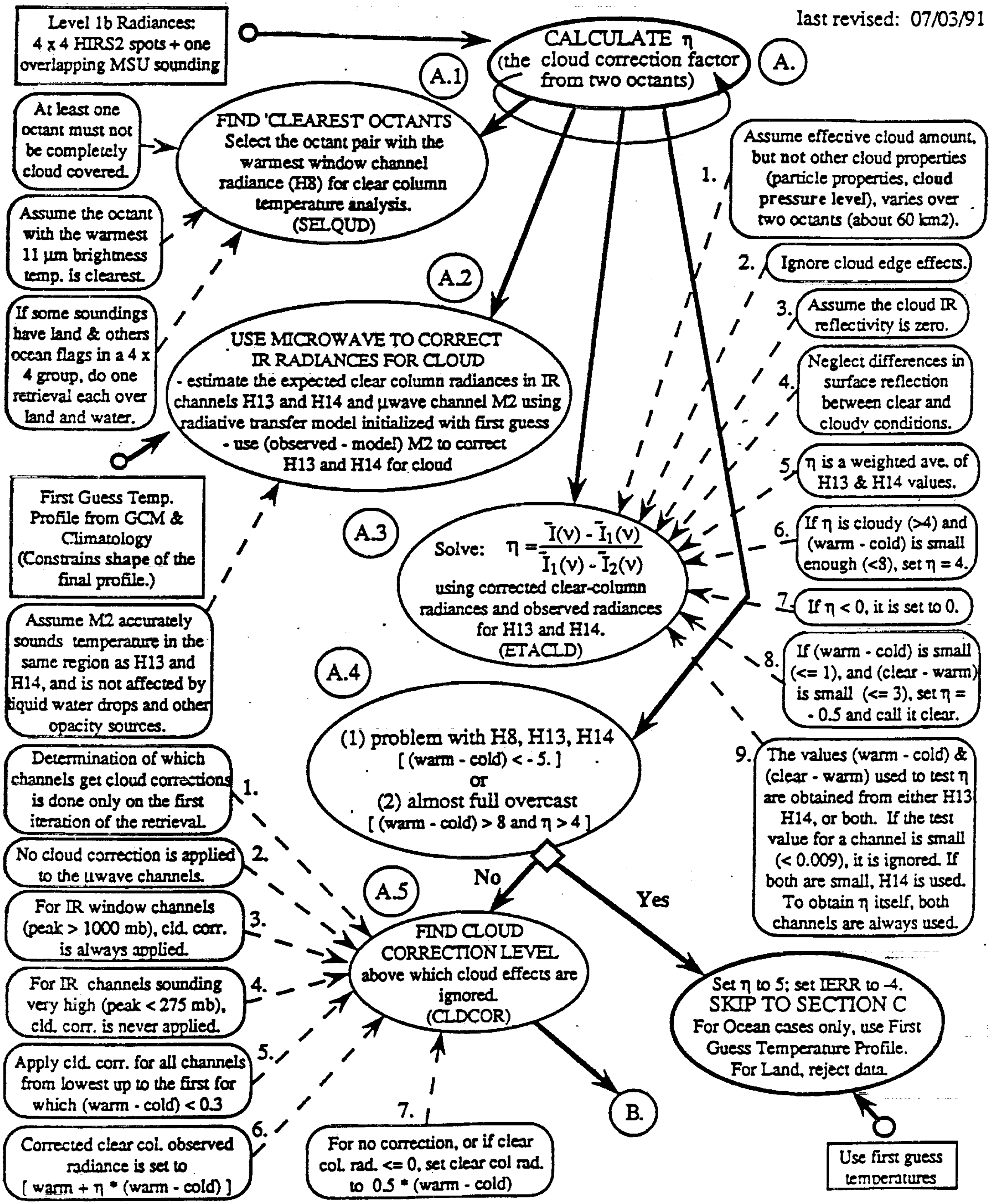


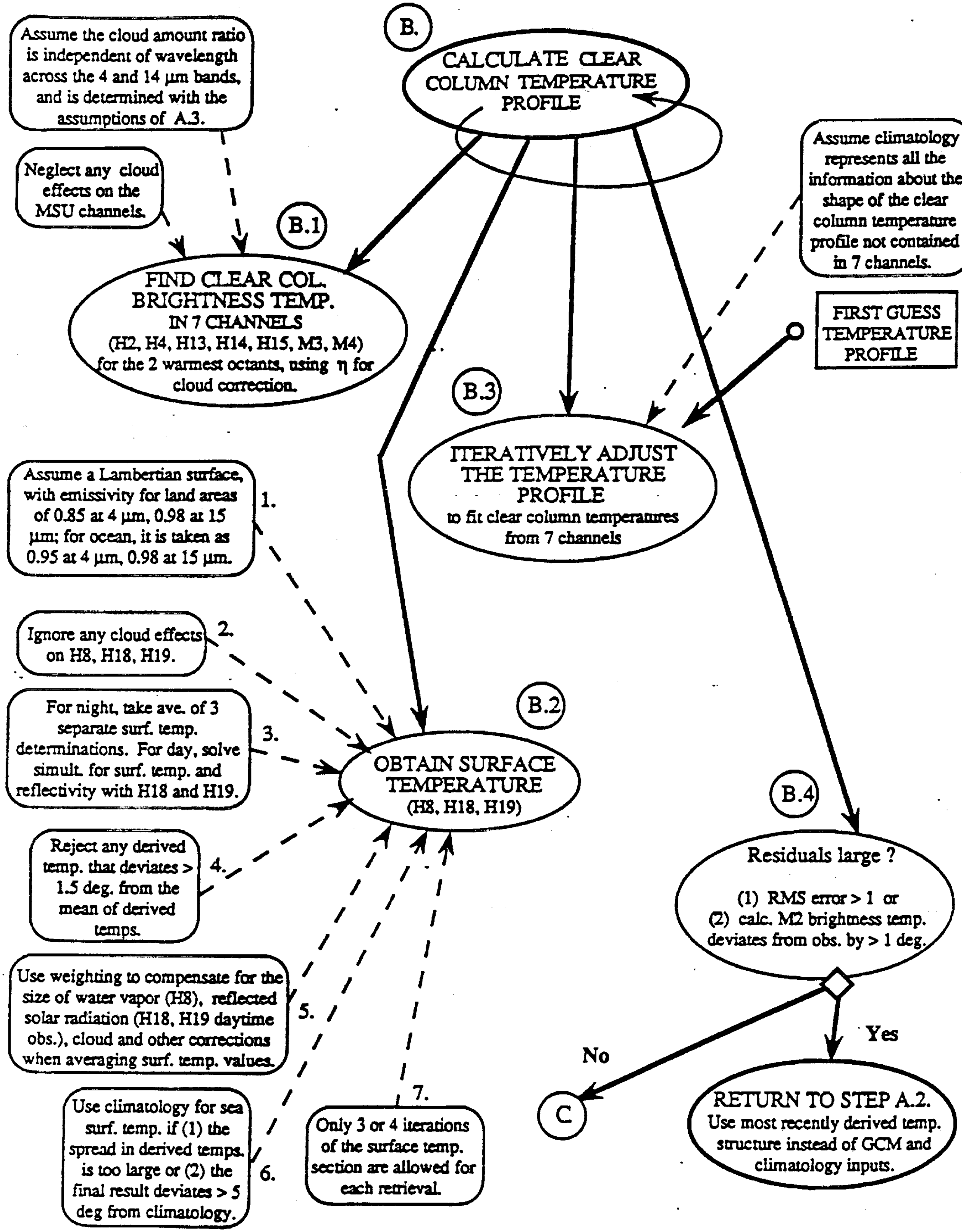


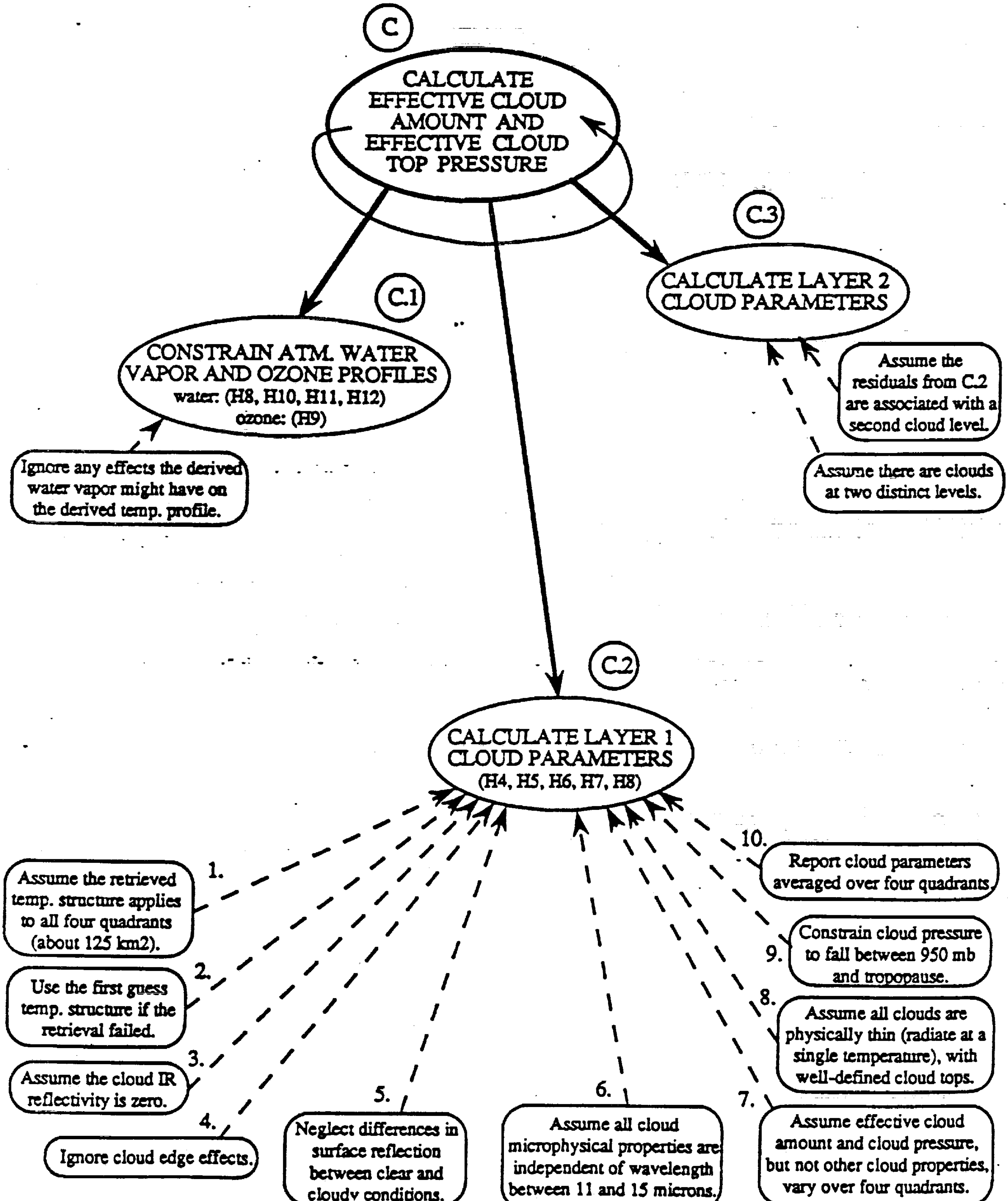
C.2.3
Limits on min and max layer thickness not implemented
Accumulate Layer Thicknesses for up to 12 layers (SUMTHK)
C.2.4
Only allow emissivity values: 0 < SEMIS < 1
Accumulate Surface Emissivity from MSU channel 1 (SUMEMI)
Reject data if water retrieval flag is bad (IWATER < 0)
C.2.5
Reject data if obs and computed Tb in Channel 8 differ by more than 3 K (ABS (DIF8) > 3)
Accumulate Column Ppt. Water, Upper Level Ppt. Water, and Surf. Pressure for Ratio (SUMPRC)
Reject data if obs and computed Tb in Channel 10 differ by more than 3 K (ABS (DIF10) > 3)
C.2.6
Limits on min and max temperatures at each level not implemented
Accumulate Temperature Profile Data at up to 20 Levels (SUMTRT)
1.
All Cloud Fractions are the sum of Cld. Frs. reported for clds. in 2 layers
C.2.7
2.
Only allow cld. fract. values in layer 1: 0 <= CLFR4 <= 1
Accumulate Day + Night, Day, and Night Total Cld. Fractions and Cloud Top Pressures, and High, Middle, and Low Cld. Frs., and albedo, in 4 quadrants (SUMCLD)
3.
Reject data if the brightness temp. fit for 5 cld. retrieval channels (FIT) > 5 K, when layers 1 and 2 are included
9.
If albedo > 1, it is set to 1 (RATIO = 100.)
4.
Reject data if the need for cld. is small (RMSNO < 5), and 2-layer cld. correction doesn't help much (FIT > 2)
8.
Skip albedo calculation if it is night (ANGSUN > 75)
5.
Set cld. = 0 if the need for cld. is small (RMSNO < 2), and 1-layer cld. correction doesn't help much (FITCK> 0.7)
6.
Set cld. = 0 if (1) cld within 150 mb of surface, (2) fit is good w/o cld (RMSNO < 3), (3) cld fract > 0.4, and (4) window channel fits well (BTD < 2)
7.
Set layer 2 cld = 0 if (1) layer 2 cld is within 100 mb of surf and layer 2 cld doesn' t help the fit (FITCK > 0.9), or (2) layer 2 cld fract > 0.4 and FITCK > 0.95

HIRS2/MSU PHYSICAL RETRIEVAL: OVERVIEW of CLOUD PARAMETER DERIVATION

last revised: 07/03/91




B.
CALCULATE CLEAR COLUMN TEMPERATURE PROFILE
Assume the cloud amount ratio is independent of wavelength across the 4 and 14 μm bands, and is determined with the assumptions of A.3.
Neglect any cloud effects on the MSU channels.
B.1
FIND CLEAR COL. BRIGHTNESS TEMP. IN 7 CHANNELS (H2, H4, H13, H14, H15, M3, M4) for the 2 warmest octants, using η for cloud correction.
Assume climatology represents all the information about the shape of the clear column temperature profile not contained in 7 channels.
FIRST GUESS TEMPERATURE PROFILE
B.3
ITERATIVELY ADJUST THE TEMPERATURE PROFILE to fit clear column temperatures from 7 channels
1.
Assume a Lambertian surface, with emissivity for land areas of 0.85 at 4 μm, 0.98 at 15 μm; for ocean, it is taken as 0.95 at 4 μm, 0.98 at 15 μm.
2.
Ignore any cloud effects on H8, H18, H19.
3.
For night, take ave. of 3 separate surf. temp. determinations. For day, solve simult. for surf. temp. and reflectivity with H18 and H19.
B.2
OBTAIN SURFACE TEMPERATURE (H8, H18, H19)
4.
Reject any derived temp. that deviates > 1.5 deg. from the mean of derived temps.
5.
Use weighting to compensate for the size of water vapor (H8), reflected solar radiation (H18, H19 daytime obs.), cloud and other corrections when averaging surf. temp. values.
6.
Use climatology for sea surf. temp. if (1) the spread in derived temps. is too large or (2) the final result deviates > 5 deg from climatology.
7.
Only 3 or 4 iterations of the surface temp. section are allowed for each retrieval.
B.4
Residuals large ?
(1) RMS error > 1 or
(2) calc. M2 brighmess temp. deviates from obs. by > 1 deg.
No
Yes
C
RETURN TO STEP A.2. Use most recently derived temp. structure instead of GCM and climatology inputs.


C
CALCULATE EFFECTIVE CLOUD AMOUNT AND EFFECTIVE CLOUD TOP PRESSURE
C.1
CONSTRAIN ATM. WATER VAPOR AND OZONE PROFILES
water: (H8, H10, H11, H12)
ozone: (H9)
Ignore any effects the derived water vapor might have on the derived temp. profile.
C.3
CALCULATE LAYER 2 CLOUD PARAMETERS
Assume the residuals from C.2 are associated with a second cloud level.
Assume there are clouds at two distinct levels.
C.2
CALCULATE LAYER 1 CLOUD PARAMETERS
(H4, H5, H6, H7, H8)
1. Assume the retrieved temp. structure applies to all four quadrants (about 125 km2).
2. Use the first guess temp. structure if the retrieval failed.
3. Assume the cloud IR reflectivity is zero.
4. Ignore cloud edge effects.
5. Neglect differences in surface reflection between clear and cloudy conditions.
6. Assume all cloud microphysical properties are independent of wavelength between 11 and 15 microns.
7. Assume effective cloud amount and cloud pressure, but not other cloud properties, vary over four quadrants.
8. Assume all clouds are physically thin (radiate at a single temperature), with well-defined cloud tops.
9. Constrain cloud pressure to fall between 950 mb and tropopause.
10. Report cloud parameters averaged over four quadrants.

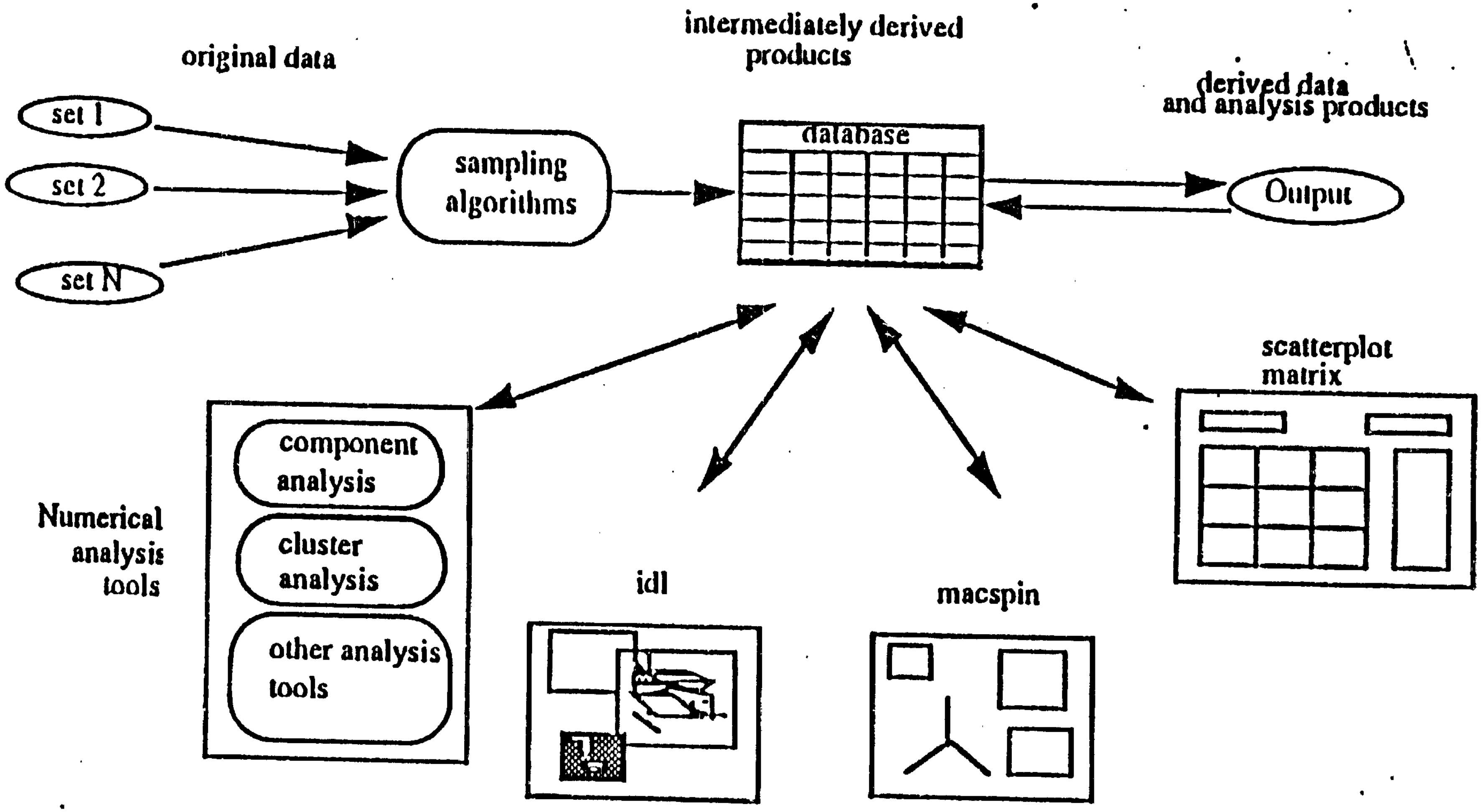


Figure 3. Interactive Data Extraction and Analysis System

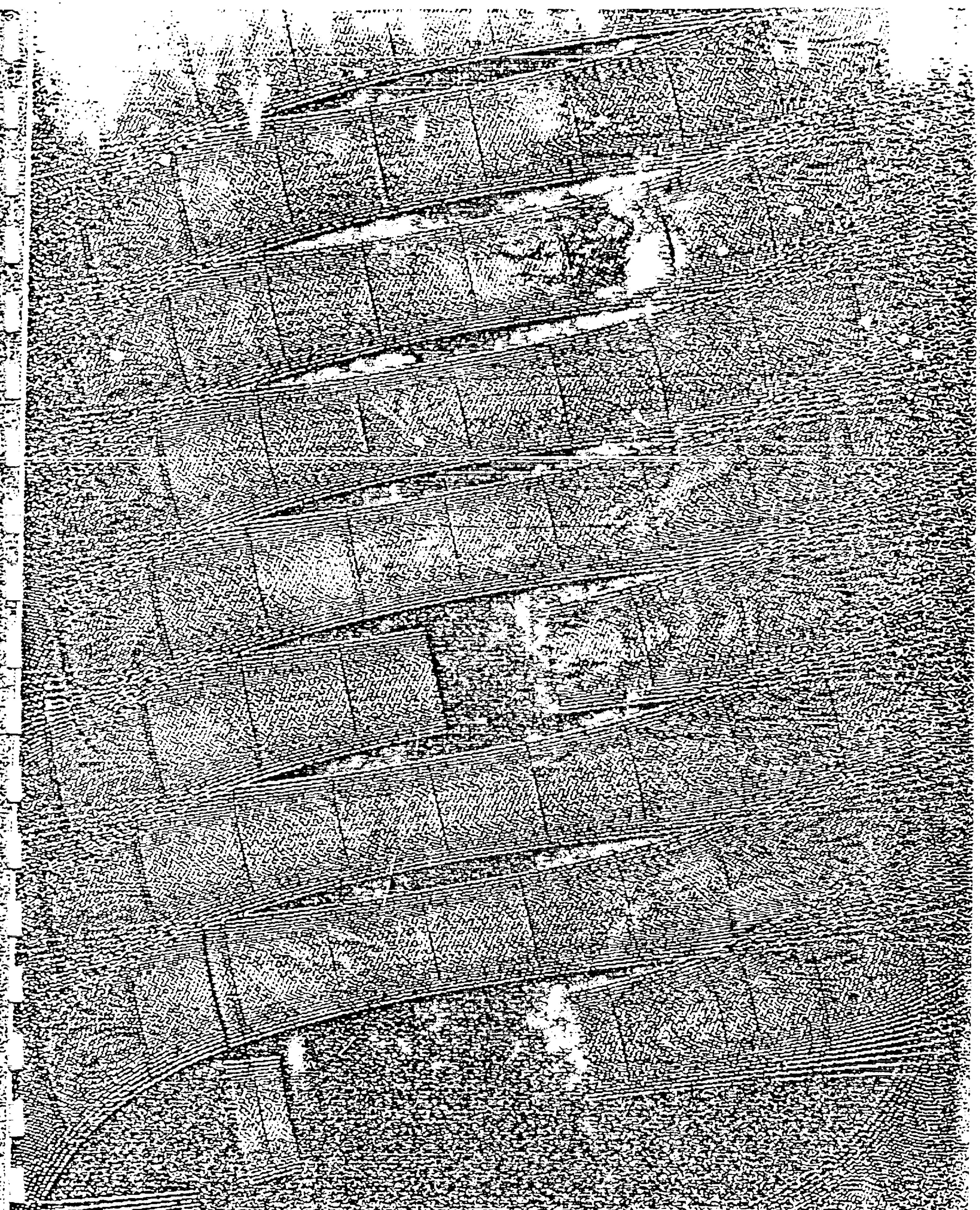
THE NEED FOR LEVEL 2 DATA

3 COLOR IMAGES SHOW HIRS2 JULY 1979 CLOUDS

(1) 2 X 2.5 DEGREE BIN,

(2) 500 X 500 KM BIN

AND (2) - (1)



ORIGINAL PAGE IS
OF POOR QUALITY

Exploratory Data Analysis Acitvities

Last Revised: 04/09/91

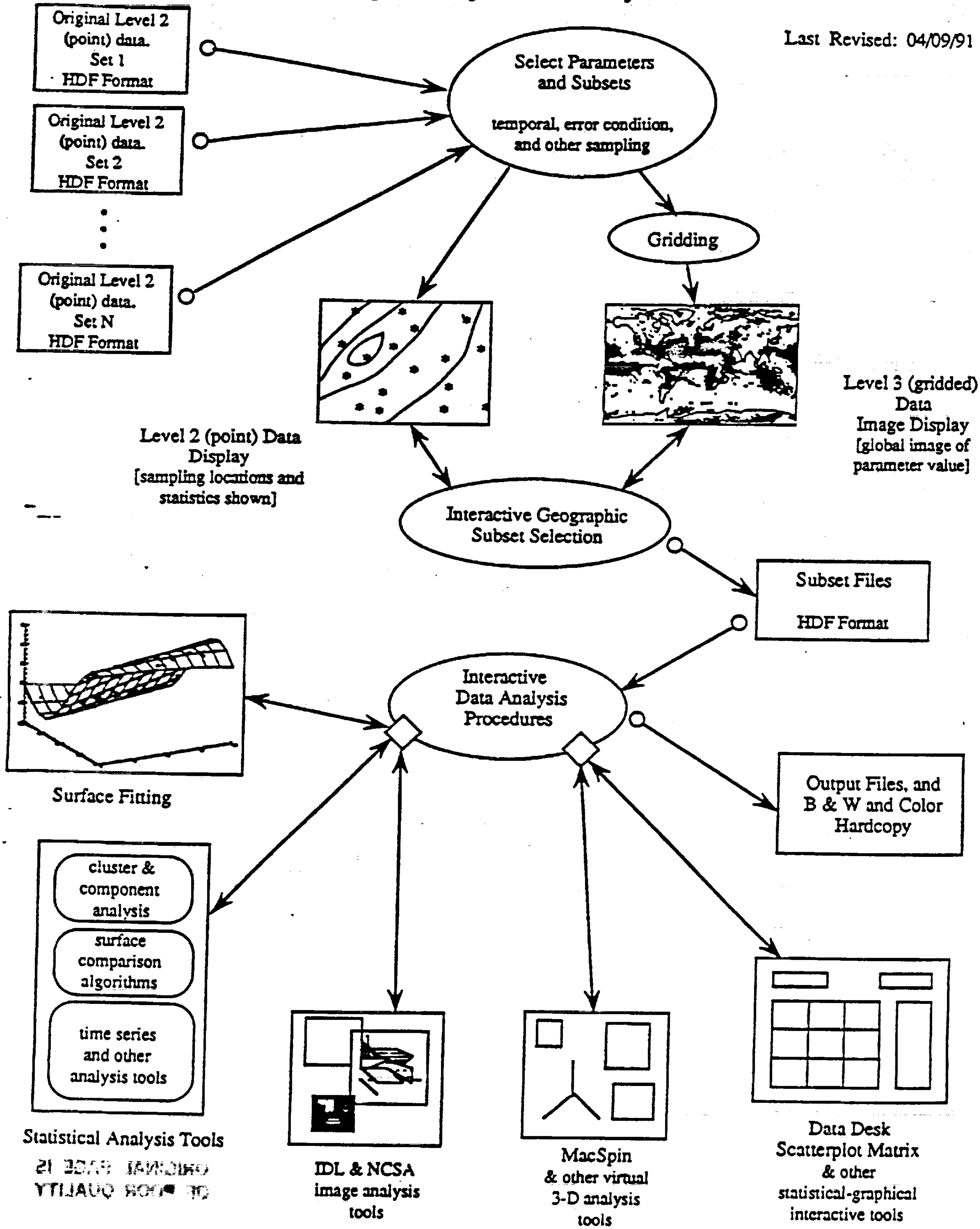


FIGURE A.6

WIGSS MENUS

last revised: 04/03/92

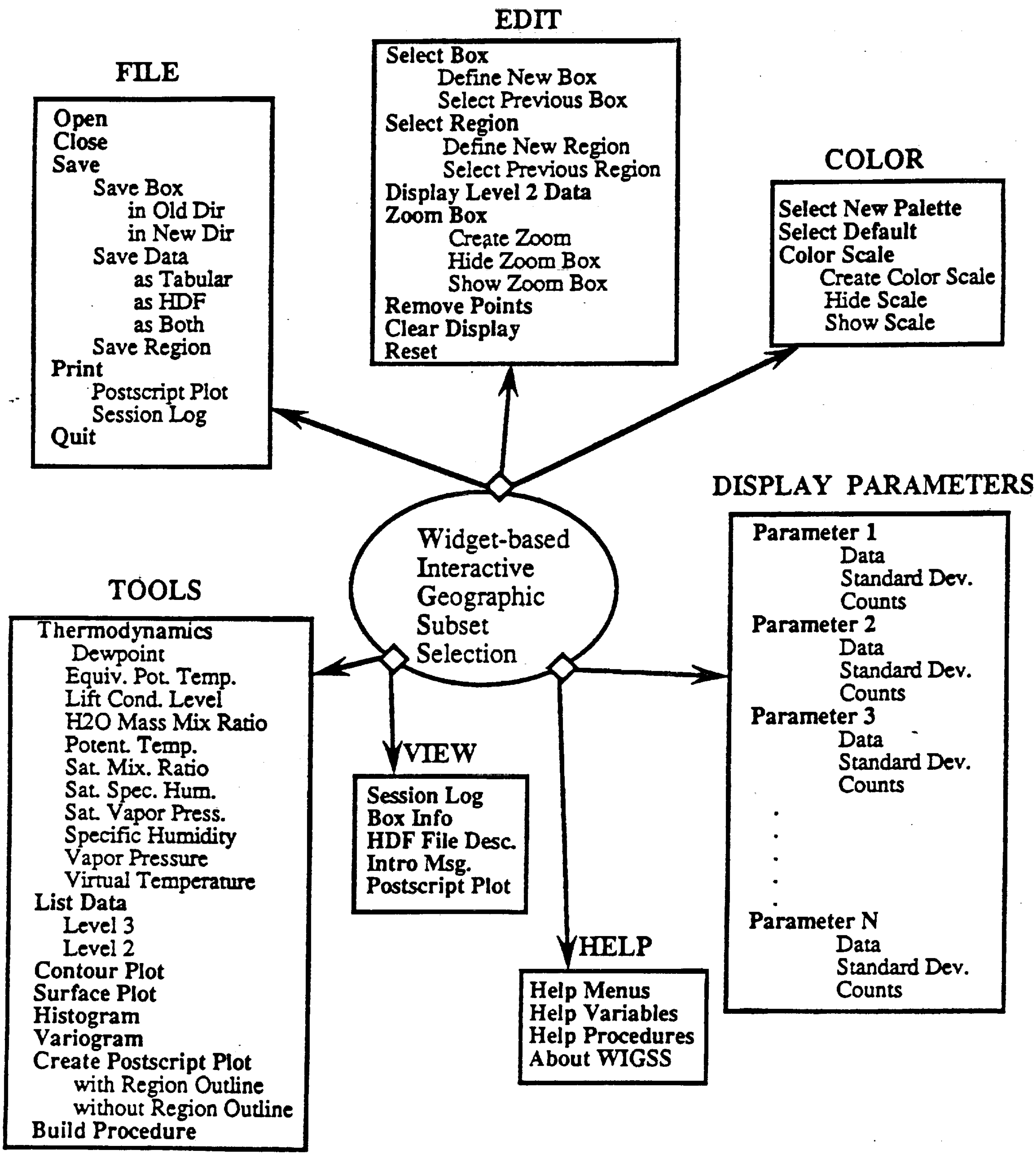
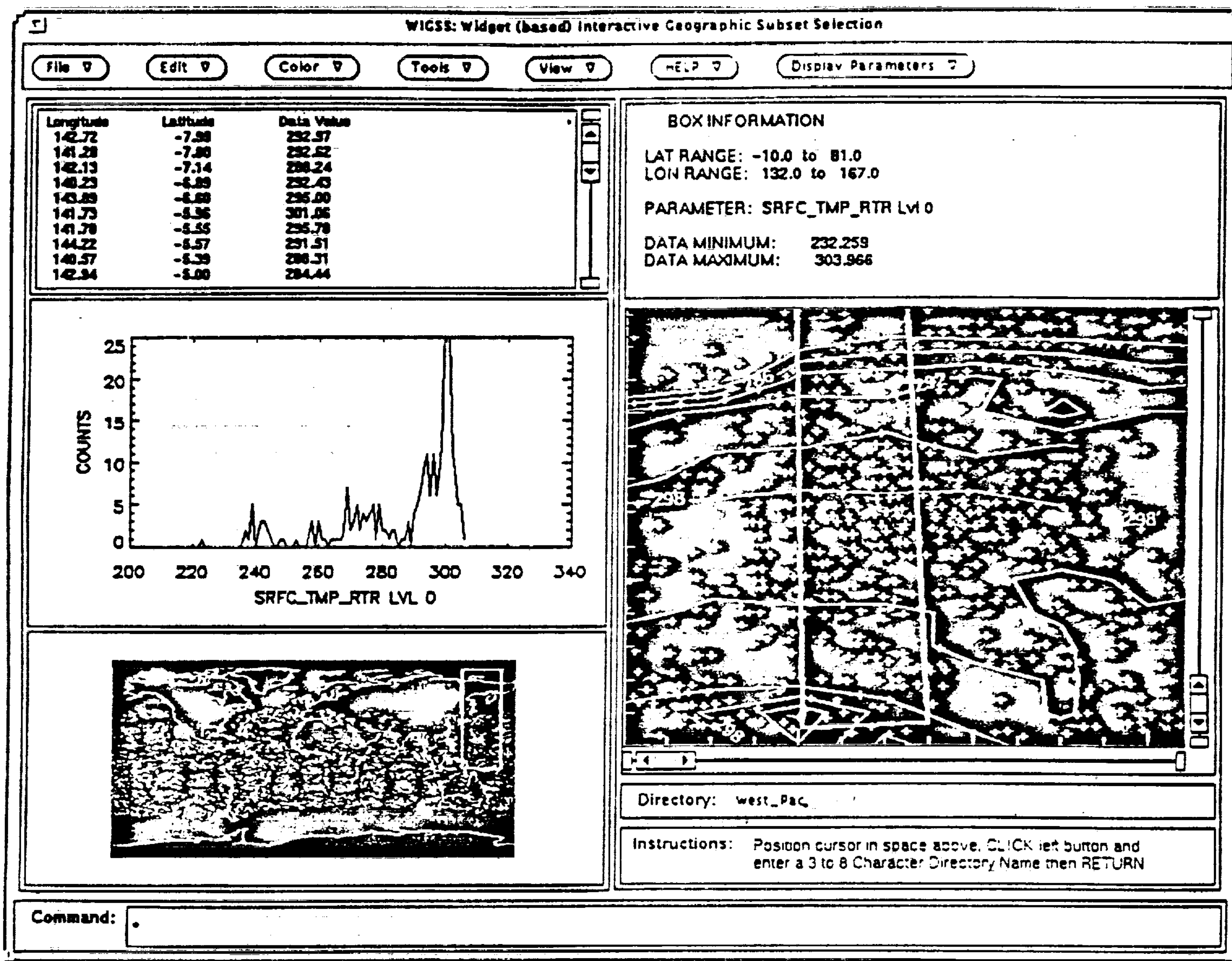


FIGURE A.7


WIGSS: Widget (based) Interactive Geographic Subset Selection
File
Edit
Color
Tools
View
HELP
Display Parameters
Longitude
Latitude
Data Value
BOX INFORMATION
LAT RANGE: -10.0 to 81.0
LON RANGE: 132.0 to 167.0
PARAMETER: SRFC_TMP_RTR Lvl 0
DATA MINIMUM: 232.259
DATA MAXIMUM: 303.966
COUNTS
0
5
10
15
20
25
200
220
240
260
280
300
320
340
SRFC_TMP_RTR LVL 0
Directory: west_Pac
Instructions: Position cursor in space above, CLICK left button and enter a 3 to 8 Character Directory Name then RETURN
Command:

EDA HIRS2/MSU STANDARD DATA FILES PROCESSING

-- Problem of multiple machine architectures

We converted our data files to Hierarchical Data Format (HDF).

[Developed at NCSA (National Center for Supercomputer Applications)]

-- Problem of data documentation
[How are the fields stored, what do they mean (units, definitions, assumptions)?]

HDF solves a part of these problems (some information about 'data objects' is stored in HDF files)

What We Have Learned About Standard Data Handling Time Scales:

-- To discover the need for HDF, learn HDF, and apply it - **~ 1 year**

-- Knowing what we now know, to rebuild from scratch - **~ 6 months**

-- To create HDF files for a different data set, of comparable complexity, in an arbitrary format -
<~ 2 months, depending on the documentation and hardware availability

-- To ingest a different data set, of comparable complexity, that is already in HDF format -
~ 2 weeks to read data, test, and to study the documentation

For data analysis, the issue of **assumptions** is a large one, not addressed in the standard data processing (discussed later).

Partial List of Software That Automatically Reads Files in HDF Format

Currently Available:

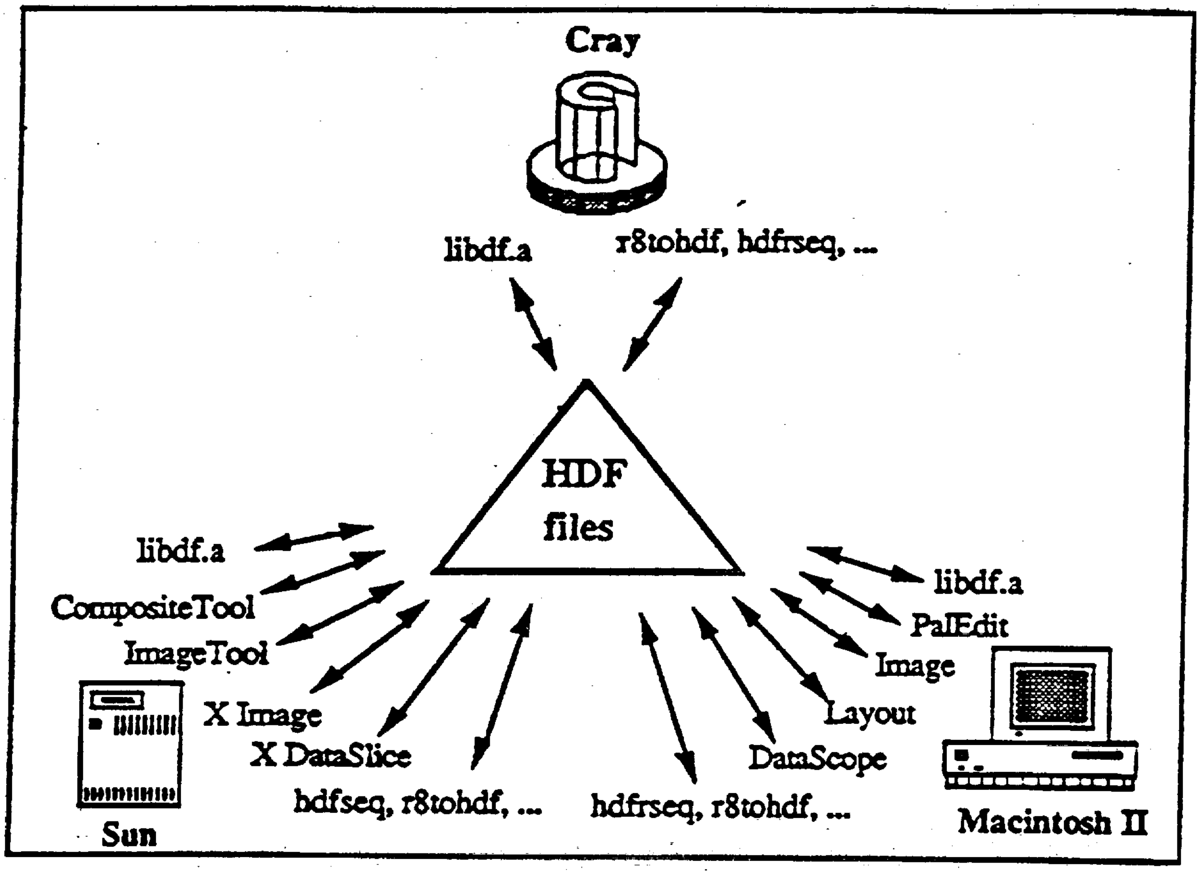
NAME	Platform	Source	Comments
Data Scope	Mac	NCSA	Display, manipulate arrays & images
Image Tool	Mac	NCSA	Display, animate image & color bar
Layout	Mac	NCSA	Create presentation from images, text
Transform	Mac	Spyglass	Combines Data Scope & Image Tool
Format	Mac	Spyglass	Similar to Layout
Dicer	Mac	Spyglass	Select & view sections of 3-D display
X-Image	Sun*	NCSA	Combines Data Scope & Image Tool
XDS	Sun	NCSA	Similar to Dicer
Reformat	Sun	NCSA	Convert FITS, TIFF, GIF, SUN, raw raster files, & x-window dumps to HDF
APE 2.0	Sun	Ohio State	Object-oriented prog. language

In Development or Testing:

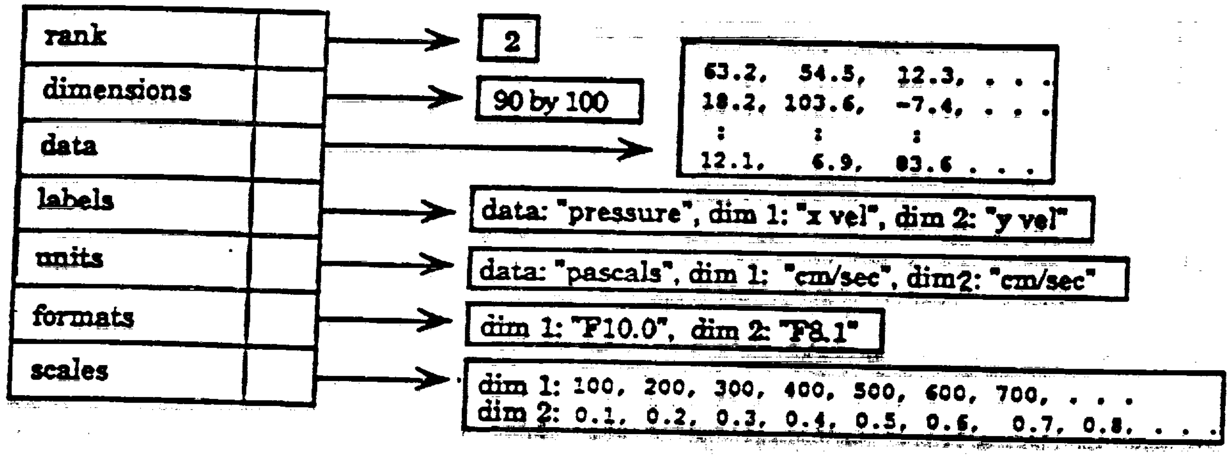
IDL	Sun	RSI	Interactive graphics prog. language
IGSS	Sun	JPL/EDA	Interactive Geographic Subset Selection
netCDF filter	Sun	NSF	Convert netCDF to HDF

* 'Sun' also runs on many other UNIX platforms, including Apollo, Alliant, Convex, Cray, DEC-ULTRIX, and IRIS Workstations.

HDF Software in an Integrated Computing Environment



HDF File with Scientific Dataset



FORTRAN:

```
INTEGER        DFSDsetdims, DFSDsetdatastrs, DFSDsetdimstrs
INTEGER        DFSDsetdimscale, DFSDputdata
REAL           press1(90,100), press2(90,100)
REAL           den1(90,100),den2(90,100)
INTEGER        shape(2), ret
REAL           xscale(90), yscale(100)

shape(1) = 90
shape(2) = 100

ret = DFSDsetdims(2, shape)
ret = DFSDsetdatastrs('pressure 1','Pascals','E15.9','')
ret = DFSDsetdimstrs(1,'x vel','cm/sec','F10.0')
ret = DFSDsetdimstrs(2,'y vel','cm/sec','F8.1')
ret = DFSDsetdimscale(1, shape(1), xscale)
ret = DFSDsetdimscale(2, shape(2), yscale)
ret = DFSDadddata('SDex4.hdf', 2, shape, press1)

ret = DFSDsetdatastrs('pressure 2','Pascals','E15.9','')
ret = DFSDadddata('SDex4.hdf', 2, shape, press2)

ret = DFSDclear()
ret = DFSDsetdatastrs('density 1','g/cm3','E15.9','')
ret = DFSDadddata('SDex4.hdf', 2, shape, den1)

ret = DFSDsetdatastrs('density 2','g/cm3','E15.9','')
ret = DFSDadddata('SDex4.hdf', 2, shape, den2)
  .
  .
  .
```

HIRS2/MSU HDF LABEL

FILE IDENTIFIER LENGTH: 5
FILE IDENTIFIER: LABEL
FILE DESCRIPTOR LENGTH: 1831
FILE DESCRIPTION:

Description: HIRS2/MSU parameters retrieved using the Goddard Laboratory for Atmospheres (GLA) Physical Inversion Algorithm Baseline 4.0. They are stored as individual objects of an HDF file. These files are the standard data source for most data analysis applications. Most of the parameters delivered on the original GSFC tapes are included. The following parameters were eliminated (either because of questions about definition, redundancy, or problems of interpretation of the values): tau; dlat; dlon; np; cldhgt; cldfrc; retwat(1); retwat(5); humret(13); rthick. Thirty_seven paramaters remain. They are listed and defined in /edal/doc/hirs_daily/rec.doc.

Level 2 data for: 06 jul 79, 0Z - 24Z. Platform: TIROS-N

Reference: Susskind, J., J. Rosenfield, D. Reuter and M. T. Chahine, 1984: Remote sensing of weather and climate parameters from HIRS2/MSU on TIROS-N. J. Geophys. Res. 89, 4677-4697.

Contact:
Robert Haskins
Jet Propulsion Laboratory
Mail Stop 183 - 301
4800 Oak Grove Dr
Pasadena, CA 91109-8001

(818)354-6893

Regional Boundaries are: Global

Number of Parameters: 37

Parameters:

YYMMDD, HHMMSS, QUADLATS, QUADLONS, DNFLAG, LANDWTR_FLAG, SAT_ZEN_ANGLE, GEOPOT_THICK, HIRS8_OBS, VIS_REFLECTANCE, SRFC_EMIS_MW, SRFC_PRES TROP_PRES_RTR, SRFC_TMP_RTR, SST_ANOMALY, TMP_PROFILE_RTR, QUAD_NUM_TMPS, QUAD_FLAG, TMP_RTR_FLAG, TB_RESIDUAL, TB_RMS_TMP, RHUM_PROF_RTR, PRECIP_WTR, WATER_FLAG, TB_RMS_WTR, HIRS8_TBDIF_WTR, HIRS10_TBDIF_WTR, CLOUD_EFRAC_L1, CLOUD_TOP_PRES_L1, CLOUD_EFRAC_L2, CLOUD_TOP_PRES_L2, RMS_ERR_INCCLD, RMS_ERR_PRECLD, CLOUD_CLEAR_PARM, HIRS8_TBDIF_CLD, OZONE_RTR, O3SENS

Comments:

Binary HDF file creation date: Mon Nov 4 16:42:31 EST 1991
Binary HDF file created on a CRAY Y-MP

SDS COUNT: 37

SDS DATA DIMENSIONS: 4 x 44821

SDS DATA LABEL: QUADLATS
SDS DATA UNITS: Degrees
SDS DATA FORMAT: F6.2
HDF OBJECT REFERENCE NUMBER: 9
HDF OBJECT DESCRIPTION:
Latitudes of four individual quadrants for cloud retrieval

--- Original Name = FLAT ---

SDS DATA DIMENSIONS: 4 x 44821
SDS DATA LABEL: QUADLONS
SDS DATA UNITS: Degrees
SDS DATA FORMAT: F7.2
HDF OBJECT REFERENCE NUMBER: 12
HDF OBJECT DESCRIPTION:
Longitudes of four individual quadrants for cloud retrieval.

--- Original Name = FLON ---

SDS DATA DIMENSIONS: 1 x 44821
SDS DATA LABEL: TMP_ERR_FLAG
SDS DATA UNITS: N/A
SDS DATA FORMAT: I3
HDF OBJECT REFERENCE NUMBER: 57
HDF OBJECT DESCRIPTION:
Error flag for temperature retrieval.
=> Positive IERR means successful temp retrieval and retrieved temp was used for water, ozone, and cloud retrieval.

=> Negative IERR means temp retrieval failed and first guess temp and moisture is used in subsequent cloud retrieval.

1000+K Converged on Kth iteration in retrieval
This parameter is always stored as 1 on the tapes that we receive from GSFC.

1100 Did not converge after 9 iterations.
This parameter is always stored as 1 on the tapes that we receive from GSFC (The information about whether or not the retrieval converged is lost.)

2 SST retrieval was not attempted over ocean, climatology SST is used.

3 Residual for HIRS2 channel 2 was large.
Ignore retrieved temperatures above 200 mb.

-4 Cloud clearing was not attempted; too cloudy to do a retrieval.

-5 Big (1 degree) RMS on Tb residual in temp sounding channels, or in MW2 channel.

-6 Not used.

.
.
.

EDA HIRS2/MSU STANDARD DATA FILES PROCESSING

Hierarchical Data Format

We have developed software that:

1. Automatically moves HIRS2/MSU physical retrieval data from the IBM tape archive to the GSFC Cray

2. Automatically converts the data into HDF format, including adding file labels and detailed parameter descriptions

3. Automatically transfers the HDF files to a user-specified remote node via the FTP utility

We also have some standard utilities, and there is software in development, that takes HDF files and

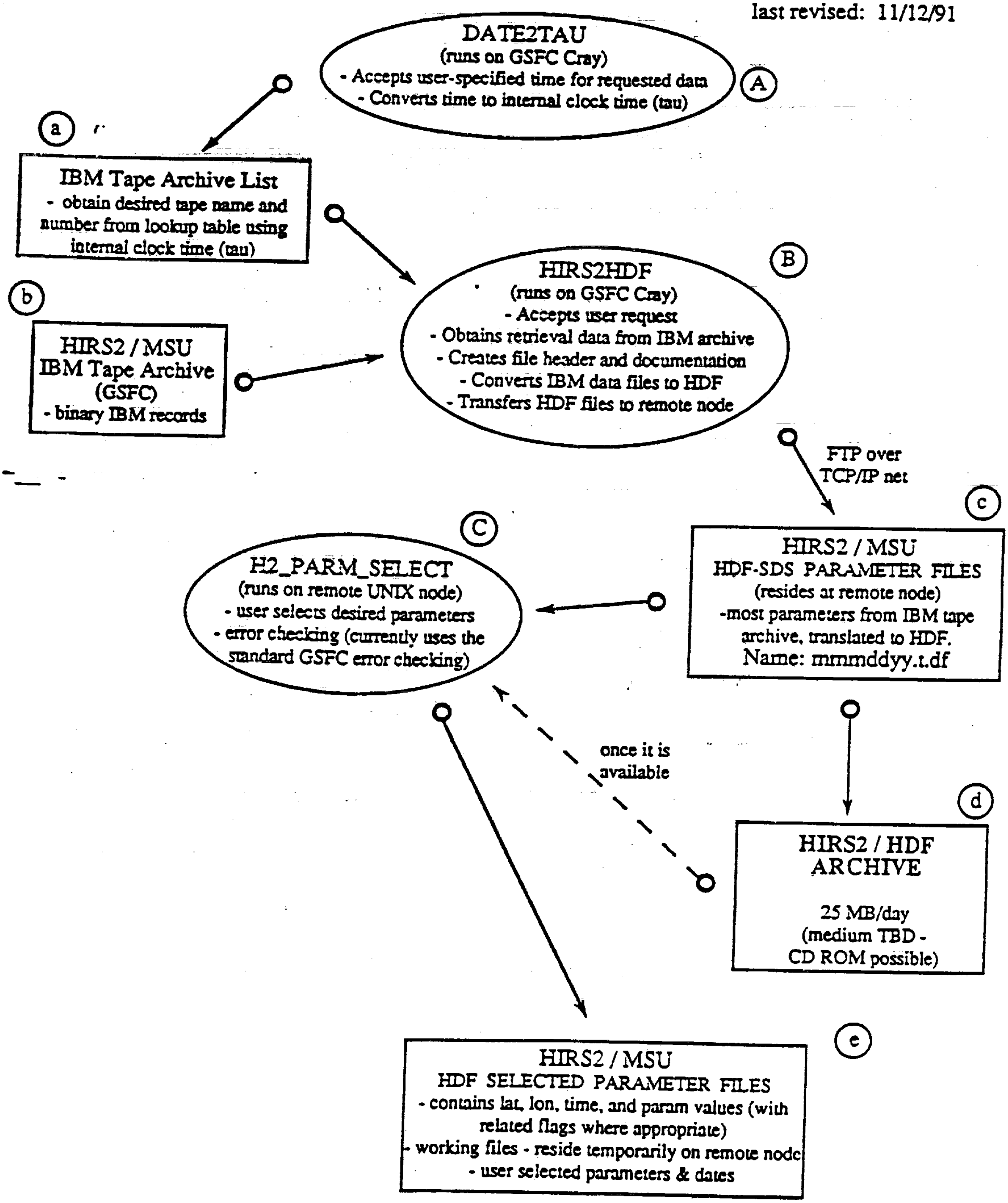
displays HDF label information

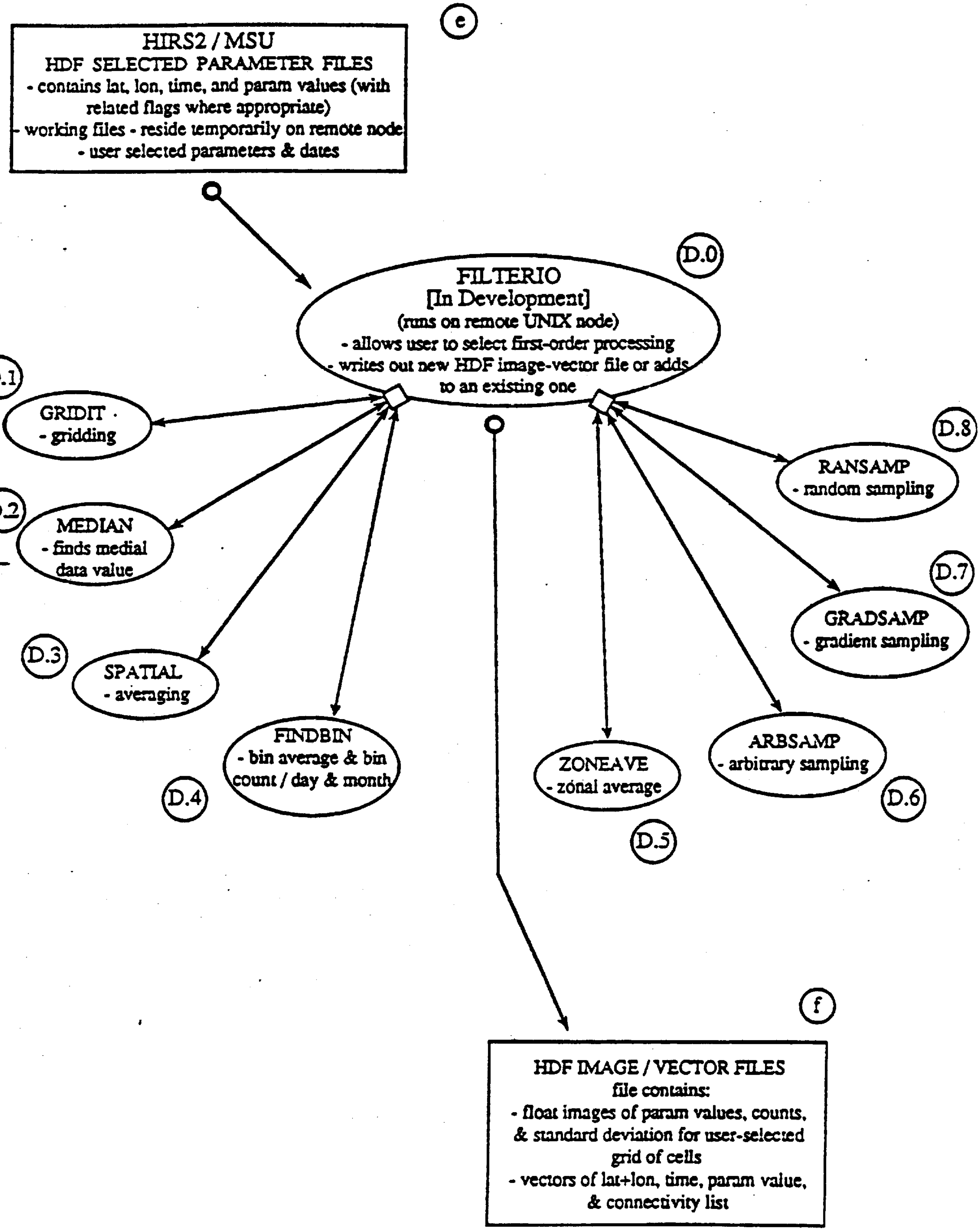
creates floating point image data from HDF vector data

displays floating point image HDF files and performs several kinds of analysis

EDA HIRS2/MSU STANDARD DATA FILES PROCESSING

last revised: 11/12/91





e
HIRS2 / MSU
HDF SELECTED PARAMETER FILES
- contains lat, lon, time, and param values (with related flags where appropriate)
- working files - reside temporarily on remote node
- user selected parameters & dates
D.0
FILTERIO
[In Development]
(runs on remote UNIX node)
- allows user to select first-order processing
- writes out new HDF image-vector file or adds to an existing one
D.1
GRIDIT
- gridding
D.2
MEDIAN
- finds medial data value
D.3
SPATIAL
- averaging
D.4
FINDBIN
- bin average & bin count / day & month
D.5
ZONEAVE
- zonal average
D.6
ARBSAMP
- arbitrary sampling
D.7
GRADSAMP
- gradient sampling
D.8
RANSAMP
- random sampling
f
HDF IMAGE / VECTOR FILES
file contains:
- float images of param values, counts, & standard deviation for user-selected grid of cells
- vectors of lat+lon, time, param value, & connectivity list

SUMMARY

Validation Issues

Statistical characterization of data sets

Finding statistics that characterize key attributes of the data sets

Defining ways to characterize the comparisons among data sets (Scale issues, statistics,...)

Selection of specific intercomparison exercises

Selecting characteristic spatial and temporal regions for intercomparisons

Impact of validation exercises on the logistics of current and planned field campaigns and model runs

Preparation of data sets for intercomparisons

Characterization of assumptions

Transportable data formats

Labeling data files

Content of data sets

Data storage and distribution (EOSDIS interface)