Statistics of the outgoing thermal radiation: Comparison between a GCM and satellite data.

Robert D. Haskins, Richard M. Goody and Luke Chen

California Institute of Technology /Jet Propulsion Laboratory, Pasadena, California

ABSTRACT

The motivation for this paper is to understand better the means available for calibrating climate models. The approach that we describe is based on data that can be obtained easily and at low cost, and it is an important candidate for testing the quality of climate models.

We compare the statistics of observed, outgoing, thermal spectra with those predicted from a climate model, for a period of approximately one year. This is a powerful approach to calibration of a model with respect to processes internal to the atmosphere. These processes, which have time scales less than a year, define the atmosphere's response to external forcing. Second-moment statistics are particularly important for testing model variability, which is central the problem of climate forcing.

We present comparisons between statistical data from IRIS, an orbiting Fourier transform spectrometer, and spectra calculated using the medium-resolution spectral code, MODTRAN, applied to the temperature and humidity profiles from a climate model, the European Center Hamburg Model (ECHAM3). Ten months of IRIS data are available, and we have compared means, standard deviations, skew and kurtosis for brightness temperatures in three tropical regions, for individual months, and for a range of time scales. The spectra are presented with a resolution of 2.8 cm^{-1} in a way that minimizes possible errors from MODTRAN.

Important differences exist in all categories of IRIS and ECHAM comparisons, demonstrating that the spectral data can provide a severe test of many aspects of the variability of a general circulation model. We discuss some of the residuals, and how they may be used to improve model performance in the context of an adjoint formalism. In the long run, the only way to have confidence in the performance of a model is to subject it to as many discriminating comparisons with data as are practicable, and this is a good candidate.

NASA's second Earth Observing System mission will have the desired characteristics; Japanese and European investigators are either flying or planning to fly Fourier transform spectrometers; and in the U.S. plans are being developed for low-cost missions with good spectrometric capabilities.

To be realistic about instrumental capabilities we study spectra from the IRIS mission that, in 1970/71, provided 10 months of data from a Fourier transform spectrometer. With modern instrumentation we can, of course, improve on the performance of IRIS.

2. THE DATA

We compare climate calculations from the European Center Hamburg Model (ECHAM3) with satellite observations (IRIS-D). The IRIS data were obtained during a 10-month period from April 1970 to January 1971, but no appropriate climate prediction is available for that time period. To investigate the possibility of validation of a GCM using satellite data we were obliged to compare IRIS data to ECHAM3 for a different epoch, chosen to be as similar as possible to the IRIS period. This means that no valid conclusions can be drawn from the comparison, but the possibilities for testing GCM calculations from this type of data are illustrated.

Comparable IRIS and ECHAM time periods were chosen on the assumption that the main source of variability in the tropical regions is the ENSO cycle. Figure 1 shows the sea-surface temperature anomalies in the tropical Pacific for 1970 and 1988, and the phasing between cycles that we have assumed.

a. IRIS-D

IRIS-D was a Fourier transform spectrometer flown during 1970 and early 1971 on Nimbus 4 (Hanel et *al.*. 1972). The spectral range was 400 to 1600 cm^{-1} and the apodized spectral resolution was 2.8 cm-l. The spectra were calibrated with an onboard black body. IRIS was a zenith-pointing sounder with an image-compensated footprint of 95 km. The orbit was sun-synchronous with equator crossings at OZ and 12Z. The signal-to-noise ratio at the mid-point of the spectrum was approximately 100 to 1. This ratio deteriorated at higher frequencies and we could only make practical use of the data below 1400 cm^{-1}. Also, much of the information in the region of 1400-1600 cm^{-1} was replicated in the region 400-600 $c\ m^{-1}$.

For the 10-month operational period of IRIS, from April 1970 through January 1971 a total of 700,000 spectra were collected. This is a large data base of calibrated spectra, a unique resource for investigating climate validation. The average brightness temperature (T_b) for the entire observation period and for the three tropical regions we have chosen to examine (see § 2.5.3) is shown in figure 2, and compared to ECHAM3 data. The spectrum is dominated by a wing of the water-vapor rotation band, 400 to 600 cm^{-1}, the carbon dioxide fundamental band, centered on 667 cm^{-1}, the window region, 800 to 1000 cm^{-1}, the ozone fundamental centered on 1042 cm-l, a wing of the water-vapor fundamental band, 1200 to 1400 cm^{-1}, and the Q-branch of the methane fundamental at 1311 cm-l.

The lower panel in figure 2 shows the pressure level of the peak of the emission to space. To a rough-and-ready degree of approximation, the outgoing radiances from opaque regions (not the ozone band) may be regarded as originating from

from our preliminary studies that the ozone observed by IRIS and that supplied by MODTRAN are different, and no conclusions can be drawn from a comparison. The carbon dioxide concentrations in 1970 may differ by a few percent from those used in MODTRAN, but calculations made by us for other purposes (Goody, *et al.*, 1995) show that the error in the outgoing radiance is negligible for the purposes of this paper.

d. Climate statistics

The mean value is only one aspect of a statistical distribution of data. Aspects of the distribution other than the mean are also of importance for judging the success of a climate prediction (Gates, 1995); for example a distribution with two well-defined maxima might indicate the presence of alternative atmospheric states for the same imposed conditions, a question that has been raised periodically in the meteorological literature. The variability, as indicated by the standard deviation, is central to climate prediction (see Polyak, 1996, for a supporting view). It is the result of atmospheric feedbacks and interactions responding to changing conditions: similar interactions and feedbacks govern the steady-state climate. Because of the spectral character of IRIS data we may observe the variability of atmospheric temperature at all relevant levels, of water vapor in the troposphere, and of ozone in the stratosphere. A successful climate model should encompass these data by accounting for the observed spectral radiances.

In addition to the standard deviation, we have evaluated the skew and kurtosis of observed brightness temperatures,

$$Sk = \frac{1}{\sigma^3}\overline{(x-\bar{x})^3},$$

$$Ku = \frac{1}{\sigma^4}\overline{(x-\bar{x})^4} - 3.$$

σ is the standard deviation. Skew and kurtosis are zero for a normal distribution.

Figure 3 shows four histograms of data, together with numerical values for the standard deviation, the skew, and the kurtosis, for four frequencies, and for both IRIS and ECHAM data. It is evident from these data that the model predictions differ from the observed data in important respects. For example, at 1339 cm-l, ECHAM fails to account for a bimodal distribution and, at 418 cm-], ECHAM has a marked skewness that is not in the data. Meteorologists and climatologists do not commonly use skew and kurtosis as climate indicators and we have little experience of their interpretation in terms of model characteristics. Skew can be increased by a non-linear process, such as temperature effects on water-vapor density; kurtosis could be caused by a positive feedback that depends upon the amplitude of a disturbance.

Skew and kurtosis are higher order moments, and they may both be strongly effected by a single faulty data point. It is partly for this reason that we employ a 3-sigma rejection criterion (see §2.5. 1). This is where spectral structure can be of great value: if the data show the structure of atmospheric gaseous absorption it is unlikely that they are the result of random events.

e. Selection of data

all data in their Western Pacific region. As might be expected, their calculation corresponds closely with ours for ATb=6 K.

In the remainder of this paper we shall use the ΔT_b=6 K selection criterion. The decrease of standard deviation in the window region if we go from ΔT_b=6 K to ΔT_b=4 K (see figure 4) suggests that we have not eliminated all influence of clouds and aerosols. However, the number of spectra available for the narrower selection is too few for the purposes of this paper.

A "cloud-free" criterion is also applied to the ECHAM data. We select data with less than 10°/0 cloudiness; the MODTRAN calculation is performed without clouds.

3) Geographic regions

The tropics is an important region in which to study the relation between observation and prediction. Here, seasonal changes are small, and the IRIS data set is more coherent than for higher latitudes. We chose three regions for study, the Indo-Pacific warm pool (WP), lying between 10N and 10S and between 90E and 1 50E, a central Pacific region (CP) lying between 10N and 10S and between 130W and 180W, and a region in the Indian Ocean (10) lying between 10N and 10S and between 50E and 90E.

Table 1 gives the number of spectra that were used in each region after the selection criteria of $$2.5.1 and 2.5.2 were applied compared to the original number of spectra. The large decrease in usable spectra in the warm pool region is caused by cloud contamination.

Table 1: Numbers of spectra used in this study.

Geographical region	number used	number available	% used
Warm Pool	723	11319	6.4
Central Pacific	5005	12234	40.9
Indian Ocean	2504	9503	26.3

3. SPECTRAL STATISTICS

a. Breakdown into geographic regions

Brightness temperature. In figure 5 the data that were displayed in figure 2 are broken down into the three regions, Warm Pool, Central Pacific and Indian Ocean, and presented as differences between whole mission mean and the regional averages of ECHAM and IRIS. This differencing eliminated MODTRAN errors and also any calibration errors in IRIS. It does so at the cost of introducing hybrid parameters that depend upon conditions in all three geographic regions.

The three regions are similar in the stratosphere, as would be expected. The amplitude of the difference of the IRIS brightness temperature is approximately twice as large in the water vapor regions as the amplitude of the ECHAM in the WP and CP regions. This indicates that the ECHAM does not contain as much water

c. Time scales.

The variance may be partitioned to reflect the principal sources of variance: geographic; weather; and seasonal. We averaged the data into bins of 1-, 5-, and 30-day duration for all data in a geographical region, and calculated standard deviation, skew and kurtosis from the binned data. The results are presented in figures 13, 14 and 15, for the Central Pacific only and for the entire data set. Also replotted for comparison are the standard deviations for individual spectra (point) from figure 6.

Standard deviation is, by definition, *largest.* for the *shortest* averaging times and vice *versa.* This order is followed in figure 13. 1-day bins may contain many spectra. As a result, the standard deviations for point are larger than the binned data sets. The point and 1-day average are similar for ECHAM and IRIS, but there are large differences for the 30-day averages. ECHAM shows very little difference between 1-day and 30-day averages in the Central Pacific. However, in the Indian Ocean and Warm Pool, this behavior is not observed.

Standard deviations in the Warm Pool differ from the Central Pacific. For IRIS, the most important differences are in the window region where they are much larger in the Warm Pool. For 1-day averages the standard deviation is 1.5K at 800 cm^{-1} and 1K at 1000 en--l. For ECHAM the most important difference is that, in the Warm Pool, the 1 -day and 30-day averages are not close; the 30-day average is typically one half to one third of the 1-day average.

There is enough similarity between the ECHAM and IRIS to show this is not just an instrument sampling issue and that information is present in the plots. Investigation of the sampled latitude vs time and longitude vs time, clearly shows the weather patterns moving though; the most visible is the Madden Julian oscillation and the movement of the ITCZ.

4. DISCUSSION

This discussion is structured in terms of a comparison between a model and satellite data, illustrating the possibilities of testing model behavior. The model data and the satellite data are for two different ENSO cycles, and geographic and temporal sampling both differ. To evaluate the magnitude of the uncertainties that these differences can create, we have compared data from ECHAM runs for 1987 and 1988, as an example of interannual variations, and we have compared 12Z and OZ data from ECHAM and AM and PM data from IRIS to estimate diurnal variations. Rather than present all of the data we have made a qualitative estimate of the amplitudes of interannual and diurnal variations, and these are presented in Table 2, together with an estimate of the precision. The precision is scarcely influenced by the instrument noise. With large data sets this contribution to uncertainties can usually be neglected, greatly simplifying the design requirements on a climate observing system.

We now comment briefly on the data in figures 5 to 15, and what they may mean for testing a numerical model. The data from the entire mission, figures 5, 6, 7 and 8, show enough similarities between ECHAM and IRIS and among the three geographic regions to suggest that our procedures are reproducible. The standard deviation for the Central Pacific, shown in figure 6, may be compared to an independent evaluation of IRIS data for the Western Pacific region of Iacono and Clough, (1996) see their figure 4b. The two regions correspond fairly well; the IRIS standard deviations agree well between the two investigations. The data show some differences between ECHAM and IRIS that differ among the three regions,

similar magnitude for most of the spectrum in the Central Pacific; the resemblance is not so great in the Warm Pool.

The "weather" variance is principally represented by the difference between the 1-day and the 30-day averages. Here we find remarkable differences between IRIS and ECHAM. ECHAM shows almost no "weather" variance in the Central Pacific, while the IRIS data show some, although it is less than the geographic variance. ECHAM is a climate program that may not treat the short-term atmospheric changes with appropriate detail. in the Warm Pool and Indian Ocean this contrast no longer holds. ECHAM shows about the same level of "weather" variance as does IRIS in the Central Pacific, but the IRIS "weather" variance is now much greater than either. These are important aspects of model behavior, but even more important is the seasonal variance, represented by the 30-day averages. IRIS and ECHAM differ considerably in the Central Pacific in the window and in groups of water lines. This is surprising, for it implies that ECHAM seasonal behavior is incorrect. Haskins *et al.* (1 996), studying the moisture flux, have shown similar results. When we look at the Warm Pool results, however, the picture looks better, The seasonal variance of ECHAM and IRIS agree fairly well, even though the shorter-period contributions to the variance do not. This feature has also been substantiated by Haskins et *al., (1* 995).

There is no simple relationship between skew, kurtosis and the time averaging that we may anticipate, as there is for the standard deviation. At first sight figures 14 and 15 appear to be very noisy, but this is not so. Most of the strong features are correlated with spectral features, e.g., the features at the 1311 cm^{-1} methane Q-branch, suggesting that they are connected to rapid changes with height and reflect real changes in the atmosphere. In addition, each curve has a low-noise region, and there is no reason why random errors should have a spectral character.

For both IRIS and ECHAM, and for both the skew and kurtosis, the numerically smallest values are associated with the statistics of individual spectra. We have indicated previously that individual spectra have large regional variance, and this result indicates that regional variance, unlike weather variance, is close to Gaussian.

One of the most striking differences between IRIS and ECHAM is shown by the kurtosis of the 1 -day averages. The strong structure shown by IRIS appears to be real, in contrast to the corresponding weak structure of the ECHAM kurtosis. As indicated by the methane Q-branch, the kurtosis of variations in the upper troposphere is approximately 8, which is only likely for a bimodal distribution.

In almost every category that we have considered there are differences or residuals between IRIS and ECHAM that exceed the uncertainties listed in table 2, suggesting that ECHAM is in error with respect to essential aspects of model variability. We now discuss the possibility of adjusting (tuning) the physical processes involved in the model to give improved performance. Before doing so should point out that the uncertainties listed in table 2 may be almost completely eliminated in a future mission, greatly increasing the discrimination of the procedures. In the future, we shall be able to run model projections for the actual period of the observations, and at the same time of day as the satellite station crossings. This should eliminate entirely the interannual and diurnal uncertainties listed in the table. Spatial scales for both model and observed data sets are on the order of 100 - 300 km, and if enough data are collected it will be possible to average the satellite footprints so that they differ only slightly.

For the following discussion we assume that significant differences between IRIS and ECHAM do in fact exist, although there are also several similarities between the two data sets. This general view is also reflected in the work of Polyak (1 996) who compared statistical second-moments (standard deviations, space- and time-lag correlations) for observed surface temperatures with 1000mb predictions from ECHAM. The data were averaged over individual months and were binned by

with errors in the ERBE data reflects the fact that it is easier to calibrate a spectrally dispersing instrument than one with band passes limited by filters.

The differences between IRIS and ECHAM data in table 3 are significant. If this were ERBE data, however, the difference would not be significant. Moreover, if the difference were detected in the ERBE data, no information as to the causes would be available.

To take an example, the contrast between the Warm Pool and the Central Pacific is, from table 3, 6.8 W m^{-2} for IRIS and 2.4 W m^{-2} for ECHAM. The breakdown into bands in table 4 suggests that a main cause of this discrepancy is the behavior of the water-vapor rotation band, and that ECHAM may be systematically wrong in its predictions of humidity.

Table 3: Integrated thermal flux, 400 cm-] to 1400 cm^{-1}, W m^{-2}.

	Warm Pool	Central Pacific	Indian Ocean
IRIS	236	242.8	238.1
ECHAM	243.5 (302.3)	245.9 (305.1)	245.1 (303.9)

Table 4: Spectral components of the data in table 3, W m^{-2}.

	400-588cm-1 water	590-756cm-1 stratosphere	758-987cm-1 window	989-1069cm-1 ozone	1071-1400cm-1 water
Warm Pool					
IRIS	60.9	34.4	78	17.4	45.4
ECHAM	63.2	36.2	82.2	15.7	46.4
Central Pacific					
IRIS	**65**	**35**	**79.1**	**17.2**	**45.6**
ECHAM	64.9	36.5	82.1	15.7	47
Indian Ocean					
IRIS	**62.1**	**34.6**	**78.6**	**17.3**	**44.7**
ECHAM	63.2	36.4	82.8	15.9	46.9

5. CONCLUSIONS

Oreskes *et al., (1994)* argue that, strict] y speaking, it is not possible to verify a climate model, in the sense of proving it to be true; and validation, when the term is used as it usually is, in the same sense, equally unobtainable. The argument is epistemological but is connected to a reality in climate research, namely that we cannot gather enough data to form a probabilistic basis for 25-year forecasts in the way that we can for 1-day forecasts. To claim that we may be able to verify or validate a climate projection is simply to overstate the case. The alternative is to aim for an engineering level of confidence by subjecting models to all of the most rigorous test that we can reasonably afford to employ.

This paper discusses an approach to testing climate models based on the outgoing spectral radiance, such as can be measured from a satellite. Radiance measurements may be made with good signal-to-noise, with absolute calibrations, and the instruments may be deployed from reasonably inexpensive spacecraft. 1 cm^{-1} is an achievable resolution that carries a great deal of information about the climate system, particularly from those levels that are most important to the

REFERENCES

Barnett, T. P., 1986: Detection of changes induced in the global tropospheric temperature field induced by greenhouse gases, *J. Geophys. Res.*, **91, 6659-6667**

Barnett, T.P. and M.E. Schlesinger: 1987, Detecting changes in global climate induced by greenhouse gases, *J. Geophys Res.*, 92, 14772-14780.

Chou, M-D, 1994: Coolness in the Tropical Pacific during an El Nino episode, *J. of Climate,* 7, 1684-1692.

Clough, S.A. and M.J. Iacono, 1995: Line-by-line calculation of atmospheric fluxes and cooling, rates II: application to carbon dioxide, ozone, methane, nitrous oxide, and the halocarbons, *J. Geophys. Res.*, **100,** 16519-16536.

Gates, W. L., 1992: AMIP: The atmospheric model intercomparison project, *Bull. Am. Met.Soc.*, 73(12), 1962-1970.

Gates, W. L., 1995: The validation of atmospheric models, Global Change. Luxemburg: Office for Official Publications of the European Communities, 218-232.

Goody, R. M., R. D. Haskins, W. Abdou, L. Chen, 1995: Detection of climate forcing using emission spectra, *Earth Observation and Remote Sensing, 5,22.-33.*

Hanel, R. A., B. Schlachman, D. Rodgers, and D. Various, 1971: The Nimbus 4 Michelson interferometer, *Appl. Opt., 10, 1376-1382.*

Hanel, R. A., B.J. Conrath, V.G. Kunde, C. Prabhakara, I. Revah, V.V. Salomonson, and G. Wolford, 1972: The NIMBUS 4 infrared spectroscopy experiment 1: calibrated thermal emission spectra, *J. Geophys. Res.*, 77, 2629-2641.

Hansen, J., A. Lacis, R. Ruedy, M. Santo and H. Wilson, 1993: How sensitive is the world's climate?, *Nat. Geogr. Res. Explor.*, 2, 142-158.

Hansen, J., A. Lacis, R. Ruedy, and M. Sante, 1992: Potential climate impact of Mount Pinitubo eruption, Geophys. *Res. Let., 19,215-218*

Haskins, R. D., T. Barnett, M Tyree, and E. Roeckner, 1995: Comparison of Cloud Products from AGCM, In Situ and Satellite Measurements, *J. Geophys. Res., 100,* D1, 1367-1378.

Haskins, R, D., E.J. Fetzer, T.P. Barnett, M. Tyree, and E. Roeckner, 1996: On the Global Transport of Moisture: Comparison of Different Estimators, submitted to *J. of Climate.*

Hall, M.C.G, and D.G. Cacuci, 1983: Physical interpretation of adjoint functions and sensitivity analysis of atmospheric models, *J. Atmos Sci.*, 40, 2537-2566.

Hasselmann, 1993: Optimal fingerprints for the detection of time-dependent change, *J. Climate, 6,1957-1971.*

Iacono, M.J. and S.A. Clough, 1996: The application of IRIS clear sky spectral radiance to investigations of climate variability y, submitted to *J. Geophys. Res.*

Wigley, T.M.L., and T.P. Barnett, 1990: Detection of the greenhouse effect in the observations, Climate Change, IPCC Scientific Assessment, 1990, *Cambridge University Press, 243-254.*

Figure 1: The sea-surface temperature anomaly in the tropical pacific during 1970-71 and 1988-89.

Figure 2. Top panel, brightness temperatures measured from IRIS and calculated from ECHAM data. The spectra are averaged over all 10 months of observations and over three tropical regions. The lower panel shows the pressure level of the maximum emission to space for each wavenumber.

Figure 3. Histograms of brightness temperature for four frequencies for both IRIS and ECHAM. $\bar{T}_b$ is the mean brightness temperature.

Figure 4. Brightness temperature, standard deviation, skew and kurtosis for all data in the Central Pacific for three clear-sky selection conditions.

Figure 5. Differences of brightness temperatures (region -3 region average) for 1 RIS and ECHAM for three geographic regions. The data are from the entire mission.

Figure 6. Standard deviations for the mission duration, observed by IRIS and calculated from ECHAM, displayed for the three geographic regions.

Figure 7. Skew, as for figure 6.

Figure 8. Kurtosis, as for figure 6.

Figure 9. T_b(mission mean)-T_b(monthly mean) in the Central Pacific for the months of April, July, October and December, 1970. IRIS and ECHAM are compared.

Figure 10. Standard deviation, as for figure 9.

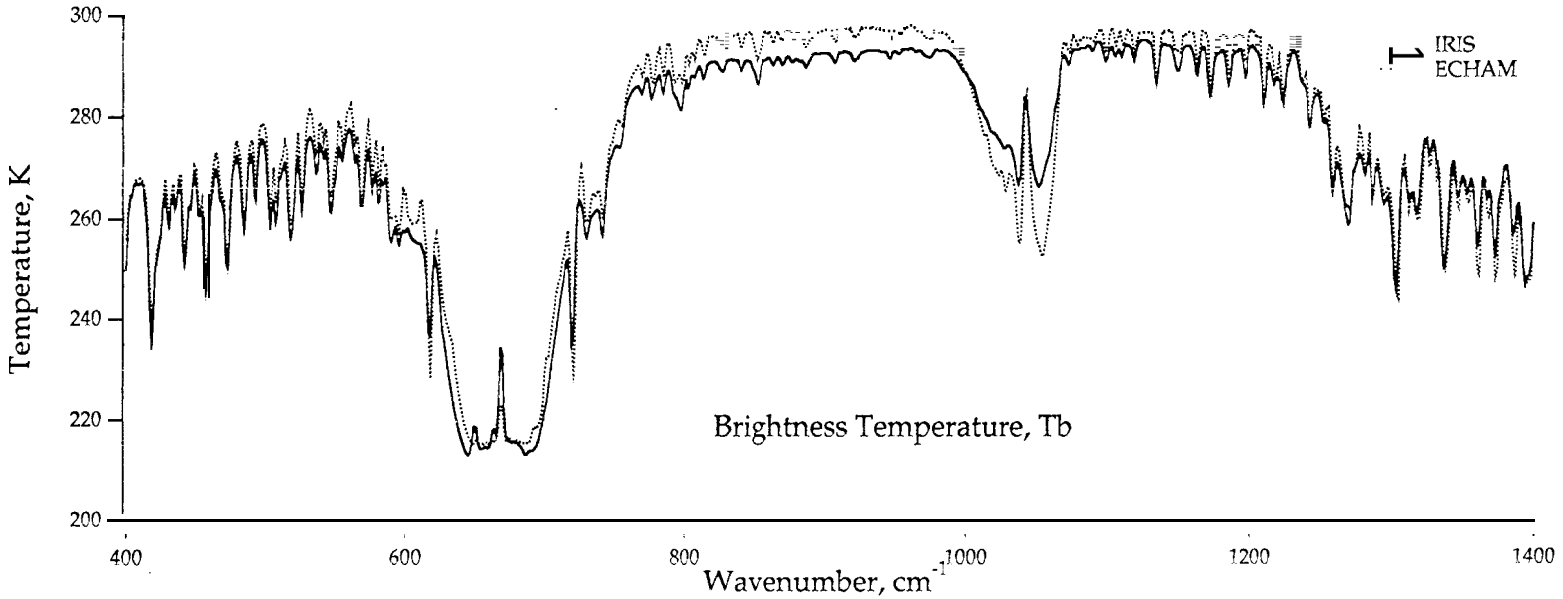
Figure 11. Skew, as for figure 9.

Figure 12. Kurtosis, as for figure 9.

Figure 13. Standard deviations for ECHAM and IRIS derived from 1-, 5-, and 30-day averages, in the Central Pacific. The whole data set is used,

Figure 14. Skew, as for figure 13.

Figure 15. Kurtosis, as for figure 13.


300
280
260
240
220
200
Temperature, K
IRIS
ECHAM
Brightness Temperature, Tb
400
600
800
1000
1200
1400
Wavenumber, cm-1

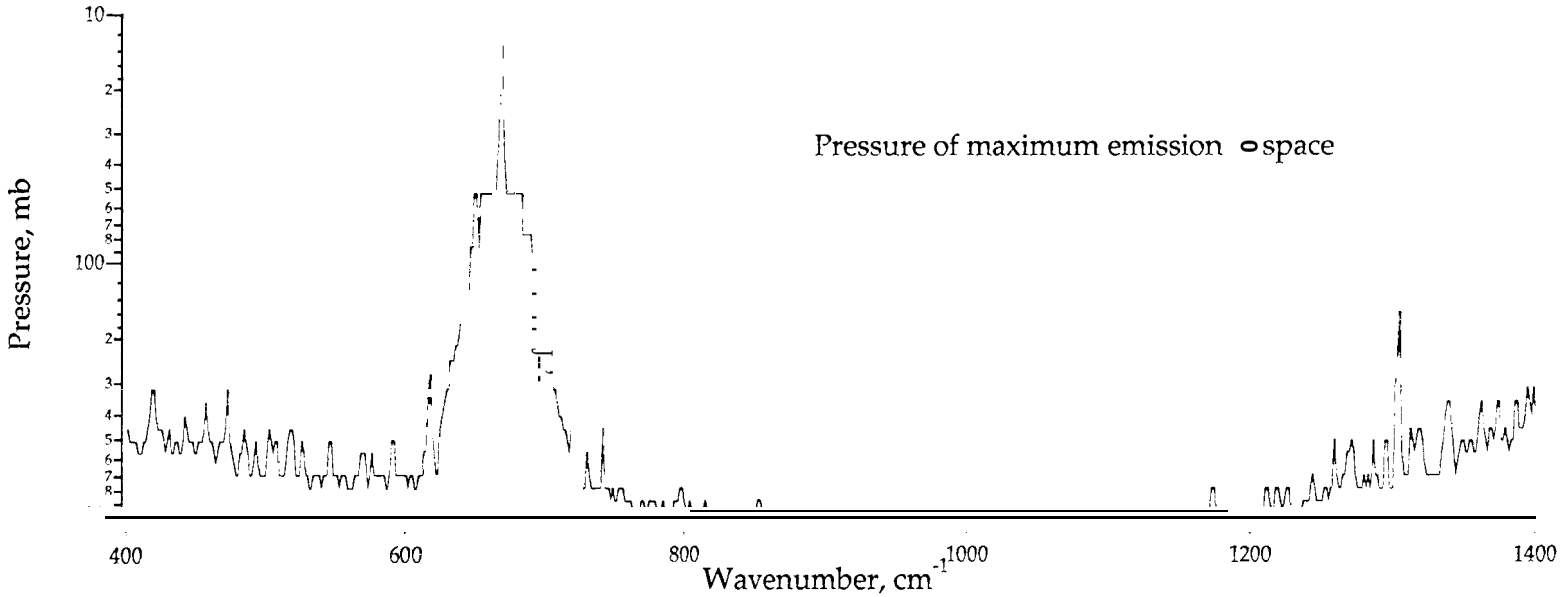
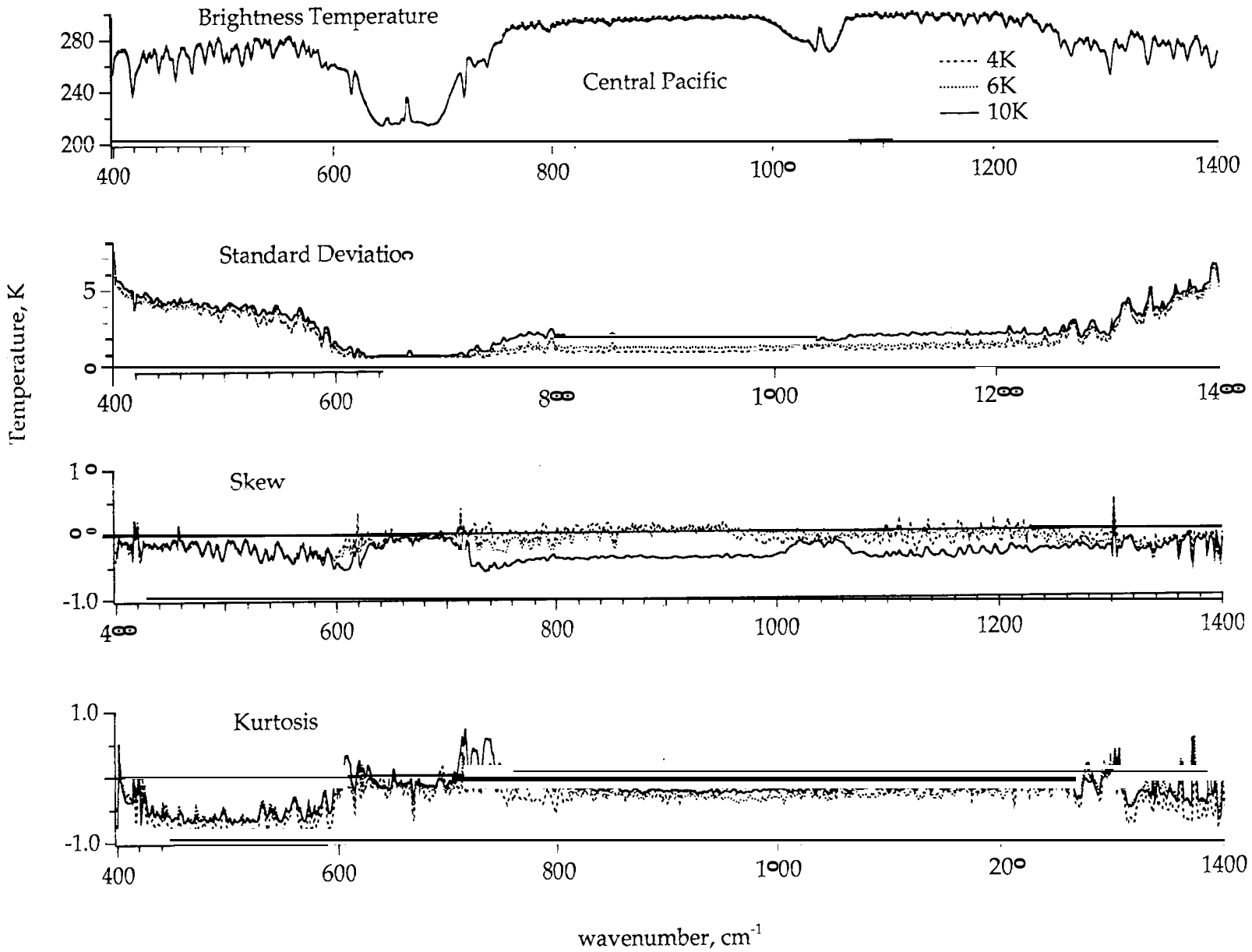

10
2
3
4
5
6
7
8
100
Pressure, mb
Pressure of maximum emission
space
400
600
800
1000
1200
1400
Wavenumber, cm-1

fig. 2


Brightness Temperature
Central Pacific
4K
6K
10K
280
240
200
400
600
800
1000
1200
1400
Standard Deviation
5
0
Skew
1.0
0.0
-1.0
Kurtosis
Temperature, K
wavenumber, cm-1

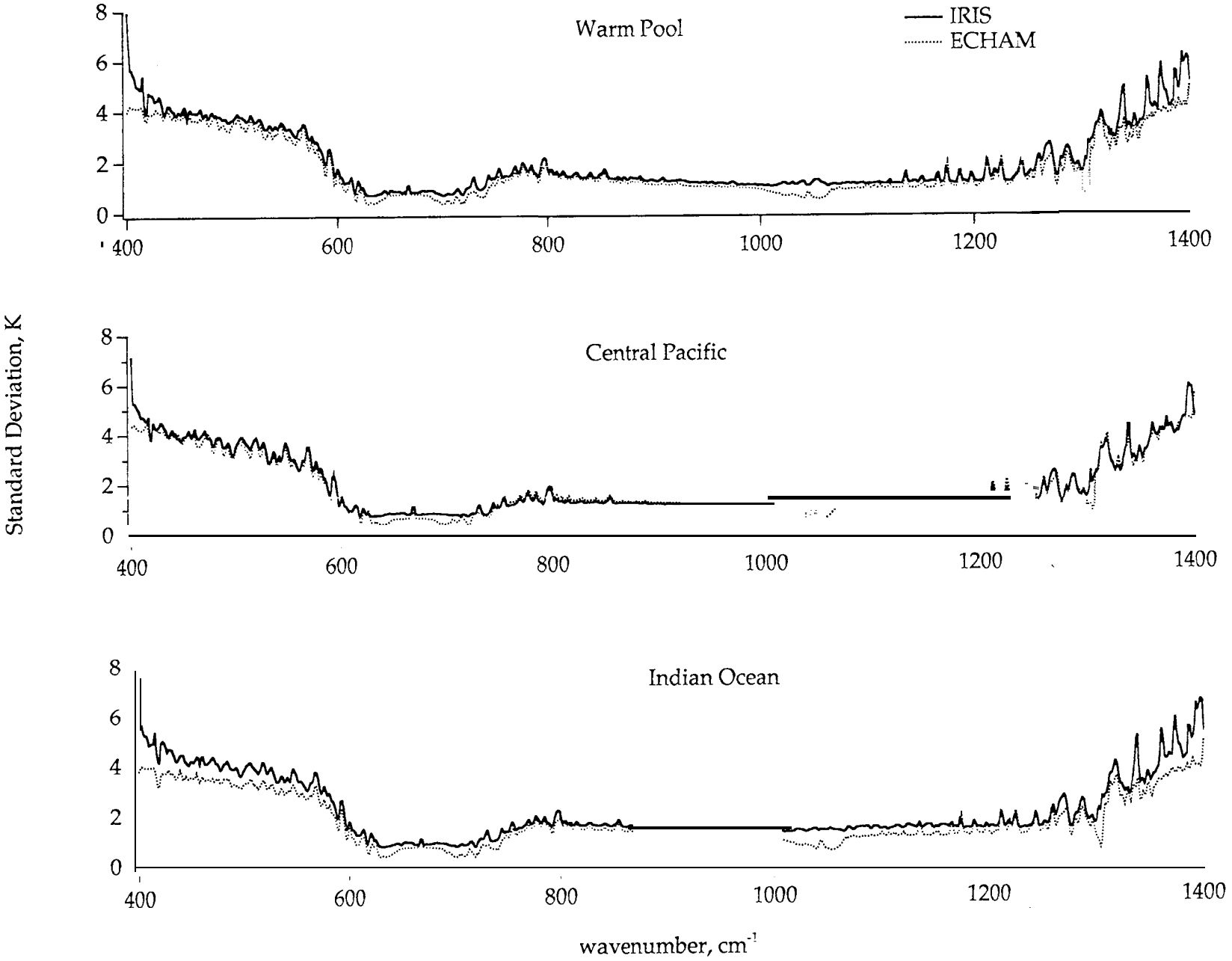

Warm Pool
IRIS
ECHAM
Central Pacific
Indian Ocean
Standard Deviation, K
wavenumber, cm-1
0
2
4
6
8
400
600
800
1000
1200
1400

fig 6

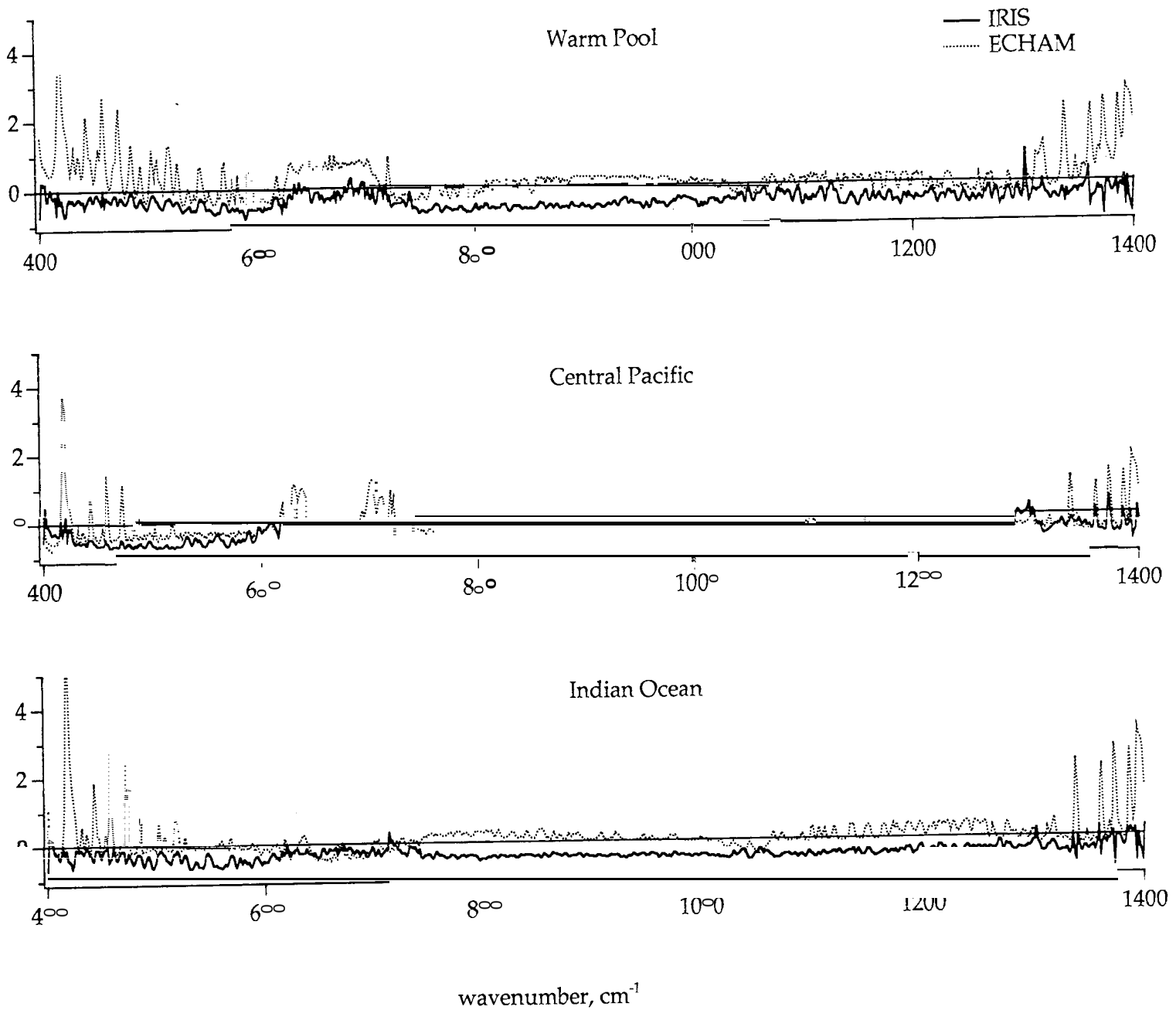

Warm Pool
IRIS
ECHAM
Central Pacific
Indian Ocean
Kurtosis, K
wavenumber, cm-1

fig 8

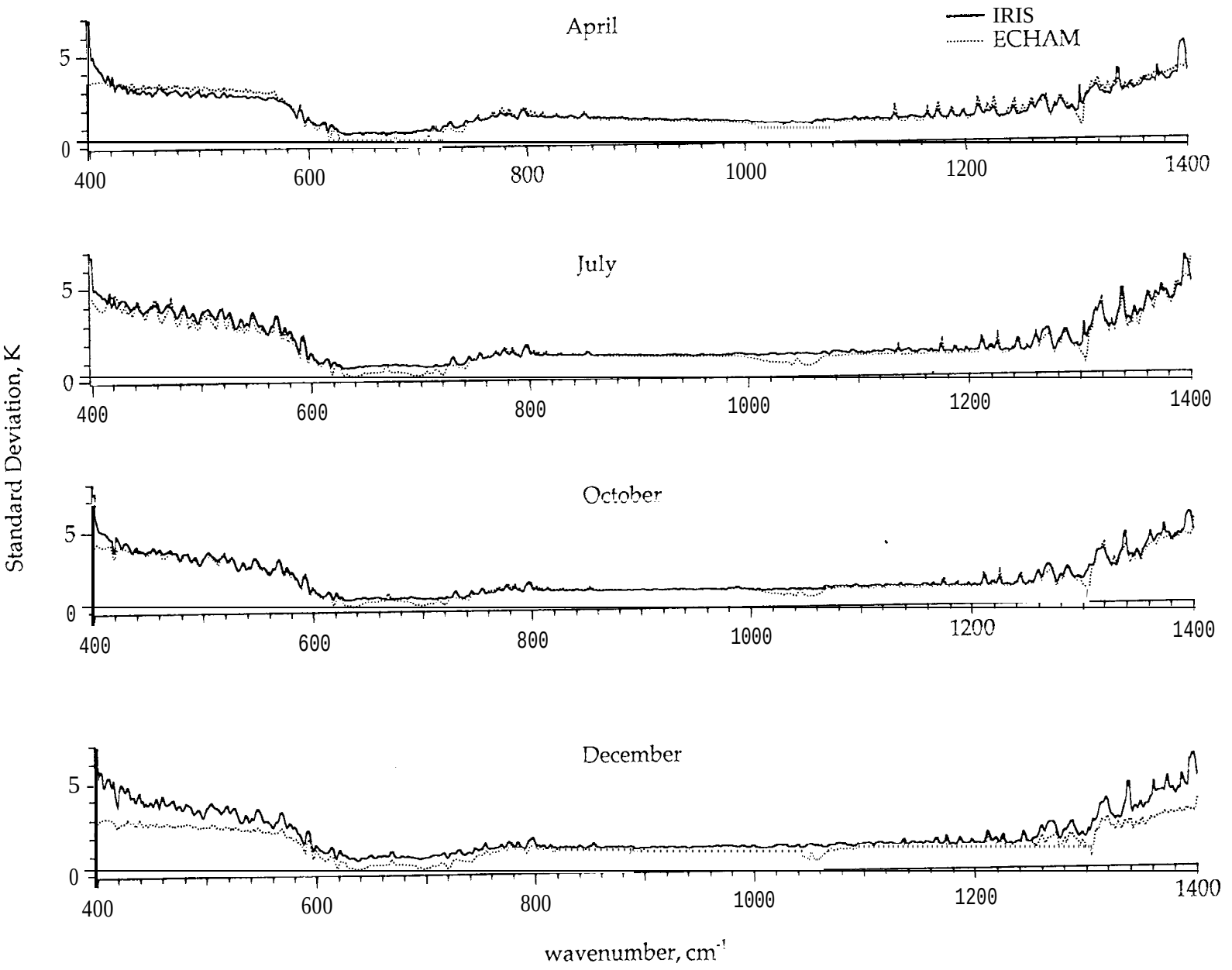

April
July
October
December
IRIS
ECHAM
Standard Deviation, K
wavenumber, cm-1
0
5
400
600
800
1000
1200
1400

Fig 10

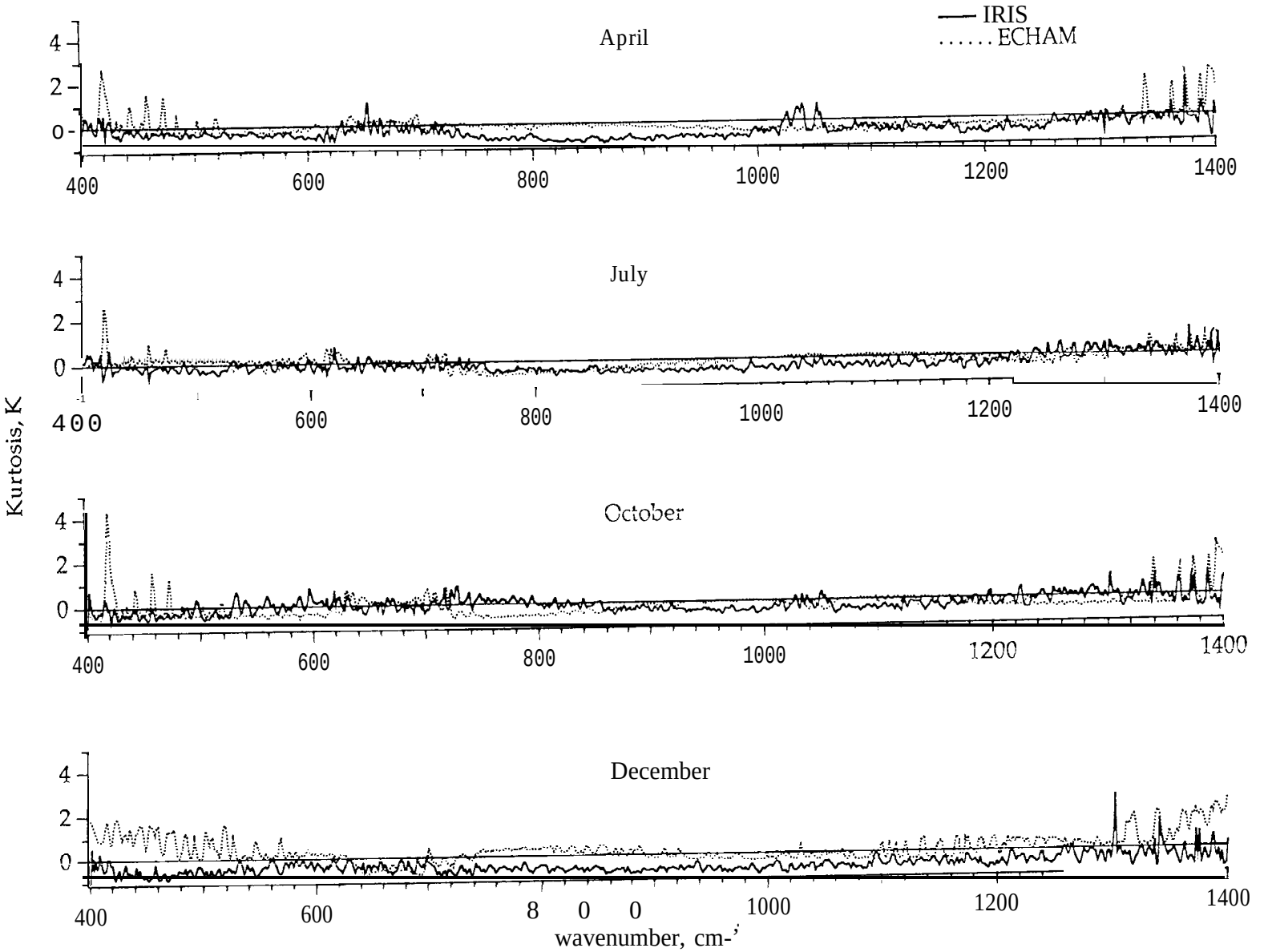

IRIS
ECHAM
April
July
October
December
Kurtosis, K
wavenumber, cm-1
400
600
800
1000
1200
1400

fig 12

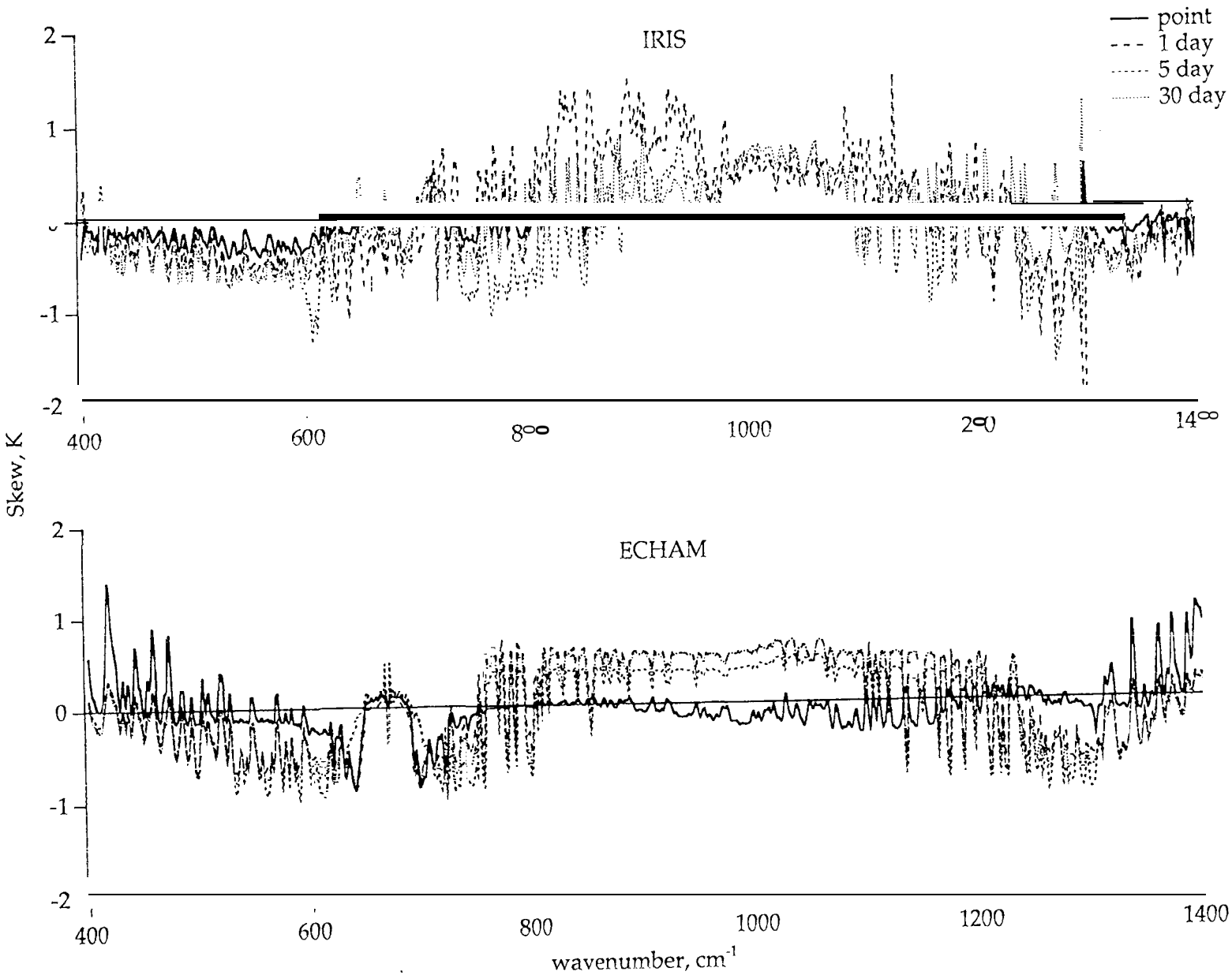

IRIS
point
1 day
5 day
30 day
Skew, K
2
1
0
-1
-2
400
600
800
1000
1200
1400
ECHAM
wavenumber, cm-1

fig 14

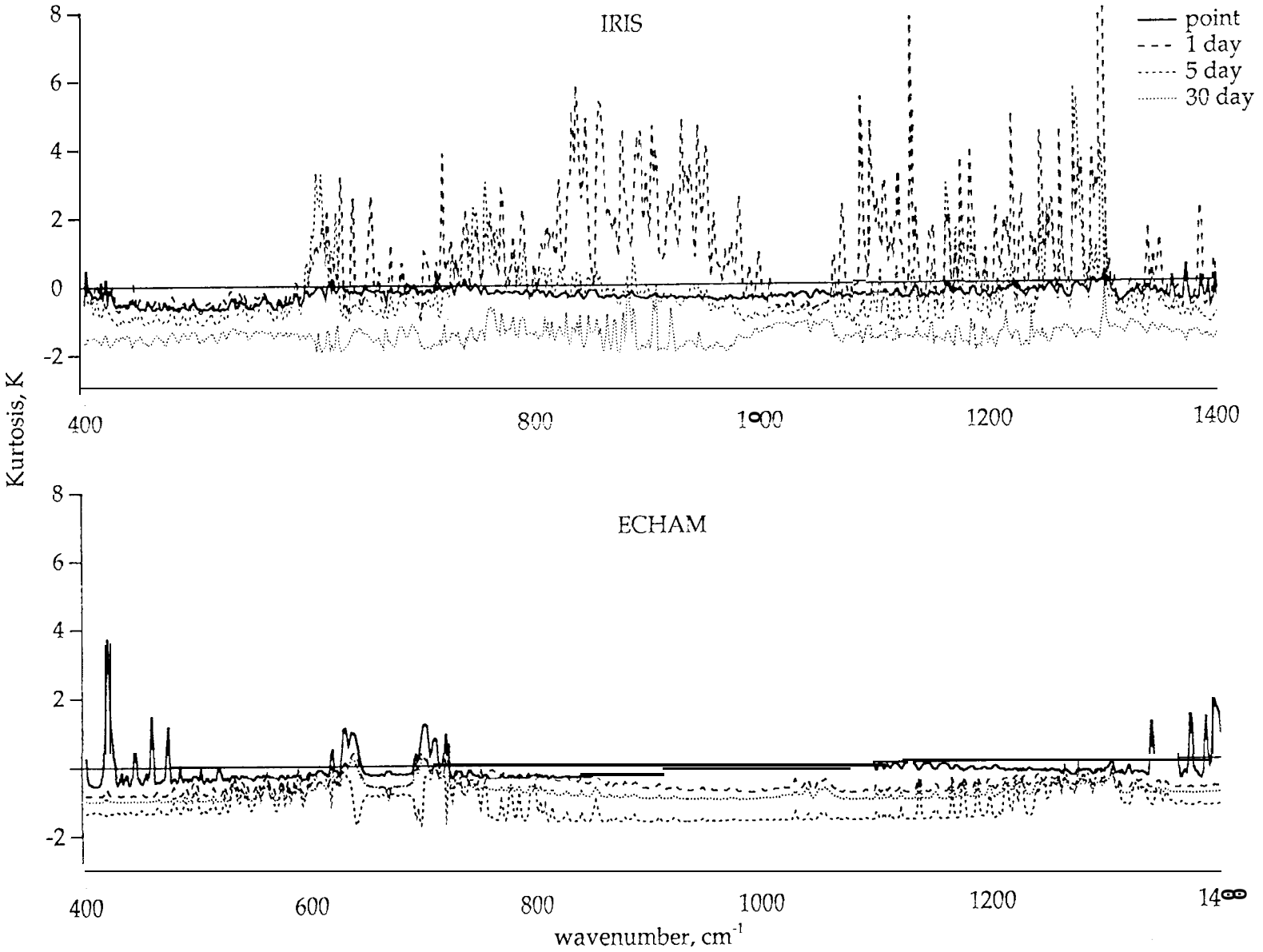

IRIS
point
1 day
5 day
30 day
Kurtosis, K
8
6
4
2
0
-2
400
800
1000
1200
1400
ECHAM
8
6
4
2
-2
400
600
800
1000
1200
1400
wavenumber, cm-1

fig 15