NASA

NATIONAL AERONAUTICS AND
SPACE ADMINISTRATION

George C. Marshall Space Flight Center
Science Directorate
Earth Science Department

ALGORITHM THEORETICAL BASIS DOCUMENT (ATBD) FOR THE LIGHTNING IMAGING SENSOR (LIS)

Principal Investigator: Hugh J. Christian
Earth Science Department, Science Directorate
NASA/Marshall Space Flight Center
Global Hydrology and Climate Center
Phone (256) 922-5828, Fax (256) 922-5723
e-mail: hugh.christian@msfc.nasa.gov

Co-Investigator: Richard J. Blakeslee
Earth Science Department, Science Directorate
NASA/Marshall Space Flight Center
Global Hydrology and Climate Center
Phone (256) 922-5962, Fax (256) 922-5723
e-mail: rich.blakeslee@msfc.nasa.gov

Co-Investigator: Steven J. Goodman
Earth Science Department, Science Directorate
NASA/Marshall Space Flight Center
Global Hydrology and Climate Center
Phone (256) 922-5891, Fax (256) 922-5723
e-mail:steven.goodman@msfc.nasa.gov

Co-Investigator: Douglas M. Mach
Atmospheric Science Department
University of Alabama in Huntsville
Global Hydrology and Climate Center
Phone (256) 922-5830, Fax (256) 922-5723
e-mail: douglas.mach@msfc.nasa.gov

May 3, 1994
Revised February 10, 1995
Revised September 10, 1996
Revised February 1, 2000

ALGORITHM THEORETICAL BASIS DOCUMENT (ATBD)
FOR THE
LIGHTNING IMAGING SENSOR (LIS)

LIST OF FIGURES

ALGORITHM THEORETICAL BASIS DOCUMENT (ATBD)
FOR THE
LIGHTNING IMAGING SENSOR (LIS)

1. INTRODUCTION

This document is the Lightning Imaging Sensor (LIS) Algorithm Theoretical Basis Document (ATBD). This document defines and describes in detail the algorithm that processes the LIS data into the basic Hierarchical Data Format (HDF) orbit granule data storage structure (a level 1B/2A data product). Section 2 of this document provides a brief overview of the contributions of LIS measurements to the Tropical Rainfall Measuring Mission (TRMM) and the Office of Earth Science's Earth Observing System (EOS), the heritage of space-based lightning measurements, and the LIS science objectives. The LIS instrument characteristics are also described at the end of section 2. Section 3 discusses the theoretical basis of the LIS data processing algorithm, defines terminology, and provides a detailed description of the operation of the LIS data processing algorithm in generating the HDF orbit granule data product. A LIS data processing scenario is given to more clearly illustrate how the LIS algorithm functions. Finally, section 4 touches on various constraints, limitations, and assumptions associated with the LIS data processing algorithm. Further information on the LIS instrument, LIS software tools, recent science results, and data products may be found at the LIS science team web site http://thunder.msfc.nasa.gov.

The LIS is an EOS instrument on the TRMM observatory designed to acquire and investigate the distribution and variability of total lightning (i.e., cloud-to-ground and intracloud) on a global basis. The addition of LIS on TRMM contributes to many of the goals and objectives of TRMM and EOS. Lightning is a unique indicator of convective intensity and represents an important means for detecting deep convection from space. When combined with other complementary measurements from the TRMM observatory, the LIS provides further insight into the nature of oceanic and continental convective clouds. The LIS data will provide the basis for the development of a comprehensive global thunderstorm and lightning climatological data base.

2. OVERVIEW

2.1 The Need for Lightning Measurements on TRMM

Lightning is a global phenomenon that can be used as a measure of some of the variables in global change. Lightning activity is associated with thunderstorms that produce heavy rainfall in many climatic regions and, therefore, can be used to increase our knowledge of the distribution, structure, and variability of thunderstorms at the local, regional, and global scale over the land and ocean. The world's major centers of organized mesoscale convection do not, in general, coincide with the 'maxima' in the thunderstorm day climatology. The orographic and environmental interactions that initiate and sustain long-lived convective activity yield a different, yet important, perspective that is difficult to appreciate when considering only the number of thunderstorm days. This occurs because thunderstorm day statistics tend to be dominated more by the diurnal cycle than by large scale dynamics and give equal weight both to a day in which perhaps only one lightning discharge occurs and to a day in which tens of thousands of discharges occur. For example, up to 25% of the entire annual lightning strikes at a given site within the Central U.S. can be accounted for by the passage of but a single mesoscale convective complex (Goodman and MacGorman, 1986).

Atmospheric teleconnections associated with naturally occurring climate variations such as ENSO (El Nino Southern Oscillation) and anti-ENSO (La Nina) events in the tropical Pacific often result in significant changes in the frequency and movement of storm tracks, precipitation patterns, and cloud cover. These climate variations will also produce changes in lightning activity in both the northern and southern hemispheres. A comparison of LIS observations during the 1997-98 ENSO and 1998-99 La Nina events showed the most significant year-to-year changes in wintertime lightning activity occurred in the Gulf of Mexico and East China Sea (more thunderstorms during the ENSO winter), and in the South Pacific Convergence Zone (more thunderstorms during the LA Nina winter). In association with a strong upper level jet anomaly over the Gulf of Mexico Basin in the ENSO winter of 1997-98, there was a 200% increase in both lightning hours and thunderstorm days year-to-year (Goodman et al., 2000).

Forest fires are often caused by lightning when the ground is dry, decayed vegetation is present, and surface winds are strong. Doubled CO_2 (i.e., global warming) climate simulations suggest a 25% increase in global lightning frequency (Price and Rind, 1990). Does this mean that naturally occurring forest fires (in northern boreal forests, in particular) will be more common in the future? Are global warming simulations of cloud amounts, distributions, and structures, limited by coarse model spatial resolution and relatively simple convective parameterizations, justified by the regional and global observations of lightning? How might sub-cloud evaporation change? An improved, long-term observation capability will help resolve these questions.

Lightning measurements provided by the Lightning Imaging Sensor (LIS) on the Tropical Rainfall Measuring Mission (TRMM) will offer an opportunity to develop combined data algorithms to investigate the electrical, microphysical, and kinematic properties of tropical thunderstorms. It is hypothesized that the type (intracloud versus cloud-to-ground discharges) and frequency of lightning are intimately related to the microphysical (e.g., ice mass, liquid water content) and kinematic properties (e.g., updraft speed) of thunderstorm systems and to the environment (e.g., available buoyant energy). Recent evidence suggests that lightning activity can pro-

vide empirical estimates or bound the range of values for some geophysical properties such as the convective rain flux and rain rate, the vertical structure and distribution of storm mass, (convective) latent heating rates, the number and distribution of thunderstorms (Goodman et al., 1988b; Goodman et al., 1989; Buechler and Goodman, 1990; Williams and Rutledge, 1990; Goodman and Christian, 1993; Williams et al., 1992).

The processes that lead to the production of lightning are tightly controlled by updraft activity and the formation of precipitation. Lightning seems to initiate soon after the onset of strong convection, after significant cloud mass and ice have formed in the mixed phase regions of the thunderstorm. Lightning activity tends to track the updraft in both amplitude and phase with rates increasing as the updraft intensifies and decreasing rapidly with cessation of vertical growth. It has been demonstrated that lightning observations from space will clearly delineate the regions of convection embedded within large stratiform cloud systems which are often obscured by cirrus anvils (Goodman and Christian, 1993). Thus the detection of lightning from space specifically identifies those regions that are of paramount importance in the convective rain formation process. This ability to uniquely identify and quantify the convective core regions of storm systems and the existence of a linear relationship between total rain volume and lightning flash rate make LIS an important part of TRMM.

Various methods have been developed to determine convective rainfall amounts remotely using visible, microwave and/or infrared observations from space borne sensors but deficiencies exist in these techniques (Barret and Martin, 1981). One problem is properly delineating the convective areas. Presently, IR thresholding for rainfall estimates is contaminated by signals from cirrus clouds. Combining the IR with VIS data is only useful during the daylight. Using the life history of the cloud to estimate rainfall amounts is hampered by the uncertainty of cloud-top features. Microwave emission and scattering both have problems with their low spatial resolution along with ambiguities between rain water and cloud signals.

A satellite measurement system that senses the amount of lightning (i.e., both intracloud and cloud-to-ground) produced by thunderstorms may overcome some of the drawbacks of current techniques and improve rainfall estimation. For example, the location of active lightning areas could be used to delineate the convective areas used by visible and infrared estimation techniques and indicate the relative amount of precipitation-sized ice (Goodman et al., 1988c). Weinman et al. (1993) propose using continuous observations from long-range sferics networks to augment the sampling gaps of the existing polar orbiting DMSP satellites. In addition, they suggest using the sferics observations to calibrate the rainfall estimates from geostationary infrared imagers (e.g., Arkin and Ardanuy, 1989), which have better temporal sampling than polar orbiting satellites, but have a weaker physical linkage to the associated rainfall, especially over land. We anticipate the lightning observations could calibrate the infrared imagers using methods similar to the combined passive microwave - infrared imager rainfall retrieval schemes (Adler et al., 1993; Kummerow and Giglio, 1992). Another possibility would be to develop a technique for estimating convective rainfall based upon the amount of lightning produced by the storm. Livingston and Krider (1978), Williams (1985), Goodman and MacGorman (1986), Cherna and Stansbury (1986), and others have developed relationships (i.e., empirical algorithms) for lightning rates as a function of storm size, height, and duration.

2.2 Historic Perspective and Heritage

Historically, the electrification of thunderstorms and the resulting lightning have been studied on the smallest of scales in an attempt to understand the underlying physical principles that drive the processes. Observation and modeling studies have strongly intertwined cloud electrification, microphysics, and dynamics, resulting in important advances in the fundamental understanding but with obvious need for further clarification of the various process interactions and feedbacks involved. Recent measurements near Darwin, Australia during the monsoon and monsoon 'break' periods have revealed dramatic changes (i.e., factor of 10) in lightning frequency associated with modest variations in surface wet bulb temperature (Williams et al., 1991). Thunderstorm wind fields, precipitation, and lightning are tightly coupled in time and space and lightning cannot occur without the others. That is, strong updrafts and precipitation are necessary (but not sufficient) conditions for the production of lightning. There is now significant effort to quantify those interactions. Also, while lightning and thunderstorms occur on local and regional scales, their effects have global consequences. For example, deep convective storms in the tropics are believed to be an important mechanism for heat transport from the surface to the upper troposphere in the Intertropical Convergence Zone (ITCZ) and contribute energy to drive both Walker and Hadley cell circulations.

Lightning has been observed from above clouds using sensors installed on airplanes (e.g., Christian et al., 1983), high altitude balloons (e.g., Holzworth and Chiu, 1982), and satellites (e.g., Turman, 1979). Airplanes are excellent platforms for making close-up, detailed measurements of lightning characteristics, and although aircraft can be vectored to specific regions of a cloud for coordinated studies of storms from above and below, it is not possible to obtain a global view using research aircraft. Nor can a global view be adequately obtained with high altitude balloon observations. Satellites, on the other hand, represent ideal platforms for investigating lightning activity on the global scale.

Over the past 30 years, more than a dozen Earth-orbiting spacecraft have flown instruments that have recorded signals from lightning. The OSO-2 (Vorpahl et al., 1970), OSO-5 (Sparrow and Ney, 1971), and DMSP (e.g., Turman, 1978; Turman and Tettelbach, 1980; Orville and Henderson, 1986) satellites observed lightning with various optical sensors. Lightning sferics have been measured by radio frequency (rf) sensors on the RAE-1 (Herman et al., 1965), ARIEL-3 (Horner and Bent, 1969), and the ISS-b (Kotaki and Katoh, 1983) satellites. The shuttle-based Night/Day Optical Survey of Lightning (NOSL) (Vonnegut et al., 1983) used a small hand-held camera package, while the shuttle-based Mesoscale Lightning Experiment (MLE) used payload bay TV cameras to obtain images of nighttime lightning activity over the Earth. The detection of lightning from some of these satellites was a primary research objective, while for others it was an unanticipated bonus. In general, however, the various instrument packages were unable to provide very precise information on lightning characteristics, spatial extent, and discharge frequency. Also, the location accuracy tended to be poor (hundreds of kilometers) due to the low spatial resolution of the detectors and the detection efficiency of the systems was generally less than 2%, resulting in severe undersampling of the worldwide lightning activity.

During the 1980's, extensive optical and electrical observations of lightning were made from a high altitude NASA U-2 airplane with the primary objective of defining a baseline design criteria for a space sensor capable of mapping lightning discharges from geostationary orbit dur-

ing day and night with high spatial resolution and high detection efficiency. The results of the U-2 investigations, parametric trade-off studies, and other research (Thomason and Krider, 1982; Norwood, 1983; Eaton et al., 1983; Christian et al., 1984, 1989; Christian and Goodman, 1987; Goodman et al., 1988a) have clearly established the feasibility of making this kind of lightning measurement from space using present state-of-the-art technology. Based on this research, an optical lightning imaging sensor has been developed for deployment in low-Earth orbit.

We can take advantage of the additional space based lightning observations being provided by the NASA Optical Transient Detector (OTD). OTD, launched in 1994 as a secondary payload on a Pegasus launch vehicle, can be used to address important validation and performance issues associated with the LIS algorithm processing software described in this document (Boccippio et al., 2000). The OTD design is based on the LIS instrument for TRMM (Christian et al., 1992) and OTD can be viewed as a LIS prototype

2.3 Experimental Objectives

The overall investigation objectives of this instrument are to fly a calibrated optical Lightning Imaging Sensor (LIS) in order to acquire and investigate the distribution and variability of total lightning over the Earth, and to increase understanding of underlying and interrelated processes in the Earth/atmosphere system. The LIS will be particularly valuable in providing observations over the data sparse oceans and tropical regions of the world. The proposed lightning measurements provided by the LIS will address primary science objectives which will not be made by any other instruments.

Lightning is one of the responses of the atmosphere to thermodynamic and dynamic forcing and, consequently, contains significant information about the atmosphere. Because of this information content, the LIS will provide unique global data sets, including:

- Detection and Location of Deep Convection Without Land-Ocean Bias
- Estimation of Precipitation Mass in the Mixed Phased Region of Thunderclouds
- Differentiation of Storms with Strong Updrafts from Those with Weak Vertical Motions
- Information on Natural NO_X and Other Trace Gas Production in the Atmosphere

Lightning measurements from LIS will contribute significantly to several important mission objectives of TRMM and EOS. Lightning activity is closely coupled to storm convection, dynamics, and microphysics, and can be correlated to the global rates, amounts, and distribution of convective precipitation and to the release and transport of latent heat. These mesoscale phenomena influence and are influenced by global scale processes. The LIS will contribute to studies of the hydrological cycle, general circulation and sea-surface temperature variations, investigations of the electrical coupling of thunderstorms with the ionosphere and magnetosphere, and observations and modeling of the global electric circuit. It will provide a global lightning climatology from which changes, caused perhaps by subtle temperature variations, will be readily detected.

2.4 Instrument Characteristics

The LIS detects and locates lightning with storm scale resolution (i.e., 5-10 km) over a large region of the Earth's surface along the orbital track of the satellite, mark the time of occurrence of the lightning with 2 ms reso-lution, and measure the radiant energy. The LIS has a nearly uniform 90% flash detection efficiency within the field of view (FOV) of the sensor, and will detect intracloud and cloud-to--ground discharges during the day and night conditions.

The LIS on TRMM views a total area exceeding 580 km × 580 km at the cloud top using a 128 × 128 charge coupled device (CCD) array. The LIS on TRMM monitors individual storms and storm systems for 80 s, a period long enough to obtain a measure of the lightning flashing rate in these storms while the storm is in the field of view of the sensor. It will be possible to estimate lightning frequency even for storms with flash rates as low as 1-2 discharges per minute. As a result of the very high temporal resolution of the LIS (i.e., events marked to the nearest 2 ms), the satellite platform will travel a distance of only 16 m in the time between successive readouts of the photodiode array. Table 2.1 gives the performance characteristics of the LIS.

Table 2.1. Performance Criteria

pixel IFOV (nadir)	5 km
total FOV	80° × 80° square FOV
wavelength	777.4 nm
threshold	4.7 μJ m^{-2} sr^{-1}
SNR	6
array size	128 × 128 pixels
dynamic range	>100
detection efficiency	>90% of all events
false alarm rate	<10% of total events
measurement accuracy	location – 1 pixel intensity – 10% time – tag at frame rate
command interface	adjust threshold record/image power on/off self test safe mode (close/open aperature cover)
weight	20 kg
power	25 watts
telemetry	data rate – 6kb/s format – PCM sample size – 12 bits
operating temperature	–25° to +40°C

Finally, it is worth stressing that the LIS represents a significant advance over any previous satellite-borne lightning detector. Lightning observations from the LIS can be readily associated with the thunderstorms that produced them, and the detection of even a single discharge is significant and provides important information (e.g., storm location, precipitation estimates, storm height, the presence of ice, lightning frequency, electric current output, etc.). The earlier satellite-borne lightning instruments could provide only a relative global distribution of lightning of a strictly sta-tistical nature which required collecting data for long periods of time before it could be meaningfully interpreted in terms of global distributions. It was not possible to determine the lightning frequency of a particular storm with the earlier instruments, nor, for that matter, to even associate an individual discharge with a particular storm.

3. ALGORITHM DESCRIPTION

3.1 INTRODUCTION

3.1.1 Lightning Signal Characteristics

The occurrence of lightning is accompanied by the sudden release of electrical energy which is converted into rapid heating in the vicinity of the lightning channel, the generation of a shock wave (which rapidly decays into an acoustic wave, i.e., thunder), and electromagnetic radiation ranging from extremely low frequency (ELF) radio waves to x-rays. One of the strongest radiation regions is in the optical wavelengths with peak power typically between 100 to 1000 MW. These optical emissions result from the dissociation, excitation, and subsequent recombination of atmospheric constituents as they respond to the sudden heating in the lightning channel. The heating is so intense (electron temperatures > 20,000 K) that the optical emissions occur primarily at discrete atomic lines with some continuum at shorter wavelengths. Measurements from a NASA U-2 airplane have shown that the strongest emission features in the cloud top optical spectra are produced by the neutral oxygen and neutral nitrogen lines in the near infrared (e.g., the OI(1) line at 777.4 nm and the NI(1) multiplet at 868.3 nm are consistently strong features).

Temporally, the optical lightning signal is comprised of a series of fast rise time, short-duration pulses associated with the energetic discharge processes occurring within the cloud. The individual pulses of cloud-to-ground lightning are generally associated with return strokes and k-changes. The optical pulse widths and rise times are highly variable and are similar for intracloud and cloud-to-ground lightning discharges; however, the interpulse intervals for intracloud flashes tend to be shorter and there are generally significantly more pulses produced during each intracloud flash.

The thundercloud is an optically thick medium and therefore strongly affects the temporal and spatial characteristics of the optical signals produced by lightning which would be observed by a satellite sensor. Although the thundercloud is optically thick, there is very little absorption at optical wavelengths. Hence, the major effect of the cloud on the optical signal is to blur the source geometry and to delay and time-broaden the pulses due to multiple scattering. Extensive measurements from an instrumented NASA U-2 aircraft have shown that the rise time of an optical pulse is typically lengthened 215 s and the pulses' widths tend to be 210 s wider as a result of this multiple scattering. The median pulse rise time and the full-width-at-half-maximum obtained from the analysis of nearly 1300 pulses produced by 79 lightning flashes are 240 s and 370 s, respectively (Christian and Goodman, 1987).

It is important to stress that, while the cloud significantly alters the temporal characteristics of the cloud top optical signals, the cloud does not block these emissions. When viewed from above, the optical lightning signals appear as a diffuse light source radiating from the cloud top. Measurements of the total optical energy radiated from the cloud top are in good agreement with ground-based measurements of cloud-to-ground flashes and support the theory that the cloud acts like a conservative scatterer, i.e., that most of the optical energy escapes the cloud (Christian and Goodman, 1987).

Of the 79 discharges referred to above, 90% produced peak radiant energy densities of 4.7 J m^{-2} sr^{-1} or greater. The region of cloud top that is illuminated by a lightning flash depends on

where the flash occurred within the cloud, the geometry and physical extent of the flash, and the characteristics of the cloud through which the lightning channel propagated and the radiation scattered. Theory, using Monte Carlo simulations of the radiation transfer of the optical lightning signals, and the NASA U-2 aircraft studies indicate that the diameter of the cloud top illumination associated with a single storm cell will typically be on the order of 10 km. Observations of large storm systems from the shuttle have shown that illuminated regions can exceed 60 km (Goodman and Christian, 1993).

Finally, it should be noted that both intracloud and cloud-to-ground lightning flashes are readily observed from above. Extensive observations with the NASA U-2 aircraft flying over the tops of thunderstorms in coordination with ground-based measurements made under the same storms (Goodman et al., 1988a) have clearly established the viability of optical detection of all lightning. Because the majority of the channel of a cloud-to-ground flash occurs within the cloud, the light emerging from the top of the cloud has undergone a similar scattering process as an intracloud flash (the portion of the channel below cloud base is essentially undetectable from above). Further, since the scattering process dominates the characteristics of the optical signature, the optical pulses from both intracloud and cloud-to-ground flashes are very similar (Guo and Krider, 1982; Thomason and Krider, 1982; Goodman et al., 1988a). We are unable at this time to distinguish between intracloud and cloud-to-ground lightning from the optical signatures alone. While this is a limitation, it is minor because, from a scientific standpoint, it is far more important to determine the total lightning rate than either just the cloud-to-ground or intracloud rate. In fact, optical measurements of lightning from above or below the cloud will detect more return strokes than ground-based detection systems (Goodman and MacGorman, 1986; Mach et al., 1986).

3.1.2 LIS Measurement Approach

The LIS is a conceptually simple device. It images the scene much like a television camera; however, because of the transient nature of lightning, its spectral characteristics, and the difficulty of daytime detection of lightning against the brightly lit cloud background, actual data handling and processing is much different from that required by a simple imager. In order to achieve the performance goals required to meet the scientific objectives, the LIS combines off the shelf components in a unique configuration. A wide field of view lens, combined with a narrow-band interference filter is focused on a small, high speed CCD focal plane. The signal is read out from the focal plane into a real-time event processor for event detection and data compression. The resulting “lightning data only” signal is formatted, queued, and sent to the satellite LAN.

The particular characteristics of the sensor design results from the requirement to detect weak lightning signals during the day. During the day, the background illumination, produced by sunlight reflecting from the tops of clouds, is much brighter than the illumination produced by lightning. Consequently, the daytime lightning signals tend to be buried in the background noise, and the only way to detect lightning during daytime is to implement techniques that increase or maximize the lightning signal relative to this bright background. These techniques take advantage of the significant differences in the temporal, spatial, and spectral characteristics between the lightning signal and the background noise. A combination of four methods are employed by the LIS for this purpose. First, spatial filtering is used which matches the instantaneous field of view (IFOV) of each detector element in the LIS focal plane array to the typical cloud-top area illumi-

nated by a lightning stroke (i.e., 5-10 km). This results in an optimal sampling of the lightning scene relative to the background illumination. Second, spectral filtering is obtained by using a narrow-band interference filter centered on a strong optical emission multiplet (e.g., OI(1) at 777.4 nm) in the lightning spectrum. This method further maximizes the lightning signal relative to the reflected daylight background. Third, the LIS employs temporal filtering which takes advantage of the difference in lightning pulse duration of the order of 400 s versus the background illumination which tends to be constant on the time scale of the order of seconds. In an integrating sensor, such as the LIS, the integration time specifies how long a particular pixel accumulates charge between readouts. The lightning signal-to-noise ratio improves as the integration period approaches the pulse duration. If, however, the integration period becomes too short, the lightning signal tends to be split between successive frames which actually decreases the signal-to-noise ratio. Since the median optical lightning pulse width when viewed from above is 400 s, an integration time of 1 ms is most appropriate to minimize pulse splitting and maximize lightning detectability. Technological limitations require that a 2 ms integration time be used in the LIS instrument design; however, this compromise will not seriously degrade the sensor's performance. Even with the three "filtering" approaches discussed above, the ratio of the background illumination to the lightning signal may still exceed 150 to 1 at the focal plane. Therefore, a fourth technique a modified frame-to-frame background subtraction is implemented to remove the slowly varying background signal from the raw data coming off the LIS focal plane. A more detailed discussion on the measurement approach adopted for the LIS is given in Christian et al. (1989).

The real time event processor generates an estimate of the background scene imaged at each pixel of the focal plane array. This background scene is updated during each frame readout sequence and, at the same time, the background signal is compared with the off-the-focal-plane signal on a pixel-by-pixel basis. When the difference between these signals exceeds a selected threshold, the signal is identified as a lightning event and an event processing sequence is enabled. The implementation of this real-time data processor results in a 10^5 reduction in data rate requirements while maintaining high detection efficiency for lightning events.

3.1.3 On-Board Signal Processing

The LIS consists of a staring imager optimized to detect and locate lightning. An imaging system, a focal plane assembly, a real-time signal processor and background remover, an event processor and formatter, power supply, and interface electronics are the six major subsystems of the sensor. The imaging system is a simple f/1.6 telescope consisting of a beam expander, an interference filter, and re-imaging optics. The 80×80 full angle LIS field of view is converged to less than 5 at the interference filter in order to minimize wavelength shifts due to non-normal incidence. The focal plane assembly, including the 128×128 element CCD array, preamplifiers, and multiplexers provides subsequent electronics with a serial data stream of sufficient amplitude. As noted earlier, if after the background removal the difference signal for a given pixel exceeds a threshold, that pixel is considered to contain an event. Once the event is identified it is time tagged, location tagged, and passed to platform telemetry via the interface electronics to the satellite LAN.

The LIS processor can extract weak lightning flashes from an intense but slowly evolving background. The daytime background varies with sun angle, clouds, ground albedo, etc., and can

reach an excess of 700,000 electrons as compared to lightning signal electrons which may be less than 5000 electrons. When a lightning stroke occurs within a single frame, a small signal is superimposed on top of the essentially constant background signal. The real-time processor continuously averages the output from the focal plane over six frames on a pixel-by-pixel basis in order to generate a background estimate. It then subtracts the average background estimate from the current signal.

The subtracted signal consists of shot noise fluctuating about a zero with occasional peaks due to lightning events. When a peak exceeds the level of the variable threshold, it is considered to be a lightning event and is processed by the rest of the circuit. The threshold must be set sufficiently high that false triggers are kept to a small percent of the total lightning rate. Clearly, the threshold must be higher during daytime when shot noise is dominated by the solar background.

The components of the real-time event processor include a background signal estimator, a background remover, a lightning event thresholder, an event selector, and a signal identifier. Analog/digital hybrid processing is used in a unique way in that it takes advantage of the strengths of each technology in order to provide high processing rates while consuming minimal power. Much of the signal processing is performed in a pipeline fashion that maximizes throughput.

The background estimator (averager) and remover (subtracter) circuits combine to perform the functions of a time domain low pass filter. The signal coming off the focal plane is fed through a buffer and clipping stage in order to ensure that a strong lightning signal does not contaminate the background estimate. This signal is then multiplied by a fractional gain (B) and added to (1-B) times the previous background estimate for the same pixel. The inverse of the fractional gain is equivalent to the number of frames used in generating the background estimate and is analogous to setting the cutoff frequency in conventional frequency domain filters. Too high a fractional gain might permit lightning events to contaminate low background estimates and would increase the processing noise. Too low a fractional gain would not allow the background estimator to respond rapidly enough to changes in background intensity.

The proper operation of the background estimator requires that the background data are clocked through the estimator synchronously with the data being clocked off the focal plane and that the number of discrete storage elements in the background memory is exactly the same as the number of pixels in the focal plane array. When data are properly synchronized, the signal appearing on the output of the delay line during a given clock cycle corresponds spatially to the signal being clocked off the focal plane. That is, it contains a history of what that specific pixel has measured over the last 1/B frames. These two signals are then subtracted using a difference amplifier in order to generate a difference signal. Since the original signal contains either background plus lightning or just background, the subtracted signal will be either a lightning signal, near zero or a false alarm as previously described.

The difference signal is then compared with the threshold level (which may be adaptive). If the signal exceeds the threshold level, a comparator triggers, which enables a switch and passes the lightning signal for further processing. In addition, the comparator output is encoded using a digital multiplexer in order to generate a row address that identifies the specific pixel that detected the lightning event. The digital outputs from the data processor represent the intensity of the lightning event and the location where the lightning occurred. These signals are then forwarded

to encoding electronics in which the data are formatted into a digital bit stream and sent to the platform LAN.

The lightning data is compressed in such a way as to minimize the telemetry bandwidth. If two adjacent pixels on the same row are illuminated by a lightning signal, only the outer pixel address is transmitted (both amplitudes are transmitted). It is up to the ground software to decompress the data and add the second adjacent address.

3.2 Definitions

The basic science data product of LIS is lightning. This product is comprised of several components, including: background images (level 1-B), events (level 2), groups (level 2), flashes (level 2), areas (level 2), viewtime data (level 2), one second data (level 1-B), browse image data (level 3), orbit statistics (level 3), and metadata. Before we can discuss the details of the various components, we must define each of the underlying data storage classes that drive the algorithm. These data storage classes are backgrounds, events, groups, flashes, areas, orbits, viewtimes, and one second data.

3.2.1 Background

A background image is a “snap shot” of the background estimate created by the LIS Real Time Event Processor (RTEP). The background data consists of 12 bit raw count amplitudes at each of the 128x128 pixel locations and the time at which the background image was taken. The background is identified as LIS02 in the Data Products Handbook, Volume 1. The background image is transmitted in the data stream along with event data to maintain the average transmission rate. When the transmission of one background is begun, the next background image is captured. New images are sent to the ground as frequently as the event load and transmission rate allow. The full set of data parameters (raw, calculated, and summary) associated with a background image will be described in Sections 3.3.1 and 3.3.2.

3.2.2 Event

An event is defined as the occurrence of a single pixel exceeding the background threshold during a single frame. In other words, each pixel output from the RTEP produces a separate event. The raw LIS instrument data consists of time, x and y pixel locations, and amplitude of the event. An event is the basic unit of data from the LIS. An event is identified as LIS03 in the Data Products Handbook, Volume 1. The full set of data parameters associated with an event will be described in Section 3.3.1.

Although an event can be thought of as a single optical pulse due to lightning, it is possible that multiple pulses occurring within the 2 ms integration window may contribute to an event. Therefore, we purposely did not use ‘pulse’ or ‘stroke’ (or other similar name) to describe the basic unit of data from the LIS (Note: an event may sometimes not be due to lightning at all. It may be produced by noise in the analog data stream exceeding the background threshold. In that case, the event is a false alarm).

3.2.3 Group

Because a single pixel will almost never correspond to the exact cloud illumination area, a lightning discharge will often illuminate more than one pixel during a single integration time. The result is two or more adjacent events at the same time frame. When these multiple events are adjacent to each other (a side or corner of the events touching), they will be placed in a single group. The formal definition of a group is one or more simultaneous events (i.e., events that occur in the same time integration frame) that register in adjacent (neighboring or diagonal) pixels in the focal plane array. A group may consist of only one event or include many events. The location data for a group will be calculated in earth-based (latitude/longitude) coordinates. This is done to provide consistent representation in the group/flash/area processing and because the ultimate goal of the analysis is to locate lightning with respect to the earth's surface. A group is identified as LIS04 in the Data Products Handbook, Volume 1. The full set of data parameters associated with a group will be described in Section 3.3.1.

Although a group may often correspond to a single lightning optical pulse, it is also possible that multiple lightning pulses occurring within the 2 ms integration window may contribute to a group. A false event due to noise at a pixel exceeding the background threshold can also contribute to a group (although noise groups often contain only one event). A discussion of the quality control and error analysis of the LIS data stream and the flagging of suspected data is contained in Section 3.5.3.

3.2.4 Flash

A lightning flash consists of one to multiple optical pulses that occur in the same storm cell within a specified time and distance. A lightning flash should then correspond to several related groups in a limited area. For the LIS algorithm, we define a flash as a set of groups sequentially separated in time by no more than 330 ms and in space by no more than 5.5 km. The algorithm uses the weighted Euclidean distance [Hartigan, 1975] between the groups to determine clustering. The weighting is set so that 330 ms is equivalent to 5.5 km. The temporal and spatial rules can be easily adjusted in the LIS algorithm processing software. We will continue to examine the rules closely during the analysis of OTD and LIS data to "fine tune" the rules defining a flash. A flash may include as few as one group with a single event or it may consist of many groups, each containing many events. Since there is the possibility that the TRMM satellite will move a significant fraction of a pixel during the time of a flash, spatial characteristics for a flash (and all higher level parameters) are calculated in ground coordinates (i.e., latitude and longitude). A flash is identified as LIS05 in the Data Products Handbook, Volume 1. The full set of data parameters associated with a flash will be described in Section 3.3.1.

We have used the term flash for this data category because we believe that, as it has been defined above, the resultant 'flash' will generally correspond to the accepted definition of a conventional lightning flash. Note that with LIS data alone, we cannot determine if a flash is a ground or cloud flash. It is possible that future versions of the LIS algorithm may incorporate data from ground flash locating systems to help interpret the LIS flashes. We do acknowledge that, on occasion, distinct conventional lightning flashes may result in a single flash being produced by the LIS algorithm (e.g., possibly in high flashing rate mesoscale convection systems). Other mismatches between algorithm flashes and actual conventional flashes will undoubtedly

also occur. Note that there is no absolute time limit to a flash. That is, as long as subsequent groups are produced in an area within the 330 ms time windows, all groups will be assigned to a single flash.

3.2.5 Area

Lightning is produced in thunderstorm cells that have dimensions of about 10 km by 10 km. Many storms, however, are multicellular and may extend over large areas and exist for many hours. Individual storms generally last much longer than the LIS will view them. Therefore, we define an area as a near contiguous region on the surface of the earth that has produced lightning (defined as a set of LIS flashes) during a single orbit of the LIS. An area thus defined consists of a set of flashes separated in space by no more than 16.5 km (approximately 3 pixels). The spatial rule can be easily adjusted in the LIS algorithm processing software if necessary after analysis of OTD and/or LIS data. An area may include many flashes or contain as few as one event (i.e., one flash consisting of one group which in turn consists of one event). There is no interflash or absolute time limit rule being imposed in the area definition since, as noted previously, the LIS viewing time is much shorter than a storm life cycle. Although there is no explicit limit to the temporal duration of an area (i.e., as long as there are events/groups/flashes in the region, all will be assigned to the area), the LIS instrument will only view any ground location within its FOV for a maximum of 80 seconds. Therefore, area duration will generally not exceed 80 seconds except possibly for spatially extensive (and very active) mesoscale storm complexes. An area is identified as LIS06 in the Data Products Handbook, Volume 1. The full set of data parameters associated with an area will be described in Section 3.3.1.

The area definition serves as a proxy for a thunderstorm, however, due to the nature of the algorithm and the spatial and temporal distribution of the data, several storms could be combined into one area. It is also possible for a single thunderstorm to be divided into more than one area. More sophisticated algorithms (with input from external ground-, airborne-, and space-based observing systems) will be needed to more precisely determine ‘thunderstorms’ in the LIS data.

3.2.6 Orbit

The data granule for TRMM has been established as one orbit. Thus, all data from the LIS is stored and summarized at the orbit granule. However, the beginning and end times of the LIS orbit granule differ from the TRMM defined orbit. Since dividing the LIS data at the equatorial crossing would often split storms, the LIS orbit granule is defined to begin and end at the latitude of the southernmost part of the orbital path. This location is often away from lightning producing tropical convection. This should lessen the probability that users will have to acquire more than one orbit to study specific lightning systems. An orbit will include every area with latitudes contained within the geographic boundaries of the orbit. All flashes, groups, and events associated each area in an orbit will be kept with the orbit regardless of where they were located. Background images occurring between the start and stop location of the orbit will also be kept with the orbit. An orbit is identified as LIS07 in the Data Products Handbook, Volume 1.

Since orbits will have a geographic start and stop at the southernmost location of the orbit, it is possible for flashes, groups and events to be on the opposite side of the orbit boundary from the parent areas. This will occur if the areas were active at the time of the orbit boundary cross-

ing. Since all of the LIS lightning data is associated with the parent area, all child data (flashes, groups, and events) will be kept in the orbit with the parent area.

3.2.7 Viewtime

Unlike many of the other instruments on TRMM, the LIS data is very dependent on how long a particular location was viewed by the LIS instrument. For a single pass, different locations on the ground can have a wide range (0 to 80+ seconds) of viewtimes. The amount of lightning in a location is not very useful without information on how long it took to produce that amount of lightning. To provide this information to the users of the LIS data, viewtime information is calculated for each point along the LIS field of view. Details of the components of the viewtime are in Section 3.3.1.

3.2.8 One Second Data

The LIS data are also very dependent on the status of the LIS instrument. The one second data provides this information as a series of one second snapshots of internal and external instrument parameters. Details of the components of the one second data are in Section 3.3.1.

3.3 LIS Data Description

The definitions described above can be used along with the EOSDIS Hierarchical Data Format (HDF) to create the data storage structure for the LIS data. The major focus of this section is to define the storage model of the data. The LIS data is stored in HDF Vgroups, Vdatas, Vsets (i.e., sets of Vdata), and Scientific Data Sets (SDSs) using the version 1 EOS HDF Standard Data Format (SDF) [Suresh *et al.*, 1994]. The software algorithm has been developed to create this storage model.

The LIS data for a single orbit is stored in two HDF files: one containing the major science data (see Figure 3.01 for the file structure) and the other the background images (Figure 3.02). This is done so users who are not interested in the background images do not have to download the large background files to get to the lightning data. The HDF file structure describes the data such that a user utilizing an HDF file reader can read and process the orbit granule data. The actual

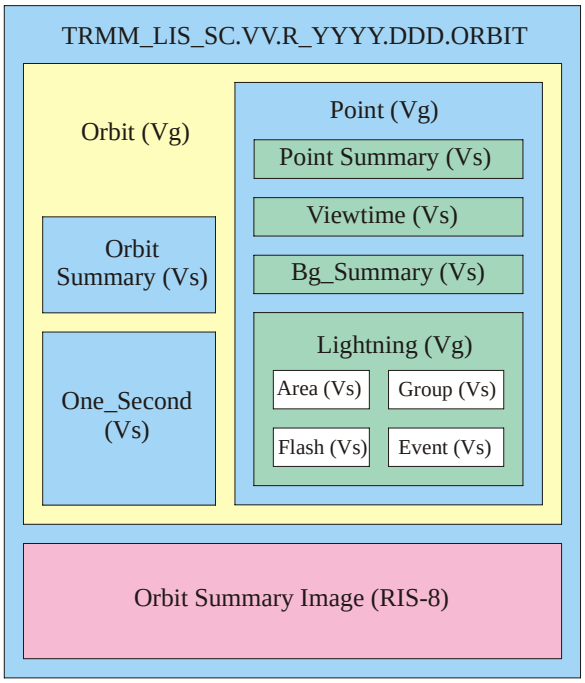

Figure 3.01 Science Data HDF File Components

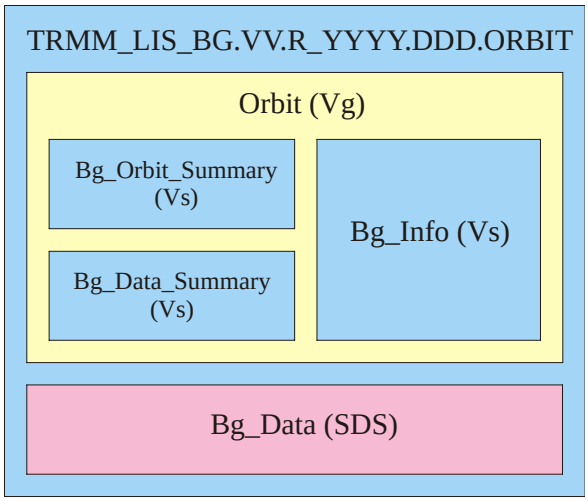

Figure 3.02 Background Data HDF File Components

data are stored in Vsets and Vgroups. Indexes are maintained within the Vgroups to link the various Vsets. The file name starts with the platform name (TRMM), followed by the instrument name (LIS), file type designator (SC for science data and BG for background image data), version number (VV), a period, and the revision number (R). After the revision number, the file name contains the year (YYYY), day of year (DDD), and the orbit number (ORBIT) of the data. For each orbit, there will be one "BG" and one "SC" file. The linkage of data within the HDF file (i.e., either via direct HDF link or by an indirect index link) and how the data storage model relates to the LIS lightning list will be described in more detail in the following sections.

3.3.1 LIS Science Data File Structures

3.3.1.1 Orbit

The orbit Vgroup contains one Vgroup and two Vsets as shown in Figure 3.01. The Vsets are the orbit_summary and the one_second Vsets. The Vgroup is the point Vgroup. Each structure will be described below.

3.3.1.1.1 Orbit_Summary (LIS07)

The Orbit_Summary (part of LIS07) is an HDF Vset that contains a summary of orbit attributes (Table 3.1). The components of the Orbit_Summary are the orbit identification number (*id_number*), the start time of the orbit in three different formats *(TAI93_start*, *UTC_start*, and *GPS_start*), the end time of the orbit in TAI93 format (*TAI93_end*), the start and end latitude locations of the orbit (*start_longitude* and *end_longitude*), information needed to find the point data (*point_data_count* and *point_data_address*), information needed to find the one second data (*one_second_count* and *one_second_address*), information needed to find the summary image data (*summary_image_count* and *summary_image_address*), the inspection status of the file (*inspection_code*), and the software configuration status (*configuration_code*). The orbit identification number is an int32 uniquely designating the LIS orbit. The times are in format appropriate for the time units (double precision for the TAI and GPS times and characters for the UTC). The longitude locations are in decimal degrees (single precison). The count and address pairs allow the user to find the other data sets (point, one second, and summary images) within the HDF file using standard HDF tools. Each orbit file contains one Orbit_Summary Vset.

Table 3.1. Orbit Attributes (LIS07)

Variable Name	Type	Order	Description
id_number	int32	1	The TRMM orbit number designation
TAI93_start	float64	1	Orbit start time (TAI93)
UTC_start	char	28	Orbit start time (UTC)
GPS_start	float64	1	Orbit start time (GPS)
TAI93_end	float64	1	Orbit end time (TAI93)
start_longitude	float32	1	Orbit start location longitude (deg)
end_longitude	float32	1	Orbit start location longitude (deg)
point_data_count	int16	1	# of point data records
point_data_address	int16	1	HDF record of first point data
one_second_count	int32	1	# of one second data records
one_second_address	int32	1	HDF record of first one second data
summary_image_count	int16	1	# of summary images
summary_image_address	int16	1	HDF record of first summary image
inspection_code	uint16	1	Inspection status
configuration_code	uint16	1	Software status

3.3.1.1.2 One_Second

The one_second Vset consists of the 16 parameters listed in Table 3.2. The first parameter (*TAI93_time*) is the time of the one_second snapshot. The next five parameters (*alert_summary, instrument_alert, platform_alert, external_alert,* and *processing_alert*) describe the status of the LIS/TRMM system. The next four parameters (*position_vector, velocity_vector, transform_matrix,* and *solar_vector*) allow the user to more accurately calculate the location of the lightning data if they desire. The next two parameters (*ephemeris_quality_flag* and *attitude_quality_flag*) are the TRMM provided status of the satellite location and orientation data. The LIS instrument thresholds are the next parameter (*thresholds*). The last parameter (*event_count*) is an array of the event counts after various stages of LIS algorithm processing.

Table 3.2. One_Second

Variable Name	Type	Order	Description
TAI93_time	float64	1	The TAI93 time of the one_second data
alert_summary	uint8	1	Bit summary of the LIS alert status
instrument_alert	uint8	1	Bit summary of the LIS instrument status
platform_alert	uint8	1	Bit summary of the LIS platform status
external_alert	uint8	1	Bit summary of the LIS external status
processing_alert	uint8	1	Bit summary of the LIS processing status
position_vector	float32	3	TRMM location vector
velocity_vector	float32	3	TRMM velocity vector
transform_matrix	float32	9	TRMM orientation matrix
solar_vector	float32	3	Sun location vector
ephemeris_quality_flag	int32	1	Bit summary of the TRMM ephemeris
attitude_quality_flag	int32	1	Bit summary of the TRMM attitude
boresight_threshold	int8	1	
thresholds	int8	16	LIS threshold values
noise_index	int8	1	
event_count	int16	6	Array of event counts

3.3.1.1.3 Point

The point structure is an HDF Vgroup with links to three Vsets and one Vgroup (Figure 3.01). The Vsets include the point_summary, viewtime, and bg_summary. The Vgroup is called lightning. Each will be described below.

3.3.1.1.3.1 Point_Summary

The point_summary consists of the 13 parameters listed in Table 3.3. This structure allows the user to quickly get to the various point datasets in the HDF. The addresses (*parent_address, event_address, group_address, flash_address, area_address, bg_address,* and *vt_address*) are the HDF addresses of the various Vsets. The counts (*event_count, group_count, flash_count, area_count, bg_count,* and *vt_count*) are the total number of each point data in the HDF.

3.3.1.1.3.2 Viewtime

To determine flashing rates on the earth, the length of observation time is required. The Vset viewtime provides this information. To determine the viewtimes, the earth is divided into 0.5^{o} latitude $\times$ 0.5^{o} longitude grid boxes. Each grid box intersected by the LIS field of view during the orbit is examined for the length of time it was within the LIS field of view. The 6 resulting parameters are stored in the viewtime Vset (Table 3.4). The first parameter (*location*) is the

Table 3.3. Point_Summary			
Variable Name	Type	Order	Description
parent_address	int32	1	HDF address of the orbit Vgroup
event_count	int32	1	Total # of events in the file
event_address	int32	1	HDF address of the first event
group_count	int32	1	Total # of groups in the file
group_address	int32	1	HDF address of the first group
flash_count	int32	1	Total # of flashes in the file
flash_address	int32	1	HDF address of the first flash
area_count	int32	1	Total # of areas in the file
area_address	int32	1	HDF address of the first area
bg_count	int32	1	Total # of background summaries
bg_address	int32	1	HDF address of first background sum
vt_count	int32	1	Total # of viewtimes in the file
vt_address	int32	1	HDF address of the first viewtime

centroid location of the viewtime grid box. Since it is possible for the LIS to "go blind" at times (due to instrument and TRMM platform limitations), the viewtime of a location can be complicated. Because of this, the viewing time for a grid box can be significantly less than the simple difference between the first and last time the location was viewed by the LIS (*TAI93_start* and *TAI93_end*). The true viewing time is stored in the effective_obs slot. The viewtime also stores a snapshot summary of the LIS and TRMM platform status (*alert_flag*) and the *approx_threshold.*

Table 3.4. Viewtime			
Variable Name	Type	Order	Description
Location	float32	2	Center of viewtime grid box (lat/lon) (deg)
TAI93_start	int32	1	Time of first view of the location
TAI93_end	int32	1	Time of last view of the location
Effective_obs	float32	1	Total viewing time of location
Alert_flag	uint8	1	LIS/TRMM summary status
approx_threshold	int8	1	Estimated threshold

3.3.1.1.3.3 Bg_Summary

The background files are stored separately from the science data since they are so large and some users may not require them. However, a user may want to know the details of the background images. The bg_summary Vset contains the background image details. The components of the bg_sumary Vset are shown in Table 3.5. *TAI93_time* is the time of the background image while *address* is the HDF address in the background file so that the user can find the actual background image. The *boresight* and *corners* parameters locate the background on the Earth.

Table 3.5. Bg_Summary			
Variable Name	Type	Order	Description
TAI93_time	float64	1	Time of background
Address	int32	1	HDF address of background image
Boresight	float32	2	Center of background image (lat/lon, degs)
Corners	float32	8	Corners of background image (lat/lon, degs)

3.3.1.1.3.4 Lightning

The Vgroup lightning contains the full instrument resolution Vector Data (LIS08). These consist of four Vsets (events, groups, flashes, and areas). The format of each of the Vsets is similar (individual differences will be noted in their own sections). The format of the Vsets is such

that every area has a set of child flashes. Each flash has a set of child groups and each group has a set of child events. Each event has one and only one parent group, each group has one and only one parent flash, and each flash has one and only one parent area. The data in each Vset can be divided into time parameters, location parameters, links to the parent/child items, and statistical parameters. The individual parameters will be described below.

3.3.1.1.3.4.1 Area (LIS06)

The area Vset (LIS06) holds the data associated with each area identified during the orbit. The 20 parameters associated with the area Vset are shown in Table 3.6. There will be one such Vset for each area recorded during the orbit. The start time of the area is stored in *TAI93_time* while the lifetime and viewing time of the area are stored in *delta_time* and *observe_time*. The start time is in TAI seconds while the view time and lifetimes are in relative seconds. The *delta_time* of the area is the time difference between the first event in the area and the last event in the area. The *observe_time* is the actual time the centroid of the area was in the field of view of the LIS instrument. The parameter *location* is the latitude/longitude of the area centroid while *net_radiance* is the calibrated total radiance of all the flashes associated with the area. The area size (relative size) is recorded in the variable *footprint*. The HDF address of the area is stored in *address*. This parameter can be used to help build the parent/child relationships between the area and its child flashes. For future use, there is a variable *parent_address* that would give the index numbers of any parent structure to this area. Since areas are at the top of the data structure, these variables are currently set to the orbit ID number. It is here to maintain symmetry in the group/flash/area levels. The links to the child flashes are stored in *child_address* while *child_count* is the number of child flashes associated with this areas. The *child_address* is the actual HDF record number of the first child flash. The parameters *grandchild_count* and *great-grandchild_count* are the numbers of groups and events associated with the area. The approximate LIS threshold active during the area is stored in the *approx_threshold* parameter. The condition of the LIS/TRMM system is stored in the *alert_flag* parameter. The roundness of the area is a ratio of the long to short dimension and is stored in the *oblong_index* parameter. The *group-*

Table 3.6. Area (LIS06)

Variable Name	Type	Order	Description
TAI93_time	float64	1	TAI start time of the area
Delta_time	float32	1	Time between first and last event
Observe_time	int16	1	LIS viewing time of the area (in seconds)
Location	float32	2	Latitude/longitude of the area (deg)
net_radiance	float32	1	Total radiance of the area
Footprint	float32	1	Size of the area (km^2)
Address	int32	1	HDF address of area
parent_address	int32	1	HDF orbit address
child_address	int32	1	HDF address of first child flash
Child_count	int32	1	# of child flashes
Grandchild_count	int32	1	# of groups in area
Greatgrandchild_count	int32	1	# of events in area
approx_threshold	int8	1	LIS threshold during area
Alert_flag	uint8	1	LIS/TRMM status during area
Cluster_index	int8	1	
Density_index	int8	1	
Noise_index	int8	1	
Oblong_index	float32	1	Eccentricity of the area
Grouping_sequence	int32	1	Time ordered sequence of the area
Grouping_status	int8	1	End status of the area

ing_sequence parameter indicates where this area goes in the time ordered series of areas. The variable *grouping_status* indicates if the area was active when it was written to the HDF file. It can indicate that excessive data rates may have terminated an area prematurely (before all possible flashes have been added to this area). In this case, data from other subsequent areas in the same HDF file may need to be appended to this area to make it complete.

3.3.1.1.3.4.2 Flash (LIS05)

The basic storage structure for the area/flash/group Vsets is the same. The major difference is that the flash and group classes have parent pointers whereas areas do not. The flash Vset (LIS05) holds the data associated with the various flashes recorded during the orbit. The various parameters associated with the flash data are shown in Table 3.7. There will be one such Vset for each flash recorded during the orbit. The flash start time is stored in *TAI93_time*, while the flash lifetime and viewing time are stored in *delta_time* and *observe_time*. The start time is in TAI seconds while the view time and lifetimes are in relative seconds. The *delta_time* of the flash is the time difference between the first and last events in the flash. The *observe_time* is the actual time the centroid of the flash was in the field of view of the LIS instrument. The parameter *location* is the latitude/longitude of the flash centroid while *radiance* is the calibrated total radiance of all groups within the flash. The flash size (relative size) is recorded in the variable *footprint*. The HDF address of the flash is stored in *address*. This parameter can be used to help build the parent/child relationships between the flash and its child groups. The parameter *parent_address* is the parent area of the flash. The links to the child groups are stored in *child_address* while *child_count* is the number of child groups associated with the flash. The *child_address* is the actual HDF record number of the first child group. The parameter *grandchild_count* is the numbers of events associated with the flash. The approximate LIS threshold active during the flash is stored in the *approx_threshold* parameter. The condition of the LIS/TRMM system is stored in the *alert_flag* parameter. The roundness of the flash is stored in the *oblong_index* parameter. The grouping_sequence parameter indicates where this flash goes in the time ordered series of flashes. The variable *grouping_status* indicates if the flash was active when it was written to the HDF

Table 3.7. Flash (LIS05)

Variable Name	Type	Order	Description
TAI93_time	float64	1	TAI start time of the flash
delta_time	float32	1	Time between first and last event
observe_time	int16	1	LIS viewing time of the flash (seconds)
location	float32	2	Latitude/longitude of the flash (deg)
radiance	float32	1	Total radiance of the flash
footprint	float32	1	Size of the flash (km^2)
address	int32	1	HDF address of flash
parent_address	int32	1	HDF orbit address
child_address	int32	1	HDF address of first child group
child_count	int32	1	# of child groups
grandchild_count	int32	1	# of groups in flash
approx_threshold	int8	1	LIS threshold during flash
alert_flag	uint8	1	LIS/TRMM status during flash
cluster_index	int8	1	
density_index	int8	1	
noise_index	int8	1	
glint_index	float32	1	
oblong_index	float32	1	Eccentricity of the flash
grouping_sequence	int32	1	Time ordered sequence of the flash
grouping_status	int8	1	End status of the flash

file. It can indicate that excessive data rates may have terminated a flash prematurely (before all possible groups have been added to this flash). In this case, data from other subsequent flashes in the same HDF may need to be appended to this flash to make it complete.

3.3.1.1.3.4.3 Group (LIS04)

The group Vset (LIS04) holds the data associated with the various groups recorded during the orbit. The various parameters associated with the group data are shown in Table 3.8. There will be one such Vset for each group that was recorded during the orbit. The start time of the group is stored in *TAI93_time* while the lifetime and viewing time of the group are stored in *delta_time* and *observe_time*. The start time is in TAI seconds while the view time and lifetimes are in relative seconds. The *delta_time* of the group is the time difference between the first event in the group and the last event in the group. The *observe_time* is the actual time the centroid of the group was in the field of view of the LIS instrument. The parameter *location* is the latitude/longitude of the group centroid while *radiance* is the calibrated total radiance of all the events associated with the group. The group size (relative size) is recorded in the variable *footprint*. The HDF address of the group is stored in *address*. This parameter can be used to help build the parent/child relationships between the group and its child events. The parameter *parent_address* is the parent flash of the group. The links to the child events are stored in *child_address* while *child_count* is the number of child events associated with the group. The *child_address* is the HDF record number of the first child event. The approximate LIS threshold active during the group is stored in the *approx_threshold* parameter. The condition of the LIS/TRMM system is stored in the *alert_flag* parameter. The roundness of the group is stored in the *oblong_index* parameter. The *grouping_sequence* parameter indicates where this group goes in the time ordered series of groups. The variable *grouping_status* indicates if the group was active when it was written to the HDF file. It can indicate that excessive data rates may have terminated a group prematurely (before all possible events have been added to this group). In this case, data from other subsequent groups in the same HDF may need to be appended to this group to make it complete.

Table 3.8. Group (LIS07)

Variable Name	Type	Order	Description
TAI93_time	float64	1	TAI start time of the group
delta_time	float32	1	Time between first and last event
observe_time	int16	1	LIS viewing time of the group (seconds)
location	float32	2	Latitude/longitude of the group (deg)
radiance	float32	1	Total radiance of the group
footprint	float32	1	Size of the group (km^2)
address	int32	1	HDF address of group
parent_address	int32	1	HDF orbit address
child_address	int32	1	HDF address of first child event
child_count	int32	1	# of child events
grandchild_count	int32	1	# of events in group
approx_threshold	int8	1	LIS threshold during group
alert_flag	uint8	1	LIS/TRMM status during group
cluster_index	int8	1	
density_index	int8	1	
noise_index	int8	1	
glint_index	float32	1	
oblong_index	float32	1	Eccentricity of the group
grouping_sequence	int32	1	Time ordered sequence of the group
grouping_status	int8	1	End status of the group

3.3.1.1.3.4.4 Event (LIS03)

The event Vset (LIS03) holds the data associated with the various events recorded during the orbit. The various parameters associated with the event data are shown in Table 3.9. There will be one such Vset for each event that was recorded during the orbit. The event time is stored in *TAI93_time*. The variable *observe_time* is how long the location of the event was within the LIS field of view. The variable array *location* stores the location of the event on the Earth. The parameter *radiance* stores the calibrated radiance of the event. The size of the event is stored in *footprint*. The HDF address of the event is stored in *address* while the parent group for the event is indicated by the variable *parent_address*. The pixel location of the event is stored in *x_pixel* and *y_pixel*. The estimated background pixel value is stored in *bg_value* while the estimate calibrated background radiance is stored in *bg_radiance*. The background data is determined either from actual the background image or an estimate. This is indicated in the variable *bg_value_flag*. The raw amplitude of the event is stored in *amplitude*. The estimated threshold level is stored in *approx_threshold*. The summary of LIS/TRMM status is stored in alert_flag. Where the event fits in the time ordered sequence of events is listed in *grouping_sequence.*

Table 3.9. Event (LIS03)

Variable Name	Type	Order	Description
TAI93_time	float64	1	Time of the event
observe_time	int16	1	Viewtime of the event
location	float32	2	Latitude/longitude of event (deg)
radiance	float32	1	Calibrated event radiance
footprint	float32	1	Size of event (km^2)
address	int32	1	HDF address of event
parent_address	int32	1	Parent group HDF address
x_pixel	int8	1	X pixel location
y_pixel	int8	1	Y pixel location
bg_value	int16	1	Estimated raw background value
bg_radiance	int16	1	Estimated calibrated background value
amplitude	int8	1	Event raw amplitude
sza_index	uint8	1	
glint_index	uint8	1	
approx_threshold	int8	1	Estimated threshold
alert_flag	uint8	1	LIS/TRMM status
cluster_index	int8	1	
density_index	int8	1	
noise_index	int8	1	
bg_value_flag	int8	1	Background value estimation method
grouping_sequence	int32	1	Time ordered sequence of event

3.3.2 LIS Background Images Data File Structure

The background images are stored in the background data file. The background data were put in a separate file so users who do not require the background data can save disk space by not ordering it. The background file contains one Vgroup and one SDS. Each (along with any subcomponents) will be described below.

3.3.2.1 Orbit

The orbit Vgroup contains three Vsets. The first is the bg_orbit_summary which contains nearly the same information as the orbit_summary Vset in the science data HDF (Section 3.3.1.1). The second Vset is bg_info and it contains nearly the same information as the one_second Vset in the science data HDF (Section 3.3.1.1). The last Vset is bg_data_summary which contains the same information as the Vset bg_summary in the science data HDF (Section 3.3.1.1).

3.3.2.1.1 Bg_Orbit_Summary

Bg_orbit_summary (part of LIS07) are orbit data stored as an HDF Vset. The bg_orbit_summary data are listed in Table 3.10. The components for the bg_orbit_summary are the orbit identification number (*id_number*), the start time of the orbit in three different formats (*TAI93_start, UTC_start,* and *GPS_start*), the end time of the orbit in TAI93 format (*TAI93_end*), the start and end latitude locations of the orbit (*start_longitude* and *end_longitude*), information needed to find the background images and their summary data (*image_summary_count, image_summary_address, image_info_count,* and *image_info_address*). The count and address pairs allow the user to find the image data sets within the background HDF file using standard HDF tools. There is one bg_orbit_summary Vdatas per background file.

Table 3.10. Bg_Orbit_Summary (LIS07)

Variable Name	Type	Order	Description
id_number	int32	1	The TRMM orbit number designation
TAI93_start	float64	1	Orbit start time (TAI93)
UTC_start	char	28	Orbit start time (UTC)
GPS_start	float64	1	Orbit start time (GPS)
TAI93_end	float64	1	Orbit end time (TAI93)
start_longitude	float32	1	Orbit start location longitude (deg)
end_longitude	float32	1	Orbit start location longitude (deg)
image_summary_count	int16	1	# of image summary data records
image_summary_address	int16	1	HDF record of first image summary data
image_info_count	int16	1	# of background image data records
image_info_address	int16	1	HDF record of first background image

3.3.2.1.2 Bg_Info

The bg_info Vset consists of the 15 parameters listed in Table 3.11. The first parameter (*TAI93_time*) is the time of the bg_info snapshot. The next five parameters (*alert_summary, instrument_alert, platform_alert, external_alert,* and *processing_alert*) describe the status of the LIS/TRMM system. The next four parameters *(position_vector, velocity_vector, transform_matrix,* and *solar_vector*) allow the user to more accuratly calculate the location of the lightning data if they desire. The next two parameters (*ephemeris_quality_flag* and *attitude_quality_flag*) are the TRMM provided status of the satellite location and orientation data. The last parameter (*event_count*) is an array of the event counts out of significant parts of the LIS algorithm.

Table 3.11. Bg_Info

Variable Name	Type	Order	Description
TAI93_time	float64	1	The TAI93 time of the bg_info data
alert_summary	uint8	1	Bit summary of the LIS alert status
instrument_alert	uint8	1	Bit summary of the LIS instrument status
platform_alert	uint8	1	Bit summary of the LIS platform status
external_alert	uint8	1	Bit summary of the LIS external status
processing_alert	uint8	1	Bit summary of the LIS processing status
position_vector	float32	3	TRMM location vector
velocity_vector	float32	3	TRMM velocity vector
transform_matrix	float32	9	TRMM orientation matrix
solar_vector	float32	3	Sun location vector
ephemeris_quality_flag	int32	1	Bit summary of the TRMM ephemeris
attitude_quality_flag	int32	1	Bit summary of the TRMM attitude
bg_value	int16	16	
noise_index	int8	1	
event_count	int16	6	Array of event counts

3.3.2.1.3 Bg_Data_Summary

The bg_data_summary Vset contains the background image details. The components of the bg_data_summary Vset are shown in Table 3.12. *TAI93_time* is the time of the background image while *address* is the HDF address in the background file so that the user can find the actual background image. The *boresight* and *corners* parameters locate the background on the Earth.

Table 3.12. Bg_Data_Summary

Variable Name	Type	Order	Description
TAI93_time	float64	1	Time of background
address	int32	1	HDF address of background image
boresight	float32	2	Center of background image (lat/lon) (deg)
corners	float32	8	Corners of background image (lat/lon) (deg)

3.3.2.2 Bg_Data (LIS02)

The Bg_data SDS (LIS02) is the actual storage location of the non-calibrated background (raw counts) image. A background image provides a “snap shot” of the sensor’s field of view when the lightning data volume permits the transmission of this image. The data from the background images are transmitted as part of the data stream. It is stored in an SDS format as 128 x 128 resolution. Although the data is stored as an int32, the actual data is only 12 bits. The SDSs are stored in time order as they are transmitted from LIS. The details of the SDS content are shown in Table 3.13.

Table 3.13. Bg_Data SDS (LIS02)

Variable Name	Type	Order	Description
-----	int32	128x128	The pixel by pixel amplitude of the background image

3.3.3 Metadata

The Metadata structure, shown in Table 3.14, is a text description of the LIS parameters unique to this orbit. There will be as many of these text structures as are needed to describe the unique LIS orbit parameters.

Table 3.14. Metadata

Variable name	type	Description
metadata	char[80]	Metadata text description

3.3.4 Intra-structure Links

In the event, group, flash, and area Vgroups/Vsets described above there are links to the parent and child items. These links are illustrated in Figure 3.03, Figure 3.04, and Figure 3.05. To keep the file overhead to a minimum, HDF tag/reference pair pointers were not used to make the links. Instead, the HDF index of each Vdata or Vgroup was recorded and used as the link. In this way, a user of the data can find the parent and child items that contributed to the item they are currently analyzing by referencing the index of the parent or child item in the HDF file. For example, if a user is looking at a flash, he can easily determine the area that contains the flash. The user can easily find and analyze other flashes in the area. The user could also look at the individual groups and events that contributed to the flash.

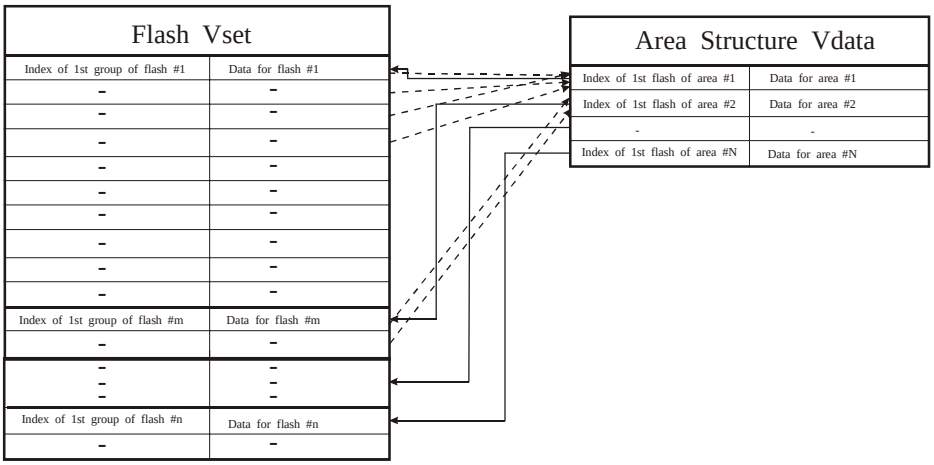

Figure 3.03 Area/Flash Data Links

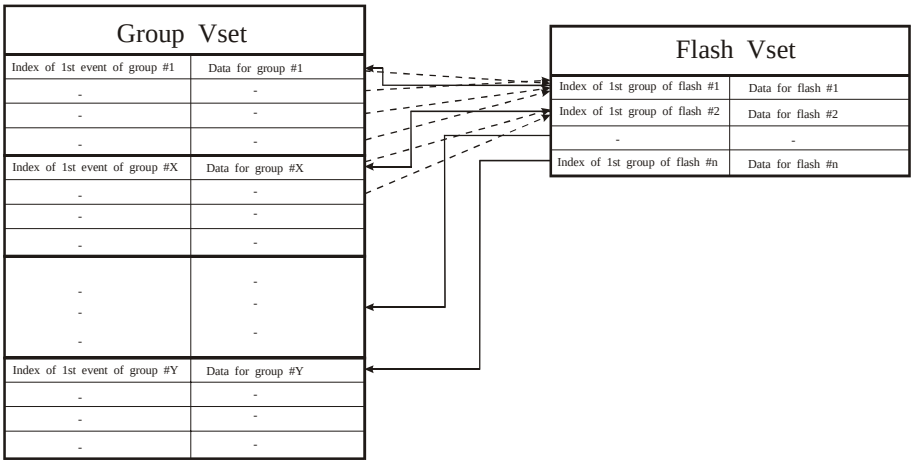

Figure 3.04 Flash/Group Data Links

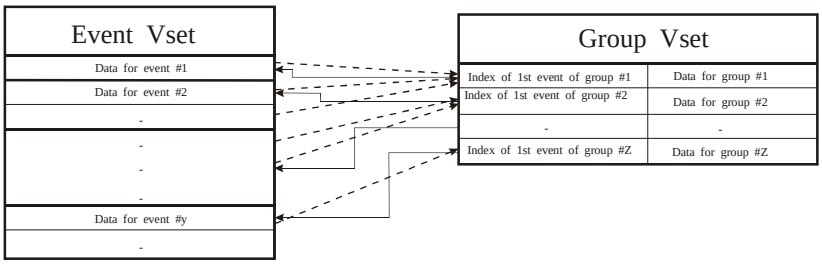

Figure 3.05 Group/Event Data Links

Other links are also provided for the user. These include links between the background images and the background image summary data in both the lightning science file and the background file. In all cases, they provide links between similar datasets.

3.4 Algorithm Mathematical Description

3.4.1 Example Data Processing Sequence

The purpose of this section is to graphically describe the algorithm that accumulate the individual LIS events into groups, flashes, and areas by "walking through" a typical LIS data scenario. In this illustrative exercise, all times indicated are times after the first event time. Numbers indicate event numbers while lowercase letters represent the groups. The flashes are designated by capital letters and the areas by Greek letters. Each subsequent section describes how the algorithm processes the events that occurred at that integration time. For the purpose of this demonstration, it is assumed that there were no events prior to the events at time 0 and that the pixel grid is 4 km wide. In general, the latitude/longitude grid in earth-based coordinates and the pixel grid will not be the same size or co-registered. In addition, the times will be time from the start of the orbit.

3.4.1.1 Time = 0 ms

The first time integration is shown in Figure 3.06. Three (**1, 2, 3**) events occur at this time integration. Since the events are simultaneous and register in adjacent (i.e., neighboring or diagonal) pixels, they are collected into a single group (**a**). The group is assigned a new parent flash (**A**) and the new flash is assigned a new parent area (α).

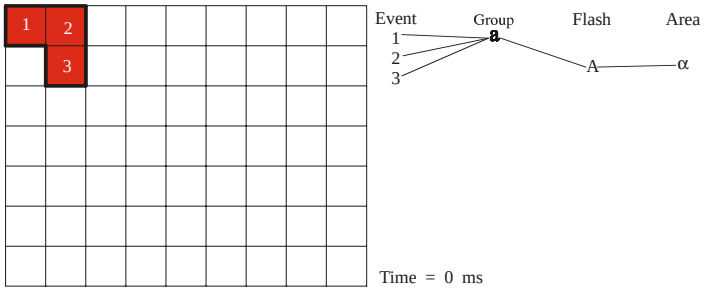

Figure 3.06 Time = 0 ms

3.4.1.2 Time = 100 ms

The next time integration with data is shown in Figure 3.07. At this time (100 ms after the first one), there are three more events (**4, 5, 6**). As in the previous case, these three new events are all assigned to a new group (**b**). These events are not assigned to group **a** since they occur at a different time. Since group **b** is within 5.5 km of group **a** (actually, they touch), and the groups occur within 330 ms of each other, group **b** is assigned to flash **A** and therefore, area α .

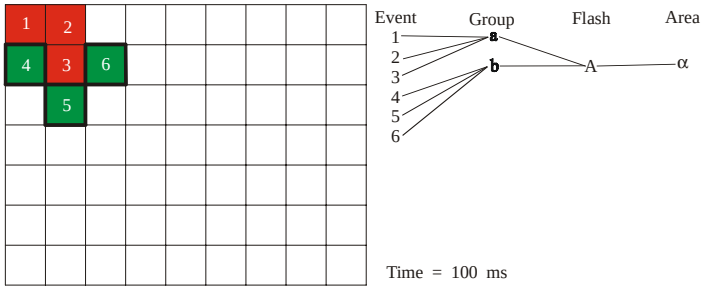

Figure 3.07 Time = 100 ms

3.4.1.3 Time = 350 ms

The next integration time with data is shown in Figure 3.08. The time is 350 ms after the time of the first events, but only 250 ms after the time of the last events. At this time there are four (**7, 8, 9, 10**) more events. Events **7** and **8** are adjacent to each other and are assigned to a new group (**c**). Events **9** and **10** are not adjacent to events **7** and **8**, but are adjacent to each other. They are assigned to another new group (**d**). Since group **c** is within 330 ms of the last group of flash **A** (250 ms) and is also within 5.5 km of the parts of flash **A**, group **c** is assigned to flash **A** and area α. Although group **d** also occurred within 330 ms of the last group of flash **A**, it is greater than 5.5 km away from any part of flash A so it is assigned to a new flash (**B**). The parts of flash **B** (i.e., group **d**) are greater than 16.5 km away from any part of area so flash **B** is also assigned a new area (β).

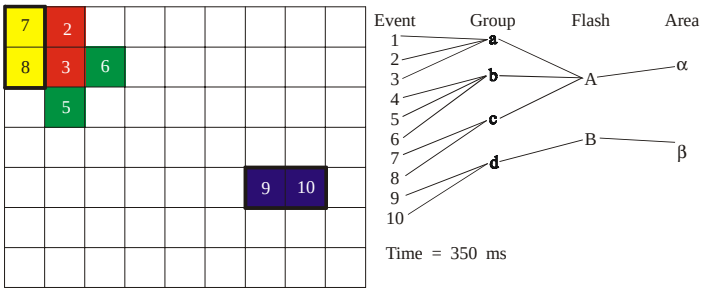

Figure 3.08 Time = 350 ms

3.4.1.4 Time = 400 ms

Figure 3.09 shows the next integration time with data. The time is 400 ms after the first events and 50 ms after the latest events. Two more events occur (**11, 12**) at this time. These two events are at the same time, but they are not adjacent to each other. They are assigned to two new groups (**e** for **11** and **f** for **12**). The two new groups are less than 330 ms (50 ms) from the time of the last group of flash **B** and are within 5.5 km (adjacent) of the parts of flash **B** so the two groups are assigned to flash **B** and area β.

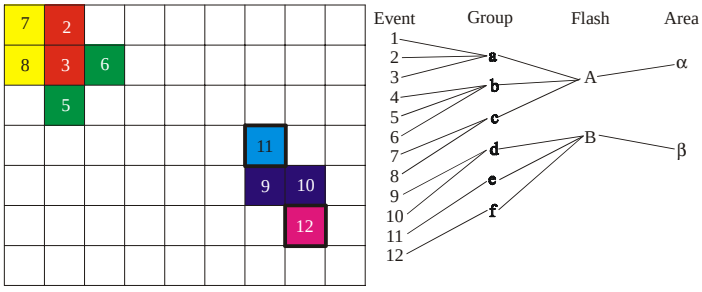

Figure 3.09 Time = 400 ms

3.4.1.5 Time = 700 ms

The last time with events (for this example) is shown in Figure 3.10. At this time integration, 700 ms after the first events and 300 ms after the last events, there are two new events (**13, 14**). The events are not adjacent, so they are assigned to two new groups (**g** for **13** and **h** for **14**). Group **g** overlaps the parts of flash **A**, however, it has now been 350 ms (greater than 330 ms) since the last group associated with flash **A**.

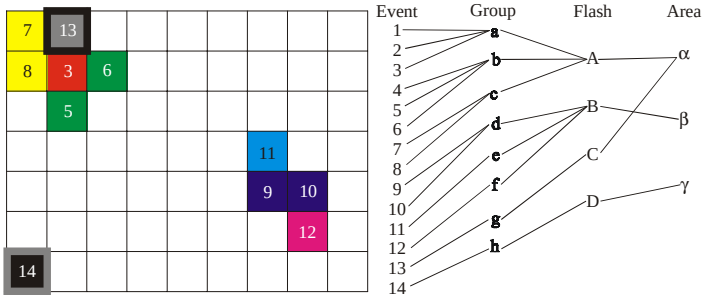

Figure 3.10 Time = 700 ms

Therefore, group **g** is assigned to a new flash (**C**). Since flash **C** overlaps the parts of area and since there is no time limit for areas, flash **C** is assigned to area . Group **h** is not within 5.5 km of any current flash, so it is assigned another new flash (**D**). Flash **D** is also not within 16.5 km of any currently active area so it is assigned another new area (γ).

3.4.1.6 Summary Data

In the example data processing sequence just described, there were fourteen events, eight groups, four flashes, and three areas. This example shows how the LIS algorithm will convert events into groups, flashes, and areas. Some of the summary data statistics that would be generated from the LIS processing algorithm are shown in Tables 3.15 (areas), 3.16 (flashes), and 3.17 (groups) for this example. During the LIS mission, the start_time is a relative time that will be counted from the beginning of each orbit.

Table 3.15. Resultant Area Data

area_id	start_time	delta_time	event_count	latlon_count	child_count	child_ids
α	0	700	7	6	2	A, C
β	350	50	4	4	1	B
γ	700	0	1	1	1	D

Table 3.16. Resultant Flash Data

flash_id	parent_id	start_time	delta_time	event_ count	latlon_ count	child_ count	child_id's
A	α	0	350	6	6	3	a, b, c
B	β	350	50	4	4	3	d, e, f
C	α	700	0	1	1	1	g
D	γ	700	0	1	1	1	h

Table 3.17. Resultant Group Data

group_ id	parent_id	group_ time	event_ count	latlon_ count	child_ count	child_ids
a	A	0	3	3	3	1, 2, 3
b	A	100	3	3	3	4, 5, 6
c	A	350	2	2	2	7, 8
d	B	350	2	2	2	9, 10
e	B	400	1	1	1	11
f	B	400	1	1	1	12
g	C	700	1	1	1	13
h	D	700	1	1	1	14

3.4.2 Algorithm Overview

There are two major products produced by the LIS software: a lightning dataset and a corresponding background dataset. To obtain these datasets, the satellite data stream needs to be decoded, filtered, clustered, and output to the appropriate HDF file. The LIS software performs this analysis in several steps as shown in Table 3.18. Note that the software unit names often do not fully reflect the tasks performed in the unit. The sections that follow will describe the tasks (see Table 3.18) needed to convert the raw data stream into the lightning and background HDF files.

Table 3.18 Software Tasks

Task	Software Unit
TRMM to Native Lightning/Background Format Converting	raw_2_stream
Pixel Based Filtering	stream_2_filtered
TRMM to Native Ephemeris Format Converting	ephemeris_2_estream
Ephemeris Filtering	estream_2_filtered
Geo-Locating	filtered_2_quicklook
Determining LIS Viewtime	viewtime
Flash Clustering	gfiltered_2_areas
Flash Based Filtering	gfiltered_2_areas
Area Clustering	gfiltered_2_areas
Area Based Filtering	areas_2_afiltered
HDF File Creation	afiltered_2_hdf

3.4.3 Significant Algorithm Details

3.4.3.1 Conversion From TRMM to Native Lightning Format

The LIS data is array (128x128) based. The details were discussed in Section 3.1. The data is formatted to the TRMM standard and sent to the ground system. The purpose of the conversion routine is to filter out and separate the lightning, background, and platform/instrument health measurements into separate data streams. The resultant lightning stream consists of time ordered time and pixel location data points. The background stream consists of time ordered time and background array value data. The instrument/platform health data is transformed into a stream of one second snapshots of the platform/instrument health. The conversion is performed in the raw_2_stream code.

3.4.3.2 Pixel Based Filtering

The lightning data stream contains many non-lightning artifacts that must be filtered. Hereafter, these non-lightning artifacts are designated as NLEs (Non-Lightning Events). Many of the artifacts are due to phenomena that occur at the LIS CCD array. In most cases, these artifacts can be removed before the data is geo-located. There are several filtering algorithms whose names reflect the general nature of the artifact being removed. Each of these filters is listed in Table 3.19 and is described below. All pixel based filtering is accomplished in the stream_2_filtered routine. Table 3.19 also includes the percentage of events removed during each stage of the filtering process for an example dataset.

Table 3.19. Pixel Based Filters					
Filter	Filter Invocation (per month)	Total Events	Removed by filter	% of original removed	% of previous step removed
Raw	0	23,132,561			
Sort	396	23,132,561	0	0.0000%	0.0000%
Dedupe	994	22,868,577	263,984	1.1412%	1.1412%
Ghost	29591	22,754,954	113,623	0.4912%	0.4969%
Lollypop	462	22,731,987	22,967	0.0993%	0.1009%
Track	33106	21,655,728	1,076,259	4.6526%	4.7346%
Blast	226	21,640,106	15,622	0.0675%	0.0721%
Total Pixel	-----	21,640,106	-----	-----	-----

3.4.3.2.1 Dedupe

The LIS system sometimes produces duplicate data points that have the same time, amplitude, and location. This filter eliminates duplicates by keeping only the first occurrence of a time/space pixel. The duplicate events are an artifact of the LIS hardware. There were 1.14% of the original events removed by the Dedupe filter (Table 3.19). A spatial plot of the removed data is shown in Figure 3.11.

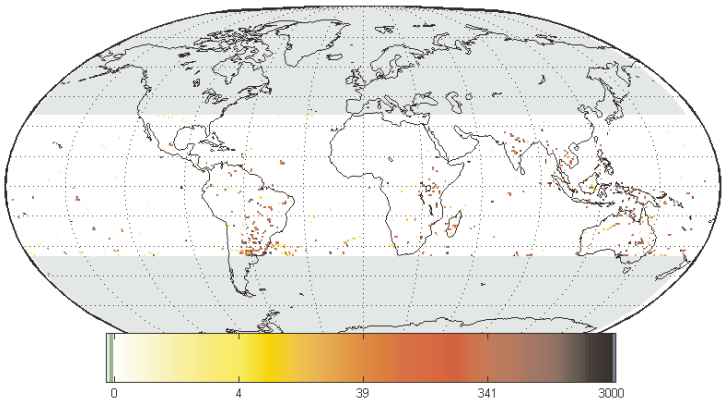

Figure 3.11 Events Removed by Dedupe

3.4.2.2.2 Ghost

The source of the ghost NLEs is the LIS CCD readout system. The CCD values are downloaded one pixel per quadrant in a round-robin fashion. If there is a bright pixel (large pulse at the readout) in one quadrant and a dim (small pulse at the readout) pixel in the next quadrant, the signal does not have enough time to settle to the correct value before the next pixel value is read. In such cases, a bright pixel in quadrant N will cause a "ghost" signal in quadrant N+1 even though there is no "real" signal in quadrant N+1. An example of a strong signal and a subsequent ghost is shown in Figure 3.12. Note that because of the readout format, the ghost signals are in "mirror image" locations in subsequent quadrants. An example of removed ghosts from a dataset

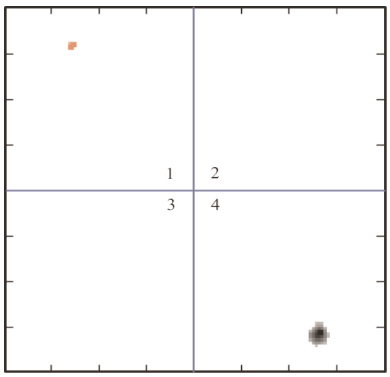

Figure 3.12 Ghost Example. The red group in quadrant 1 is a ghost of the group in quadrant 4

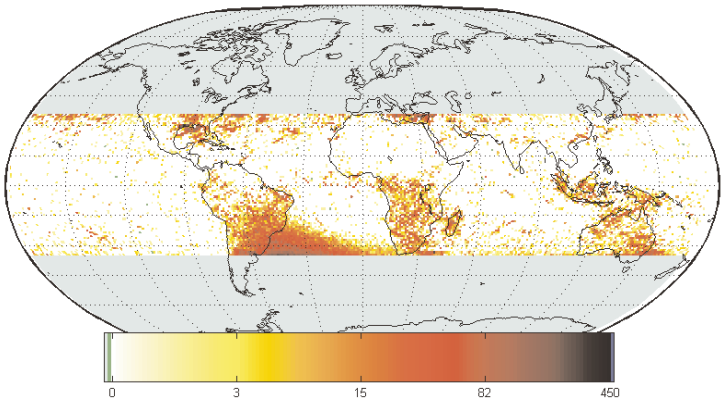

Figure 3.13 Events Removed by Ghost Filter

is shown in Figure 3.13. These removed ghosts represent 0.49% (Table 3.19) of the original event count.

3.4.3.2.3 Lollypop

The lollypop NLEs are due to bright pulses overloading the CCD array. They are similar in appearance to blooming on a TV monitor. When there is a bright lightning pulse, vertical streaks (i.e., a set of pixels with the same "y" coordinate) are attached to the top and bottom of the bright pulse. Although the large and bright pulse is due to lightning, the extended upward and downward streaks are not. This artifact occurs when optical energy is deposited during the read-out phase of the CCD array timing cycle. The lollypop is different from the streak in that the central part of the lollypop is actually due to lightning and that the "tail" is always vertical. Figure 3.14 shows two examples of lollypop pulses. All events of a Lollypop occur at the same frame time.

Lollypops occur with a large percentage of the groups that contain over 100 events. In most cases the maximum amplitude of these large groups is saturated at 127 counts. Lollypops are also associated with raw amplitudes greater than about 100 counts. The filter distinguishes between the stick and the Lollypop. The filter first identifies pulses with more than 9 events that have a major/minor axis ratio of at least 1.5 and that are vertically oriented. The filter then determines the maximum width of the vertical streak (1/2 of the minor axis) that will be removed. The filter then finds the vertical extent of the pulse and deletes the upper and lower sections until the width of the sections is greater than 1/2 of the total width of the pulse. The remaining events in the pulse are left alone. The effects of the lollypop filter on the example dataset are shown in Figure 3.15. Only 0.1% of the events in the example dataset were removed by the Lollypop filter (Table 3.19)

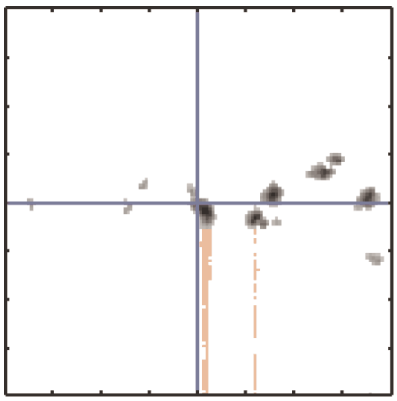

Figure 3.14 Lollypop Examples. The red 'sticks' extending from the bright groups are the artifacts to be removed by the Lollypop filter.

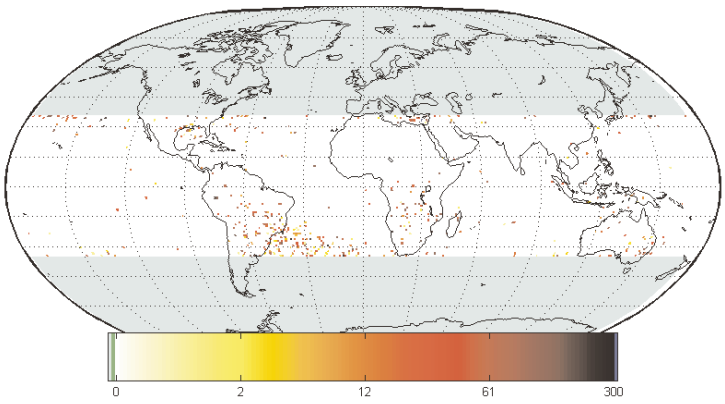

Figure 3.15 Events Removed by the Lollypop Filter

3.4.3.2.4 Track

Energetic particles impacting the CCD array trigger events. Particle paths out of the CCD array plane produce a compact group of events centered on one location. When the particle

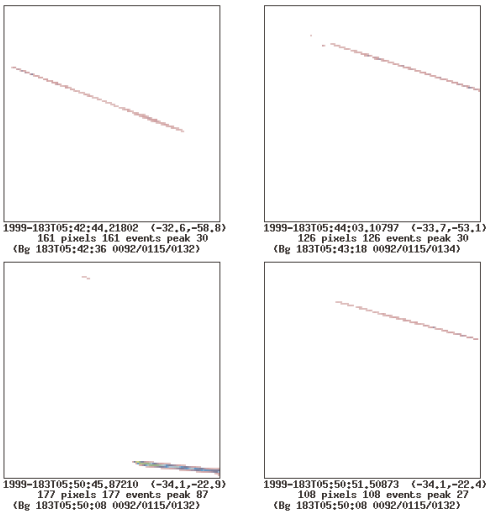

Figure 3.16 Examples of Tracks

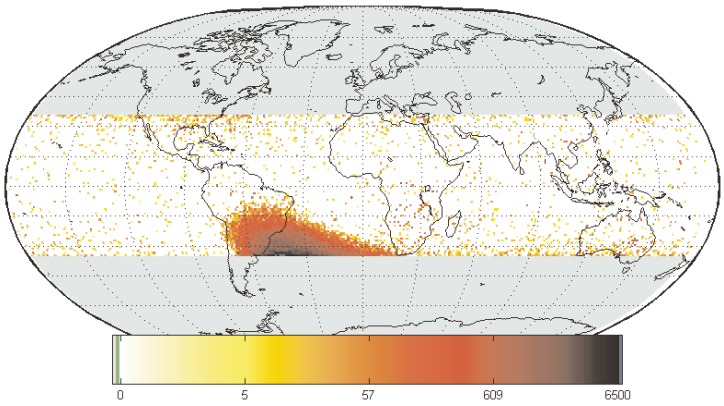

Figure 3.17 Events Removed by Track Filter

impacts the CCD array in the plane of the detector, a track of events is detected at a single frame time. The Track filter identifies and removes these tracks. A track is identified when the source produces a long linear group of pixels in a single time frame. It is very unlikely that lightning would cause a narrow streak over several hundred kilometers in a single 2 ms CCD time frame. Examples of tracks are shown in Figure 3.16. Events removed by the Track filter are shown in Figure 3.17. The Track filter removed 4.65% of the original events in the example data set (Table 3.11).

3.4.3.2.5 Blast

The source of these NLEs is unknown. The characterization of this NLE source is that occasionally a significant fraction of the whole LIS array will have events during the same 2 ms data frame (Figure 3.18). Since it is very unlikely that more than 20 lightning storms will each have a stroke in the same 2 ms time frame, any frame with more than 20 groups will be eliminated. For the example dataset, the original event count was reduced by 0.07% (Table 3.19). Figure 3.19 shows the data removed from the example dataset by the blast filter.

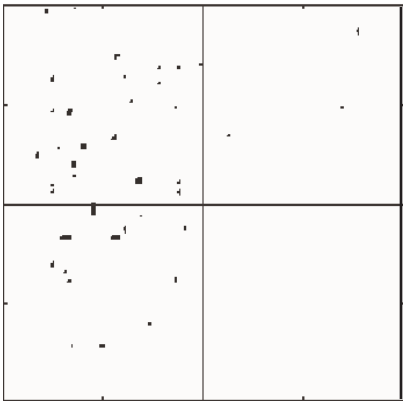

Figure 3.18 Example of a Blast NLE

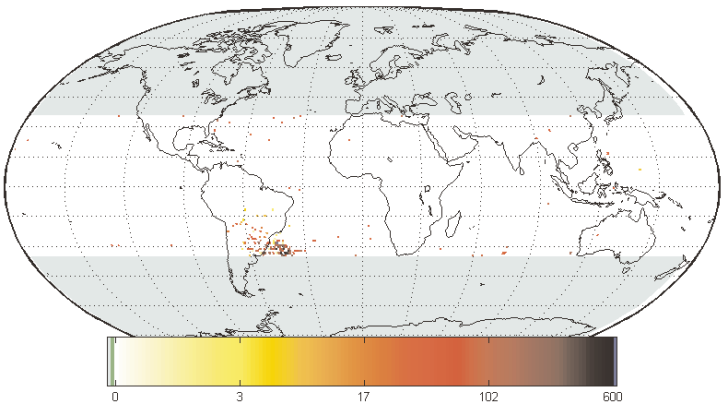

Figure 3.19 Events Removed by Blast Filter

3.4.3.3 Conversion From TRMM to Native Ephemeris Format

The TRMM ephemeris data are processed by code similar to core SDP Toolkit functions, but adapted for LIS processing. The first step in geo-location is to convert the TRMM native format ephemeris into a stream of satellite locations and satellite orientation vectors. These two streams are then later used to geo-locate the lightning and background data to Earth coordinates. The routine that creates the ephemeris/attitude data is ephemeris_2_estream (Table 3.18).

3.4.3.4 Ephemeris Filtering

Anomalous artifacts in the TRMM ephemeris are identified and removed by the ephemeris filtering routine called ephemeris_2_estream (Table 3.18).

3.4.3.5 Geo-Location

The filtered lightning, background, and ephemeris data can now be combined to produce lightning and background data projected to Earth coordinates. This is the geo-location process. The pixel location of the LIS data is combined with the TRMM satellite location (ephemeris), the TRMM satellite orientation (attitude), and the relative orientation of the LIS instrument on the TRMM platform (look vector) to determine the Earth location of the lightning pulse. Note that since the LIS data represents lightning output at the cloud top, the geo-locations are those of the approximate tropopause and not ground locations. This distinction will not matter for data directly below the TRMM platform, but for angles near the edge of the LIS field of view, the difference can be as large as 15 km. Although the SDP Toolkit has functions to directly calculate the Earth location from the TRMM data, we have elected to streamline and customize the code to our needs. The geo-location is accomplished in the routine filtered_2_quicklook (Table 3.18).

3.4.3.6 Viewtime Determination

Because the LIS instrument views a particular location for only a short period of time during an overpass, the “viewtime” is a very important parameter. Two lightning events occurring during a one second viewing window indicate a more intense storm than two lightning events over an 80 second viewing window. This routine computes viewtimes for $0.5° \times 0.5°$ latitude/longitude grid boxes within the field of view of LIS during an orbit (see Section 3.3.1.1.3.2). The routine name that determines the viewing window for each Earth location is called viewtime (Table 3.18).

3.4.3.7 Flash Clustering

The flash clustering routine (described in detail in Section 3.4.1) is called gfiltered_2_areas (Table 3.18). The routine first clusters the data to the flash level (i.e., producing groups and flashes) and then uses statistical information to filter the flash data. The flash based filtering will be described in the next section.

3.4.3.8 Flash Based Filtering

The next step after applying the event based filters is to geolocate the data. This is done in the Geo routine. In this routine, removal of data due to geo-location errors effectively serves

Table 3.20. Flash/Area Based Filters

Filter	Filter Invocation (per month)	Total Events	Removed by filter	% of original pixel-filtered events removed	% of previous step removed
Geo	9729	21,422,274	217,832	0.9417%	1.0066%
Particle	615358	10,346,834	11,075,440	47.8781%	51.7006%
Putback	19632	10,430,955	-84,121	-0.3636%	-0.8130%
Contrast	16490	10,251,091	179,864	0.7775%	1.7243%
Glint	9236	6,388,003	3,863,088	16.6998%	37.6847%
Jumper	214	6,370,501	17,502	0.0757%	0.2740%
Guilt	621	6,368,505	1,996	0.0086%	0.0313%
Single	330	6,362,633	5,872	0.0254%	0.0922%

to “filter” the data. This is illustrated in Table 3.20 which reveals that 0.94% of the original event count was discarded due to geo-location error in the Geo routine.

After applying event based filtering and geolocating the data, non-lightning artifacts remain in the data stream. These remaining NLEs are filtered based on the local temporal and spatial characteristics of lightning compared with the characteristics of the noise artifact being filtered. To simplify processing, this filtering is done after the data has been clustered into groups, flashes and/or areas. Artifacts to be filtered (and the amount of data discarded in each step) are shown in Table 3.20 and are described in more detail in the following sections.

The flash-based filters are based on the concept of persistence. That is, if there is at least one verified flash due to lightning in a region (passed all filters), then additional pulses in the same region are also likely due to lightning. If no verifiable lightning flashes are identified in a region, then there should continue to be no lightning pulses in the region. For example, if there is only one group in a region, there is no confirming information to verify that this event is a lightning event. The more groups in a region, the more likely it is that they are due to lightning.

3.4.3.8.1 Clustering Algorithm

The data is clustered into flashes before the data is further filtered to allow additional NLEs to be differentiated from the lightning events. The filtering is done by exploiting the differences in the spatio-temporal characteristics between NLEs and lightning events. The algorithms used to cluster the data into groups, flashes and areas are described in Section 3.4.

3.4.3.8.2 Particle Filter

When averaged over the whole orbit, the major source of NLEs passed by the pixel-based filters is due to random energetic particles impacting on the CCD array. The effect of these particles is to produce temporally and spatially random groups of events. Seldom is a single pixel activated by a particle impact. The average rate of these random events varies over the globe reaching a maximum in the South Atlantic Anomaly (SAA). Because of the shape and orientation of the Earth’s magnetic field, the trapped energetic particles penetrate to LEO satellite altitudes near the Atlantic coast of South America.

Although lightning can also be considered a random occurrence, lightning events cluster in both time and geo-located space. Energetic particle impacts occur in a single frame time. It is only by chance that two independent impacts (on the CCD) map to the same earth location in a moving reference frame. Lightning discharges are a complex sequence of optical events that pro-

duce internal illumination in a cloud. The spatial extent of the cloud determines the maximum size of the illuminated region. Therefore, the particle filter theory is to keep the flashes that cluster in time and space and to eliminate the flashes that do not cluster.

A weakness of this method is that, by random chance, noise will sometimes produce clusters that look like lightning clusters and these will be included in the lightning data set. The number of noise clusters increases as the noise rate increases. In addition, some very small lightning clusters with few groups will resemble randomly generated data and may be removed from the lightning data set. The selection criteria for accepting random noise as data and rejecting lightning as data are not independent of one another. Restricting the amount of noise that will be accepted into the data set will increase the amount of lightning data rejected. Restricting the amount of lightning data eliminated will increase the amount of noise passed. Thus, before processing, the parameters must be set to maximize the retention of actual lightning and minimize the acceptance of noise. The processing algorithm retains the flashes rejected at this stage for possible reinsertion during execution of the Putback routine (Section 3.4.3.10.1). The Particle filter itself consists of gfiltered_2_areas and areas_2_afiltered (Table 3.20), which will be described in the following sections.

To be able to characterize how the algorithm clusters random noise, we simulate a noise data set and use the clustering algorithm to determine how many noise clusters are produced at a given input noise rate. This is accomplished through a Monte-Carlo simulation technique.

3.4.3.8.2.1 Monte-Carlo Simulation

To estimate how noise will cluster with this algorithm, we first determine what the noise actually looks like in the real data set. This was done by examining the LIS data during times when the data stream was dominated by noise. Any groups that looked like true lightning were removed to produce pure noise data sets. This examination was the conceptual basis for the simulation of a noise data set. Simulated noise data sets were created at various noise levels (noise event rates).

3.4.3.8.2.2 Data Set Construction

Random number generators were used to create simulated noise pulse data sets that had a uniform spatial distribution and a Gaussian temporal distribution. The location of the simulated noise pulse was simulated by uniform random values in both the X and Y dimensions. The temporal spacing between pulses was simulated by a Gaussian distribution with an average interpulse time corresponding to noise rates from 0.0625 to 512 groups per second. This covers the range of data rates found in the LIS data.

The resultant data sets were considered to be “group” data and not event data. Because the sources of the random noise often fires more that one pixel at a time, a set of random single pixels would not simulate the random noise data very well. In addition, treating the simulated data as groups allows the X, Y locations to be other than integer pixel values (again, better simulating randomly distributed groups).

The resulting data sets were then clustered with the LIS clustering algorithm to produce noise flashes at the various noise rates. One possible problem with the simulated random noise

data sets was the influence of the pseudo- random number sequence on the clustering of the data. Several data sets with the same time/space parameters but different starting points in the random number sequence were created and compared. The statistics for the different runs were very similar and therefore the pseudo-random number sequence was not a factor in the outcome of the simulations.

The resulting noise flash rates are plotted versus the noise data rates in Figures 3.20, 3.21, and 3.22 for single, double, and triple group flashes, respectively. Only about 1.5 noise flashes are expected at noise event rates at the highest reported by LIS (about 500 per second). There were very few noise flashes with four or more groups. Therefore, for the noise data rates observed by LIS, we assume a flash is real lightning if it contains more than 3 groups.

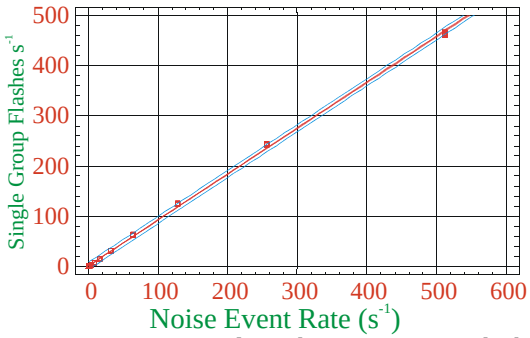

Figure 3.20 Expected Single Group Noise Flash Rate

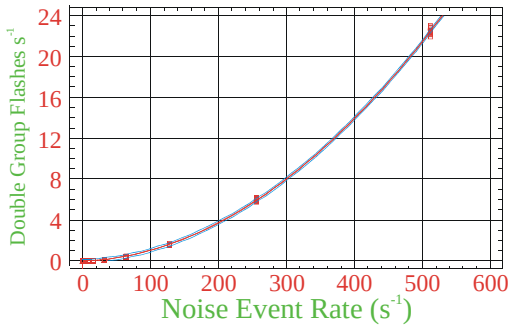

Figure 3.21 Expected Double Group Noise Flash Rate

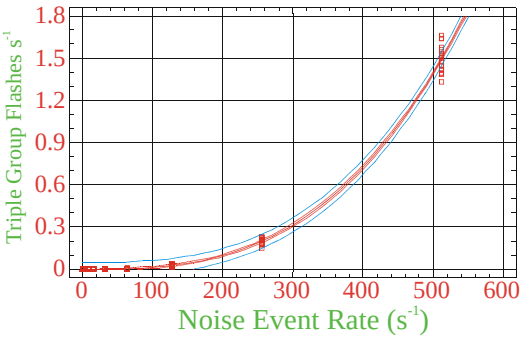

Figure 3.22 Expected Triple Group Noise Flash Rate

3.4.3.8.2.3 Results

Since most noise pulses do not overlap with other noise pulses, the noise single group flash rate (Figure 3.19) increases linearly with increasing noise rates. Interestingly, the double noise group expected flash rate (Figure 3.20) increased quadratically with increasing noise rate. The curve of triple noise group flashes versus noise rate fit a cubic curve (Figure 3.21). There were not enough four group noise flashes to produce a reasonable regression curve.

Consequently, at some point, the number of expected double/triple/quad group noise flashes would exceed the single group noise flashes. This phenomenon is a consequence of increasing arrival rates, which increases the chance for subsequent pulses to overlie the location of a previous pulse. Thus, at very high noise rates (higher than observed by LIS), the noise would group into more multiple group flashes than single group flashes.

3.4.3.8.2.4 Estimation of the Noise Rate

Estimation of the actual noise rate is complicated by the fact that we do not measure the noise rate. All we have is the noise plus data rate (N + D). Therefore we have to estimate the noise rate by making some basic assumptions about the data stream. The first assumption is that

the unfiltered LIS data stream is dominated by noise data. This assumption is true during most time intervals. The second assumption is that lightning will be a burst-like signal added onto the more uniform noise data.

Using these assumptions, we can eliminate the largest one-second pulse rate values from a section of raw data to estimate the background noise level. First, a time series of one second data rates are computed from the LIS data. This time series is then smoothed by employing a 120 second average centered at each one second interval. Next, the rates are sorted and the bottom 70% are averaged to get the estimated noise rate. During high lightning rate periods, overestimates of the noise rate may occur and cause rejection of some actual data. However, these flashes are retained by the program for possible reinsertion into the data stream during the Putback procedure. The FIFO full periods are not used in the background noise level determination.

3.4.3.8.2.5 Acceptance/Rejection of Flashes

Once the group noise rate is estimated, the results of the Monte-Carlo simulation can be used to make an estimation of the number of expected noise clusters of each group count (1, 2, 3). The expected number of noise clusters for each group count is compared to the number of LIS flashes for each group count found during the time of interest. If the expected number of noise clusters (of the group count) is significantly less than the actual number of flashes, then all flashes of that group count are passed as actual lightning. For example, if the expected number of 2-group noise clusters is 1, and the number of LIS 2-group flashes is 10, all of the 2-group flashes are accepted at a noise acceptance level of 10% or less. In this example, 10% of the LIS 2-group flashes are expected to be NLEs (one of the ten 2-group LIS flashes is noise). The current noise acceptance level is 10%.

3.4.2.8.2.6 Rejected Flashes

If the number of expected noise clusters is greater than the maximum fraction allowed, all flashes of that group count are rejected for that second. For example, consider a one second interval with a noise rate of 51 groups per second. At this noise rate, we would expect 0.5 two group noise flashes per second. If 3 LIS 2-group flashes are observed in the 1 second interval, then all 3 flashes would be rejected because the fraction of flashes due to noise (17%) is greater than the 10% noise acceptance rate. The rejected flashes are retained for possible reintroduction to the data set during the Putback procedure.

3.4.3.8.2.7 Filtering Results

The effect of noise filtering is shown in Figure 3.23 (before) and Figure 3.24 (after). The flash based filtering is accomplished by the routine gfiltered_2_areas. This procedure removes 47.9% of the events in the event filtered test data set (Table 3.20).

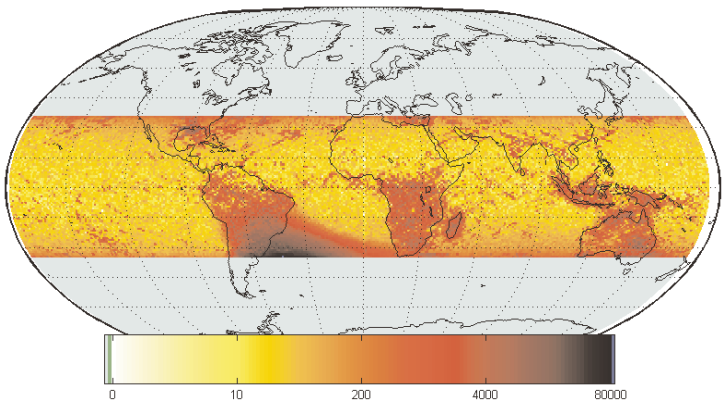

Figure 3.23 LIS Flash Density Before Noise Filtering

3.4.3.9 Area Clustering

The accepted flashes are then clustered into areas by gfiltered_2_areas. These areas are used to seed the Putback algorithm (Section 3.4.3.10.1) that re-examines the discarded flashes and reinserts those likely to be lightning.

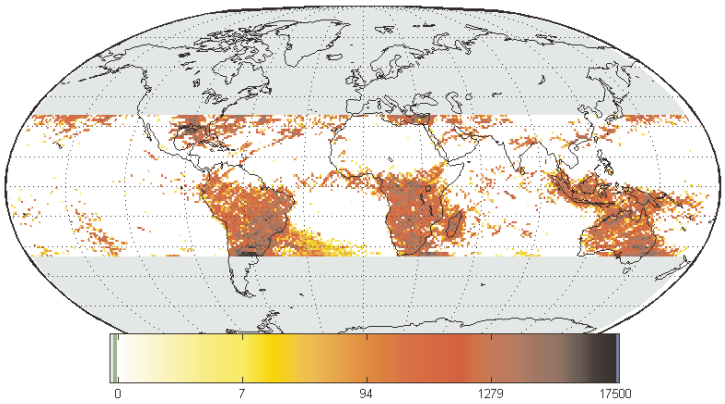

Figure 3.24 LIS Flash Density After Noise Filtering

3.4.3.10 Area Based Filters

The routine that contains the area based filters is called areas_2_afiltered. The filters fall into two very general categories. The first is the Putback algorithm that returns previously rejected noise data to the output stream based on the recalculated noise rates. The second type removes flashes from the data stream based on their very non-lightning characteristics. Each will be described in the sections below.

3.4.3.10.1 Putback Algorithm

The putback algorithm attempts to reintroduce any flashes previously rejected by the first part of the particle filter. The flashes were originally rejected based on the assumption that they were randomly produced and thus were likely not lightning. More information is available at this stage to determine whether the 1, 2, and 3 groups flashes rejected earlier are likely to be lightning. We have a better estimate of the noise rate (the flashes rejected previously) and we now have geo-located regions where we are reasonably sure there is lightning (the flashes we did not reject previously). These are the lightning areas defined in Section 3.2.5. We can now re-examine any flashes that were rejected on the basis they were random groups over the whole LIS FOV and see if they are still random over the footprint of each area of "known flashes."

First, the algorithm uses the rejected flash database to re-estimate the background noise rate using the same time window, sorting, and averaging technique as in the rejection algorithm for the entire LIS FOV. Next, a background noise rate for each lightning area is calculated by multiplying the entire LIS FOV noise rate by the view time of the area divided by the area footprint size. This usually increases the probability that a rejected flash close to or within the lightning area footprint is actually lightning. The probability that a rejected flash is actually lightning is now recomputed. If the probability that this rejected flash is lightning, it is

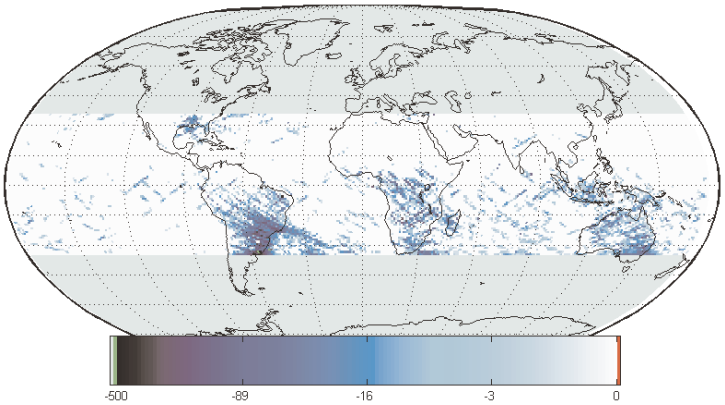

Figure 3.25 Events Reinserted Into the Data by Putback. Note that the negative numbers represent events added back to the dataset

added back into the area. An example of the Putback data for the January 1998 dataset is shown in Figure 3.25. In this case, the Putback filter reinserted 0.76% of the flashes rejected by the Particle filter (Table 3.20).

3.4.3.10.2 Contrast Filters

Additional filters are needed to identify non-random complex artifacts. These features may be correlated with brightness contrasts within the LIS FOV. Regions that may exhibit high brightness contrast include regions with snow cover, coastlines, cloud edges, or solar reflections. These contrast filters were created to remove artifacts initially identified through Quality Assurance inspection. Examples of the artifacts were analyzed, and a strategy designed to identify and remove them from the LIS data. These artifacts are related to non-random events such as solar illumination and may be modified by properties of the measurement system. Because some of these artifacts are not related to events on the surface of the earth, the geolocation calculations can produce strange "flashes". The two major types of contrast filters are called Contrast and Glint (Table 3.20).

3.4.3.10.2.1 Contrast

The contrast filters examine the LIS data and determines if a flash may be due to the LIS FOV moving across a bright dark boundary. Examples of these boundaries include edges of snowfields, coastlines, and cloud edges. Because lightning is relatively uncommon over the oceans, a few NLEs incorrectly identified as lightning can significantly bias the statistics. Unfortunately, regions of strong contrast (as viewed by the red LIS optical filter) between sunlit clouds and the water surface occur over the tropical oceans. A diagnostic test for these artifacts is to examine the event amplitudes and look for an extreme departure from a log normal distribution, or to examine the interpulse time interval distribution to look for a non Poisson distribution. For weak flashes composed of 2 or three groups of events there is not enough information to use amplitude or timing distributions, so other summary parameters are used to identify these artifacts in these cases. Data removed by the Contrast filter is shown in Figure 3.26. This filter removed 0.78% of the data sent to the flash/area filters (Table 3.20).

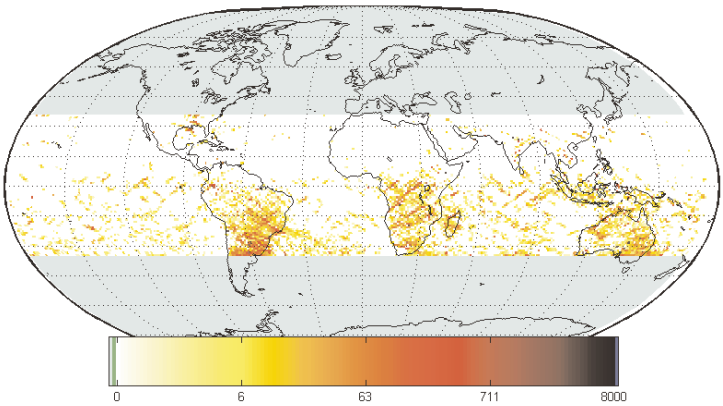

Figure 3.26 Events Removed by the Contrast Filter

3.4.3.10.2.2 Glint Filters

Two other filters remove spurious events related to bright solar reflections. These artifacts are produced by solar radiation strongly reflected or scattered back toward the platform. It is not known whether the sources are bodies of water or some other apparently bright object. The geometrical conditions for specular reflection are computed for every event, i.e., the cosine of the angle between the line of sight to the event and the direction of specular reflection.

The first glint filter (case 1) is based on the characteristics of some glint to have a lage number of events, quasi-continuously generated within a small area. The filter criteria for the first glint filter are daytime and a flash count rate greater than 250 groups per second for a duration greater than 0.2 second with a standard deviation of the event pixel y values less than three. The Y-axis was selected because it is directed perpendicular to the direction of motion of TRMM.

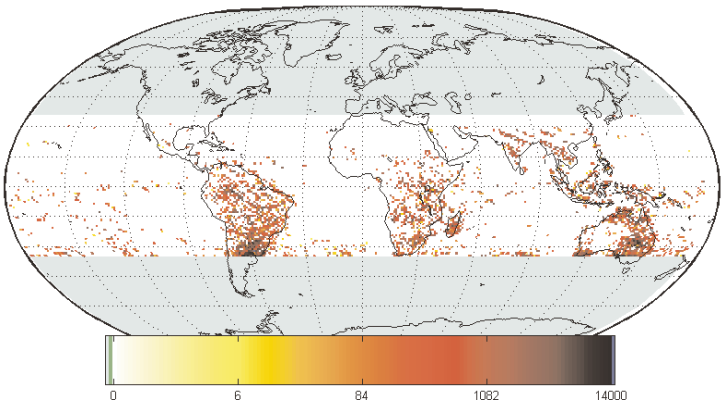

Figure 3.27 Events Removed by the Glint Filter

The second glint filter (case 2) assumes that the energy near the solar specular reflection angle is so large that it overwhelms most lightning signals. This filter removes all groups that are within 5 degrees of the specular reflection angle provided they occur in a flash with more than 5 groups of events. The data removed by the glint filters is shown in Figure 3.27. The glint filters remove 16.7% of the events that passed the event filters (Table 3.20).

3.4.3.10.3 Johnny Jumper

There is one geo-location-based filter designed to remove an instrument artifact that produces a stream of spurious events in the last row of the pixel array. The cause of this artifact is unknown. It occurs less than once per day. It does not appear to be related to solar radiation or background contrast. The artifact may be related to a real lightning event in the last few rows of the pixel array. This artifact shows up as a series of pixels triggered in the 127th row of the pixel array. The spurious pixels appear in frame after frame. The locations of the spurious pixels are variable, often in alternate positions. When these artifacts are geolocated on a map, a narrow line of "flashes" is observed. The filter identifies and removes flashes that contain only events in the 127th row. The data removed by the Johnny Jumper filter is shown in Figure 3.28. The Johnny Jumper filter removes about 0.08% of the events after event filtering (Table 3.20).

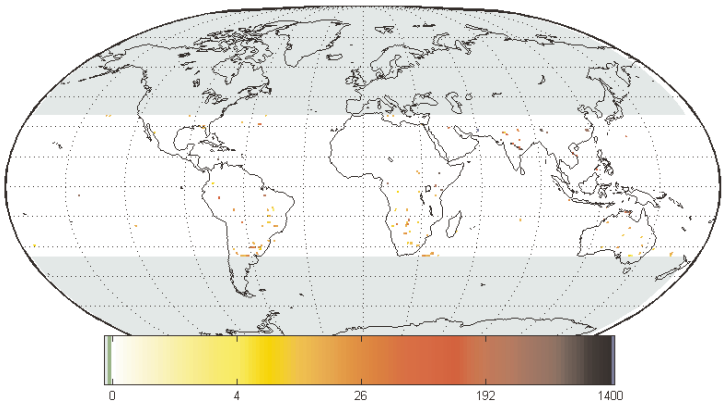

Figure 3.28 Events Removed by the Johnny Jumper Filter

3.4.3.10.4 Guilt By Association Filter

If 95% of events in an area are identified as artifacts, we assume that the remaining 5% of the events are not lightning. This filter is designed to remove the remaining pixels that escape the

fixed criteria of the previous filters. Even if actual lightning is in this area, the large number of artifacts already removed indicates that the data would be seriously contaminated. If the other filters have removed 95% of the pixels from a "flash" and there are less than six events left, the remaining events are removed via the Guilt filter (Figure 3.29). This step removes 0.009% of the event filtered events (Table 3.20).

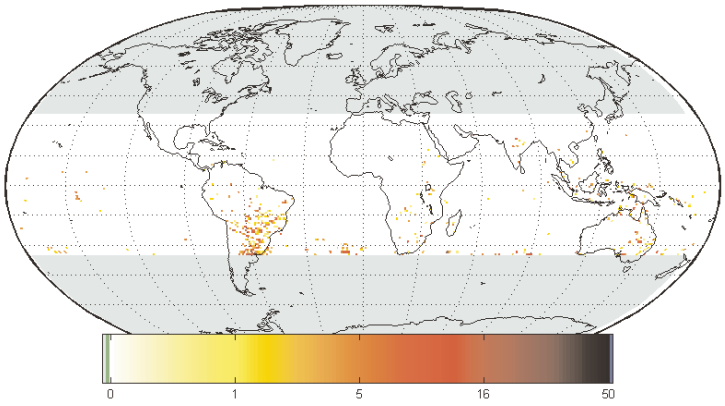

Figure 3.29 Events Removed by the Guilt By Association Filter

3.4.3.10.5 Single Group Areas Filter

There is a final filter that removes any remaining single group areas. These single pulse areas can not be distinguished from random noise. Because of this, they are eliminated. The Single filter removes about 0.25% of the events that passed the event filters (Table 3.20). A map of events removed by this filter is shown in Figure 3.30.

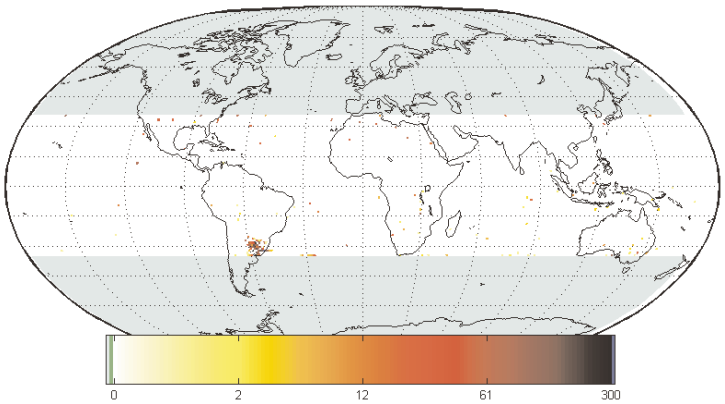

Figure 3.30 Events Removed by the Single Group Flash Filter

3.4.3.11 HDF Output

The final step is to convert the data into HDF and write it to the two HDF files. The format we use is described in Section 3.3 above. Each orbit of data is put into its own set of background and lightning HDF files.

3.5 Calibration, Validation, and Quality Control

3.5.1 Calibration

LIS calibration can be divided into two general categories. The first calibration category is to obtain an absolute radiometric calibration of the LIS sensor. This calibration is primarily a pre-launch activity, although efforts will be undertaken to determine and monitor how the absolute calibration changes during the lifetime of LIS. The second calibration category is a performance calibration of the LIS that can be fully obtained only after LIS is placed into orbit. Performance calibration is extremely important for the utilization and interpretation of LIS data.

3.5.1.1 Pre-launch Calibration (Absolute Radiometric Calibration)

The pre-launch calibration primarily addresses LIS radiometric calibration. The pre-launch calibration activities and procedures are described in detail in the *LIS Calibration*

Procedures Document. These activities include: (1) D.C. uniformity, linearity and false alarm rate tests, (2) field-of-view (FOV) test, (3) A.C. response test, (4) detection efficiency, and (5) spectral test.

The D.C. uniformity and linearity test involves exposing the entire LIS FOV to a steady, isotropic optical source and varying the source amplitude level. The D.C. response for each pixel is fully characterized in this test. In addition, the number of false alarms (i.e., transient events such as those produced by photon shot noise in LIS) is counted for different illumination levels.

In the FOV test set-up, the LIS is illuminated with a highly collimated light source whose azimuth and elevation incidence angles are precisely known relative to the LIS boresight. An Euler angle analysis of LIS output data from this test provides a precise mapping between illuminated pixel and associated light source incidence angles. This test also determines the extremities of the LIS FOV.

The A.C. response test involves illuminating selected LIS pixels with a transient optical pulse. This test provides a very precise radiometric calibration of the selected pixels to transient optical signals as a function of D.C. background illumination. During this test several pixels across the CCD array are individually calibrated and characterized (only one pixel is tested at a time). The transient log-amp response of the Real-Time-Event-Processor (RTEP) of LIS is determined from this test.

The A.C. response test can provide an initial estimate of the LIS detection efficiency when the results of the test are coupled with the energy distribution statistics derived from cloud top optical lightning observations taken during high altitude U-2 aircraft storm overflights. Ground truth calibration/validation studies conducted after the LIS is in orbit will be needed to more precisely determine the detection efficiency (i.e., see section 3.7)

The narrow pass-band filter of LIS is measured using a monochromator as part of the spectral test set-up. Center wavelength and full-width at half power are characterized in the LIS spectral test.

Finally, a lightning simulator is used to exercise the LIS instrument. The simulator employs an acousto-optical modulator and a mirror scanner to externally modulate a laser light signal to generate simulated lightning transients. In this test LIS is illuminated by several thousand simulated lightning transient waveforms. The test provides an end-to-end verification of the operation of LIS subsystems prior to launch. It should be noted that, in the present design, the lightning simulator does not produce signals of sufficient quality and stability to be used for radiometric calibration purposes.

3.5.1.2 Post-launch Calibration (Performance Calibration)

Once the LIS is launched and becomes operational the performance of the LIS will be characterized and key performance parameters calibrated. A high priority will be placed on quantifying the LIS performance early in the mission and then monitoring this performance throughout the life of the instrument. Changes in LIS performance during the mission might be an indication of drifts in the radiometric calibration of the instrument. This would be precluded if the natural variability of the LIS performance parameters are found to be high. Initial estimates of the performance parameters will be made during the pre-launch calibration activity.

The key performance parameters are the detection efficiency and the false alarm rate. The detection efficiency is defined as the percentage of lightning flashes occurring in the FOV of the instrument that are detected by the sensor. False alarm rate is defined as the percentage of total detected flashes that are not attributable to lightning. These performance parameters may display significant dependence on the conditions under which the observations are obtained. These conditions including LIS threshold setting, background intensity, observation time (e.g., time of day, time of year), storm characteristics (e.g., continental vs. maritime, large vs. small, developing vs. decaying, high flash rate vs. low flash rate) and geographical location. The effects of these conditions may be very interdependent and the responses nonlinear.

The performance calibration of LIS will be accomplished through intensive ground truth field experiments. From these field experiments we will produce data bases consisting of coincident observations from the LIS with detailed ground based lightning observations at the TRMM ground truth sites in Florida and elsewhere, and augmented with radar/rain gauge networks, geostationary satellites, and other ground based lightning detection systems. In addition, coincident ground truth measurements will be made using the high altitude ER-2 aircraft (simultaneous lightning, infrared, passive microwave, radar observation will also be obtained with the ER-2). The ground truth lightning data that will be collected and analysed includes, but is not limited to, regional ground based lightning networks, long range sferics networks, interferometry, VHF time-of-arrival, optical and electric field sensors and satellites (e.g., OTD).

The correlative or ground-truth observations made in support of the calibration/validation of the LIS instrument and data processing code will be archived in the MSFC DAAC, as well as a record of any changes in calibration parameters and/or software used during the life of the instrument.

The coincident databases that are assembled for the calibration/validation efforts will also support the LIS science activities. For example, these data bases will contribute to the development of combined instrument precipitation algorithms for TRMM that incorporate lightning observations. These databases will be used to investigate the ratio of in-cloud to cloud-to-ground lightning. The reader is referred to chapter 2 for an overview and discussion of the scientific contributions of LIS.

3.5.2 Validation

Validation is the process of verifying and tuning the performance of both the data processing algorithms described in this document and the LIS hardware. This process will include (1) remotely adjusting threshold settings to maximize detection and minimize false alarm rate, (2) verifying the true amplitude, time of occurrence, and location of lightning events detected, and (3) verifying background image brightness and alignment. The same databases being assembled and used for LIS performance calibration will be applied in the validation process. The *Science Data Validation Plan for the Lightning Imaging Sensor (LIS)* (http://spso.gsfc.nasa.gov/validation/valpage.html) discusses the pre- and post-launch validation (and post-launch calibration) activities.

3.5.2.1 Pre-launch Validation

As noted in section 2.2, in the TRMM pre-mission period, we plan to take advantage of the space lightning observations that will be provided by the launch of OTD in 1995. To validate the LIS data processing algorithm described in this document, data from the OTD instrument will be used. The OTD is a prototype LIS instrument (actually an engineering test version of LIS). Therefore, OTD will detect lightning in the same way and produce the same data stream as LIS. We will verify the LIS data processing algorithm using data from the OTD, various ground-based lightning detecting systems and other ground truth observations (ground-, airborne-, and satellite-based).

3.5.2.2 Post-launch Validation

After launch, the LIS data processing algorithm will be validated using data from OTD and coincident ground truth lightning observations. The post-launch validation process will determine how well the data algorithm processes the LIS data stream and how accurately the defined data structures correspond with the observed lightning. The LIS data processing algorithm will be adjusted or tuned to provide the best correspondence with the observations. The capability for easily adjusting the higher order data structures has been designed into the code. Any changes that are made in the LIS data processing algorithm will be reflected in updates to this document.

3.5.3 Quality Control and Diagnostics

There are several levels of quality assurance information provided with the LIS data. For the lightning data, the events, groups, flashes, and areas are all assigned quality tags that indicate whether the data element is positioned relative to events having high lightning probabilities or associated with solar glint, high noise rates, or randomly spaced events. These quality tags are based on the output of the computer algorithms and are automatically generated. Because there can be temporary variations in the behavior of the LIS instrument that can lower the effective quality of the data, it is necessary for a LIS team scientist to inspect the data for anomalies. Once the data is assembled into data granules (orbits), the data will be scrutinized for anomalies in the number and location of events, the view time calculations associated with the instrument field-of-view, the ephemeris used to geo-locate the lightning data, and the time stamps provided by the satellite clock. In addition, the data will be inspected for irregularities in the information reported in the HDF structure, such as an improper orbit start and stop time designations, data extending beyond the orbit start and stop times, and the possibility of having no lightning data or background images to record in the orbit granule.

3.5.4 Exception Handling

3.5.4.1 Out of Range Data

In nearly all cases, the possible valid range of the input data will fill the space (bit-wise) allocated to the data. Therefore, it will be nearly impossible for values to exceed the normal range of data. In any case where a data value does exceed the normal input range, the data will be flagged for later user intensive analysis.

3.5.4.2 Data Packet Errors

Any data packet errors that pass through the TRMM packet checking will probably result in valid looking data. The only way to correct such errors is for the user to view any suspicious data packets and make a case by case determination of the possible presence of an error. In most cases, the error will be unrecoverable, and the data will have to be thrown out.

3.5.4.3 Granule Division Errors

Since all events, groups, and flashes associated with an area are kept with that area, granule division errors are not generally possible. The one exception is the processing at the end of a day. Until the data from the next day arrives, any flashes and areas active at the time of an orbit change will be artificially truncated. The file structures for the truncated sets (flash, and area) will have a flag set that indicates the item was active when the orbit was terminated. It is possible that data in the next orbit should be associated with the old sets. This error will be corrected when the next day's data arrives. Any orbits with possible split data will be reprocessed and the HDF files overwritten to eliminate any truncated structures.

4. CONSTRAINTS, LIMITATIONS, AND ASSUMPTIONS

4.1 Telemetry Link Data Rate Constraints

The LIS instrument data flow from the TRMM platform is limited by the LIS hardware to 8 kbits s^{-1}. If the pulse (event) data rate is greater than this, the data will be buffered on-board to maintain the maximum 8 kbits s^{-1}. If the on-board buffers are not sufficiently large to handle some high pulse data rate incidents, the buffers will overflow. This will result in unrecoverable loss of data. A buffer overflow flag will be set in the LIS telemetry stream and whatever data that can be transmitted will be sent to the ground. Because of how the overflowing buffers affect the whole data stream, it will sometimes result in contamination of the pulse data stream with background data. Since the bad pulse data will be in the correct pulse format, there is no way to automatically correct the stream or identify and throw out the bad data in the stream.

Recovery from these data errors will be automatic, however. Once the overflow condition is eliminated, correct data flow from the LIS instrument will return. Correction of the data during the overflow condition will take user intervention. The level 0 algorithm will simply mark the sections of the data stream that might contain errors and then let the users correct the problem by viewing the data and making intuitive assumptions. After the thorough examination of the test data from the OTD instrument, some of the user intervention steps may be automated.

4.3 Limitations

4.2.1 Maximum/Minimum Values

The pulse data is a digital seven bit number. The minimum possible amplitude for data is 0x01, however, programmable (real-time user set) threshold settings will normally be set higher than the 0x01 level. So that the data compression method used in the data stream will work, there can be no data pulses with an amplitude of 0x00. The maximum value of a pulse is seven bits of 1's, or 0x7f (127 decimal counts). This maximum count value corresponds to an optical energy that is several times larger than the largest lightning pulse energy of 181 $\mu J\ m^{-2}\ sr^{-1}$ reported by Goodman et al. (1988). Any event more energetic than the optical energy associated with the maximum 0x7f reading will be tagged 0x7f, i.e., saturation will occur.

The background data is a twelve bit number. The minimum valid value is 0x000 and the maximum value is 0xfff (4095 decimal counts). The mean (and standard deviation) background count value across the OTD pixel array for the maximum expected diffuse spectral cloud-top solar reflection of 375 $Wm^{-2}\ sr^{-1}\ \mu m^{-1}$ is only 2622 ± 114 counts, substantially less than the 4095 saturation level. Solar specular reflection from a water surface will, however, account for a larger background amplitude value. However, any illumination values that correspond to 4095 or greater will be reported as 4095.

4.2.2 LIS Alignment Criteria

In accordance with MSFC-SPEC-2028, the instrument design shall be compatible with the spacecraft provided mounting plane, hole pattern, and alignment adjustment ranges and procedures. The instrument boresight and focal plane grid lines shall be nominally parallel or perpendicular to the TRMM spacecraft reference axes, and shall define the instrument reference axes.

The sensor boresight direction shall be within 12 arcmin of nominal, and the focal plane grid lines within 24 arcmin of nominal, both with respect to the TRMM spacecraft reference axes. The LIS boresight shall be perpendicular to the instrument base plane within 12 arcmin. Furthermore, it is desired that the rotation of the instrument reference axes about the boresight be within 12 arcmin of nominal with respect to a corresponding set of instrument base plane/hole pattern derived reference axes.

4.3 Assumptions

4.3.1 Lightning Signal Assumptions

Calibrated, aircraft-mounted, wide field-of-view optical pulse sensors designed to detect the prominent oxygen emission triplet in the lightning spectrum were used to observe the transient lightning optical emissions from cloud-top [Christian and Goodman, 1987; Goodman et al., 1988]. These studies have provided the basis for much of the sensor requirements provided in Table 2.1. The distribution of the number of optical pulses in a flash was found to be exponential in character with about 12 optical pulses per flash as a mean (Christian and Goodman, 1987). As indicated in section 3.1.1, about 90% of the lightning in this study produced peak radiant energy densities of 4.7 $\mu J\ m^{-2}\ sr^{-1}$ or greater. Hence, in the design of the LIS this energy value has been assumed the appropriate threshold requirement for LIS to achieve 90% detection efficiency. Although many anomalous waveforms were not included in many of the statistical analyses of optical pulse sensor data, the peak radiant energy density cited is considered conservative.

4.3.2 Background Illumination Assumptions

The same calibrated, aircraft-mounted, wide field-of-view optical pulse sensors described in the previous paragraph were also used to observe the diffuse optical emissions from cloud-top [Christian and Goodman, 1987; Goodman et al., 1988]. From these studies, the range of background illumination should range from near 0 (for night ocean) to near 375 $W\ m^{-2}\ sr^{-1}\ \mu m^{-1}$. Sun glint off of the ocean will cause much greater signal amplitude that will most likely saturate the LIS instrument.

REFERENCES

Adler, R.F., A.J.Negri, P.R. Keehn, I.M. Hakkarinen, 1993. Estimation of monthly rainfall over Japan and surrounding waters from a combination of low-orbit microwave and geosynchronous IR data, J. Appl. Meteor., **32**, 335-356.

Arkin, P. A., and P. E. Ardanuy, 1989. Estimating climatic-scale precipitation from space: A review, J. Climate, **2**, 1229-1238.

Barrett, E.C., and D.W. Martin, 1981. The Use of Satellite Data in Rainfall Monitoring, Academic Press, New York, 340 pp.

Boccippio, D. J., K. Driscoll, W. Koshak, R. Blakeslee, W. Boeck, D. Mach, D. Buechler, H.J. Christian, and S.J. Goodman, The Optical Transient Detector (OTD): Instrument characteristics and cross-sensor validation, J. Atmos. Oc. Tech, 17, 2000 (in press).

Buechler, D. E., and S. J. Goodman, 1990. Echo size and asymmetry: Impact on NEXRAD storm identification, J. Appl. Met., **29**, 962-969.

Cherna, E.V., and E.J. Stansbury, 1986. Sferics rate in relation to thunderstorm dimensions, J. Geophys. Res., **91**, 8701-8707.

Christian, H. J., R. L. Frost, P. H. Gillaspy, S. J. Goodman, O. H. Vaughn, M. Brook, B. Vonnegut, and R. E. Orville, 1983. Observations of optical lightning emissions from above thunderstorms using U-2 aircraft, Bull. Am. Meteor. Soc., **64**, 120.

_____, W. W. Vaughn, and J. C. Dodge, 1984. A technique for the detection of lightning from geostationary orbit, Preprints, 7th Int. Conf. Atmos. Elect., Albany, New York, 452.

_____ and S. J. Goodman, 1987. Optical observations of lightning from a high altitude airplane, J. Atmos. Ocean. Tech., **4**, 701.

_____, R. J. Blakeslee, and S. J. Goodman, 1989. The detection of lightning from geostationary orbit, J. Geophy. Res., **94**, 13329-13337.

_____, _____, and _____, 1992. Lightning imaging sensor (LIS) for the earth observing system, NASA TM-4350, available from the National Technical Information Service, Springfield, VA 22161-2171, 36 pp.

Eaton, L. R., C. W. Poon, J. C. Shelton, N. P. Loverty, and R. D. Cook, 1983. Lightning mapper sensor design study, NASA CR-170909, Marshall Space Flight Center, AL.

Goodman, S. J., and D. R. MacGorman, 1986. Cloud-to-ground lightning activity in mesoscale convective complexes, Mon. Wea. Rev., 114, 2320.

_____, H. J. Christian, and W. D. Rust, 1988a. Optical pulse characteristics of intracloud and cloud-to-ground lightning observed from above clouds, J. Appl. Meteor., **27**, 1369-1381.

_____, D. E. Buechler, P. D. Wright, and W. D. Rust, 1988b. Lightning and precipitation history of a microburst producing storm, Geophys. Res. Lett., **15**, 1185-1188.

_____, _____, and P. J. Meyer, 1988c. Convective tendency images derived from a combination of lightning and satellite data, Weather and Forecasting, **3**, 173-188.

_____, D. E. Buechler, P. D. Wright, W. D. Rust, and K. E. Nielsen, 1989. Polarization radar and electrical observations of microburst producing storms during COHMEX, Preprints, 24th Conf. on Radar Meteorology, Tallahassee, FL, Mar. 27-31, Amer. Meteor. Soc., Boston.

_____, and _____, 1990. Lightning-Rainfall Relationships, Preprints, Conf. on Operational Precipitation Estimation and Prediction, Anaheim, CA, Feb. 7-9, Amer. Meteor. Soc., Boston.

_____, and H. J. Christian, 1993. Global observations of lightning. Appears as chapter in Atlas of satellite observations related to global change, R. J. Gurney, J. L. Foster, and C. L. Parkinson, eds., Cambridge University Press, New York, 191-219.

Goodman, S. J., D. E. Buechler, K. Knupp, K. Driscoll, and E. W. McCaul, The 1997-98 El Nino event and related wintertime lightning variations in the southeastern United States, Geophys. Res. Lett., 27, No. 4, Feb. 15, 2000, 541-544.

Guo, C. and E. P. Krider, 1982. The optical and radiation field signatures produced by lightning return strokes, J. Geophys. Res., **87**,8913.

Hartigan, J. A., 1975. Clustering Algorithms, John Wiley and Sons, New York, 351 pp.

Holzworth, R. H. and Y. T. Chiu, 1982. Sferics in the stratosphere, CRC Handbook of Atmospherics, Vol. **2**, H. Volland, ed., 1.

Horner, F. and R. B. Bent, 1969. Measurements of terrestrial radio noise, Proc. Roy. Soc. **A**, 311, 527.

Kotaki, M. and C. Katoh, 1983. The global distribution of thunderstorm activity observed by the ionosphere sounding satellite (ISS-b), J. Atm. Terr. Phys., **45**, 833.

Kummerow, C.D., and L. Giglio, 1992. A passive microwave/infrared technique for overcoming sampling limitation of TRMM, presented at the International Workshop on the Processing and Utilization of Rainfall Data Measured From Space, Tokyo, Japan, 3/16-20.

Livingston, J. M. and E. P. Krider, 1978. Electric fields produced by Florida thunderstorms, J. Geophys. Res., **83**, 385.

Mach, D. M., D. R. MacGorman, W. D. Rust, and R. T. Arnold, 1986. Site errors and detection efficiency in a magnetic direction-finder network for locating lightning strikes to ground, J. Atmos. Ocean. Tech., **3**, 67.

Norwood, V., 1983. Lightning mapper sensor study, NASA CR170908, Marshall Space Flight Center, AL.

Orville, R. E. and R. Henderson, 1986. Global distribution of midnight lightning: September 1977 to August 1978, Mon. Wea. Rev., **114**, 2640.

Price, C., and D. Rind, 1990. The effect of global warming on lightning frequencies, Preprints, 16th Conf. on Severe Local Storms, Oct. 22-26, Kananaskis Park, Alberta, Canada, Am. Meteor. Soc., Boston, 748-751.

Sparrow, J. G. and E. P. Ney, 1971. Lightning observations by satellite, Nature, **232**, 540.

Suresh, R., D. Ilg, K. Whalen, and T. Meyer, 1994. EOSDIS Version 0 Data Product Implementation Guidlines, Draft Version 1.0, GSFC 50-003-04.

Thomason, L. W. and E. P. Krider, 1982. The effects of clouds on the light produced by lightning, J. Atmos. Sci., **39**, 2051.

Turman, B. N., 1978. Analysis of lightning data from the DMSP satellite, J. Geophys. Res., **83**, 5019.

_____, 1979. Lightning detection from space, Am. Sci., **67**, 321.

_____ and R. J. Tettelbach, 1980. Synopic-scale satellite lightning observations in conjunction with tornadoes, Mon. Wea. Rev., **108**, 1978.

Vonnegut, B., O. H. Vaughan, and M. Brook, 1983. Photographs of lightning from the space shuttle, Bull. Am. Meteor. Soc.. **64**, 150.

Vorpahl, J. A., J. G. Sparrow, and E. P. Ney, 1970. Satellite observations of lightning, Science, **169**, 860.

Weinman, J.A., J.L. Schols, and L.S. Chiu, 1993. Precipitation distributions from combined airborne radar and multi-frequency radiometric measurmeents and from ground-based VLF sferics measurements in support of NASA's tropical rianfall measuring mission, presented at the NATO Advanced Research Workshop on Global Precipitation and Climate Change, Toulon, France, 27 September - 1 October.

Williams, E. P., 1985. Large-scale charge separation in thunderstorms, J. Geophys. Res., **90**, 6013.

_____, and S. A. Rutledge, 1990. Studies of electrification and lightning in deep tropical precipitation systems, Preprints, AMS Conf. on Cloud Physics, San Francisco, CA, July 23-27, Amer. Meteor. Soc., Boston, 5 pp.

_____, E.R., S.A. Rutledge, S.G. Geotis, N. Renno, E. Rasmussen, and T. Rickenbach, 1992. A radar and electrical study of tropical "hot towers". J. Atmos.Sci., **49**, No.15, 1386-1395.