INT. J. REMOTE SENSING, 1991, VOL. 12, NO. 5, 921-952

Effects of satellite data resolution on measuring the space/time variations of surfaces and clouds

GENEVIEVE SÈZE

Laboratoire de Météorologie Dynamique du CNRS,
91128 Palaiseau Cedex, France

and WILLIAM B. ROSSOW

NASA Goddard Space Flight Center, Institute for Space Studies,
New York 10025 U.S.A.

(Received 12 June 1989; in final form 15 January 1990)

Abstract. The correlated distributions of satellite-measured visible and infrared radiances, caused by spatial and temporal variations in clouds and surfaces, have been found to be characteristic of the major climate regimes and can be described by the attributes of bidimensional and monodimensional histograms and time-composite images. Most of the variability of both the surfaces and clouds is found to occur at scales larger than the minimum resolved by satellite imagery. Since satellite imaging data sets are difficult to analyse because of their large volumes, many studies reduce the volume by various sampling or averaging schemes. The effects of data resolution and sampling on the radiance histogram statistics and on the time-composite image characteristics are examined. In particular, the sampling strategy used by the International Satellite Cloud Climatology Project is tested. This sampling strategy is found to preserve the statistics of smaller cloud variations for most regions, with the exception of very rare events, if they are accumulated over large enough areas (at least 500 km in dimension) and long enough time periods (at least one month).

1. Introduction

Meteorological satellites allow for observations of clouds and surfaces over a large range of space and time scales (Rossow 1989). Complete global coverage over long time periods also provides better comprehension of the larger scale cloud and surface characteristics associated with different climate regimes and the larger scale interactions that occur among these regimes. Sèze and Rossow (1991)—henceforth SR91)—have shown that the shape and related statistics of the bidimensional radiance histograms, obtained from visible (VIS) and infrared (IR) satellite images, provide a stable and distinctive descriptor of the main characteristics and variations of cloud cover and surfaces in a region during a period by demonstrating (1) the rapid convergence of the histogram statistics to well defined values as either region size or cumulation time period is increased and (2) the relative insensitivity of the statistics to the choice of cumulation time and space scales at larger scales. They have also shown that analysis of time-composite images, in the same way as the original images, provides a way to separate the contributions of the surface and clouds and their spatial and temporal variability to the measured variations of visible and infrared radiances.

The aim of this study is to investigate the effects of changing data characteristics on the determination of the spatial and temporal variability of these radiances.

Specifically, satellite imagery data are limited explicitly by their space/time resolution, as well as their radiometric resolution. In addition, it is common to reduce the volume of satellite data by various averaging or sampling schemes. Here the emphasis is on the latter topic: we do not examine the variations at scales less than 5 km nor time scales less than one day. However, in SR91, we have already shown that, with the exception of diurnal variations (Hartmann and Short 1980, Minnis and Harrison 1984), the radiance variability is larger at scales ⩾ 30 km and ⩾ 1 day than at smaller scales (Wielicki and Welch 1986, Welch *et al.* 1988, Gifford 1989, Cahalan and Joseph 1989).

Up to now, averaging to reduce the spatial and temporal resolution of satellite imagery data has been the more common data-volume-reduction scheme in climate studies, where the scale for averaging the data can be as large as daily–monthly at 2·5° × 2·5° up to 5° × 5° (Hartmann and Short 1980, Cahalan *et al.* 1982, Hartmann and Recker 1986, Duvel 1988). However, relationships between radiance variations at different wavelengths, especially in the visible and infrared, are not always linear (Coakley and Baldwin 1984); thus, averaging can produce radiance values and relationships that are not actually representative of the most common situation in some regions (see figure 4 in SR91). Although some variation of cloud properties occurs on very small scales (Wielicki and Welch 1986, Kuo *et al.* 1988, Minnis and Wielicki 1988, Welch *et al.* 1988, Gifford 1989, Sèze and Smith 1989, Cahalan and Joseph 1989), it is not practical to survey cloud properties on global scales for long time periods using data of very high resolution, such as that from Landsat or SPOT. Compromises must be made to keep enough information about the smaller scale cloud structures and variations and, at the same time, to obtain a good representation of the global distribution and temporal evolution of cloud systems. However, reducing the spatial resolution by averaging eliminates all information about smaller scale variations.

The International Satellite Cloud Climatology Project (ISCCP) reduces the satellite data volume by space/time sampling (Schiffer and Rossow 1983, 1985), instead of by averaging. This procedure retains the original image pixel spatial resolution of between 4–8 km, but samples them at a spacing of about 30 km. In this study we compare the effects on the climatological statistics of this spatial sampling procedure with the effects of spatial averaging to about 30 km resolution (Rossow *et al.* 1985). The effects of time sampling have been examined by Harrison *et al.* (1983) and Brooks *et al.* (1986).

The climatology of surface and cloud, space and time variations is represented by the characteristics of the bidimensional and monodimensional radiance histograms cumulated over spatial scales of about 300 km and time scales of about one month (see SR91). These histograms can represent the variation of the VIS and IR spectral characteristics of the clouds and the surface (1) in space (Desbois *et al.* 1982, Platt 1983), (2) in time (daily histograms: Desbois and Sèze 1984, histograms of daily spatial averages: Hartmann and Short 1980, Hartmann and Recker 1986); or (3) both (time-cumulated histograms, Sèze and Desbois 1987, SR91). The full resolution METEOSAT data used for this study and the two reduced versions of this data that are tested are described. The bidimensional and monodimensional radiance histograms and their associated statistics for the full resolution data, examined in SR91, are also summarized.

For the two reduced-resolution data sets and the seven special study regions, the same bidimensional and monodimensional radiance histograms and all of the associated statistics have been calculated. Comparison of these results with those from the full-resolution data set shows the effects of dat sampling and averaging on the climatological statistics.

To clarify the relative contributions of space and time variability in these cumulated histograms, time-composited images are also constructed. These images are formed by calculating a particular statistic of the time series of radiances for each location. For example, the time mean, mode, maximum or minimum radiance for each location can be displayed as and analysed in the same fashion as the original images (SR91). The effects of data sampling on the radiance statistics obtained from these images are examined.

2. Satellite data and radiance statistics

2.1. *Data*

The full-resolution data used for this study are METEOSAT images at 1200 UTC from 15 July to 10 August 1983. The resolutions of the radiance measurements at the subsatellite point are 2·5 km for visible (sampled to 5 km) and 5 km for infrared data. The visible (VIS) measurements are made at an effective wavelength of 0·7 μm with a radiometric resolution of 64 levels (counts), which is equivalent to a precision of 2 per cent in reflectivity. The infrared (IR) measurements are made at 11 μm with 256 levels (counts), which is a precision of 0·4 K in brightness temperature for the warmer desert regions, of 0·5 K at about 270 K of brightness temperature, and of 1 K for high (cold) clouds. Because the atmosphere is nearly transparent at these two wavelengths, the distribution of VIS and IR radiances is produced primarily by the space and time variations of clouds and the surface (ocean or land).

METEOSAT is a geosynchronous weather satellite operated by the European Space Agency (ESA) that provides observations covering a region ±70° of geocentric angle centred on the equator and the Greenwich meridian (figure 1 of SR91), encompassing all of Europe and Africa, plus a portion of Brazil, and most of the North and South Atlantic Ocean. ESA re-maps these observations on to a fixed grid by simulating a geosynchronous satellite situated at a fixed position (the real satellite moves with time around this fixed position). Therefore one scene element (or pixel) corresponds to one fixed geographical location.

Seven areas have been specially studied (figure 1 of SR91). They were chosen to represent the main types of radiance distributions observed from METEOSAT (figure 3 (*a*) of SR91): (1) North Africa (32° N, 6° E), (2) Sahel (19° N, 4° W), (3) south Subtropical Atlantic (16° S, 6° E), (4) tropical Atlantic (9° N, 20° W), (5) central tropical Africa (6° N, 27° E), (6) Europe (49° N, 13° E), and (7) North Atlantic (52° N, 21° W). These regions will be referred to as N. Africa, Sahel, S. Atlantic, Tr. Atlantic, Tr. Africa, Europe and N. Atlantic, respectively.† Bidimensional radiance histograms are constructed for regions with (60 × 60) pixels (2·5° × 2·5°), (60 × 120) pixels (2·5° × 5°), (90 × 180) pixels (3·75° × 7·5°), and (120 × 240) pixels (5° × 10° degrees), centred on these areas. We refer to region sizes throughout this paper by the number of FULL resolution pixels within them.

From these images, which we call FULL resolution data, two reduced data sets have been constructed: (1) data sampled to 30 km spacing (called SAMPLED, similar

---

†17 days of data were available for Sahel, S. Atlantic, Tr. Africa and Tr. Atlantic; 20 days of data were available for N. Africa, Europe and N. Atlantic.

to the ISCCP B3 reduced-resultion radiance data set) by taking one line out of six and one pixel of six and (2) data averaged to 30 km resolution (called AVERAGED) by averaging radiances on 6 × 6 pixel subregions. Thus, the number of pixels for SAMPLED/AVERAGED data is $\frac{1}{36}$ that of the FULL data.

We also explore the effects of differing radiance resolutions. Most of the statistics reported use VIS/IR radiances with 64/128 levels, equivalent to about 2 per cent/1 K precision. In the figures and tables presented, these radiances are re-scaled to count values from 0 to 255. This means that changes in VIS and IR, produced by changes in clouds and the surface and expressed in these count values, represent roughly the same magnitude changes in total radiation (i.e. a few watts per square metre). Other radiance resolutions are studied: 4 per cent/2 K and 4 per cent/4 K. The IR scale is reversed, so that the smaller counts represent the highest brightness temperatures and the lagest counts the lowest. Later in the discussion, when we refer to the minimum/maximum in infrared (IR MIN/MAX), we are referring to the lowest/highest temperatures, not to the minimum/maximum count values. To avoid confusing radiances in the two spectral bands, we will refer to the lowest/highest VIS as darkest/brightest and the lowest/highest IR as coldest/warmest.

2.2. *Histograms and radiance statistics*

For all radiance resolutions for the FULL, SAMPLED and AVERAGED data sets, bidimensional and monodimensional histograms are constructed for all four region sizes and each day (single image); these results are then cumulated for time periods of 2, 3, 4 ... up to 17 or 20 days.

For each histogram, associated statistical quantities are computed (table 1): minimum (MIN), maximum (MAX), average (AVG), mode (MOD), mode frequency (PEAK) , and the standard deviation (SD). These quantities are calculated for both the two-dimensional (2D) (indicated by prefix of VIS–IR) and one-dimensional (1D) (prefix VIS or IR) VIS and IR distributions (sometimes we use the suffix 2 and the suffix 1, respectively). The relation of the 1D and 2D radiance distributions is evaluated by calculating the correlation (COR) between VIS and IR. All of these quantities are computed for time-cumulated histograms (called CUM parameters for the 17 or 20 day period and CUM*N* parameters for *N* day periods) and for each daily histogram to show the evolution of these parameters with time. The average and standard deviation of the daily values (called AVG parameters and SD parameters) are also computed (table 2).

A quantitative measure of radiance variation or histogram dispersion is provided by the percentage of the total possible radiance pairs occurring in each bidimensional histogram (called AREA for 'surface area'). To define the high-frequency part(s) in each bidimensional histogram, a frequency is chosen such that the portion of the population inside the isolines is, for example, 80 per cent, and the portion outside is 20 per cent (radiances inside the isoline have a higher frequency than radiances outside). The 80 per cent isoline is the intermediate contour in figures 1 and 2 and the 50 per cent isoline is the innermost contour. To determine these isolines, the cumulative fraction of the total population as a function of the frequency is calculated (see figure 2 in SR91). The fractions of the total possible surface area contained inside the 80 per cent isolines are called 80 per cent CUM-AREA and 50 per cent CUM-AREA for the time-cumulated histograms.

The separate contributions of time and space variations can be determined from the regional variations of time-series statistics at each location. This is done by calculating the same histograms and associated statistics for time-composite images (SR91). In these images, the value associated with a scene element is a statistic extracted from the radiance time series at that location (e.g. MIN, MAX or MOD). In SR91, the spatial variability of these images is examined and the statistics associated with the regional histograms (mainly the MOD) are compared with the regional CUM parameters.

2.3. *Interpretation of histogram shapes and statistics*

The characteristic shapes of the VIS–IR histograms are illustrated in figures 1 and 2 and figure 3 (*a*) of SR91. Figure 3 (*a*) of SR91 shows the range of histogram shapes that characterize the main climate regimes observed by METEOSAT. Figure 1 shows the mono- and bidimensional radiance distributions accumulated over 17 days in July 1983 for the seven smaller-sized regions defined in figure 1 of SR91; figure 2 shows some daily histograms. The daily histograms indicate the range of variability of the histogram shapes from one day to another, mainly due to changes in cloud properties; on the other hand, figure 1 and figure 3 (*a*) of SR91 show the well defined shapes of the time-cumulated histograms.

Inside these shapes the higher frequencies are indicated by two contours, representing the 80 per cent and 50 per cent isolines. In the daily histograms, the extent of the high-frequency portions is highly variable; but the time-cumulated histograms, in spite of their larger overall size (CUM-AREA), are characterized by a relatively small 80 per cent CUM-AREA. This behaviour is magnified for the 50 per cent zones. The shapes of these high frequency regions describe the more frequent radiance pairs recorded in the region during the period, usually associated with the characteristics of the surface (except for S. Atlantic). By construction the most frequent VIS-IR pairs (CUM-MOD) are in these high-frequency zones (figure 1).

The seven regions can be gathered into three groups by their shapes in the radiance plane (figure 1): Group 1 (N. Africa and Sahel), Group 2 (S. Atlantic), and Group 3 (Tr. Africa, Tr. Atlantic, Europe and N. Atlantic). The characteristics of these three groups are summarized very briefly here; a much more detailed description is given in SR91.

Group 1: histograms with very small VIS radiance variation, an elongated distribution along the IR axis, and a compact cluster of points and MOD2 at warm and (relatively) dark radiance values (North Africa; figure 1 (*a*)). These histograms are characteristic of generally clear desert areas with homogeneous surface properties where occasional cirrus clouds occur. The Sahel region (figure 1 (*b*)) histograms have the same kind of shape, but with two surface clusters, one colder than the other; the more frequent presence of thin clouds and occasional thicker cirrus or altostratus also causes a sparse distribution of colder and brighter radiances.

Group 2: histograms with a very small IR radiance variation, but a significant VIS radiance variation (figure 3 (*a*) of SR91), which is characteristic of low-level stratocumulus cloud layers over ocean (S. Atlantic; figure 1 (*c*)). This region is one of only a few regions in the METEOSAT view that is so dominated by large amounts of persistent cloudiness that the MOD values represent clouds, rather than the surface.

Group 3: histograms with a large dispersion in both the VIS and IR radiances, which is characteristic of a mixture of different kinds of clouds at different levels (e.g.) high convective clouds and a low cloud layer over the sea (Tr. Atlantic; figure 1 (*d*)), a mixture of high-, middle- and low-level clouds in a convective regime over land (Tr. Africa; figure 1 (*e*)), and high, middle and low clouds over land (Europe; figure 1 (*f*))

Table 1. Definitions of symbols used to represent various statistical quantities calculated from the monodimensional and bidimensional visible (VIS) and infrared (IR) radiance distributions cumulated over various time and space scales.

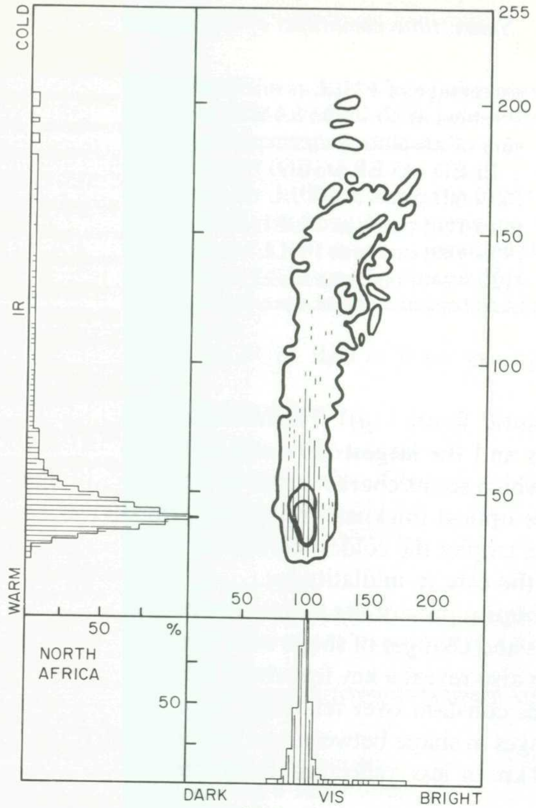
VIS	visible radiance values (0·7 μm) as counts from 0–255, representing small to large values with a precision of 2 per cent.
IR	infrared radiance values (11 μm) as counts from 0–255, representing large to small values with a precision of 1°C near 20°C.
FULL	data set with 2·5 km resolution in VIS sampled to 5 km spacing and with 5 km resolution in IR.
SAMPLED	data sampled from the FULL data to 30 km spacing (ISCCP B3 data set).
AVERAGED	data averaged to 30 km resolution.
Statistics computed from monodimensional radiance distributions (VIS/IR means VIS and/or IR, also indicated by suffix of 1)	
VIS/IR MIN	minimum, darkest/coldest, radiances.
VIS/IR MAX	maximum, brightest/warmest, radiances.
VIS/IR AVG	average radiances.
VIS/IR	standard deviation of radiances.
VIS/IR MOD	most frequent radiance value (mode).
VIS/IR PEAK	largest frequency in the histogram, expressed as a fraction of the total image pixel population.
r.m.s.	root mean square of radiance differences between any pair of image pixels in the distribution.
Statistics computed from bidimensional radiance distributions (VIS–IR means VIS and IR, also indicated by a suffix of 2)	
VIS/IR MOD	most frequent pair of radiance values (mode), not generally equal to individual MOD1 values.
VIS–IR PEAK	largest frequency in the histogram, expressed as a fraction of the total image pixel population
VIS–IR COR	correlation between the VIS and IR values.
AREA	area occupied by the bidimensional histogram as a percentage of the total possible area in the VIS–IR space.
80 per cent isolines	frequency isolines in the bidimensional histogram defined by the frequency chosen such that the portion of the population inside the isolines is 80 per cent and the portion outside is 20 per cent; radiances inside the isoline have a higher frequency than the radiances outside.
80 per cent AREA	area occupied by the higher frequency portion of the bidimensional histogram and defined by the 80 per cent isolines.
50 per cent isolines	frequency isolines in the bidimensional histogram defined by the frequency chosen such that the portion of the population inside the isolines is 50 per cent and the portion outside is 50 per cent; radiances inside the isoline have a higher frequency than radiances outside.
50 per cent AREA	area defined by the 50 per cent isolines.
Cumulation statistics	
CUM parameters	prefix for quantities obtained from histograms cumulated over whole time period (17 or 20 days) covered by imaging data.
CUMN parameters	prefix for quantities obtained from histograms cumulated over *N* days of data (one image per day).
AVG parameters	prefix for quantities representing the average of values obtained from histograms cumulated from single images (one day).
SD parameters	prefix for quantities representing the standard deviation of values obtained from histograms cumulated from single images (one day)
IM parameters	prefix for time-composite images—these images represent histogram quantities, such as MIN, MAX, AVG, MOD, taken from the time series of radiance values at each image pixel location.
AREA-DIF	percentage of FULL pixels having radiance values that are not represented at all in the SAMPLED (averaged) bidimensional histograms.
DIF-SUM	sum of absolute frequency differences between the FULL and SAMPLED (AVERAGED) bidimensional histograms.
Region size 1	(60 × 60) pixels at FULL resolution (2·5° × 2·5°)
Region size 2	(60 × 120) pixels at FULL resolution (2·5° × 5°)
Region size 3	(90 × 180) pixels at FULL resolution (3·75° × 7·5°)
Region size 4	(120 × 240) pixels at FULL resolution (5° × 10°)

and ocean (N. Atlantic; figure 1 (*g*)). The histograms in this group have the brightest and coldest clouds and the largest SD both in VIS and IR. The VIS/COR is also highest (0·8–0·9), which seems characteristic of regions with high and/or middle cloud cover with variable optical thickness (reflectivity/emissivity) and/or partial coverage of the pixels. In the tropics the coldest clouds correspond to the brightest clouds, but this is not always the case in midlatitudes.

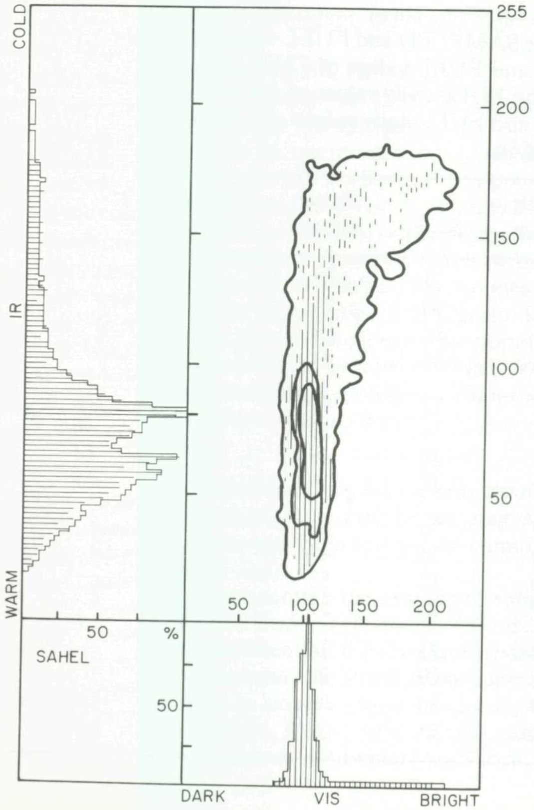
These three groups represent the major climate regimes revealed by the particular histogram patterns and changes of shape with location, shown in figure 3 (*a*) of SR91. The shape changes also reveal a key feature of the distribution of climate regimes: not only are the shapes constant over relatively large regions of about 2000 to 5000 km scale, but the changes in shape between two regions are rather abrupt, occurring on a scale of about 500 km or less, reflecting the dominant control of the cloud structures

Table 2. Ratio of the SAMPLED and FULL values of CUM-AREA (lines 1 to 4), ratio of the AVERAGED and FULL values of CUM-AREA (lines 5 to 8), average of the ratio of SAMPLED and FULL daily values of AREA (lines 9 to 12), and average of the ratio of AVERAGED and FULL daily values of AREA (lines 13 to 16) for four regions sizes in each study region.

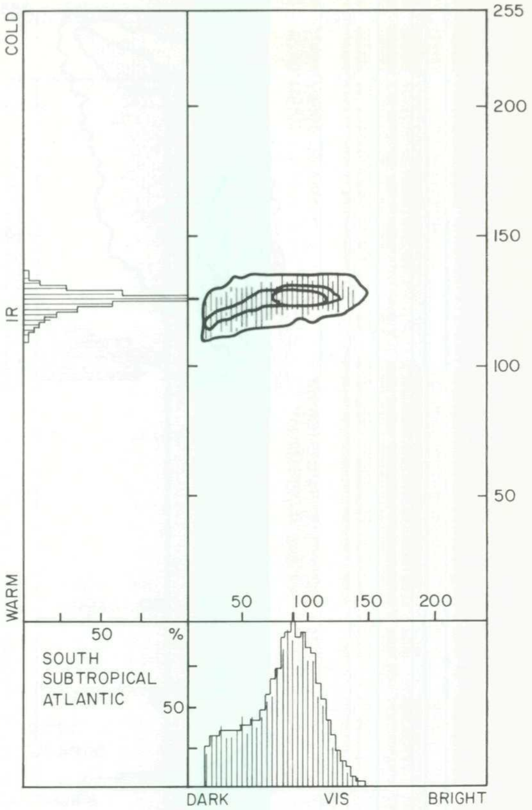
Regional size	North Africa	Sahel	South Atlantic	Tropical Atlantic	Tropical Africa	Europe	North Atlantic
60 × 60	25	34	65	36	37	43	55
60 × 120	25	40	71	51	49	57	66
90 × 180	29	46	68	68	63	65	74
120 × 240	38	53	54	75	69	69	75
60 × 60	24	32	47	35	32	38	51
60 × 120	25	35	50	47	40	47	60
90 × 180	27	42	46	58	51	54	65
120 × 240	34	47	39	63	55	59	66
60 × 60	37	26	41	11	12	16	16
60 × 120	39	30	47	14	16	19	21
90 × 180	30	32	53	19	21	23	30
120 × 240	32	35	57	26	26	28	38
60 × 60	33	25	24	10	12	14	15
60 × 120	38	29	28	13	15	17	20
90 × 180	27	30	33	18	19	21	27
120 × 240	30	33	37	24	23	25	34


COLD
IR
WARM
255
200
150
100
50
50
100
150
200
50
%
NORTH
AFRICA
50
DARK
VIS
BRIGHT

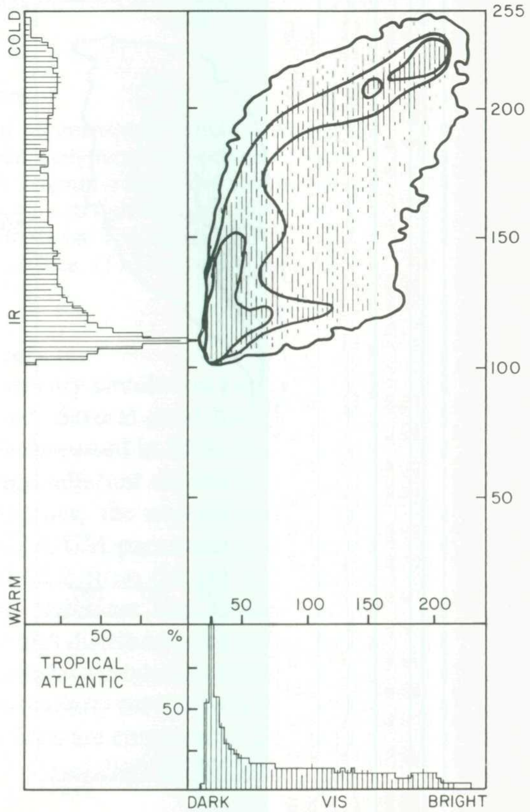
1 (*a*)


COLD
IR
WARM
255
200
150
100
50
50
100
150
200
50
%
SAHEL
50
DARK
VIS
BRIGHT

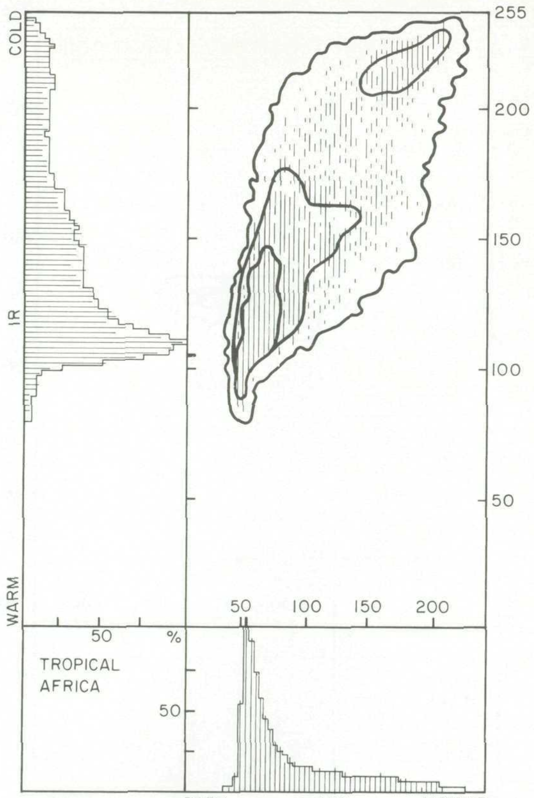
1 (*b*)


COLD
IR
WARM
255
200
150
100
50
50
100
150
200
50
%
SOUTH
SUBTROPICAL
ATLANTIC
50
DARK
VIS
BRIGHT

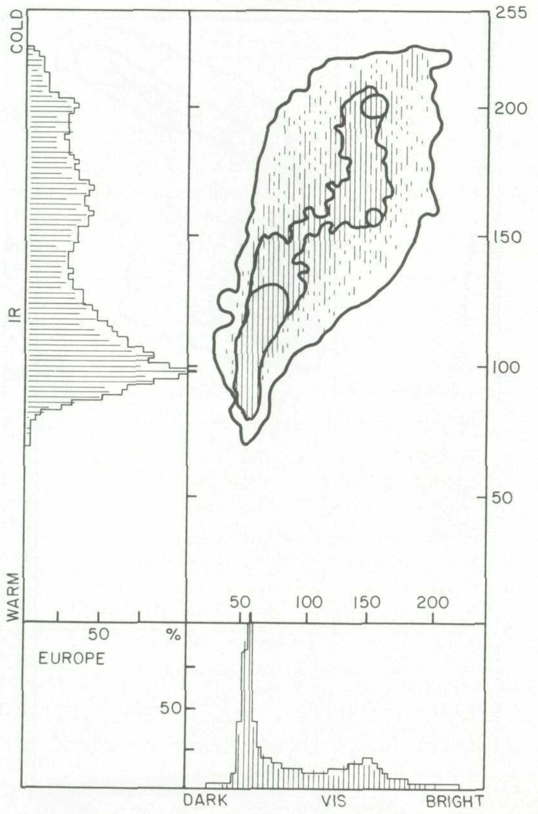
1 (*c*)


COLD
IR
WARM
255
200
150
100
50
50
100
150
200
50
%
TROPICAL
ATLANTIC
50
DARK
VIS
BRIGHT

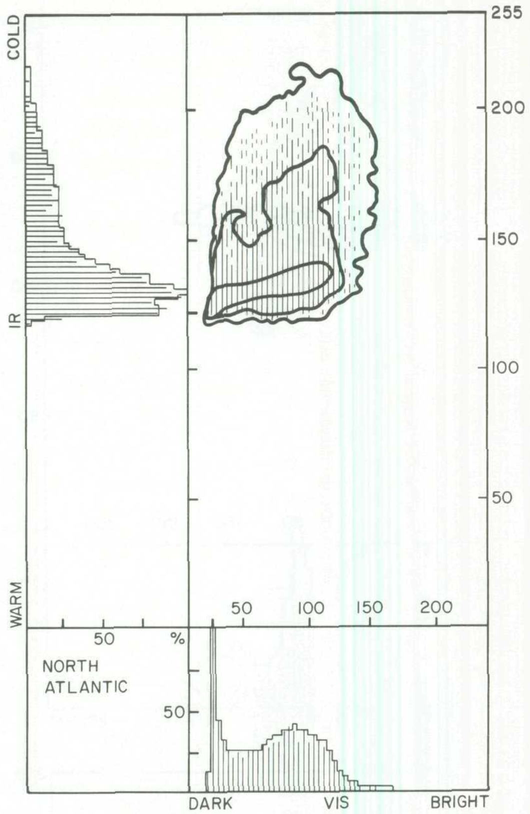
1 (*d*)



1 (*e*)



1 (*f*)

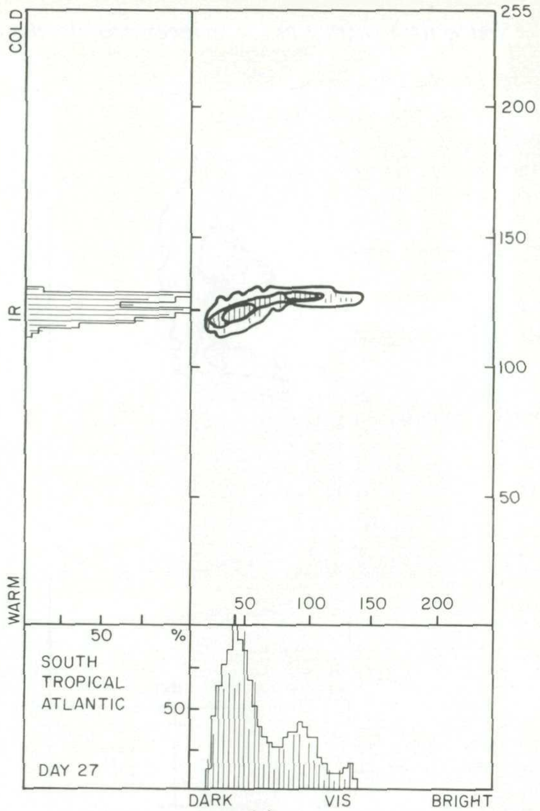


1 (*g*)

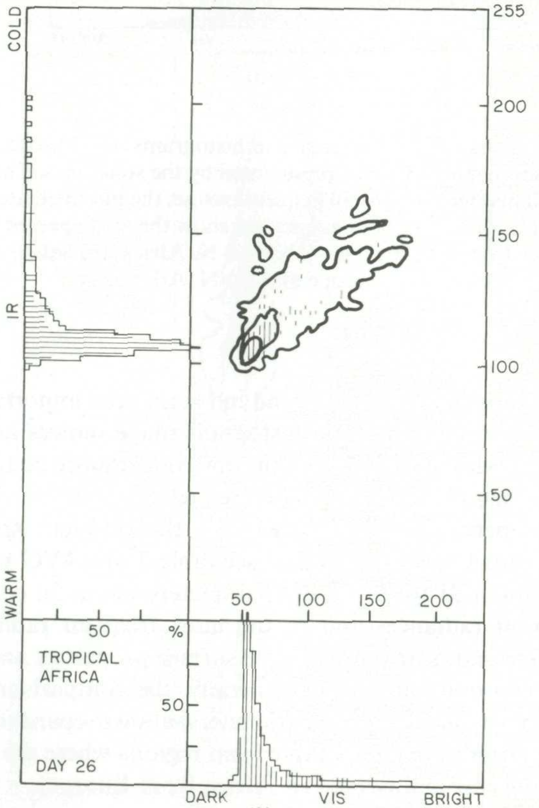
Figure 1. Mono- and bidimensional radiance histograms accumulated over 17 days or 20 days for the seven study regions, represented by the solid lines. The inner contour in the bidimensional histogram is the 50 per cent isoline; the intermediate contour is the 80 per cent isoline. The thin straight line segments show the area covered by the corresponding SAMPLED histograms. The regions are (*a*) N. Africa, (*b*) Sahel, (*c*) S. Atlantic, (*d*) Tr. Atlantic, (*e*) Tr. Africa, (*f*) Europe and (*g*) N. Atlantic.

by the large-scale planetary circulation and indicating the important influence of the land/water distribution. Several good histogram shape indices and spatio-temporal radiance statistics are discussed in SR91 that provide a more compact description of the differences between adjacent climmate regimes.

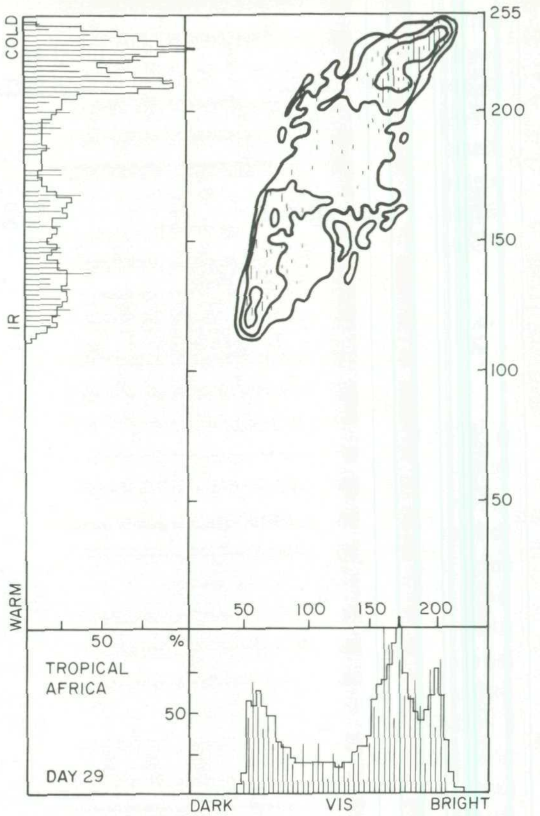
For surface properties, the comparison of the different spatio-temporal and temporal statistics (for CUM parameters, see table 3; for AVG parameters, table 4; and for statistics obtained from the IM parameters, table 5), corresponding to the darkest and warmest radiances and to the most frequent radiances, provides an estimate of the mode and distribution of the surface properties and the magnitude of their spatial and temporal variability. Generally, the comparison of the CUM and AVG parameters (particularly the SD parameters) allows separation of regions where time and space variations are comparable from regions where the space variability is smaller than the time variability (i.e. Tr. Africa from Europe).



(*a*)



(*b*)



(*c*)

Figure 2. Daily mono- and bidimensional histograms showing the correspondence of the FULL and SAMPLED resolution radiance distributions. The contours indicate the FULL resolution distribution with the same meaning as in figure 1. The SAMPLED distribution is shown as an overlay of short line segments. Days are in July 1983.

For cloud properties, the VIS- and IR-SD and the histogram AREA indices separate multi-layer cloud-covered regions from regions characterized by one very frequent event (e.g. clear sky or stratocumulus cloud layer). For the multi-layer cloud-covered regions, these indices and the MOD time-composite image give an indication of the different types of events associated with that region for the time period (e.g. clear sky and thin or small broken cloud, high thick cloud, middle thick cloud, low thick cloud). The presence of rate events, such as cirrus or altostratus over desert regions (e.g. Sahel or N. Africa), is indicated by comparing the coldest and brightest radiances (IR CUM-MIN/VIS CUM-MAX) with the average value of the daily histogram parameter (AVG-MIN/AVG-MAX) and the more frequent, coldest and brightest values for single pixels (VIS IM-MAX and IR IM-MIN). For regions characterized by relatively frequent occurrences of cloud (e.g. areas that are not deserts), the IR IM-MIN/VIS IM-MAX, VIS/IR MOD values, the percentage of pixels colder or brighter than a certain value and the correlation between these coldest and brightest values characterize the coldest and brightest clouds.

Table 3. VIS and IR CUM parameters for SAMPLED data for the seven regions (region size 60 × 120 pixels). IR MIN/MAX represent the warmest and coldest temperatures respectively. One count is approximately equivalent to 0·5 per cent for VIS and 0·5°C for IR.

				VIS						
REG	MIN	MAX	AVG	MOD1	PEAK1	MOD2	PEAK2	SD	COR	AREA
N. Africa	72	152	90	92	33	92	5	6	0·1	3
Sahel	72	196	95	96	25	96	2	11	0·4	6
South Atlantic	12	132	72	76	7	76	3	27	0·7	3
Tropical Atlantic	8	208	64	16	14	16	3	54	0·9	14
Tropical Africa	32	200	74	44	13	44	2	37	0·9	12
Europe	28	196	79	48	17	44	2	40	0·9	13
North Atlantic	12	144	62	16	12	16	4	32	0·5	10
				IR						
	MAX	MIN								
North Africa	24	176	42	38	14	38	5	11	0·1	3
Sahel	20	177	71	62	5	78	2	26	0·4	6
South Atlantic	111	133	124	124	32	124	3	4	0·7	3
Tropical Atlantic	101	231	144	108	8	108	3	38	0·9	14
Tropical Africa	83	232	136	108	6	104	2	33	0·9	12
Europe	72	220	132	96	5	96	2	37	0·9	13
North Atlantic	117	210	140	126	8	120	4	18	0·5	10

A key characteristic of these different indices and parameters is that they are almost independent of location within relatively large regions, making them good descriptors of climate regimes. However, in larger regions temperature contrasts between ocean and land or with altitude in mountain ranges may be rather high. To avoid such situations, time-composite images can be used to test the spatial homogeneity of surface properties (and mean cloud properties) and to select statistically homogeneous regions. Thus, the effects of changes in data resolution or of data sampling schemes on climate statistics can be assessed by how these changes affect the histogram characteristics.

3. Histograms with different space and time resolutions

3.1. *Comparison of associated histogram statistics*

Comparison of the SAMPLED histograms with the FULL resolution histograms shows differences that depend on the histogram shape and, consequently, the climate regime. Visual comparisons can be made using figures such as figures 1 and 2. In these figures the FULL resolution histogram (represented by the solid contours) is overlaid by the reduced-resolution histogram (thin line segments which only show which radiances occur at least once in the reduced data). Note that the number of positions in the VIS–IR radiance plane occupied in the reduced-data histograms is smaller than in the FULL resolution histograms. From the monodimensional histograms, we also see that the frequency of occurrence of each radiance value differs significantly between the FULL resolution and reduced-resolution DAILY histograms (figure 2: the frequencies are scaled by the individual mode frequencies), but that these differences are smaller for the time-cumulated histograms (figure 1). To quantify the differences between the two reduced-data histograms and the FULL resolution histograms, we compare first the radiance statistics (discussed in SR91) extracted from these distributions.

For the histograms cumulated over the whole time period (figure 1), the radiance pairs occurring at FULL resolution, but not at the SAMPLED resolution, belong generally to the lowest frequency parts of the histograms (outside the 80 per cent isoline). This is reflected in the ratio of the values of CUM-AREA shown in table 2, an underestimation of VIS CUM-MAX, corresponding to the brightest cloud, and an overestimation of IR CUM-MIN, corresponding to the coldest cloud. The difference in behaviour between the VIS CUM-MAX and IR CUM-MIN is due to the larger variability of the cloud reflectance compared with the tendency for clouds to form in layers, where their temperatures are constrained by the altitude of the tropopause in the ITCZ or the height of the boundary layer in the S. Atlantic region. On the other hand, the much smaller spatial/temporal variability of surface properties (see SR91) reduces the differences of the extreme radiances coresponding to the surface (VIS CUM-MIN and IR CUM-MAX) to nearly zero for the oceanic regions and to <2° and 4 per cent for the land regions. We note, however, that the 95th and 5th percentile radiance values for the FULL and SAMPLED VIS and IR distributions, respectively, are equal.

The other time-cumulated statistics, such as the CUM-AVG, MOD and SD, are nearly identical for both FULL and SAMPLED data sets. Since the FULL and SAMPLED MODs have the same value, we compare the differences between the FULL and SAMPLED PEAK values with the differences expected for a random sample instead of our systematic sample. For all regions, the frequency differences are

Table 4. AVG and SD parameters (first and second line) for SAMPLED data and the average and SD of the differences between the FULL and SAMPLED daily parameters during the 17 (or 20) day period (third and fourth line) for the seven regions (60 × 120 pixels). For IR the lowest values correspond to MAX temperatures and the highest values to MIN temperatures.

	VIS									
REG	MIN	MAX	AVG	MOD1	PEAK1	MOD2	PEAK2	SD	COR	AREA
North Africa										
	76	110	90	91	36	91	9	5	−0·1	0·6
	2	11	1	2	4	3	3	1	0·4	0·2
	−6	10	0	0	−3	0	−1	1	0·0	1·4
	2	11	0	1	1	1	1	0	0·0	1·5
Sahel	77	125	96	97	28	96	6	8	0·2	1·2
	2	25	5	3	6	3	2	6	0·4	0·2
	−2	20	0	0	1	0	−1	0	0·1	3·2
	2	15	1	1	2	2	1	1	0·1	1·7
South Atlantic	24	117	71	69	12	73	8	22	0·6	0·8
	15	9	15	27	3	26	3	6	0·3	0·2
	−7	6	1	1	−2	1	−2	0	0·0	0·9
	9	5	2	5	1	3	1	1	0·1	0·4
Tropical Atlantic	14	173	65	34	17	34	7	41	0·7	1·7
	4	44	31	48	14	51	7	15	0·2	0·5
	−1	21	−1	3	0	12	0	0	0·0	11·4
	1	14	3	12	2	41	1	2	0·0	4·2
Tropical Africa	38	172	74	59	16	60	4	29	0·8	1·7
	4	22	22	30	7	39	3	10	0·1	0·4
	−5	20	0	−1	−1	1	−1	0	0·0	9·1
	3	9	2	6	2	8	1	1	0·0	3·4
Europe	41	151	79	68	22	79	4	22	0·6	1·5
	14	35	31	37	12	41	2	11	0·3	0·5
	−11	18	0	4	−1	7	−1	0	0·0	7·2
	8	11	1	19	2	26	1	1	0·1	3·7
North Atlantic	17	119	61	49	16	38	7	27	0·4	1·6
	6	14	17	35	13	30	9	5	0·2	0·4
	−2	10	1	−4	−1	2	−1	0	0·0	6·0
	4	5	1	18	1	4	1	1	0·0	1·7

Table 4.—*continued*

			IR							
	MAX	MIN								
North Africa	30	78	41	38	20	37	9	8	−0·1	0·6
	4	41	6	6	7	4	3	6	0·4	0·2
	−2	13	1	0	2	0	−1	0	0·0	1·4
	1	14	1	4	2	2	1	1	0·0	1·5
Sahel	40	137	72	65	12	66	6	19	0·2	1·3
	14	32	17	16	5	14	2	6	0·4	0·2
	−5	17	−1	1	−2	0	−1	0	0·1	3·2
	3	11	1	5	2	3	1	1	0·1	1·7
South Atlantic	116	128	123	125	47	124	8	3	0·6	0·8
	114	2	2	2	20	4	3	1	0·3	0·2
	−3	1	0	0	0	0	−2	0	0·0	0·9
	2	1	0	1	3	1	1	0	0·1	0·4
Tropical Atlantic	108	211	144	129	14	121	7	28	0·7	1·7
	7	29	24	37	11	34	7	11	0·2	0·5
	−2	3	0	−1	−1	6	0	0	0·0	11·4
	2	4	2	9	2	23	1	1	0·0	4·2
Tropical Africa	100	204	136	122	10	121	4	23	0·8	1·7
	8	21	22	31	4	34	3	8	0·1	0·4
	−4	8	0	2	−1	0	−1	0	0·0	9·1
	2	10	2	9	1	12	1	1	0·0	3·4
Europe	97	185	132	127	10	125	4	19	0·6	1·5
	18	30	30	42	4	4	2	8	0·3	0·5
	−6	7	1	5	−2	5	−1	0	0·0	7·2
	4	7	1	15	2	22	1	1	0·1	3·7
North Atlantic	121	187	139	127	16	126	7	15	0·4	1·6
	3	18	9	5	10	6	9	5	0·2	0·4
	−2	8	1	4	−1	0	−1	0	0·0	6·0
	1	5	1	11	2	1	1	1	0·0	1·7

Table 5. CUM parameters (first line), the MOD of the IM parameters (second line), and the percentage of pixels with a radiance within a range, ±delta, of the MOD value for the six of the seven regions (60 × 120 pixels) for SAMPLED data. The parameters shown are the MIN, MAX, AVG and MOD. Delta values in VIS are 4 per cent for MIN, 6 per cent for AVG and MOD, and 8 per cent for MAX; in IR they are 3°C for MAX and 4°C for AVG and MOD. In the fourth line, the fraction of pixels in the IR IM-MIN image colder than some value or pixels in the VIS IM-MAX image higher than some value are also reported. IR MIN/MAX refers to warmest/coldest temperatures.

	VIS				IR			
REGION	MIN	MAX	AVG	MOD	MAX	MIN	AVG	MOD
N. AFR.	72	152	90	92	24	176	41	38
	88	96	88	92	29	56	43	36
	94	94	99	98	90	—	97	97
					1·5% pixels colder than 160			
SAHEL	72	196	95	96	20	177	72	62
	92	104	96	96	46	168	78	80
	90	70	100	99	42	—	50	38
					23% pixels colder than 160			
S. ATL.	12	132	72	76	111	133	124	124
	12	108	68	76	115	128	124	125
	65	93	86	55	100	—	100	100
					94% pixels = MOD ± 1°			
TR. ATL.	8	208	65	16	101	231	144	108
	12	188	64	16	102	220	146	108
	100	72	68	88	100	—	80	78
	80% pixels brighter than 150				94% pixels colder than 200			
TR. AFR.	32	200	74	44	83	232	136	108
	40	164	76	44	100	222	136	106
	100	60	91	91	83	—	85	68
	80% pixels brighter than 150				77% pixels colder than 200			
EUROPE	28	196	79	48	72	220	132	96
	44	152	76	48	86	200	133	98
	99	84	95	98	88	—	90	64
	77% pixels brighter than 150				92% pixels colder than 180			

found to be less than the standard deviation of a theoretical error distribution for random sampling (Saporta 1978).

Increasing the cumulation region size for the histograms from 60 × 120 to 120 × 240 pixels (if no land/sea or climate boundary are encountered) improves the correspondence between the FULL and SAMPLED resolution histograms (table 2) and decreases the deviations in all the parameter values. However, although VIS CUM-MAX does increase in the SAMPLED data set for larger cumulation regions, the difference between FULL and SAMPLED VIS CUM-MAX does not decrease where high thick clouds are present (e.g. the tropics).

Comparison of the SAMPLED and FULL daily statistics, represented by the values of AVG and SD parameters (table 4), again shows good agreement for the quantities dominated by the higher frequency portions of the distributions, namely AVG-AVG, AVG-MOD and AVG-SD; but differences are larger for the IR AVG-MIN, VIS AVG-MAX and the AVG-AREA values. On some days, much larger differences occur, even for the MOD and PEAK values (e.g. the Tr. Atlantic, where complex cloud systems produce such large radiance dispersions that they do not have a well defined MOD). Except for regions where large radiance excursions are rare, the average daily differences of VIS-MIN/IR-MAX are larger than the differences found for the VIS CUM-MIN/IR CUM-MAX.

The evolution of the time-cumulated histogram parameters for the SAMPLED data with increasing accumulation time period is generally similar to the evolution of the same parameters found at FULL resolution (see SR91), namely the values converge towards the FULL values obtained over the longest cumulation time. However, the MIN and MAX parameters attain their long-term values more slowly than the other parameters.

The deviations of the CUM and AVG parameters for the AVERAGED data from those of the FULL resolution parameters are slightly larger than for the SAMPLED CUM and AVG parameters. For the extreme radiances (especially IR MIN/VIS MAX) the differences are larger than the SAMPLED differences by factors of 1·5 to 2; usually, the largest differences are for VIS CUM-MAX. The radiance distribution variability (represented by AVERAGED CUM-SD and AVG-SD) is smaller than the SAMPLED and FULL variability, which is the expected effect of averaging the daily values (Shih *et al.* 1984). In the ITCZ, this underestimation of the total variability can reach 5 K in IR and 10 per cent in VIS. The larger effect on the VIS is due to the larger variability of cloud reflectivity at the smallest scales (affecting VIS MAX); IR MIN for the coldest and thickest clouds does not vary as much at smaller scales because the clouds are organized into layers.

Comparing the AREA ratios (table 2) for the AVERAGED and FULL histograms to those of the SAMPLED and FULL histograms emphasizes the difference in behaviour between the SAMPLED and the AVERAGE data set. Although these ratios are very similar for the two data sets for one day and small regions, as the time period and region size increase, the AVERAGED/FULL ratio does not grow as rapidly as the SAMPLED/FULL ratio. In other words, the SAMPLED data give a better representation of the FULL resolution histograms as the SAMPLED population size grows, particularly in the multi-level cloud regions, while in regions where rare events can occur neither reduced data set represents these rare events well.

3.2. *Histogram area*

3.2.1. *Regional behaviour*

The variation of the ratio of the SAMPLED and FULL histogram AREAs as the time period and the region size for accumulation are increased and the average ratio obtained for the daily histograms reveals three patterns, associated with the three types of histogram shapes that occur (figure 1; tables 2 and 6).

Group 1: in regions characterized by low cloud-cover amounts (e.g. N. Africa and Sahel), the evolution of the AREA ratio with time shows both a period of stability and a period of decrease (figure 3), despite a very compact distribution (80 per cent CUM-AREA ≪CUM-AREA). The unchanging ratio is observed during the period of clear sky and can be explained by the fact that the SAMPLED data are selected from the full resolution METEOSAT images at constant locations; increasing the population of pixels does not improve the correspondence to the FULL data. For other satellites, where the relation of scene element and location is variable, time cumulation of the histograms would produce convergence of this ratio, because the image-to-image 'motion' of the sampled pixels eventually covers more of the region. The infrequent appearance of sparse cirrus, some of which is missed by the SAMPLED data on a particular day, causes significant increases in CUM-AREA

Table 6. Ratio of the average daily values of AREA for the SAMPLED and FULL data histograms and of the values of AREA for histograms cumulated over 5 days, 10 days and 17 days (CUM-AREA) for 60 × 120 pixels and three radiance resolutions, (1) 2 per cent–1°C, (2) 4 per cent–2°C, and (3) 4°–6°C. Results are shown only for four of the seven regions.

DAY	TROPICAL ATLANTIC			CENTRAL AFRICA		
Resolution	1	2	3	1	2	3
Avg. daily	0·14	0·27	0·45	0·16	0·30	0·44
5	0·27	0·54	0·71	0·29	0·53	0·67
10	0·39	0·66	0·77	0·40	0·65	0·75
17	0·51	0·73	0·81	0·49	0·72	0·78
DAY	SAHEL			SOUTH TROPICAL ATLANTIC		
Resolution	1	2	3	1	2	3
Avg. daily	0·30	0·43	0·52	0·47	0·60	0·73
5	0·36	0·46	0·51	0·62	0·68	0·78
10	0·39	0·49	0·53	0·69	0·72	0·80
17	0·40	0·52	0·59	0·71	0·73	0·83

without altering 80 per cent CUM-AREA and leads to a decrease in the SAMPLED/FULL AREA ratio on certain days. The ratio at the end of the cumulation period is especially low (25 per cent for N. Africa with 60 × 120 pixels). This effect also leads to large differences (25 counts ≈30–50 K) between the FULL and SAMPLED IR CUM-MIN. In contrast, due to the high frequency of clear sky and the homogeneity of surface properties compared with cloud properties, the average ratio of the daily AREA values is relatively high compared to regions with multi-level cloudiness. (The decrease in this ratio with increasing region size is produced solely by the introduction of land/ocean boundaries to the larger regions.)

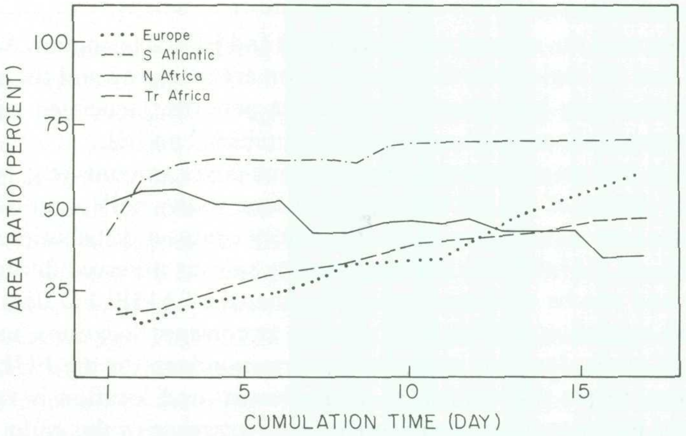


Figure 3. Evolution of the ratio of the bidimensional histogram areas for SAMPLED and FULL resolution data with cumulation time period for N. Africa (solid), S. Atlantic (dot–dashed), Tr. Africa (dashed) and Europe (dotted).

Group 2: in regions characterized by a persistent homogeneous cloud cover over the whole period (S. Atlantic), the AREA ratio increases in the first few days then stays constant (figure 3). When no cirrus occur, this ratio is largest (71 per cent) for the largest region size. Comparison of the CUM-AREA ratios with the average ratio of the daily AREA values (table 6) shows, that for an area of homogeneous cloud cover, accumulation of the data over time causes a significant increase (20 per cent) of the ratios. However, as in the desert regions, the occasional appearance of sparse cirrus during the period (more probable in the larger regions) can cause a decrease in the AREA ratio. The average daily AREA ratio is higher compared to the multi-level cloud cover regions, due to the small, but uniform dispersion of the radiances (80 per cent CUM-AREA ≈ CUM-AREA). Because of the very low frequency of clear sky pixels in this oceanic region, the differences between the daily FULL VIS-MIN and SAMPLED VIS-MIN are often large, much larger than for other regions.

Group 3: for regions characterized by multi-level cloud cover (Tr. Atlantic, Tr. Africa, Europe, N. Atlantic) AREA ratios increase rapidly with increasing cumulation time period and region size, reaching a maximum value around 70–75 per cent for a 17 day period for the largest region size. The average daily AREA ratio also increases with region size; but the maximum value is only about 30 per cent for Europe, the Tr. Atlantic, and Tr. Africa and about 40 per cent for the N. Atlantic because of the small number of pixels used to represent the large spatial variability that typically occurs on a daily basis. For example, for Tr. Atlantic and a region of 120 × 240 pixels, a FULL daily histogram occupies 1720 distinct (VIS–IR) radiance positions, on average, but the total number of pixels forming one daily SAMPLED histogram is only 800.

There are some differences in the evolution with time of FULL and SAMPLED AREA ratios between Europe and the tropical regions. For Europe the average daily AREA ratio (table 2) is larger due to the smaller daily spatial variability (table 4: smaller AVG-AREA and AVG-SD) and generally larger spatial scales of clouds (see SR91); but, because of the time variations in this region, the SAMPLED/FULL AREA ratio for Europe converges to a similar value as the ITCZ regions with increasing cumulation time period. This characteristic also means that the increase in AREA ratio with cumulation time period is more dependent on the particular days included in the data set for midlatitudes than for the tropics.

Thus, in regions with spatially and temporally variable cloud cover, the correspondence of the SAMPLED histograms to the FULL histograms gets better as the region size and/or the accumulation time period increases for the space and time scales that we have studied. For the homogeneous cloud cover regions (S. Atlantic) the same improvement occurs, especially in the first part of the cumulation period, as long as increasing the region size or time period does not introduce another kind of event. For the almost completely clear regions (N. Africa), the same kind of behaviour is found, although improvements as the time period increases are smaller and changes in surface properties can decrease the correspondence between the FULL and SAMPLED histograms in larger regions.

For the seven regions the maximum CUM-AREA ratios reached only about 70 per cent, due to accumulation over a relatively short time period: the convergence toward the 17 day cumulated value of AREA, observed after only 5 to 10 days for FULL resolution data (except in the case of rate events, see discussion in SR91), does not appear in the behaviour of the SAMPLED (or AVERAGED) resolution data. Instead, the SAMPLED histogram area continues to increase with time to the end of the 17 day period, implying that accumulation of the SAMPLED histograms for longer periods (e.g. one month) is required to improve the convergence.

3.2.2. *Radiance resolution effect*

Reduction of the radiance resolution generally decreases the differences between the SAMPLED and FULL histograms as shown in table 6. The CUM-AREA ratio increases, with the largest increase (about 20 per cent for a change of VIS/IR radiance resolution from 2 per cent/1 K to 4 per cent/2 K) occurring for the multi-level cloud covered regions for the smallest region size. Other increases in CUM-AREA ratio are smaller, because the differences between the FULL and SAMPLED data are associated primarily with rare events. For the multi-level cloud covered regions, the large dispersion of radiances (large AREA with low frequencies) guarantees that any particular radiance pair that does not occur in the SAMPLED data set is very near some other radiance pair that does. A further decrease of radiance resolution to 4 per cent/4 K, equivalent to 32 levels in VIS and IR, does not change the CUM-AREA ratio as much, indicating that most of the 'rarer' radiance values have already been combined with other more frequent radiances.

3.3. *Radiance frequency comparison*

To determine the significance of the radiance pairs in the FULL resolution histograms that are not represented in the SAMPLED (AVERAGED) histograms, to determine the accuracy of the radiance frequencies obtained at reduced resolution, and to quantify the discrepancy between the FULL and SAMPLED (AVERAGED) bidimensional histograms, we calculate: (1) the percentage of FULL pixels having radiance values that are not represented at all in the SAMPLED (AVERAGED) bidimensional histograms (called AREA-DIF), (2) the average FULL frequency of occurrence of these radiance values, (3) the relative frequency differences at each position in the radiance plane. The latter are compared to the standard deviation of the difference distribution expected for a random sampling of a normally distributed quantity. We also calculate the sum of the squares of the relative frequency differences (chi-square), and the sum, of absolute frequency differences (called DIF-SUM).

3.3.1. *SAMPLED data set*

For the daily histograms, as we saw in the previous section, the values of AREA-DIF for the FULL and SAMPLED histograms are large. For the smaller region sizes, they can reach 80 per cent for the very disperse histograms in the multi-level cloud regions (Group 3). However, for the larger region sizes, AREA-DIF decreases to only about 30 per cent. Reducing the radiance resolution from 64/128 levels in VIS/IR to 32/64 levels decreases AREA-DIF further to about 10 per cent. The most rapid decrease in AREA-DIF is produced by lengthening the time period of accumulation from 1 to 5 days. Increasing the number of pixels in any manner, by enlarging the region size or by increasing the cumulation time period, produces a similar decrease in AREA-DIF. For the 17 day (or 20 day) time-cumulated histograms and in 60×60 pixels regions, the maximum AREA-DIF is 27 per cent for the Tr. Atlantic, but only 2 per cent for the compact histograms associated with the persistent low cloud layers in S. Atlantic. For 120×240 pixel regions, the maximum value of AREA-DIF for time-cumulated histograms falls to 2 per cent, even for the multi-layered regions.

The improved correspondence between the SAMPLED and FULL histograms with increasing region size or cumulation time period is due (e.g. in the Tr. Atlantic) both to an increase in the ratio of SAMPLED and FULL CUM-AREA and to a decrease in the frequency value equivalent to one FULL resolution pixel. On the average, the number of FULL pixels with each radiance pair that is not represented in SAMPLED (AVERAGED) histograms is less than 20 for the more cloud-covered regions and less than 10 in the rarely cloudy desert regions, independent of the region size. For daily histograms, this average number is less than 10. At the particular SAMPLED/AVERAGED resolutions we have examined, a frequency of 20 FULL pixels is reduced to a frequency of about 0·5 pixel (20/36) (i.e. the unrepresented radiances in the SAMPLED histogram are those with a FULL pixel number that is equivalent to less than one pixel at reduced resolution). Thus, as the total pixel number (population of the sample) increases with increasing region size or time period, the FULL frequency (equivalent to 20 pixels) of the unrepresented radiances becomes smaller and smaller, equivalent to rarer and rarer events. Reduction of the radiance resolution from 64/128 levels to 32/64 in VIS/IR does not change the average pixel number at each radiance pair, but does reduce the number of discrete pairs that must be represented in the histograms, as we saw in the previous section.

Looking at the frequency differences as a function of the frequencies themselves, shows that the largest differences occur in the lowest frequency part of the histograms; this is particularly true for the multi-level cloud cases, where the large histogram dispersion (figure 1) induces a large number of very low frequencies (reflected in values of 80 per cent CUM-AREA $\ll$ CUM-AREA). For each of the radiance pairs present in the SAMPLED daily and time-cumulated histograms, the relative differences at each frequency between the FULL and SAMPLED histograms (frequency differences normalized by the SAMPLED frequency) are calculated. We then compute the percentage of radiance positions with a frequency difference less than twice the standard deviation of the theoretical error distribution expected for random sampling (Saporta 1978): for a random sampling this percentage is expected to be close to 95 per cent. In the present case, for all the regions, more than 95 per cent of the differences between our systematically SAMPLED and FULL daily[†] and time-cumulated, mono- and bidimensional histograms are within the expected range. As the pixel population increases by cumulation over larger regions or longer times, this percentage decreases toward the theoretically expected value (95 per cent).

To understand this effect, we also calculate the frequency differences for a RAMDOM-SAMPLED data set produced by a specially random selection from the FULL resolution data, as opposed to the SYSTEMATIC-SAMPLED set considered above: for the RANDOM-SAMPLED data the percentage of frequency differences in the expected range is generally lower than or equal to those for the SYSTEMATIC-SAMPLED data set and their variability is larger. This occurs because of the strong correlation of radiances on small spatial scales; systematic sampling is more likely to capture some rare radiances, especially if they are spatially clustered. This conclusion is supported by the one exception to this behaviour, which is N. Africa, where for the VIS time-cumulated histogram, the lower percentage (about 87 per cent of frequency differences are smaller than expected) in the SYSTEMATIC-SAMPLED case comes from the constant location of the SAMPLED pixels. Thus, for a given SYSTEMATIC-SAMPLED histogram, the relative difference in frequency with the FULL histogram at a particular radiance pair has an equal or higher probability to be

---

[†] For the daily histograms, we do not consider each histogram separately, but rather the frequency set formed from all of the daily histograms for a given region.

Table 7. Half the average of the absolute frequency difference between SAMPLED and FULL daily histograms and for histograms cumulated over 5, 10 and 17 days for three region sizes. Half the average value represents the fraction of the total FULL pixel population with radiances that are not represented at SAMPLED resolution.

DAY	TROPICAL ATLANTIC			CENTRAL AFRICA		
Region size	60 × 60	60 × 120	120 × 240	60 × 60	60 × 120	120 × 240
Avg. daily	60	53	41	62	55	38
5	49	38	22	45	36	21
10	41	29	16	37	28.	15
17	33	23	12	29	21	12
	SAHEL			SOUTH TROPICAL ATLANTIC		
Region size	60 × 60	60 × 120	120 × 240	60 × 60	60 × 120	120 × 240
Avg. daily	35	30	27	27	22	15
5	23	19	14	16	13	8
10	19	14	11	14	11	6
17	17	13	9	11	9	5

smaller than that expected for a completely random sample. As the size of the sample population grows, this difference in sampling disappears.

To quantify the overall agreement between the FULL and SAMPLED time-cumulated distributions, the chi-square values are computed for the time-cumulated histograms. (The chi-square values[†] are the sums of the squares of the frequency differences; see Saporta (1978) and Snedecor and Cochran (1980)). All the histograms show good agreement between the FULL and SAMPLED distributions; in other words, all the chi-square values are smaller than the reference value and the differences are not significant at the 50 per cent level (Snedecor and Cochran 1980), except for the N. Africa VIS histogram. We note that larger chi-square values are found for the VIS histograms than for the IR histograms, related to the larger spatial variability at smaller scales of the VIS radiances than the IR radiances. A similar effect on the agreement between FULL and SAMPLED radiance distributions produced by differences in spatial variability was also noted by Duvel (1983) for daily IR distributions in 2·5° × 2·5° boxes (in his study the radiance resolution was reduced to 15 levels).

The final measure of the discrepancy between the FULL and SAMPLED radiance histograms for the different regions is the sum of absolute radiance frequency differences between the FULL and SAMPLED (AVERAGED) distributions (called DIF-SUM); half of this value indicates the fraction of the total population not represented by the same radiance pairs in the two histograms. Examples of these differences are given for different time periods, region sizes and radiance resolutions in tables 7 and 8. Their magnitude again depends on the particular climate regime (histogram shape) and decreases with increasing time period, region size, and decreasing radiance resolution, except for the 'rare event' cases.

3.3.2. *AVERAGED data set*

Daily histograms for 60 × 60 pixel regions for AVERAGED and SAMPLED data exhibit very similar values of DIF-SUM when compared to the FULL histograms; however, the differences for the AVERAGED histograms decrease less rapidly than for the SAMPLED histograms as the population increases (tables 7 and 8). Moreover, the maximum frequency of the radiances not represented in the AVERAGED data set can be twice that found for the SAMPLED set. The percentage of frequency differences smaller than twice the standard deviation of the theoretical difference is also less than for the SAMPLED case and decreases as the number of pixels increases. For the time-cumulated monodimensional histograms, which have a small number of low frequencies, this percentage can be as low as 65 per cent in VIS. For the time-cumulated histograms, the chi-square value is often larger than the theoretical value. In other words, the differences between AVERAGED and FULL histograms are systematic and significant at the 99·9 per cent level. Thus, averaging the data changes the shape of the frequency distribution and permanently eliminates some radiance values in the lower frequency portions of the distributions. Even for the high-frequency parts of the histograms, the differences between the FULL pixel percentage and AVERAGED pixel percentage situated inside the 80 per cent isolines of the FULL histogram reaches 12 per cent for the stratocumulus area and 7 per cent for the multi-level cloud regions. As the time period or region size increases, the AVERAGED data become less representative of the FULL resolution data set relative to the SAMPLED data set.

4. Comparison of time-composite image statistics

The time-composite images (see figure 12 in SR91) represent for each location some of the main characteristics of the time variations of the surface and cloud cover

Table 8. Same as table 7, but for the AVERAGED and FULL histograms.

DAY	TROPICAL ATLANTIC			CENTRAL AFRICA		
Region size	60 × 60	60 × 120	120 × 240	60 × 60	60 × 120	120 × 240
Avg. daily	58	53	43	64	58	42
5	51	43	25	52	43	27
10	44	32	20	44	34	22
17	36	27	16	25	28	19
	SAHEL			SOUTH TROPICAL ATLANTIC		
Region size	60 × 60	60 × 120	120 × 240	60 × 60	60 × 120	120 × 240
Avg. daily	38	33	30	40	34	25
5	28	22	18	28	24	17
10	22	18	14	25	21	15
17	22	17	13	19	18	13

[†] For a random sample, the chi-square values exhibit a specific probability distribution. Then, depending on the number of samples and the number of radiance intervals in the histogram, we can determine a reference value of chi-square such that the probability of the chi-square from a specific sample being greater than the reference value is 50 per cent.

for the considered period, and they provide some indication of the relative spatial and temporal variability of the surface properties, of the frequency of occurrence of clouds over the region, and of the main characteristics of the most common cloud types that occur. In this section, the effect of data sampling on the information contained in these images is examined.

In SR91 the same histogram analysis was performed on these time-composite images as on the original images to obtain measures of the spatial variability of the point time statistics. The spatial variability of these images was also examined by calculating radiance differences as a function of distance between locations. The statistics associated with time-composite images (table 5) were compared with the regional CUM parameters: CUM-MIN, CUM-MAX, CUM-AVG, and CUM-MOD. Here, the quantities associated with mono- and bidimensional histograms at SAMPLED resolution are compared with those obtained at FULL resolution. As in the comparison of the daily and time-cumulated histograms above, we calculate the sum of the absolute frequency differences between the FULL and SAMPLED bidimensional histograms (DIF-SUM) for each time-composite image.

The DIF-SUM values are generally smaller than the average difference for the daily histograms, since a large part of the spatial variation is removed by the compositing. Exceptions are the VIS IM-MAX and IR IM-MIN in the desert regions due to the higher spatial variability of these quantities. Differences in the time-composite images with increasing separation distance are generally very small, except for VIS IM-MAX (SR91). It is this spatial homogeneity of the radiance statistics inside a region and the relatively large scale of that homogeneity that explain why increasing the cumulation time period or the region size improves the correspondence of the SAMPLED and FULL histograms.

The FULL/SAMPLED histogram frequency differences for the original images are strongly related to the spatial variability of the associated composite image: the larger the spatial variability in the time-composite image, the larger the differences between the FULL and SAMPLED histograms. For example, VIS IM-MIN, IR IM-MAX and the VIS/IR IM-AVG, which have the lowest SD and the lowest spatial variability (less than 2 per cent and 2 K for a 5 pixel separation), the DIF-SUM values between the FULL and SAMPLED distributions are generally small compared to the daily histogram differences: less than 20 per cent of the total population for the VIS IM-MIN and IR IM-MAX and less than 30 per cent of the population for the VIS and IR IM-AVG. For IR IM-MOD, despite the larger spatial variability of the IR, the differences between the FULL and SAMPLED distributions are still not very large. The differences for the VIS IM-MOD are also small, except for regions with a high percentage of stratocumulus (e.g. S. Atlantic) where the IM-MOD represents cloudy and not clear radiances. For the VIS IM-MAX and IR IM-MIN the DIF-SUM values are the largest and are similar to those for the daily histograms, reflecting the large spatial and temporal variability of the coldest and brightest clouds compared, for example, to surface properties (see SR91). As with the time-cumulated histograms, a large part of these differences come from the low-frequency part of the histogram and generally decrease as the region size increases.

The FULL and SAMPLED parameters associated with the time-composite image histograms (table 5, also table 7 in SR91) are also very similar, even in cases where some differences appear in the histogram frequencies. The exceptions generally involve the extreme values. Some small differences can also appear in the SD (which is usually less than 1 per cent except for the highest variability images) and the PEAK

(usually 2 to 3 per cent). The MOD values are very close (differences less than 2 per cent and less than 2 K) and the percentage of pixels around the MOD, plus or minus a certain amount, are also very similar. The Sahel is different; there the regional mode of IR IM-MIN lacks a distinct PEAK due to the very large spatial variations of the radiances in that region.

Generally, the relationships found at FULL resolution between the time-composite parameters and the VIS and IR CUM-MIN, MAX, AVG and MOD are preserved in the SAMPLED data set, showing its ability to describe the climatology of these regions. That the SAMPLED data set can replicate the main characteristics of the FULL time-composite images, in particular their spatial variability, demonstrates again the relative homogeneity of the radiance spatial variations as compared to the time statistics for a given region (see SR91). Even though we sampled the data at a larger spatial scale (30 km), the statistics of variations at smaller scales (5 km) have been preserved.

5. Discussion

5.1. *Scales of space/time variability*

That the SAMPLED and AVERAGED histograms both coverge towards the FULL histograms indicates that the variability at smaller (less than 30 km) spatial scales is smaller than the larger scale variability and that it is 'homogeneous' (i.e. the statistics are the same) over the larger scales. The direct test of this proposition (differences as a function of separation distance) shows that the magnitude of the spatial variations is smaller on the smaller scales and that the rate of increase of the differences is larger on scales less than about 30 km and generally decreases from 30 to 350 km (see SR91). If the increase in distance does not introduce a new type of surface or cloud, this conclusion can be extended to scales of 350 km to 1200 km in longitude and 700 km in latitude. The very small differences between locations in the time-composite images also illustrate the homogeneity. That the SAMPLED data reproduce the FULL data statistics somewhat better than the AVERAGED data indicates that some significant smaller scale variations are lost by averaging the data at 30 km scale, but are almost completely retained by the sampling. In other words, as long as the sample population is large enough, sampled data can retain information, in a statistical form, about variations on the scale of the actual instrument resolution.

The increase in radiance variability with spatial scale is due mainly to the variability of the cloud cover and not to surface properties. For example, differences as a function of separation distance in the VIS MIN and IR MAX time-composite images (usually representing surface properties) shows that the increase in variability on scales from 5 to 350 km is very small (<2 per cent and 6 K), even for the highly variable Sahel region (SR91). However, it must be noted that large variations of surface properties can occur on small scales (i.e. if coastlines, maintain ranges or large lakes are included). The VIS MAX and IR MIN time-composite images show, on the other hand, that the variability of clouds is generally larger at larger scales; but the fact that the maximum value for radiance differences is almost independent of the spatial scale indicates that the largest changes in cloud properties can still occur on very small spatial scales (see SR91). Thus, these results also support the conclusion that, while sampling at original resolution can preserve information about these smaller scale variations, averaging destroys it.

The homogeneity of small scale variability statistics over scales of about 500–2000 km and relatively long time periods (cf. discussion in SR91) indicates that the population size of a sampled data set can be made larger by increasing the size of the cumulation region or time period or both. This is also shown by the fact that, as region size increases, the radiances occurring on a certain day are more comparable to the time-cumulated radiance distribution for that region (SR91). However, some regions (e.g. Europe) show somewhat more variability from day to day than they exhibit spatially (i.e. increasing region size does not produce as rapid an improvement in the correspondence of the daily histograms and the cumulated histogram). Convergence of the time-cumulated SAMPLED and AVERAGED histograms towards the FULL resolution histogram is also not as rapid in this case.

5.2. *Measurement strategies and methods*

5.2.1. *Measurement resolution and sampling*

The importance of using a very high spatial resolution for satellite observations has long been argued (Lo and Johnson 1971); however, this produces very large data sets that may be impractical to analyse. That the SAMPLED histograms converge more closely to the FULL histograms as the sample size increases suggests that the 'statistical' resolution of the SAMPLED data is equivalent to that of the FULL data. Although the actual magnitude of the variability at scales <30 km was found to be small, the AVERAGED data did lose some information about the smaller scale variations. Hence, if the STATISTICS of the smaller scale variations are all that need to be retained, the resolution of individual image pixels can be maximized, but data volume can be reduced to a manageable size by sampling.

The sampling frequency must be selected to provide a large enough population to capture the rarer events over sufficiently small spatial and temporal scales. Our results suggest that the significant variations (i.e. those not caused solely by a local change in cloud cover properties inside the same cloud system) occur on scales greater than 100 km. However, rare events such as cirrus or isolated convection can take place at smaller scales. The time sampling must resolve the diurnal cycle, but our results suggest that this may only require sufficient observations at the same time of day without observations every day (Harrison *et al.* 1983, Brooks *et al.* 1986). Our results are not adequate to judge the predominant time scales of variation.

The dependence of the histogram differences on radiance resolution, together with the fact that the differences occur primarily in the very-low-frequency portions of the histograms, suggests that much of the difference between SAMPLED and FULL histograms may have little significance for climatological averages. However, high radiance resolution may also be necessary to detect the presence of very thin cirrus or low broken clouds. Moreover, the divergence of some of the AVERAGED statistics from the FULL statistics suggests that some of the low-frequency component represents specific, but relatively rare, events (e.g. cirrus over Sahara) that are better captured in the SAMPLED data than in the AVERAGED data. Only if the lower frequency radiances occurred by themselves would the AVERAGED data be able to represent them. The nonlinear relationship between the VIS and IR for certain cloud types is also distorted by averaging (SR91). Thus, the validity of specific analysis methods must be judged against the amount and scales of radiance variability and the desired accuracy of the measurements. Particular note should be taken of the effect of rare events on the rate of convergence of some parameters.

5.2.2. *Cloud and surface statistics*

The number of pixels in the SAMPLED resolution daily histograms is small; if the radiance variability is high, many radiance values will not be represented. This means that small differences in observation location, time, resolution, and sampling strategy can produce large variations in the results, in regions where the radiance dispersion is large. For example in Tr. Africa, the average number of different VIS/IR radiance pairs occurring in the FULL daily histograms (for 60×120 pixels) is 900, whereas the number of pixels available in the SAMPLED data to represent these 900 distinct radiance pairs is only 200. Even if the radiance resolution is reduced to 2 K and 4 per cent, the number of distinct radiance pairs in the FULL data is still 300.

Comparison of the differences in all the statistics between the SAMPLED/AVERAGED time-cumulated histograms and those obtained from the FULL data shows that these differences are situated in the lowest-frequency part of the histograms. This is illustrated in figure 1 for the SAMPLED data. As the SAMPLED histograms are accumulated for larger region sizes and/or time periods, the differences become smaller because the number of new radiance values occurring in the larger regions or time periods does not grow as rapidly as the total number of pixels (compare figures 1 and 2). This effect is a consequence of the statistical stability of the radiance variations at these scales, as represented by the stability of the histogram areas (SR91). As the number of samples (pixels) increases by accumulation, the frequency of pixels at each radiance value is larger, reducing the errors in the corresponding SAMPLED histogram frequencies and statistics.

Our results reveal three types of behaviour: regions where rare events occur (clouds in deserts, clear sky in the marine stratus), regions where the spatial variations of cloudiness on a daily basis are generally similar to the time cumulated variations (Tr. Africa, Tr. Atlantic, N. Atlantic), and regions where the daily spatial variations are smaller than the time variations (Europe). In the first case, the reduced data sets do not represent the radiance variations associated with the 'rare' events very well and cumulation of the data does not improve the representation as rapidly. Obtaining a good description of these regions may require very long cumulation periods over large regions. In the second case, the cumulation of the data over either larger regions or time periods readily improves the reduced data representation of the variations. In the third case, although the time cumulation of the reduced data can accurately portray the time-averaged characteristics of the radiances, the only way to reconstruct the daily statistics is to make use of the spatial homogeneity of the daily statistics over large regions to obtain a large enough sample or to classify each different type of event and cumulate radiance distributions over many examples of like events.

The large temporal and relatively large spatial radiance variability indicate that the comparison of different types of data (for example, measurements from two different satellites with different resolutions or point measurements by lidars with spatially extended satellite data) must be effected carefully to get a meaningful result. Since the radiance variability that occurs in a small region at one time can be fairly large, observations of the same scene by two different satellites can result in significant differences in the inferred situation for individual measurements. These large differences will usually occur in the inferred cloud properties, because they depend on the lower frequency radiances included in the observations (e.g. Tr. Africa), whereas the clear radiances, associated with the higher frequency part of the distributions, are more accurately represented with a sampled data set. The S. Atlantic provides a significant counter-example. However, the same convergence behaviour, discussed in this paper, should occur in such comparisons (i.e. cumulation of the comparisons over larger regions and time periods is necessary to detect any systematic differences between the observations). The relatively good correspondence of the time-series statistics at a point to the regional time statistics and the conservation of these statistics in a SAMPLED data set also show that point measurements and spatially extended measurements of clouds, sampled or not, can be compared with good success if done in a statistical way. The same success may not occur for measurements of surface properties in heterogeneous locations, unless the observations are statistically random in location.

5.2.3. *ISCCP statistics*

These results show specifically that the ISCCP sampled data can provide a good measure of the original radiance variations on special scales down to the scale of the original image pixel sizes (4–8 km) if a sufficiently large population is used to calculate the statistics. If cloud and surface variations on scales as small as 250 km are to be described accurately, then the ISCCP results must be aggregated over at least a month (this is the stated objective of ISCCP, Schiffer and Rossow (1983)). If the variations on time scales as short as 1 day are to be described, then the results must be aggregated on spatial scales of 500 to 1000 km, being careful not to cross between different climate regimes (figure 3 in SR91).

5.3. *Conclusions*

If cloud and surface properties vary more in time than in space (e.g. Europe), then spatial averaging to reduce data volume does not remove much information; however, such a reduction would eliminate information about the more spatially variable clouds in the tropics. Such regions could be studied with time-averaged data, however. Sampled data can preserve the variation statistics of the full-resolution data for both types of regions, including giving some indications of rarer, more extreme events, especially if the sampling is systematic. Sampling also preserves other relationships in the data, such as radiance correlations, that are not always captured by averaged data. In other words, sampled data can describe the original data distribution shape, if a sufficiently large population is obtained, whereas averaging data changes the data distribution shape. It is the radiance distribution shapes that provide a distinct characterization of the cloud and surface types that are part of the definition of a climate regime (SR91).

Acknowledgments

This study was made possible by support from the NASA Climate Program (managed by Dr Robert A. Schiffer) for one of us (GS) to visit NASA Goddard Institute for Space Studies. We also benefited from conversations with colleagues; we thank M. Desbois, L. Garder, E. Matthews and I. Fung. Figures were drawn by L. Del Valle and final word-processing was done by E. Devine.

References

Brooks, D. R., Harrison, E. F., Minnis, P., Suttles, J. T., and Kandel, R. S., 1986, Development of algorithms for understanding the temporal and spatial variability of the Earth's radiation balance. *Reviews of Geophysics*, **24,** 422–438.

Cahalan, R. F., and Joseph, J. H., 1989, Fractal statistics of cloud field. *Monthly Weather Review*, **117,** 261–272.

Cahalan, R. F., Short, D. A., and North, G. R., 1982, Cloud fluctuation statistics. *Monthly Weather Review*, **110,** 26–43.

Coakley, J. A., and Baldwin, D. G., 1984, Towards the objective analysis of cloud from satellite imagery data. *Journal of Climate and Applied Meteorology*, **23,** 1065–1099.

Desbois, M., and Sèze, G., 1984, Use of space and time sampling to produce representative satellite cloud classifications. *Annuals of Geophysics*, **2,** 599–606.

Desbois, M., Sèze, G., and Szejwach, G., 1982, Automatic classification of clouds on METEOSAT imagery: Application to high-level clouds. *Journal of Applied Meteorology*, **21,** 401–412.

Duvel, J. P., 1983, Étude de facteurs influencant le bilan radiatif de la terre et sa détermination à partir de données satellitaires notamment l'anisotropie et la variation diurne tellurique. Diplôme de Docteur de Troisième cycle de l' Université Pierre et Marie Curie, Paris VI.

Duvel, J. P., 1988, Analysis of the diurnal, interdiurnal and interannual variations during northern hemisphere summers using Meteosat Infrared channels. *Journal of Climate*, **1,** 471–484.

Gifford, F. A., 1989, The shape of large tropospheric clouds, or 'Very like a whale.' *Bulletin of the American Meteorological Society*, **70,** 468–475.

Harrison, E. G., Minnis, P., and Gibson, G. G., 1983, Orbital and cloud cover sampling analyses for multisatellite earth radiation budget experiments. *Journal of Spacecraft and Rockets*, **20,** 441–445.

Hartmann, D. L., and Recker, E. E., 1986, Diurnal variation of outgoing longwave radiation in the tropics. *Journal of Climate and Applied Meteorology*, **25,** 800–812.

Hartmann, D. L., and Short, D. A., 1980, On the use of the earth radiation budget statistics for studies of clouds and climate. *Journal of Atmospheric Science*, **37,** 1233–1250.

Kuo, K. S., Welch, R. M., and Sengupta, S. K., 1988, Structural and textural characteristics of cirrus clouds observed using high spatial resolution LANDSAT imagery. *Journal of Applied Meteorology*, **27,** 1242–1260.

Lo, R. C., and Johnson, D. R., 1971, An investigation of cloud distribution from satellite infrared radiation data. *Monthly Weather Review*, **99,** 599–605.

Minnis, P., and Harrison, E., 1984, Diurnal variability of regional cloud and clear sky radiative parameters derived from GOES data. Part II: November 1978 cloud distributions. *Journal of Climate and Applied Meteorology*, **23,** 1012–1031.

Minnis, P., and Wielkicki, B., 1988, Comparison of cloud amount derived using GOES and LANDSAT data. *Journal of Geophysical Resarch*, **93,** 9385–9403.

Platt, C. M. R., 1983, On the bispectral method for cloud parameter determination from satellite VISSR data: separating broken cloud and semitransparaent cloud. *Journal of Climate and Applied Meteorology*, **22,** 429–439.

Rossow, W. B., 1989, Measuring cloud properties from space: A review. *Journal of Climate*, **2,** 201–213.

Rossow, W. B., Mosher, F., Kinsella, E., Arking, A., Desbois, M., Harrison, E., Minnis, P., Ruprecht, E., Sèze, G., Simmer, C., and Smith, E., 1985, ISCCP cloud algorithm intercomparison. *Journal of Climate and Applied Meteorology*, **24,** 877–903.

Saporta, G., 1978, *Théories et methodes de la satatistiques.* Publication de l'Institut Français du Peétrole. (Paris: Société des Éditions Technip).

Schiffer, R. A., and Rossow, W. B., 1983, The International Satellite Cloud Climatology Project (ISCCP): The first project of the World Climate Research Programme. *Bulletin of the American Meteorological Society*, **64,** 779–784.

Schiffer, R. A., and Rossow, W. B., 1985, ISCCP global radiance data set: A new resource for climate research. *Bulletin of the American Meteorologivcal Society*, **66,** 1498–1505.

Séze, G., and Desbois, 1987, Cloud cover analysis from satellite imagery using spatial and temporal characteristics of the data. *Journal of Climate and Applied Meteorology*, **26,** 287–303.

Sèze, G., and Rossow, W. B., 1991, Time-cumulated visible and infrared radiance histograms used as descriptors of surface and cloud variations. *International Journal of Remote Sensing*, **12,** 877–920.

Sèze, G., and Smith, L. A., 1989, On quantifying the geometry of stratocumulus radiance fields. *IRS '88: Current Problems in Atmospheric Radiation*, edited by J. Lenoble and J. F. Geleyn (Hampton, Virginia: A. Deepak Publishing).

Shih, C., Campbell, G. G., and Vonder Haar, T. H., 1984, Effect of data resolution on satellite cloudiness estimation. XXVI COSPAR, Graz, Austria, July 1984. *Advances in Space Research*, **5,** 398–401.

SNEDECOR, G. W., and COCHRAN, W. G., 1980, *Statistical Methods* (Iowa State University Press).

WARREN, S. G., HAHN, C., and LONDON, J., 1985, Simultaneous occurrence of different cloud types. *Journal of Climate and Applied Meteorology*, **24,** 658–667.

WELCH, R. M., KUO, K. S., WIELICKI, B. A., SENGUPTA, S. K., and PARKER, L., 1988, Marine stratocumulus cloud fields off the coast of southern California observed using LANDSAT imagery. Part I: Structural characteristics. *Journal of Applied Meteorology*, **27,** 263–278.

WIELICKI, B., and WELCH, R. M., 1986, Cumulus cloud properties using Landsat satellite data. *Journal of Climate and Applied Meteorology*, **25,** 261–276.