NASA
NATIONAL AERONAUTICS AND
SPACE ADMINISTRATION

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

SCIENCE DATA VALIDATION PLAN
FOR THE
LIGHTNING IMAGING SENSOR (LIS)

Hugh J. Christian\textsuperscript{1}, LIS Principal Investigator
Richard J. Blakeslee\textsuperscript{1}, LIS Validation Coordinator
Dennis J. Boccippio\textsuperscript{1}
William J. Koshak\textsuperscript{1}
Steven J. Goodman\textsuperscript{1}
Douglas M. Mach\textsuperscript{2}

\textsuperscript{1}Earth Science Department, Science Directorate
NASA/ Marshall Space Flight Center
Global Hydrology and Climate Center
Huntsville, AL 35812

\textsuperscript{2}Atmospheric Science Department
University of Alabama in Huntsville
Global Hydrology and Climate Center
Huntsville, AL 35899
<table>
<thead>
<tr>
<th>TABLE OF CONTENTS</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. INTRODUCTION</td>
<td>3</td>
</tr>
<tr>
<td>1.1 Science and Measurement Objectives</td>
<td>3</td>
</tr>
<tr>
<td>1.2 Missions</td>
<td>6</td>
</tr>
<tr>
<td>1.3 Science Data Products</td>
<td>6</td>
</tr>
<tr>
<td>2. VALIDATION CRITERION</td>
<td>7</td>
</tr>
<tr>
<td>2.1 Overall Approach</td>
<td>7</td>
</tr>
<tr>
<td>2.2 Sampling Requirements and Trade-offs</td>
<td>8</td>
</tr>
<tr>
<td>2.3 Measures of Success</td>
<td>9</td>
</tr>
<tr>
<td>3. PRE-LAUNCH ACTIVITIES</td>
<td>10</td>
</tr>
<tr>
<td>3.1 Laboratory Radiometric Tests</td>
<td>10</td>
</tr>
<tr>
<td>3.2 Laboratory Performance Tests</td>
<td>11</td>
</tr>
<tr>
<td>3.3 Laboratory Test Results</td>
<td>11</td>
</tr>
<tr>
<td>3.4 Completed OTD VALIDATION STUDIES</td>
<td>12</td>
</tr>
<tr>
<td>4. POST-LAUNCH ACTIVITIES</td>
<td>13</td>
</tr>
<tr>
<td>4.1 Overview</td>
<td>13</td>
</tr>
<tr>
<td>4.2 Post-launch Validation Goals and Objectives</td>
<td>13</td>
</tr>
<tr>
<td>4.3 Post-launch Validation Activities</td>
<td>15</td>
</tr>
<tr>
<td>INTERNAL CONSISTENCY CHECKS (Bootstrapped Validation)</td>
<td>15</td>
</tr>
<tr>
<td>CROSS-SENSOR VALIDATION</td>
<td>17</td>
</tr>
<tr>
<td>CROSS-SENSOR CALIBRATION</td>
<td>20</td>
</tr>
<tr>
<td>5. IMPLEMENTATION OF VALIDATION RESULTS IN DATA PRODUCTION</td>
<td>22</td>
</tr>
<tr>
<td>5.1 Approach</td>
<td>22</td>
</tr>
<tr>
<td>5.2 Role of LIS SCF and EOSDIS</td>
<td>22</td>
</tr>
<tr>
<td>5.3 Plans for Archival of Validation Data</td>
<td>23</td>
</tr>
<tr>
<td>6. SUMMARY</td>
<td>24</td>
</tr>
<tr>
<td>7. REFERENCES</td>
<td>27</td>
</tr>
<tr>
<td>APPENDIX</td>
<td>30</td>
</tr>
<tr>
<td>A.1 Instrument Description and Characteristics</td>
<td>30</td>
</tr>
<tr>
<td>A.2 List of Acronyms</td>
<td>34</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

1.1 SCIENCE AND MEASUREMENT OBJECTIVES

Science Objectives

Interest in lightning as a remote sensing measurement and variable of global change has grown with the recognition that lightning conveys useful information about many atmospheric processes [Christian, 1992; Davis, 1983]. For example, since lightning activity is closely linked to storm dynamics and microphysics, it can be related to the global rates, amounts, and distribution of convective precipitation [Goodman, 1986; Goodman, 1990; and Petersen, 1998] and the release and transport of latent heat. The location and distribution of latent heating associated with convection, in turn, influences larger scale atmospheric circulations and weather patterns [Chang, 1999; Goodman, 1996; Goodman, 2000]. Williams [1992a] has hypothesized that global lightning activity may provide a very sensitive measure of temperature change associated with climate variability. Climate simulations of doubled CO$_2$ concentrations (i.e., global warming) suggest a 25% increase in global lightning frequency [Price, 1990]. The high temperatures attained within lightning channels provide a mechanism for the production of nitrous oxides and other trace gases [Chameides, 1986]. However, additional details of the global lightning occurrence are needed in order to properly assess the impact of this natural production of trace gases [Levy, 1996]. Lightning relationships are also being sought with atmospheric electrical processes such as the global electric circuit [Blakeslee, 1989; Driscoll, 1993].

The Lightning Imaging Sensor (LIS), in combination with its predecessor, the Optical Transient Detector (OTD) are now providing the first nearly unbiased climatology on the rates, distributions, and variability of lightning activity on the global scale. The LIS is a NASA Earth Observing System (EOS) instrument on the Tropical Rainfall Measuring Mission (TRMM) satellite, as depicted in Fig. 1-1. It is specifically designed to acquire and investigate the distribution and variability of total lightning (i.e., cloud-to-ground and intracloud discharges) in the data sparse tropical regions of the Earth.

Lightning measurements provided by LIS will offer a unique 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 provide 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, 1988a; Goodman, 1989; Buechler, 1990; Williams, 1990; Williams, 1992b; Goodman, 1993].

The LIS is one of five instruments on the TRMM platform. The other instruments include the TRMM Microwave Imager (TMI), the Visible and Infrared Scanner (VIRS), the Precipitation Radar (PR), and the Cloud and Earth Radiant Energy System (CERES) [Kumerow, 1998]. Each instrument provides a unique set of data that can be used to learn more about the atmospheric circulation, weather, and climate. By comparing the LIS data with data from other instruments on TRMM, it is possible to examine relationships between lightning activity and several important climatological parameters, including rainfall, latent heat release, and convection.

The processes that lead to the production of lightning are tightly controlled by the cloud updraft 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 upper 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, 1993]. Thus, the detection of lightning from space specifically identifies those regions that are of paramount importance in the 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 addition on TRMM.

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, 1988b]. Weinman [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, 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, 1993; Kummerow, 1992]. Another possibility would be to develop a technique for estimating convective rainfall based upon the amount of lightning produced by the storm. Livingston [1978], Williams [1985], Goodman [1986], Cherna [1986], and others have developed relationships (i.e., empirical algorithms) for lightning rates as a function of storm size, height, and duration.

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).

Measurement Objective

The LIS is a small, solid state optical imager. The overall measurement objective of the LIS is to detect and locate lightning during day and night with high detection efficiency and location accuracy, mark the time of occurrence, and measure the radiant energy. During the 1980’s, extensive optical and electrical observations of lightning were made from a high altitude U-2 aircraft with the primary goal of defining baseline design criteria for space sensors capable of meeting this measurement objective. The results of the U-2 investigations, parametric trade-off studies, and other research [Norwood, 1983; Eaton, 1983; Christian, 1987; Goodman, 1988c, Christian, 1989] clearly established the feasibility for making this kind of lightning measurement from space using present state-of-the-art technology.

The successful launch in 1995 of the OTD, an engineering prototype of the LIS, demonstrated that the LIS measurement objective would be easily met. The OTD launch ushered in a new era of space based lightning sensors specifically designed to address the deficiencies of earlier measurements and provide accurate statistics on the frequency and distribution of lightning worldwide. The prior lightning observations from space had been severely limited by one or more problems including low or unknown detection efficiency, poor spatial and temporal resolution, a limited number of observations or brief period of observation, and incomplete sampling of the diurnal cycle [Christian, 1989]. The OTD was launched into a nearly circular orbit of 740 km with a 70º inclination, providing observations of lightning activity over most parts of the world. Since its launch, it has optically detected lightning flashes (both intracloud and cloud-to-ground discharges) that occurred within its 1300 × 1300 km field-of-view with a high, uniform detection efficiency (~50%) and storm scale spatial resolution (~10 km) during both day and night.

Figure 1-2. The LIS field-of-view and the orbital track of the TRMM satellite during a 24 hour period.
In November 1997, LIS joined the OTD in orbit when it was launched as a component of the TRMM. The LIS is basically the same instrument as OTD except that its sensitivity has been improved by a factor of three and it is in a 350 km altitude, 35° inclination orbit as shown in Fig. 1-2. The increased sensitivity results in a detection efficiency approaching 90%, while the lower orbit results in an improved spatial resolutions of between 3.9 and 5.4 km (3.9 at nadir and 5.4 at limb). However, the improved LIS spatial resolution is obtained at the cost of a decreased field-of-view of 580 × 580 km. Since the LIS travels around the Earth with a velocity greater than 7 km/s, the instrument can monitor individual storms and storm systems for lightning activity for 82 s (nominal), 80-85 s (typical) as it passes overhead. The LIS has a sampling rate of slightly greater than 500 frames per second resulting in a 2 ms time resolution. In addition to the lightning event data, the LIS also periodically captures an image of the background scene (~every 30-40 seconds) which is very useful for validation of navigation and processing algorithms.

A more detailed description of the LIS instrument, characteristics and operation and the requirements for the calibration facility are given in Section A.1.

1.2 MISSIONS

As noted previously, the LIS is a scientific payload on the TRMM. The TRMM platform was launched on November 28, 1997 and the mission is now expected to continue 4 to 5 years from launch (until 2001, 2002). The pre-launch LIS calibration and validation activities are reviewed in Section 3. The post-launch validation activities (both present accomplishments, activities underway and planned efforts) are described in detail in Sections 4. The OTD, the engineering prototype of LIS, was launched in April 1995 aboard a Pegasus rocket on the MicroLab-1 satellite. The OTD has both proved (validated) the LIS design concept and served as the initial test bed for the science data validation activities outlined in this document. The OTD mission is now drawing to a close. Therefore, validation efforts will now focused almost entirely on the LIS mission except in cases of cross calibration/normalization of LIS and OTD data sets.

1.3 SCIENCE DATA PRODUCTS

The basic science data product of LIS is lightning as described in the 1997 MTPE EOS Data Products Handbook [Wharton, 1997], 1999 EOS Reference Handbook [King, 1999], and the LIS Algorithm Theoretical Basis Document (ATBD) [Christian, 2000]. This product is comprised of several components, including: raw data (level 1-A), background image (level 1-B), events (level 1-B), groups (level 2), flashes (level 2), areas (level 2), one-second vector data (level 2), a browse image (level 3) and orbit statistics (level 3). A detailed description of these components, the LIS HDF data structure and the LIS data processing algorithms can be found in [Christian, 2000].
2. VALIDATION CRITERION

2.1 OVERALL APPROACH

In providing guidance on the content and format of the EOS instrument team validation plans, the EOS Panel on Data Quality has stated:

Validation is the responsibility of the *algorithm providers* (i.e., the LIS Science Team). It involves specifications of the transformations required to extract estimates of high-level geophysical quantities from calibrated basic instrument measurables and specification of the uncertainties in the high-level geophysical quantities. Validation requires detailed knowledge of the relationship between measurables and geophysical quantities of interest over the full range of possible conditions. Pre-launch activities include determination of algorithms and characterization of uncertainties resulting from parameterizations and their algorithmic implementation. Post-launch activities include refinement of algorithms and uncertainty estimates based on near-direct comparisons with correlative data and selected, controlled analyses.

The LIS Science Team is pursuing a validation plan consistent with this approach. *Validation*, in this document, is defined as the process of verifying and/or tuning the performance of the sensor performance parameters, data processing algorithms, and sensor hardware. *Calibration*, in this document, shall generally refer to any activity whereby a known sensor stimuli is used to determine the geophysical value of, judge or improve the sensor output.

The LIS validation will address *Sensor Performance*, *Data Processing Algorithms*, and *Scientific Retrieval Algorithms*. Observations of lightning from a broad range of lightning detector technologies shall be used to judge the correctness or quality of the LIS data, and the overall sensor and algorithm performance. The independent lightning observations (referred to as *ground-truth* or *truthing* data in this document) will be used to characterize existing random and/or systematic sensor errors that might exist. With knowledge of such errors, it will be possible to improve the LIS data product by removing a portion (perhaps most) of the possible error. In addition to cross-sensor truthing, significant validation is possible based upon verification of the internal consistency of standalone LIS data, and the data's consistency with both expected physics and the nominal functions of various applied noise filters. This approach is discussed in more detail in Section 4. Finally, while best estimates shall be made of true sensor performance based upon cross-sensor validation, some instances exist in which other sensors have comparable uncertainty (e.g., the OTD, some surface networks at far ranges). In these instances, relative cross-sensor normalization will be performed to allow the respective data sets to at least be intercomparable, and hence to both extend data baselines and bridge the gap between asynoptic TRMM sampling and continuous, ground-based monitoring.

*Sensor Performance*

The overall on-orbit performance parameters of LIS to be evaluated during validation include the following parameters: detection efficiency, false alarm rate, geolocation accuracy, signal
amplitude (background and event intensity) accuracy, timing accuracy, and bias, either spatial or temporal. The quality and variability of these parameters over time shall also be monitored. In addition the validation process will include remotely adjusting the LIS threshold settings to maximize detection and minimize 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 detections that are not attributable to lightning. These performance parameters may display significant dependence on the conditions under which the observations are obtained. These conditions include 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), geographical location, and the version of the data processing software employed. The effects of these conditions may be very interdependent and the responses nonlinear.

Data Processing Algorithms

In the TRMM pre-mission period, we have taken advantage of the space lightning observations provided by the April 1995 launch of the OTD. The LIS data processing algorithms described in detail in the ATBD [Christian, 2000] were initially validated by processing data from the OTD instrument, various ground-based lightning detecting systems and other ground truth observations (ground-, airborne-, and satellite-based). The OTD, being a prototype of the LIS instrument, detects lightning in a similar way and produces a similar data stream as the LIS.

Following its launch, the LIS data processing algorithm has been validated using data from LIS 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 periodically adjusted or tuned during the mission to provide the best correspondence with the observations. Based on analysis of the first two years of data, a major revision of the processing code (version 5) will be implemented in the second quarter of 2000 and a reprocessing of all the data collected to date will be made. These changes (and any subsequent changes) that are made in the LIS data processing algorithm will be reflected in updates to the ATBD.

Scientific Retrieval Algorithms

The LIS Science Team will participate in the development, testing, and validation of scientific retrieval algorithms that incorporate LIS data. The process will be similar to that described for validation of the data processing algorithms but will undoubtedly include modeling studies in addition to field observations.

2.2 Sampling Requirements and Trade-Offs

A broad range of ground-truth observations will be utilized to validate the LIS data. These include a variety of lightning measurement systems and networks as well as ancillary
observations (e.g., radar data and products, satellite data and products). The specific temporal, spatial, radiometric sampling requirements depend upon the particular LIS data process being validated. In some cases, the ground-truth lightning measurements provide only cloud-to-ground detections (e.g., National Lightning Detection Network (NLDN)) or have uncertain detection efficiencies (e.g., NLDN, long range sferics networks). In these cases, a cross normalization approach will be taken. Every effort will be made to sample different portions of the globe (during different seasons and time of day) with equivalent and/or similar ground truth observations in order to fully validate the performance of LIS. Broadly sampled data over extended time periods and wide geographical regions are particularly important for accurately characterizing the LIS detection efficiency. The validation of the accuracy of LIS timing, geolocation, and signal intensity, while fundamental to the scientific applications of LIS data, probably do not require as broad of a sampling requirement as that associated with detection efficiency validation.

2.3 Measures of Success

Validation shall be considered successful when it can be shown that correlative ground-truth data sets scientifically establish or prove the accuracy of the LIS data product. Following the launch of LIS, each facet of the LIS sensor performance and data processing algorithms shall be analyzed. These analyses shall be submitted for publication in scientific journals (such as the Journal of Geophysical Research). If these analyses, being independently reviewed by the standard peer-review process for scientific publication, are accepted for scientific publication in credible journals, the validation process shall be deemed successful.
3. PRE-LAUNCH ACTIVITIES

3.1 LABORATORY RADIOMETRIC TESTS

Optical laboratory experiments determined the radiometric response of LIS on a pixel by pixel basis. The pre-launch radiometric tests and procedures are described in detail in LIS Calibration Procedures Document [1995], and included: a D.C. uniformity and linearity test, field-of-view (FOV) test, A.C. response test, and a spectral test. These activities were completed for LIS in December 1996 prior to delivery of the instrument to GSFC for integration on the TRMM satellite. In order to conduct accurate radiometric tests, a class 10,000 clean room of the Space Sciences Laboratory, NASA-Marshall Space Flight Center, was used. Calibration of test instrumentation and traceability of test equipment back to secondary or primary standards was maintained at all times. In particular, the optical sources employed were traceable to National Bureau of Standards (NBS) maintained services or materials (e.g., NBS calibrated silicon photodetector in the spectro-radiometer system). The following summarizes each of the radiometric tests.

D.C. Uniformity and Linearity Test

The D.C. uniformity and linearity test involved exposing the entire LIS FOV to a steady, isotropic optical source (output from an 8” integrating sphere) and varying the source amplitude level. The D.C. response for each pixel was fully characterized in this test. The response of a pixel to various D.C. stimulus levels defines the pixel linearity. The variability in D.C. response from pixel to pixel is a measure of pixel uniformity.

Field-of-View (FOV) Test

In the FOV test set-up, the LIS was illuminated with a highly collimated light source whose azimuth and elevation incidence angles were precisely known relative to the LIS boresight (lens axis). An Euler angle analysis of LIS output data from this test provided a precise mapping between illuminated pixel and associated light source incidence angles. As such, this test is fundamental to the LIS lightning geolocation process. This test also determined the extremities of the LIS FOV.

A.C. Response Test

In the A.C. response test, a pixel was illuminated with a steady background signal while simultaneously stimulated with a transient optical pulse. By repeating this test for several pixels throughout the FOV, a very precise radiometric calibration of the transient piece-wise linear response of the Real-Time-Event-Processor (RTEP) of LIS was obtained.

Spectral Test

The narrow pass-band filter of LIS was scanned using a monochromator as part of the spectral test set-up. Center wavelength and full-width at half power were characterized in the LIS spectral test, as well as filter response as a function of off-boresight incidence angle.
3.2 LABORATORY PERFORMANCE TESTS

In this document, performance test refers to those tests that include, but are not limited to, determining the LIS lightning detection efficiency and false alarm rate. The lightning detection efficiency is defined as that fraction of flashes in the LIS FOV that are actually detected by LIS. The false alarm rate is defined as the number of false event detections by LIS per second. All performance tests were completed in the same (class 10,000) clean room described in mentioned section 3.1 above.

The A.C. response test described above provided an initial estimate of the LIS detection efficiency. The results of this test were correlated with lightning optical energy distribution statistics derived from U2 aircraft thunderstorm overflights [Christian, 1987, Goodman, 1988c]. With these data and knowledge of the lower limits of detectability from the A.C. test, an initial estimate of LIS detection efficiency on-orbit was obtained.

A preliminary indication of LIS false alarm rate was obtained by simply illuminating a region of the LIS FOV with the 8” integrating sphere that was used in the D.C. response test described above. For a given sphere output level, and a programmed LIS “threshold profile”, the number of false alarms could be determined. For example, a brighter sphere output corresponds to more shot noise, and a greater probability of LIS false alarm. By increasing the programmable LIS threshold values, false alarm rate is reduced to acceptable levels (i.e., < 10 %).

LIS performance was also evaluated in a limited way using a Lightning SIMulator (LSIM) optical test set-up that is described in in LIS Calibration Procedures Document [1995]. The simulator employed an acousto-optic modulator and a mirror scanner to externally modulate a laser light signal to generate simulated lightning transients. LSIM also employed a halogen lamp illuminated slide (of a satellite cloud field) to simulate a realistic background. In the LSIM test, LIS was illuminated by several thousand simulated lightning transient waveforms on top of various background brightness levels. The fraction of these events that were detected was logged, as well as the number of false event detections. However, the LSIM signal quality and stability was found not nearly as accurate as in the radiometric calibration tests.

3.3 LABORATORY TEST RESULTS

A formal and detailed description of radiometric and performance test results is provided in Koshak [2000a]. Standard radiometric response curve data derived from the laboratory tests were forwarded to the LIS Science Computing Facility (SCF) software development team for incorporation into the LIS Data Processing Algorithm.

In summary, it was found that the DC response of each pixel had linear correlation coefficients that were better than 0.98, and pixel non-uniformity varied by < 4.5%. The FOV was approximately square with angular dimensions 78.5° X 78.5° and the lens transfer function was highly linear over much of the FOV. The piecewise-linear nature of the LIS transient response was verified; the slope of the response curve for the first (most used) piece was about 0.5 millijoules per meter squared per steradian per nanometer per 7-bit event count. The spectral bandwidth of LIS was < 1 nm and within tenths of nanometers from the oxygen emission line.
triplet of the lightning spectrum. Prelaunch (laboratory) estimates of LIS lightning detection efficiency were 95% (about 25% better than the associated estimate obtained for OTD).

3.4 COMPLETED OTD VALIDATION STUDIES

The launch of OTD in April 1995 has provided the LIS Science Team with the unique opportunity to develop, test, and refine the procedures, analysis tools and investigations needed for LIS validation. A detailed empirical (cross-sensor) validation study of the OTD has been completed and is currently in press [Boccippio, 2000a]. This study included estimates of the intrinsic variance in reported OTD radiances due to the applied calibration technique, as well as flash localization (spatial and temporal) accuracy, and CG detection efficiency estimates based upon a large statistical intercomparison with the NLDN. As a result of this study, the radiance calibration technique was revised for LIS (documented in Boccippio, 2000b, in review) yielding more than a 50% reduction in calibrated radiance uncertainty. The algorithms utilized in the detection efficiency portion of this study are reusable and will be applied directly to LIS validation upon availability of v5 LIS data (see Section 4 for details on v4 vs. v5 data).
4. POST-LAUNCH ACTIVITIES

4.1 OVERVIEW

This section documents the motivation, scope and implementation of post-launch LIS validation activities. Section 4.2 describes the basic goals of post-launch validation, including assessment of LIS data integrity, flash counting, detection efficiency and false alarm rate, accuracy, bias, derived products, and inferred physical properties. Section 4.3 describes the actual validation activities, which can broadly be categorized as internal consistency checks, cross-sensor validation and cross-sensor normalization. Section 4.4 summarizes the mapping between objectives and activities, and outlines the current implementation schedule.

Some context is required for the current (and planned) state of LIS science data distribution. LIS “version 4” (v4) data files were distributed shortly after launch and continue to be the reference science data product. Validation activities described below identified specific weaknesses in the v4 data, including minor processing bugs, suboptimal tuning of some noise filters, and improved algorithms for other filters (many of these filters first required a statistically large data set to be collected on-orbit before baseline signal and noise properties could be established). A revised version of the LIS production code (v5) has been completed as of 2000/Q1 and an internal data set reprocessing is due by 2000/Q2. (This reprocessing is an implementation schedule-limiting factor for many of the activities described below; preliminary assessments have been performed using these techniques but final assessment is contingent upon the availability of two variants of the science data). Cross-comparison of the two data sets (using approaches described below) is planned for 2000/Q2-Q3, and a decision on whether to publicly release v5 data, or to implement an already-in-development v6 set of algorithms, is expected by 2000/Q4.

4.2 POST-LAUNCH VALIDATION GOALS AND OBJECTIVES

Seven broad categories of LIS sensor performance and science data assessment motivate the post-launch validation activities:

Data integrity

This most basic category includes basic correctness of the science data (HDF format) files, including verification that all reported data are within physical bounds and are correctly linked (in the clustering hierarchy from CCD pixel event, to “group” (lightning stroke), to “flash” (lightning flash, or contiguous set of channels), to “area” (loosely a thunderstorm cell). Additionally, verification that all sensor, platform, environmental and algorithm warning flags are set correctly (to allow appropriate data filtering during postprocessing and analysis) is a key component of this goal.

Flash counting

Since the LIS (like the OTD) has an intrinsic spatial resolution which is much larger than individual lightning channels (components of flashes) and may sometimes be on the scale of
individual storm cells, a clustering algorithm must be employed which makes best-possible guesses at how to assemble individual CCD pixel illuminations into nominal strokes, flashes and storm areas. Especially during high flash rate storms, this algorithm may be significantly challenged by the input data density and rate. For historical purposes and cross-sensor comparisons, the flash level clustering is of primary importance. Assessment of how correctly the LIS production code assembles pixel events into flashes is thus a key goal of post-launch validation activities. In particular, estimates of the frequency in which the algorithm incorrectly fragments (or amalgamates) truly contiguous (or disparate) channel structures are needed.

**Detection efficiency and false alarm rate**

As discussed in Section 3, estimates of the sensor lightning detection efficiency and false alarm rate are required to convert measured flash rates to absolute flash rates, to establish cross-calibrations with other lightning detection systems, and to estimate bounds on the minimum flash rate detectable during finite (80 second) TRMM overpasses of individual ground locations.

**Accuracy**

This includes the accuracy with which the LIS is able to localize actual lightning occurrences, both in space and time. Spatially, it is affected by the correctness of the transformation matrices used to map the mounted LIS position to geolocated earth coordinates, as well as the stability of the TRMM platform itself. Temporally, it is affected by the accuracy of the onboard TRMM clock, the internal LIS millisecond ‘tick’ counter, and the correct interaction between the two. Because the spatio-temporal resolution and desired accuracy of the LIS are both higher than most other TRMM sensors, independent confirmation of nominal lightning localization is required.

**Bias**

The LIS, like the OTD, is designed to be relatively bias-free, especially in comparison with most surface-based (RF) lightning networks, which often have intrinsic range- or physics-dependent biases. Nonetheless, two factors warrant more detailed examination of realized LIS biases. First, the use of variable thresholding (i.e., CCD sensitivity dependent on background radiance levels) to maximize detection efficiency introduces the possibility of a modest net day/night bias in the DE of the final data. Second, the high noise rates (both radiation and optical) in raw LIS observations require significant filtering to be performed in the LIS production code. Many of these filters utilize the known spatio-temporal clustering characteristics of true lightning illuminations (both at the sub-flash level and the storm level) to distinguish lightning from noise. However, since realized lightning exhibits a very wide dynamic range of characteristics (both total illumination and actual storm flash rates), the possibility exists that these filters may preferentially eliminate either lightning with low information content (few pixel illuminations) or storms with low flash rates (few overlaid flashes). This effect only translates to a bias if it covaries with coherent regional or physical variability in the underlying lightning spectrum. Specific examples would be high optical depth storms (which might preferentially attenuate many candidate pixels below the LIS threshold) or low flash rate storms (as might be expected over open ocean). While this bias is difficult to quantify with cross-sensor
comparisons, it can be mitigated significantly by robust design of the noise filters themselves (i.e., design which keeps the possibility of bias in mind).

Derived products

LIS science data include several ‘second-order’, or derived, quantities, including radiance and spatial footprint. Verification of these derived quantities is thus an important (albeit secondary) goal of post-launch validation.

Inferred physics

This goal begins to bridge the divide between validation studies and science application. It includes assessment of the robustness of various LIS observables (e.g., the scientific usefulness of group-level vs. flash-level aggregate products, the meaningfulness of area-level “storm/cell” delineations, the physical utility of measured radiance). It also includes assessment of physical inferences from LIS data (e.g., assessment of storm electrical energetics based on counted flashes). Better understanding of the robustness of LIS data will help constrain future analyses and guide broader scientific application of the data (e.g., to storm microphysics or dynamics).

4.3 POST-LAUNCH VALIDATION ACTIVITIES

The goals discussed above are addressed through specific post-launch validation activities, described in this section. These post-launch validation activities may be classed into three broad categories. *Internal consistency checks* include basic assessment of the science data integrity and physical consistency of the applied algorithms. This class also includes heuristic analyses which “bootstrap” unknowns from the science data set itself. *Cross-sensor validation*, of course, utilizes existing surface validation sensors, platforms or networks to provide objective truth at resolution higher than and accuracy greater than that of the LIS sensor itself. *Cross-sensor normalization* includes determining normalization adjustments to integrate LIS data with earlier (and concurrent) OTD data (hence extending the baseline of tropical lightning measurements). It also includes use of the validated LIS data to calibrate surface-based instruments which may have unknown range-dependent detection efficiency. This activity thus helps extend (locally) total lightning estimates past the TRMM end-of-mission. Also, since many of these surface sensors are continuously monitoring, they help bridge the gap between asymptotic (and effectively instantaneous) LIS flash rate measurements and continuous storm (and lightning) evolution. This in turn helps establish demonstration studies for possible future continuous (geostationary) space-based lightning detection missions.

**INTERNAL CONSISTENCY CHECKS (Bootstrapped Validation)**

*Manual and automatic quality assurance (QA)*

Each LIS science data (HDF) file undergoes both an automatic quality assurance check (testing for values-in-bounds, etc.) and a manual (man-in-the-loop) examination. This process guarantees basic data set integrity. The process also ensures that known sensor, platform, environment, and algorithm issues (e.g., platform maneuvers, SAA noise, etc) are properly
flagged in the science data, and that new or unexpected variations in sensor performance do not propagate into the distributed data (as yet no significant such variations have been documented).

**Algorithm tuning**

Many instances exist where the v4 LIS data (and rejected noise) can be used to bootstrap estimates of algorithm accuracy and robustness. One significant example involves assessment of the accuracy of the lightning clustering algorithm. By examining the statistical likelihood of true (LIS-identified) flashes to occur concurrently within the field of view (but at spatial separations large enough not to challenge the algorithm), a baseline estimate can be derived which predicts whether the LIS clustering algorithm is significantly incorrectly fragmenting or amalgamating true flashes at scales close to the pixel resolution. This is a zeroth-order estimate, but can identify the most glaring inconsistencies and provide upper bounds on severity of such algorithmic error. The approach has been successfully tested on OTD data [Boccippio, 2000a] and will be used as one benchmark for v4 vs. v5 production code assessment (expected completion 2000/Q2).

A second example of bootstrapped validation includes examination of the logical consistency of applied noise filters. The heuristic (but physically based) tests here examine the consistency of filter performance; e.g., radiation noise filters should reject (outside the SAA) a geographically uniform “D/C” noise field, optical artifact filters should be inactive at night, and the subset of true lightning flashes likely to be falsely rejected by optical artifact filters can be estimated using nighttime data. These approaches have already been used to implement the v5 production code, and further investigation using such approaches have led to recommended v6 algorithms (not yet implemented in a full production code suite).

**Statistical distributions.**

A further implementation of the internal / bootstrap approach allows estimates of sensor bias in either geographic regions (the SAA) or temporal subsets (daytime scenes) in which bias is either expected or suspected. Within the SAA, adaptive noise filters run ‘more aggressively’ due to the higher ambient radiation noise rates. The net effect is to selectively eliminate low-information-content flashes and low-flash-rate storms (which are indistinguishable from radiation noise). The incremental reduction in net detection efficiency will be estimated (to zeroth order) by examining the overall distribution of either lightning optical properties or storm flash rates both within and outside the SAA; the relative “depletion” of low-information-content lightning within the SAA serves as an estimate of algorithm impact (local bias). This bootstrap approach invokes an assumption that true geographic variability is secondary to algorithm effects within this severely noise-impacted region.

A second application of this approach is assessment of day/night variability (bias) in flash detection efficiency (due to background radiance-dependent variable thresholding in the LIS). Here, the full dataset will be postprocessed and an artificial threshold (corresponding to peak daytime thresholds) will be applied. This will serve as a zeroth-order estimate of nighttime to daytime detection efficiency loss (i.e., based on the amount of nighttime data rejected). The estimate will be imperfect as (1) as low-amplitude events are removed from individual flashes, the possibility will arise that these flashes would not have survived noise filtering in the full production code, (2) true day/night variability in flash optical properties may exist in nature.
 Nonetheless, it will provide a baseline estimate which can be further confirmed (or refuted) by cross-sensor validation, as described below. The approach has already been used to identify an optical software filter bug which effectively negated the benefits of variable thresholding in the v4 data (i.e., nighttime and daytime flash DE were effectively made the same by this bug).

**Cross-Sensor Validation**

*Surface TOA/DF CG detection networks.*

A number of regional or local RF surface lightning detection networks exist which locate CG lightning using time-of-arrival and direction-finding approaches at low frequency. These include the National (now North American) Lightning Detection Network (NLDN or NALDN) operated by Global Atmospherics, Inc. (GAI), which covers the CONUS, the GAI Long Range Network, which utilizes the same receivers but applies more tolerant detection algorithms to provide offshore data, and the NASA/MSFC-operated Brazilian Lightning Detection Network (BLDN) deployed in Rondônia under TRMM/LBA (which will operate in that location for at least two years). Each of these networks provides important validation data for LIS.

The NLDN is a robust network, with stationary configuration and continuous operation during the TRMM mission. Analytic estimates of NLDN detection efficiency are available as well as a more limited subset of ground truthed DE estimates. The NLDN data have been used to estimate the CG detection efficiency of the OTD [Boccippio, 2000a] as well as its location accuracy and day/night bias. The same analysis will be performed on LIS data upon v5 data reprocessing (for both v4 and v5 data sets). The large size of this truthing data set will allow quantitative estimation of LIS DE as a function of exact threshold (as opposed to a net day/night bias); this can be compared with bootstrapped estimates as described above. As with the OTD-NLDN analysis, the accuracy of these results will be fundamentally limited by: (1) the fact that NLDN detects CGs only, and (2) the ambiguity in pairing optically observed flashes (which occur over a finite time window) and NLDN observed flashes (which are assigned a single nominal time corresponding to the first return stroke).

The GAI Long Range network is an experimental implementation of alternative detection algorithms using the same receivers as the NALDN. It exhibits a rapid (nearly exponential) drop-off in sensitivity with range-to-network, and is thus of limited utility for statistical DE validation. However, it should exhibit a negligible FAR, and thus can be used to explicitly identify false negatives in the LIS data set in individual storms (possible candidate flashes which may have been improperly rejected by the noise filters). Preliminary analyses of this network have been performed using OTD data, identifying and quantifying known issues with differential day/night sensitivity (due to ionospheric height) and land/ocean sensitivity (due to surface scattering). With this knowledge in place, the long range data can be appropriately subsetted to generate internally consistent validation data sets for comparison with LIS data (i.e., enough knowledge is now available to properly design controlled validation experiments using this network).

The BLDN is a shorter range network of four receivers. Its performance should be nearly constant over approximately one-two network diameters. Efficient retrieval have been specially
developed to optimize the flash/stroke location solutions from small network configurations such as employed in the BLDN [Koshak, 2000b]. Solutions can be obtained using Time-of-arrival (TOA) and magnetic direction finding techniques with as few as two sites detecting an event. During 1999, site corrections for each of the receivers were derived which will enable much more accurate location of observed flashes; additionally, a database of network/receiver uptime and performance were constructed for the deployment to date. Real-time and archival products have been generated merging BLDN, GAI-Long Range, LIS and GOES-IR data. A reprocessing of the entire BLDN data set to date is planned for 2000/Q2. Once available, reprocessed BLDN data will enable a confirmation of NLDN-derived results in a different convective regime. Additionally, the BLDN exists on the periphery of the effective SAA region for LIS; this region is a “ramp-up/ramp-down” zone for the adaptive noise filters described above. The BLDN can thus serve as one of the few truthing sources for the SAA region, and for intercomparison against bootstrapped estimates of bias introduced by the adaptive noise filters.

Additional surface TOA/CG networks include the ALDF network deployed by NASA/MSFC near Darwin, Australia and the Australian equivalent of the NLDN. The above analyses may be extended to include these data on an as-needed and as-capable basis.

**Surface VHF/TOA IC+CG detection networks.**

Total (intracloud and cloud to ground) lightning can be mapped using time-of-arrival networks operating at VHF frequencies. These networks typically have stable flash detection efficiencies at short range (up to 100 km from the receivers) [Boccippio, 2000c,d]. Two such networks are operational in the United States (the NASA LDAR network, permanently installed at KSC, and the New Mexico Tech LMS network, deployed in field programs of opportunity). A third (the NASA LMA, being deployed for MSFC in the Tennessee Valley region) will be operational by 2000/Q2, a fourth (an operational prototype by GAI) will be eventually be deployed at the Dallas/Ft Worth Airport, and a fifth (again constructed by NMT) will be deployed near the National Severe Storms Laboratory in Oklahoma.

At close range to these networks, total lightning DE is nearly 100%, and these systems thus provide some of the best direct truth for LIS available. Analysis is slightly more complicated than with NLDN data as automated flash separation (clustering) algorithms have not yet been validated for VHF/TOA data. In addition to DE, cross-sensor comparisons with these networks provide estimates of LIS location accuracy, and allow validation of the LIS calculated flash footprints. Finally, since these networks provide 3-D flash channel mapping, they can indirectly contribute to understanding of optical attenuation in the cloud (i.e., better quantification of what specific channel components LIS optimally sees).

A complete diagnostic analysis of the KSC LDAR network has recently been performed [Boccippio, 2000c,d]. This preliminary analysis has enabled an assessment of the maximum useful ranges for cross-sensor comparisons, and a better understanding of the network-specific issues. As with the GAI-LR network, this “validation of the validation sensor” approach now allows LIS validation experiments to be designed with greater control and accuracy. Overpass-by-overpass browse products of LIS, NLDN, LDAR and NEXRAD data are routinely generated by the GHRC. LDAR channel maps have already been used to provide preliminary estimates of
LIS location accuracy, and have been combined with EFM charge center retrievals (see below) in validation studies.

Detailed cross-comparison with the NMT network (when deployed in Oklahoma, 1998) has been performed by [Thomas, 2000]. This analysis demonstrated the tendency of LIS to preferentially detect the upper branches of CG and IC flashes, narrowed location accuracy estimates to about 5 km, and allowed an assessment of the incremental DE losses introduced by both v4 and v5 LIS data (in comparison with raw, unfiltered data).

Most extensive VHF/TOA-LIS intercomparisons will be performed with the MSFC LMA network, currently under construction and operational in 2000/Q2. With direct control over network operations and direct access to raw data, much more highly controlled intercomparisons will be undertaken both on a case-by-case and statistical basis. The LMA network will undergo shakedown and validation from 2000/Q2-2000/Q3, and a large enough data set for cross-sensor comparisons should be available by 2001/Q3.

**Surface Electric Field Mill network**

The surface electric field mill network at NASA/KSC provides dipole or tripole electric charge center locations and amounts of charge transferred for individual flashes, within a very short (10-20 km) radius of the network. This network thus provides some of the only quantitative information on flash energetics available from a LIS truthing sensor (the only other exception is the ER-2 LIP, below). This information (in combination with more highly resolved LDAR channel mapping) has direct relevance to physical inferences on storm electrical energetics from LIS flash counts (the storm electrical energetics provide the coupling between lightning flash counting and the underlying storm microphysics/dynamics). Complete EFM flash inversions have been performed for a limited number of EFM/LDAR/NLDN/LIS observed flashes, and a processing algorithm & interface software have been developed to rapidly process new overpass data. This cross-sensor database will be continued through the TRMM mission and statistical results summarized at end-of-mission (many LIS overpasses are required to view a statistically interesting number of flashes within the comparatively small FOV of the EFM network).

**Aircraft studies (e.g., ER-2, DC-8, North Dakota Citation, UAV)**

We have compared LIS-derived event amplitudes with the existing lightning optical energy statistics derived from NASA U2 aircraft flights above thunderstorms in the 1980's. This has helped verify that LIS results are indeed within reasonable bounds. We will continue to use the ER-2 Lightning Instrument Package (LIP) or similar packages on other aircraft to provide validation for the LIS sensor and investigate lightning-storm relationships. The validation data sets obtained during the airborne field campaigns will be used to confirm the on-orbit performance of the LIS instrument, the data processing algorithms and the scientific retrieval algorithms.

Measurements above thunderstorms of lightning and storm electrification are still needed to develop, test, and refine relationships dependent on total lightning observations. Ground-based measurements of lightning are often not sufficient to support these efforts because the ground-based lightning location networks either only detect cloud-to-ground lightning or detect intracloud
lightning with unknown detection efficiency. Detecting intracloud lightning is important since intracloud discharges often represent the dominant contribution to total lightning. The ER-2 lightning instrumentation will detect total storm lightning and differentiate between cloud-to-ground and intracloud discharges. In many cases, the ER-2 provides the only lightning and electrical measurements that are available for the storms overflown.

Until now, the ER-2 aircraft has provided the best means to fly above thunderstorms since the ER-2 can be easily and quickly vectored over regions of interest. The ER-2 lightning instrumentation is flown with other sensor systems to provide unique and detailed observations of storm and precipitation structure as well as moisture and temperature fields in the surrounding environment. Besides the lightning instrumentation, the ER-2 storm payload usually includes infrared and visible sensors, passive millimeter and microwave radiometers, Doppler radar, and interferometers. Since the ER-2 platform provides a cloud top perspective similar to that viewed by a space sensor, the multi-parameter data sets obtained during the field experiments will support detailed satellite simulations of storm measurements and specifically provide important validation and calibration of TRMM. In addition, comprehensive data sets using a variety of ground-based, aircraft, balloon-borne, and satellite platforms will be collected during these experiments. During the latter period of the TRMM mission, a storm instrument package may be flown on a UAV aircraft.

**Cross-Sensor Normalization**

**OTD-LIS cross-normalization**

Concurrent (albeit asynoptically undersampled) tropical lightning observations from the OTD sensor are available from 1997/Q4-2000/Q1 (after which the OTD mission was terminated). This data sample is too small to allow detailed cross-calibration of the two sensors’ detection efficiencies at anything less than the global (tropical) composite scale (i.e., the stochastic underlying lightning population is too undersampled to allow quantitative comparison of either detailed maps or highly time resolved seasonal cycles). Nonetheless, this bulk adjustment factor will serve as a starting point for merging the two data sets (to extend the tropical record back to 1995, or equivalently to extend the OTD tropical record forward to TRMM end-of-mission). It will be used as a benchmark against which to check independent direct (cross-sensor validation based) detection efficiency estimates for each sensor. In theory, these direct estimates should yield the same net result as the cross-calibration; in practice, sampling-induced variance may complicate interpretation. Such “climatological” cross-calibration will be performed using both v4 and v5 LIS data as soon as reprocessed v5 data are available.

**OTD-LIS-GAI/LR cross-normalization**

This activity is only indirectly related to LIS validation. The OTD and LIS sensors can be used to provide a climatologically-based estimate of the GAI Long Range network’s (unknown) detection efficiency vs. range, and as a function of day/night or land/ocean RF signal transit paths. This activity has already been completed using OTD data and will be re-verified using LIS data. By providing a spatio-temporally varying DE normalization factor for the GAI-LR network, the spatial domain over which continuous (albeit higher variance) lightning observations are available
is greatly increased. The indirect relevance to LIS validation arises from the greatly improved
capacity to now represent the true spatio-temporal stochastic variability (especially over oceanic
domains) of the “true” lightning distribution, which LIS subsamples dramatically. This will
enable indirect variance estimates for regional LIS estimates of total flash rate, and perhaps storm
cell flash rates, and will allow investigation of the effects of aliasing of travelling (lightning
producing) synoptic wave disturbances with the precessing TRMM orbit. The approach is thus:
(1) Use OTD and LIS bulk data to assess the GAI-LR network performance, (2) “calibrate” the
continuous GAI-LR network data, (3) use the continuous record for variance studies. Steps (1)
and (2) have been completed as of 2000/Q1; step (3) is anticipated to be completed by TRMM
end-of-mission.

LIS-VHF/TOA cross-normalization

As with the GAI-LR network, until recently surface VHF-TOA total lightning mapping
networks had essentially unknown detection efficiency as a function of range. As discussed
above, these performance characteristics have now been assessed both analytically and
empirically (using LIS data). From the empirical cross-comparison, the predicted maximum
ranges at which VHF-TOA networks maintain nearly constant detection efficiency have been
verified. Since this is somewhat farther range (90-100 km for the KSC LDAR) than
conventionally and conservatively assumed, the domain over which detailed case-study cross-
comparisons can be performed is greatly increased. The approach is thus: (1) Use theory to
predict VHF-TOA network flash DE vs. range, (2) Use LIS to confirm the maximum range at
which VHF-TOA network flash DE is stable (effectively invariant at near-100%), (3) Use the
expanded cross-comparison domain to gather a larger sample of cross-sensor LIS validation
data, as discussed above. Steps (1) and (2) have been completed as of 1999/Q2; step (3) will be
completed upon reprocessing of the v5 LIS data.
5. IMPLEMENTATION OF VALIDATION RESULTS IN DATA PRODUCTION

5.1 APPROACH

The validation data sets are being used as confirmation of the on-orbit performance of LIS and the algorithms. Corrections and refinements to the algorithms will be implemented with periodic reprocessing of the data planned at intervals throughout the mission commensurate with overall TRMM plans. As noted in Section 4, the LIS “version 4” (v4) data files were distributed shortly after launch and continue to be the reference science data product. A reprocessing (v5) is planned for the first quarter of calendar year 2000. Following the reprocessing, the products will be initially analyzed and quality assured by the LIS science team members to determine whether the v5 data will be deemed acceptable for release to the scientific community or whether a v6 code revision will be necessary.

5.2 ROLE OF LIS SCF AND EOSDIS

The role of the LIS SCF and EOSDIS is to ingest, process, archive and distribute the massive amounts of Earth science data that will be collected throughout the next decade. EOSDIS anticipates that over 300 terabytes of data will be archived during these years.

The initial paradigm was that data would be collected and archived at Distributed Active Archive Centers (DAACs). However, experience with the EOSDIS program has led to the proposal of a new direction for information management of the EOS data in the form of a more distributed federation of information service providers and their users.

As one of the first members of the EOSDIS Federation, the LIS SCF is a precursor to the Earth Science Information Partners (ESIP) concept, which calls for a strong working partnership between a data center and a science team in order to provide support to the research community for a specific set of products. With a goal of focusing data management around primary science research areas, the LIS Science Computing Facility (SCF) will augment its original roles of science research and algorithm production with the additional roles of such data center functions as producing, archiving and distributing data. This effort provides continued critical science data support within the Global Hydrology and Climate Center (GHCC) for lightning and ancillary data. This "enhanced" LIS SCF represents a new direction for EOSDIS, namely, with science data management and operations under the control of the instrument Principal Investigator.

An important aspect of data distribution is providing user accessibility to the data through common electronic means. The LIS SCF utilizes the EOSDIS Version Zero (V0) Information Management System protocols. All LIS SCF data sets available to the public are visible and may be ordered through EOSDIS Data Gateway web interface. In addition, the LIS SCF is a member of the Federation of Earth Science Information Partners (as an ESIP-1) and is collaborating with other ESIPs to further the dissemination of data and products to a wider audience.
The LIS SCF will provide users on-line data access through locally developed World Wide Web (WWW) data search and order applications and dataset-specific "home pages". It will build on Marshall Space Flight Center's (MSFC) EOSDIS Version 0 heritage by providing on-line access to its data through EOSDIS search and order systems as well. Data catalog interoperability with the ECS is also planned, although the LIS SCF will not have the full complement of ECS Version 1 hardware and software. The LIS SCF will have to address many on-line data access issues including bandwidth limitations, multimedia displays, definition and generation of both test data and other sample data, and interoperability with other systems. The experience gained as a partner in the EOS Baseline Federation will provide lessons learned for other SCFs and science data producers as the EOSDIS Federation grows.

5.3 PLANS FOR ARCHIVAL OF VALIDATION DATA

The LIS SCF will archive data retrieved from the Lightning Imaging Sensor, various lightning on-orbit calibration and validation data sets, and ancillary data. Table 5-1 lists the planned calibration and validation data set holdings of the LIS SCF. This list should not be considered exhaustive, and is likely to be expanded. The most current information about these data sets can be accessed via the home page: http://thunder.msfc.nasa.gov. Additional data sets may be obtained on an as needed basis through collaboration with other scientists.

<table>
<thead>
<tr>
<th>Lightning Data</th>
<th>Source</th>
<th>Period of Record</th>
<th>Annual Vol</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIS Raw Data</td>
<td>LIS Science Team</td>
<td>11/97 - present</td>
<td>38.3 GB</td>
</tr>
<tr>
<td>LIS Lightning Products</td>
<td>LIS Science Team</td>
<td>11/97 - present</td>
<td>22.2 GB</td>
</tr>
<tr>
<td>OTD Raw Data</td>
<td>LIS Science Team</td>
<td>4/95 - present</td>
<td>22.2 GB</td>
</tr>
<tr>
<td>OTD Lightning Products</td>
<td>LIS Science Team</td>
<td>4/95 - present</td>
<td>11.7 GB</td>
</tr>
<tr>
<td>OLS Analog Derived Lightning</td>
<td>NGDC / LIS Science Team</td>
<td>6/73 - 12/91 (intermittent)</td>
<td>0.01 GB</td>
</tr>
<tr>
<td>Lightning Ground Strike data (restricted)</td>
<td>GAI NLDN</td>
<td>1/94 - present</td>
<td>0.3 GB</td>
</tr>
<tr>
<td>US Composite Lightning Products from NLDN (rstr)</td>
<td>LIS Science Team</td>
<td>1/94 - present</td>
<td>0.3 GB</td>
</tr>
<tr>
<td>GAI Long Range Lightning (restricted)</td>
<td>GAI</td>
<td>8/96 - present</td>
<td>1.3 GB</td>
</tr>
<tr>
<td>LIS Ground Truth (Aircraft, miscellaneous)</td>
<td>LIS Science Team</td>
<td>11/95 - present</td>
<td>0.4 GB</td>
</tr>
<tr>
<td>LDAR</td>
<td>KSC / Lincoln Labs</td>
<td>1997 - present</td>
<td>1.0 GB</td>
</tr>
<tr>
<td>Alabama Lightning Mapping Array (LMA)</td>
<td>LIS Science Team</td>
<td>4/00 - present</td>
<td>est. 100.0 GB</td>
</tr>
<tr>
<td>Brazil Lightning Detection Network (BLDN)</td>
<td>LIS Science Team</td>
<td>1/99 - present</td>
<td>0.3 GB</td>
</tr>
<tr>
<td>Field Mill</td>
<td>KSC</td>
<td>Xx - present</td>
<td>25.0 GB</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Radar Data</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>US Composite Rainfall Products from NWS Radars</td>
<td>WSI / LIS Science Team</td>
<td>1/94 - present</td>
<td>14.6 GB</td>
</tr>
<tr>
<td>Melbourne US Nexrad site</td>
<td>NWS / Lincoln Labs</td>
<td>4/95 - present</td>
<td>13.0 GB</td>
</tr>
<tr>
<td>Hytop US Nexrad site</td>
<td>NWS / LIS Science Team</td>
<td>2/00 - present</td>
<td>est. 100.0 GB</td>
</tr>
</tbody>
</table>

Table 5-1. LIS SCF Calibration and Validation Data Holdings. The data holdings are held in compressed format (providing about a factor of 3.5 in overall compression)
6. SUMMARY AND IMPLEMENTATION SCHEDULE

Figure 6-1 illustrates the conceptual mapping between class of validation activity, specific activity, and end validation goals. The important feature of this graph is its redundancy: each validation goal is achieved not only through multiple activities, but also spanning both main categories of activity (while the most important goal, detection efficiency assessment, is achieved through activities from all three categories). This ensures both high confidence in final goal assessments and high return from each validation activity.

Table 6-1 illustrates the planned implementation schedule for all validation activities. Most activities have already passed the “preliminary analysis” stage, have well established analysis methodologies and algorithms prepared, and await only v5 re-processed data. Publication of the results of these activities will be contingent on the v5 data analyses and whether a v6 code revision is deemed necessary (or whether the v5 data will be deemed acceptable for community release).
Figure 6-1. Conceptual mapping between class of validation activity, specific activity, and end validation goals.
<table>
<thead>
<tr>
<th>Validation Activity</th>
<th>Validation Yield</th>
<th>Data Window</th>
<th>Preliminary Analysis</th>
<th>Full Study Possible</th>
<th>Expected complete</th>
</tr>
</thead>
<tbody>
<tr>
<td>QA</td>
<td>error-free files</td>
<td>97/Q4-EOM</td>
<td>97/Q4</td>
<td>ongoing</td>
<td>EOM</td>
</tr>
<tr>
<td>Clustering algorithm</td>
<td>flash counting accuracy</td>
<td>97/Q4-EOM</td>
<td>99/Q1</td>
<td>00/Q1</td>
<td>00/Q3</td>
</tr>
<tr>
<td>Filter consistency</td>
<td>higher DE, lower regional bias</td>
<td>97/Q4-EOM</td>
<td>98/Q3</td>
<td>00/Q1</td>
<td>00/Q4</td>
</tr>
<tr>
<td>SAA distributions</td>
<td>estimate of DE loss in SAA</td>
<td>97/Q4-EOM</td>
<td>00/Q2</td>
<td>00/Q1</td>
<td>00/Q4</td>
</tr>
<tr>
<td>D/N distributions</td>
<td>estimate of DE loss in daytime</td>
<td>97/Q4-EOM</td>
<td>00/Q1</td>
<td>00/Q1</td>
<td>00/Q4</td>
</tr>
<tr>
<td>NLDN</td>
<td>land CG DE, location accuracy</td>
<td>97/Q4-EOM</td>
<td>96/Q4</td>
<td>00/Q1</td>
<td>00/Q3</td>
</tr>
<tr>
<td>NLDN-LR</td>
<td>oceanic CG false negative rate</td>
<td>97/Q4-99/Q1</td>
<td>99/Q4</td>
<td>00/Q1</td>
<td>00/Q3</td>
</tr>
<tr>
<td>BLDN</td>
<td>tropical CG DE, SAA edge analysis</td>
<td>99/Q1-EOM</td>
<td>99/Q3</td>
<td>00/Q1</td>
<td>01/Q1</td>
</tr>
<tr>
<td>LDAR</td>
<td>land IC+CG DE, location acc, footprints, attenuation</td>
<td>97/Q4-99/Q2</td>
<td>99/Q2</td>
<td>00/Q1</td>
<td>01/Q1</td>
</tr>
<tr>
<td>LMS</td>
<td>filter analysis, footprints, emission physics</td>
<td>98/Q3,99/Q3</td>
<td>99/Q1</td>
<td>00/Q1</td>
<td>00/Q4</td>
</tr>
<tr>
<td>LMA</td>
<td>land IC+CG DE, location acc, footprints, attenuation</td>
<td>00/Q2-EOM</td>
<td>00/Q4</td>
<td>00/Q1</td>
<td>01/Q4</td>
</tr>
<tr>
<td>EFM</td>
<td>inferred physics, false positive, false negative</td>
<td>97/Q4-EOM</td>
<td>99/Q2</td>
<td>00/Q1</td>
<td>01/Q4</td>
</tr>
<tr>
<td>ER2</td>
<td>inferred physics, false positive, false negative</td>
<td>97/Q4-EOM</td>
<td>98/Q2</td>
<td>00/Q1</td>
<td>EOM</td>
</tr>
<tr>
<td>OTD</td>
<td>long-term (1995-EOM) calibration</td>
<td>97/Q4-00/Q2</td>
<td>99/Q2</td>
<td>00/Q2</td>
<td>00/Q4</td>
</tr>
</tbody>
</table>

Table 6-1. Summary of validation activities and tentative implementation schedule.

All quarterly dates are for calendar year. **Validation activity** indicates the category of investigation discussed earlier in this section (grouped into the three main areas, Internal Consistency, Surface/Airborne Sensors and Cross-Satellite Calibration. **Validation yield** indicates the primary benefit of the activity. **Data window** indicates the period of available validation data; EOM indicates End-of-Mission. Data window is constrained either by LIS data window, validation sensor deployment, or non-stationarity in validation sensor dataset (e.g., configuration changes). **Preliminary analysis** includes date or expected data of primary validation algorithm development, validation sensor diagnostics, or preliminary intercomparisons (used to refine and tune the LIS production code). **Full study possible** date is when a full, statistically robust analysis was or will be possible; in most instances this is set by the date of first internal LIS data reprocessing (00/Q1). **Expected complete** dates are tentative and contingent on level of effort available within the LIS team, as well as future field deployments of opportunity.
7. REFERENCES


(Also web access at http://eospso.gsfc.nasa.gov/eos_homepage/scipubs.html)

Driscoll, K. T., A time-averaged analysis of the electric currents in the vicinity of isolated thunderstorms, Ph.D. dissertation, Auburn University, 1993.


(Also web access at http://eospso.gsfc.nasa.gov/eos_homepage/scipubs.html)


(also web access at http://eospso.gsfc.nasa.gov/eos_homepage/scipubs.html)


APPENDIX

A.1 INSTRUMENT DESCRIPTION AND CHARACTERISTICS

The LIS is a small, solid state optical imager that detects lightning from low Earth orbit with high detection efficiency and location accuracy, marks the time of occurrence, and measures the radiant energy. 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 comprise the major elements of the sensor. Physically, these optical and electrical elements are combined into a cylindrical sensor head and telescope assembly (20 × 37 cm) and an electronics assembly (31 × 22 × 27 cm) as shown in Fig. A-1. The LIS has a total mass of approximately 20 kg, consumes less than 30 W of power, and generates a telemetry data rate of only 8 kb/s. The wide angle lens in the LIS imaging system requires an 80º × 80º nadir field-of-view for unobstructed observations of the thunderstorms being overflown. Table A-1 summarizes the overall instrument parameters and performance criteria. Table A-2 summarizes the LIS sensor design requirements and calibration facility requirements that were based on the measured properties of cloud-top lightning [Christian, 1987; Goodman, 1988] and the background signals.

Being a staring imager, the LIS views a scene much like a television camera. However, the actual data handling and processing are much different than that required by a simple imager because of the transient nature of lightning, its spectral characteristics, and the difficulty of daytime detection of lightning against brightly-lit cloud backgrounds. In order to achieve the

Table A-1. LIS Parameters and Performance Criteria.

<table>
<thead>
<tr>
<th>Field-of-View (FOV)</th>
<th>80º × 80º</th>
<th>Measurement Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pixel IFOV (nadir)</td>
<td>4 km</td>
<td>location</td>
</tr>
<tr>
<td>Interference Filter</td>
<td></td>
<td>intensity</td>
</tr>
<tr>
<td>wavelength</td>
<td>777.4 nm</td>
<td>time</td>
</tr>
<tr>
<td>bandwidth</td>
<td>1 nm</td>
<td>tag at frame rate</td>
</tr>
<tr>
<td>Detection Threshold</td>
<td>4.7 J m⁻² sr⁻¹</td>
<td>Dimensions</td>
</tr>
<tr>
<td>Signal to Noise Ratio</td>
<td>6</td>
<td>sensor head assembly</td>
</tr>
<tr>
<td>CCD Array Size</td>
<td>128 × 128 pixels</td>
<td>electronics box</td>
</tr>
<tr>
<td>Dynamic Range</td>
<td>&gt; 100</td>
<td>20 × 37 cm</td>
</tr>
<tr>
<td>Detection Efficiency</td>
<td>~ 90 %</td>
<td>Weight</td>
</tr>
<tr>
<td>False Event Rate</td>
<td>&lt; 5 %</td>
<td>20 kg</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Power</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30 Watts</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Telemetry</td>
</tr>
<tr>
<td></td>
<td></td>
<td>data rate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8 kb/s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>format</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PCM</td>
</tr>
</tbody>
</table>
Table A-2. Sensor design and calibration requirements.

<table>
<thead>
<tr>
<th>SENSOR DESIGN REQUIREMENTS</th>
<th>CALIBRATION REQUIREMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detection efficiency</td>
<td>&gt; 90% of all events</td>
</tr>
<tr>
<td>False alarm rate</td>
<td>&lt; 10% of total events</td>
</tr>
<tr>
<td>Background</td>
<td></td>
</tr>
<tr>
<td>Radiance</td>
<td>331 W/m²/ster/µm</td>
</tr>
<tr>
<td>Accuracy</td>
<td>10%</td>
</tr>
<tr>
<td>Lightning signal</td>
<td></td>
</tr>
<tr>
<td>Threshold</td>
<td>4.7 µJ/m²/ster</td>
</tr>
<tr>
<td>Dynamic range</td>
<td>&gt; 100:1</td>
</tr>
<tr>
<td>Accuracy</td>
<td>10%</td>
</tr>
<tr>
<td>Spectral</td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.1 nm</td>
</tr>
<tr>
<td>Resolution (FWHM)</td>
<td>1.0 ± 0.1 nm</td>
</tr>
<tr>
<td>Spatial</td>
<td></td>
</tr>
<tr>
<td>Wavelength</td>
<td>777.4 nm</td>
</tr>
<tr>
<td>Resolution</td>
<td>0.625° IFOV</td>
</tr>
<tr>
<td>Angular range</td>
<td>80° x 80° FOV</td>
</tr>
<tr>
<td>Temporal</td>
<td></td>
</tr>
<tr>
<td>Resolution</td>
<td>2.0 ms</td>
</tr>
<tr>
<td>Accuracy</td>
<td>2.0 ms</td>
</tr>
</tbody>
</table>

performance goals required to meet the scientific objectives, the LIS combines off the shelf components in a unique configuration. The design employs an expanded optics wide field-of-view lens, combined with a narrow-band interference filter that focuses the image on a small, high-speed CCD focal plane. The signal is read out from the focal plane at 500 images per second 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 spacecraft for transmission to ground stations.

In the case of LIS, the imaging system is a f/1.6 lens consisting of a beam expander, an interference filter and re-imaging optics. With this imaging system, 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 narrow-band interference filter is centered on the strong oxygen emission line (i.e., the oxygen multiplet at OI(1) at 777.4 nm) in the lightning spectrum. The focal plane assemble, including the 128 × 128 element CCD array, preamplifiers, multiplexers, and clock and drive electronics provides subsequent circuits with an analog data stream of appropriate amplitude. The overall signal flow through the LIS instrument is depicted in Fig. A-2.
The sensor design has been considerably influenced by the requirement to detect weak lightning signals during the day when the sunlight reflecting from the tops of clouds is much brighter than the illumination produced by lightning. This requirement was met by implementing special filtering techniques in the instrument hardware to take advantage of the significant differences in the temporal, spatial, and spectral characteristics between the lightning signal and the background noise. The final step in this process is to apply a frame-to-frame background subtraction to remove the slowly varying background signal from the raw data coming off the LIS focal plane.

Figure A-3 schematically illustrates how the RTEP extracts weak lightning events from an intense but slowly varying background (the heights of the optical transients are exaggerated for clarity). The RTEP continuously averages the output from the focal plane over a number frames on a pixel-by-pixel basis in order to generate a precise but evolving image of the background scene. It then subtracts the average background estimate from the current off-the-focal-plane signal (again on a pixel-by-pixel basis) to extract the candidate optical lightning transients (Fig. A-3a). With the background removed, 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

Figure A-3. (a) Extraction of weak lightning signals from background, and (b) thresholding of subtracted signal.
circuit (Fig. A-3b). 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 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.

Consideration of the background signal strength also was an important element in determining sensor design requirements. As the background signal increases, random photon shot noise in the sensor increases and the probability of false lightning event detections (or false alarms) increases. During the day, diffuse cloud reflectance of solar radiation constitutes the primary background signal and it is generally steady in comparison to the transient lightning pulses discussed above. At night, electronic noise tends to dominate background sources. The maximum background expected due to diffuse cloud-top reflection of solar radiation is about $331 \text{ Wm}^{-2}\text{µm}^{-1}\text{ster}^{-1}$. Much larger values are obtained from specular reflection of solar radiation, e.g., reflection from quiescent water surfaces.

Background sources that change suddenly, i.e., within the 2 ms integration time of LIS, may also introduce false triggers. These sources may include surface glint (e.g., direct solar reflection from ocean or lakes), and radiation or orbit-induced changes in the background. For example, the passage of LIS over a changing cloud field tends to reduce the system signal-to-noise ratio because of the filter response lag. This can lead to higher false event rates. This type of false alarm will be simulated in the calibration laboratory by yawing the LIS sensor head as it is illuminated by an 8" integrating sphere output (see LIS Calibration Procedures Document [1995] for additional details).
A.2 List of Acronyms

ALDF  Advanced Lightning Direction Finder  
ATBD  Algorithm Theoretical Basis Document  
BLDN  Brazil Lightning Detection Network  
CCD  Charge Coupled Device  
CG  Cloud-to-Ground (lightning discharge)  
CERES  Cloud and Earth Radiant Energy System  
CONUS  Continental United States  
DAAC  Distributed Active Archive Center  
DE  Detection Efficiency  
DF  Direction Finder  
DMSP  Defense Meteorological Satellite Platform  
ECS  EOSDIS Core System  
EFM  Electric Field Mill  
EOS  Earth Observing System  
EOSDIS  Earth Observing System Data and Information System  
ENSO  El Nino Southern Oscillation  
ESIP  Earth Science Information Partners  
FOV  Field-Of-View  
GAI  Global Atmospherics, Inc.  
GHCC  Global Hydrology and Climate Center  
GHRC  Global Hydrology Resource Center  
GOES  Geostationary Operational Environmental Satellite  
GSFC  Goddard Space Flight Center  
HDF  Hierarchical Data Format  
IC  Intracloud (lightning discharge)  
KSC  Kennedy Space Center  
LBA  Large-scale Biosphere-Atmosphere Experiment in Amazonia  
LDAR  Lightning Detection and Ranging  
LIP  Lightning Instrument Package  
LIS  Lightning Imaging Sensor  
LMA  Lightning Mapping Array  
LR  Long Range  
LSIM  Lightning Simulator  
MSFC  Marshall Space Flight Center  
NBS  National Bureau of Standards  
NDN  No Document Number  
NLDN  National Lightning Detection Network  
NALDN  North American Lightning Detection Network  
OLS  Optical Line Scan  
OTD  Optical Transient Detector  
PR  Precipitation Radar (TRMM sensor)
A.2 List of Acronyms (Continue)

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>RF</td>
<td>Radio Frequency</td>
</tr>
<tr>
<td>RTEP</td>
<td>Real-Time Event Processor</td>
</tr>
<tr>
<td>SAA</td>
<td>South Atlantic Anomaly</td>
</tr>
<tr>
<td>SCF</td>
<td>Science Computing Facility</td>
</tr>
<tr>
<td>TOA</td>
<td>Time Of Arrival</td>
</tr>
<tr>
<td>TM1</td>
<td>TRMM Microwave Imager</td>
</tr>
<tr>
<td>TRMM</td>
<td>Tropical Rainfall Measuring Mission</td>
</tr>
<tr>
<td>v4</td>
<td>Version 4 (LIS data release)</td>
</tr>
<tr>
<td>v5</td>
<td>Version 5 (LIS data release)</td>
</tr>
<tr>
<td>VHF</td>
<td>Very High Frequency</td>
</tr>
<tr>
<td>VIRS</td>
<td>Visible and Infrared Scanner</td>
</tr>
<tr>
<td>WSI</td>
<td>Weather Services, Inc.</td>
</tr>
</tbody>
</table>