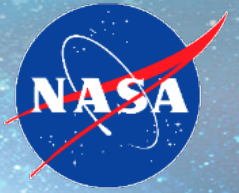
Evolution of the GLM Lightning Cluster Filter Algorithm

Douglas Mach

Universities Space Research Association (USRA)



Introduction

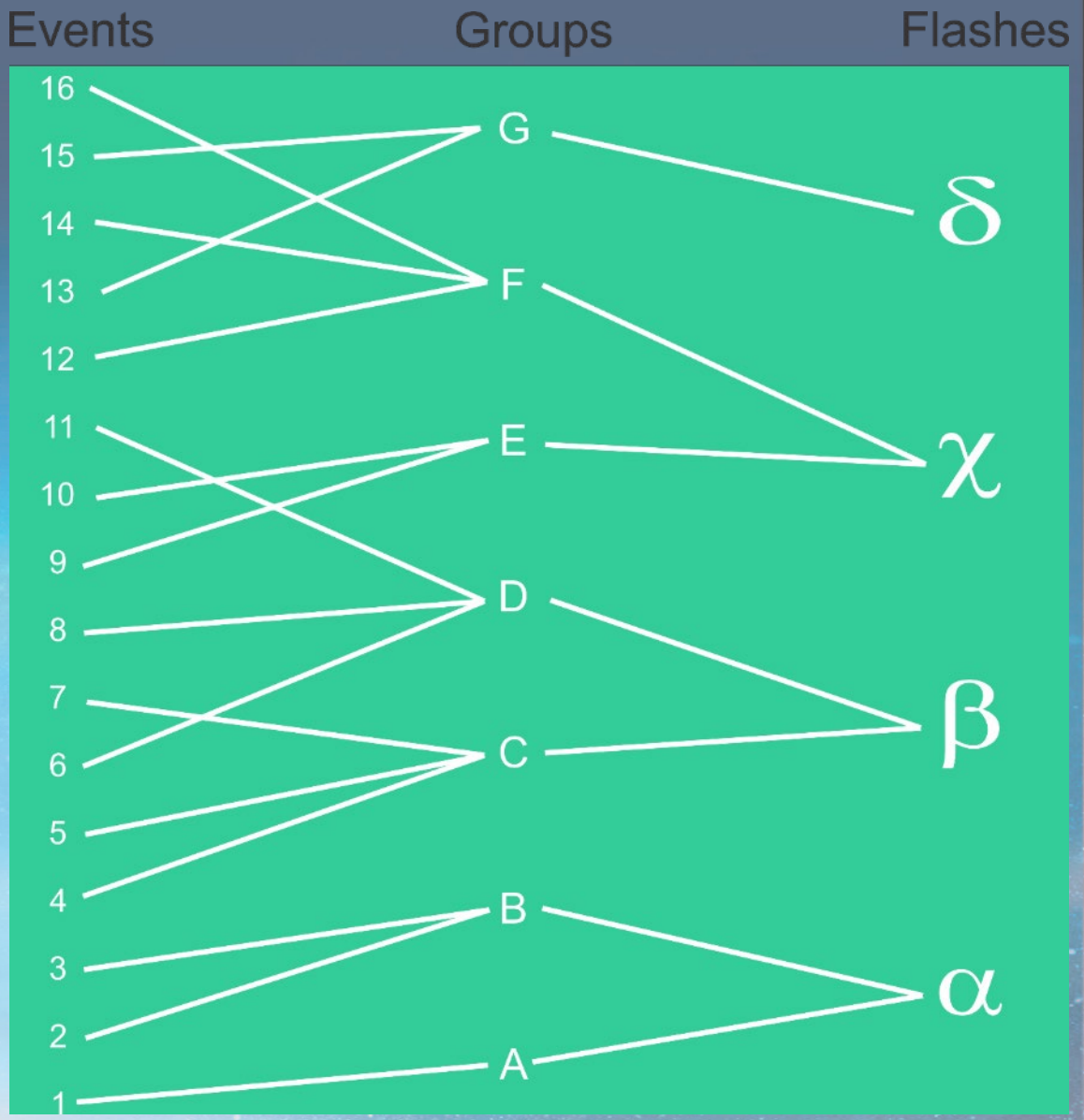
- In 1995 we launched our first instrument dedicated to the optical detection of lightning from space
  - Nearly continuous measurements of lightning from orbit for 25 years!
- The algorithms to process that lightning data have changed (improved?) over the same 25 years
  - Many of the changes were in response to artifacts found in the instrument data
  - Few of the changes were to improve the actual lightning data
  - Initial clustering algorithm proved to be quite robust
- Now that lightning is considered an essential climate variable, it is time to concentrate on improving the utility of the multiple lightning datasets
  - Facilitate comparisons and combinations of the available lightning data
- This presentation will trace the evolution of the current GLM algorithm from the beginnings through today, and into the future

- Clustering Definitions
- Legacy Algorithms
  - Optical Transient Detector
  - TRMM Lightning Imaging Sensor
  - ISS Lightning Imaging Sensor
- Current Algorithm (Operational)
  - Level 0 to Level 1b
  - Level 1b to Level 2
  - GLM Algorithm Changes
  - Future GLM Operational Algorithm Changes
  - Limitations
- New Non-Operational Algorithm Effort
  - Introduction to the Common Scientific Algorithm
  - New Clustering Levels
  - Basic Proposed Algorithm Concept
  - Level 3 Example Gridded Products
  - Cross-Sensor Products Concept
  - Advantages of the Common Scientific Algorithm
  - Action Items
- Questions

CLUSTERING DEFINITIONS

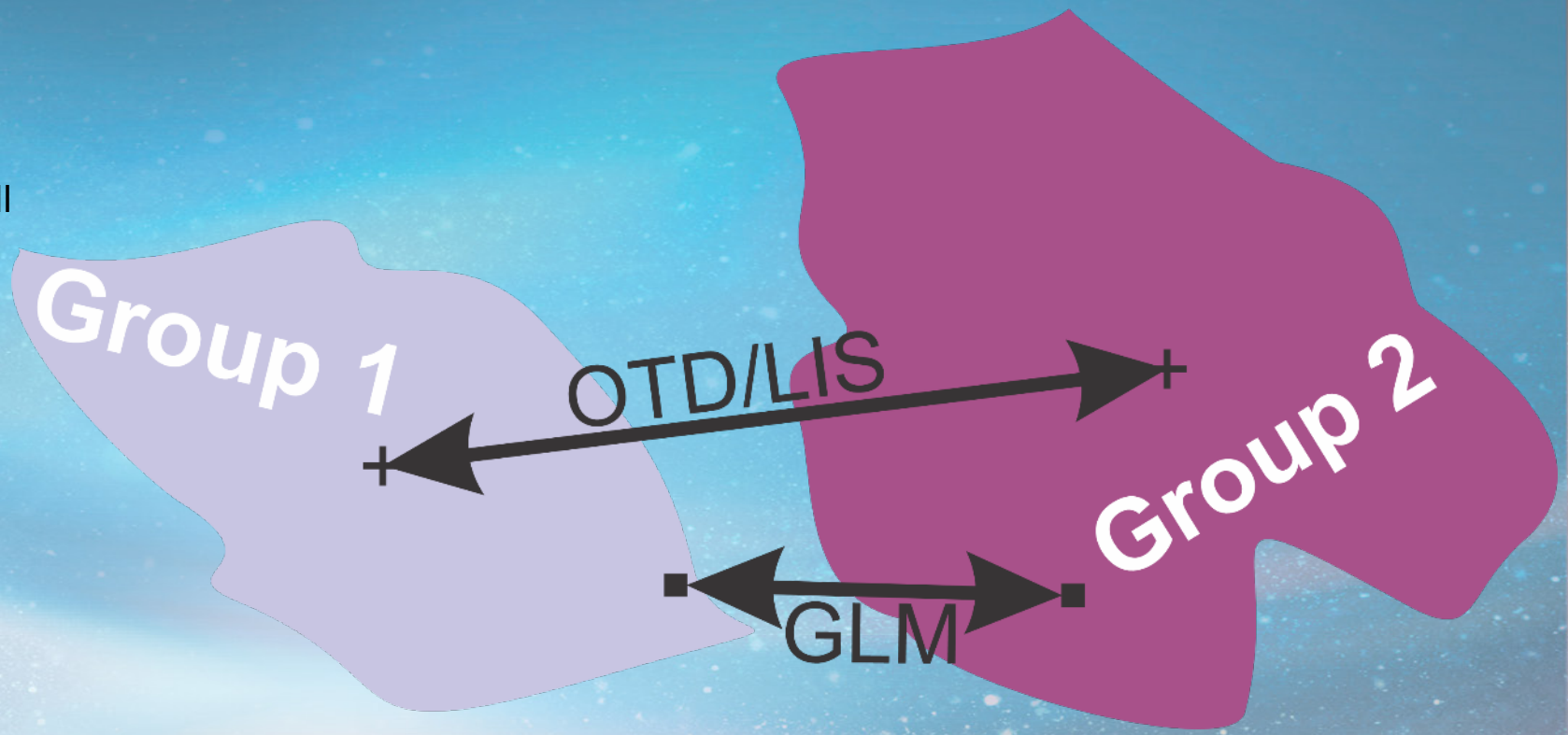
Clustering Linkage

- Hierarchical, single-parent, multiple-child clustering system
  - One to many child events are combined into a single parent group
  - One to many child groups are combined into a single parent flash
  - …
- No “fuzzy sets”
- No cross-element linkage



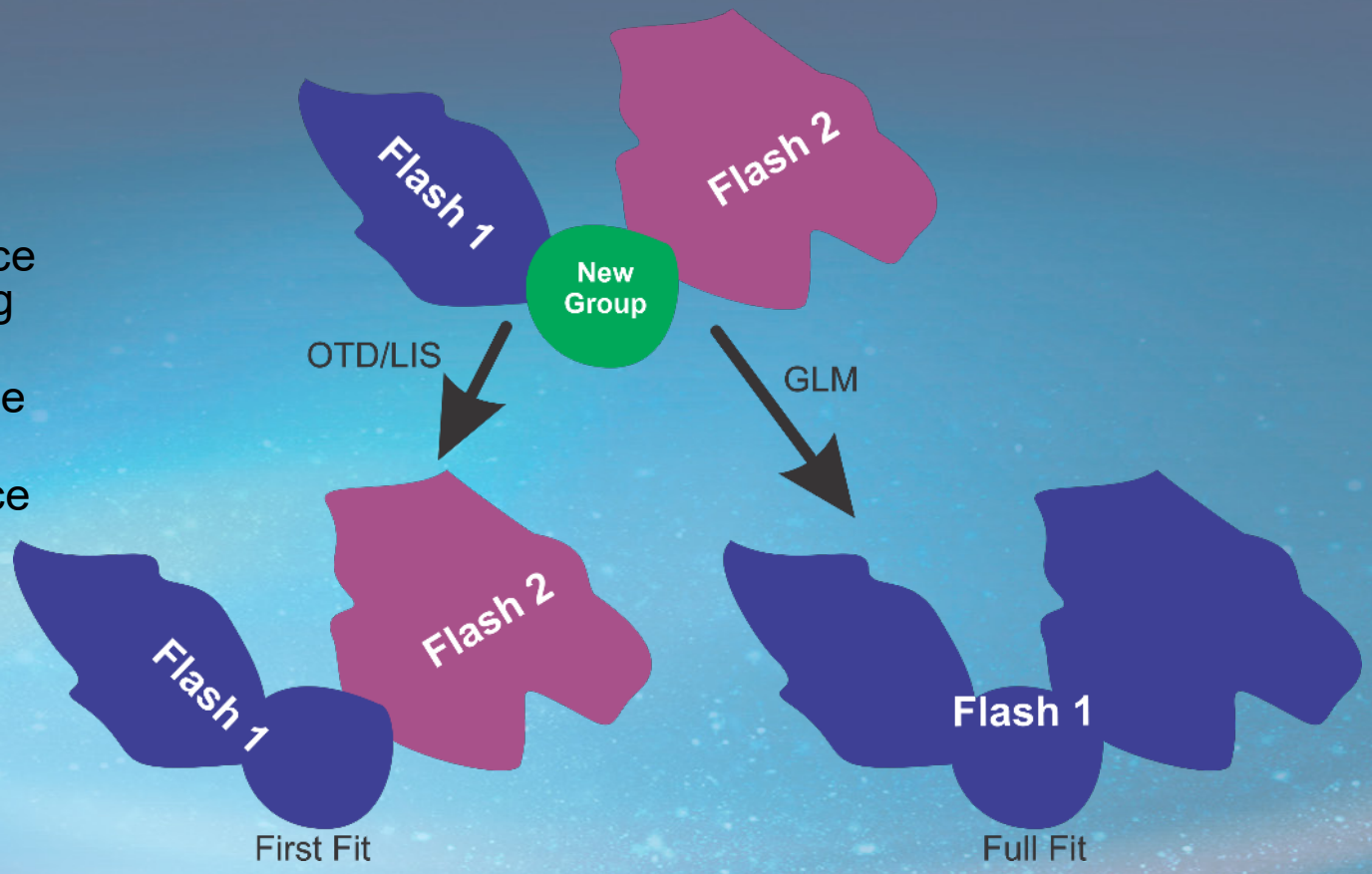
Clustering Distance Comparisons

- Compare Centroids
  - When determining if two groups are part of the same flash, the centroids of the groups are compared
  - Saves memory and computer time
  - Large groups can overlap but still not be combined into the same flash
- Compare Events
  - When determining if two groups are part of the same flash, the individual events in each group are compared
  - Utilizes more memory and computer time
  - Overlapping groups will be combined into the same flash



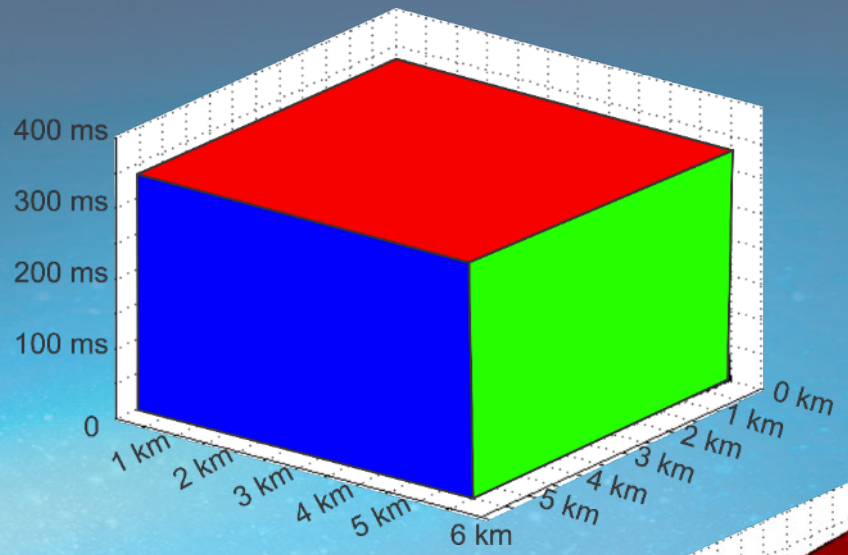
First Fit vs. Full Fit

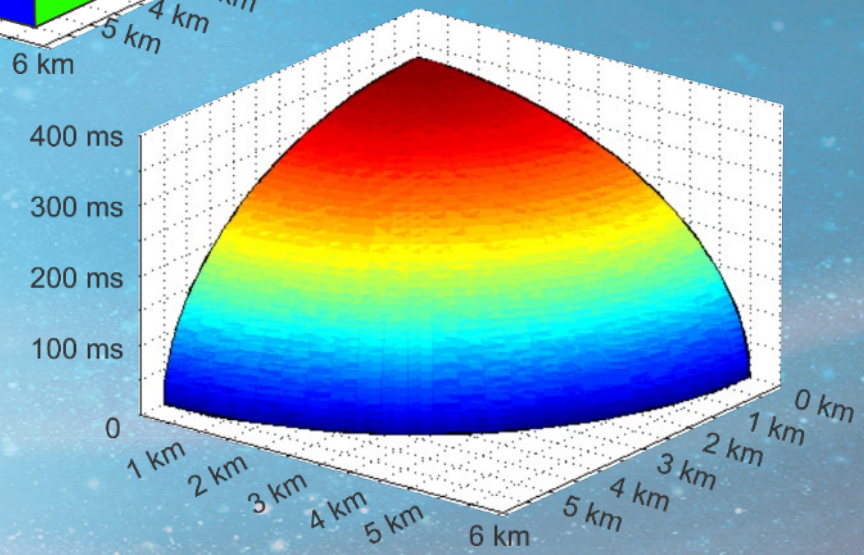
- First Fit
  - Events added to first group within time/space range (can create separate, but overlapping groups)
  - Groups added to first flash within time/space range (possible overlapping flashes)
  - Flashes added to first area within time/space parameters (overlapping possible)
  - OTD, TRMM-LIS, ISS-LIS
- Full Fit
  - If an event can be clustered with more than one group, the groups are combined (no overlapping groups)
  - If a group can be clustered with more than one flash, the flashes are combined (no overlapping flashes)
  - GLM



Boxcar vs. WED

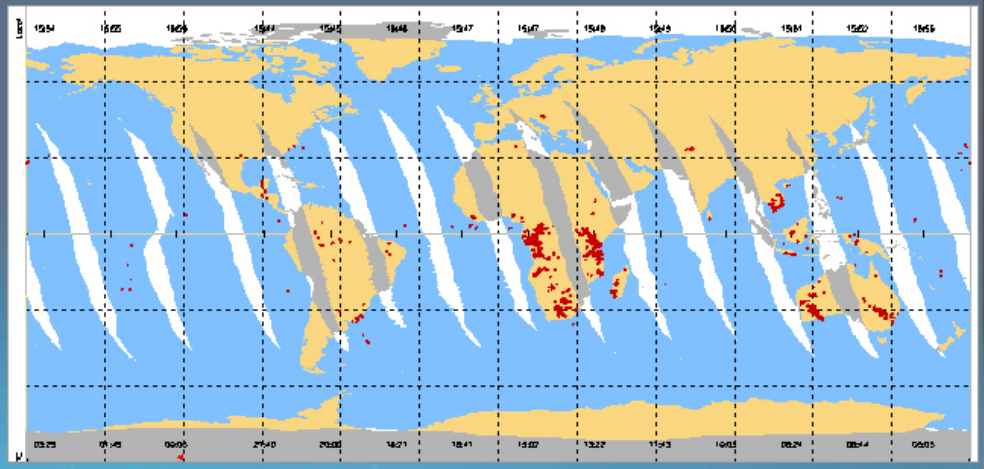
- Boxcar
  - Groups are added to a flash if the distance and the time parameters are individually met
  - Flashes are added to an area if the distance and the time parameters are individually met
  - Two items at the limit of all three parameters (lat, lon, time) will be clustered
  - Used on OTD
- Weighted Euclidean Distance (WED)
  - Groups are added to a flash if the scaled (square root of the sum of squares) distance parameter is met
  - Flashes are added to an area if the scaled (square root of the sum of squares) distance parameter is met
  - Two items at the limit of all three parameters (lat, lon, time) will NOT be clustered
  - Used on TRMM-LIS, ISS-LIS, GLM





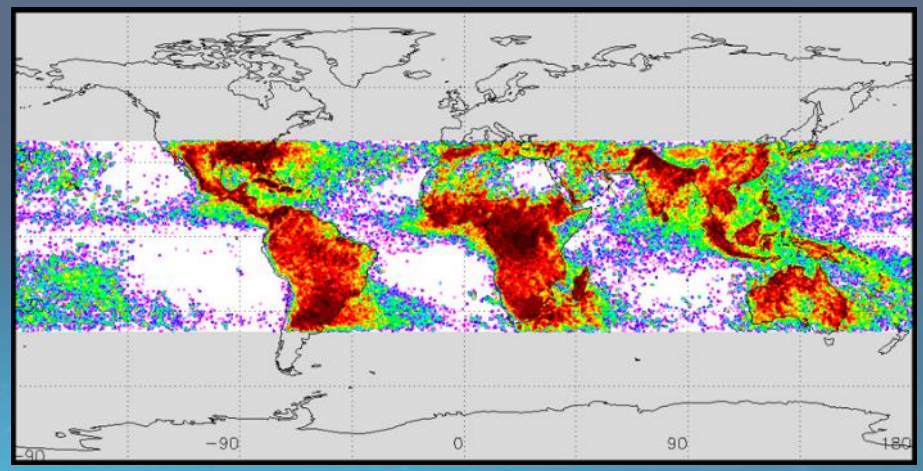
LEGACY ALGORITHMS

Optical Transient Detector (OTD)



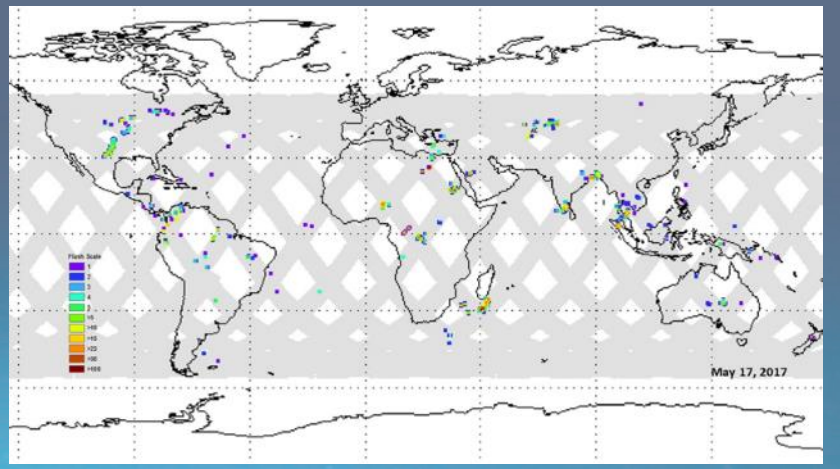
- Part of the Orbview-1 (formerly Microlab-1) spacecraft
- Launched April 1995
- Mission ended March 23, 2000
- Event Filters
  - Dedupe
  - Time fix
  - Ghost
  - Lollypop
  - Track
  - Blast
  - Jumper
  - Boom mask
  - **Filtered events removed from processing stream**
- Geolocation Method
  - SPICE
- Clustering Parameters
  - Events-Groups
    - First fit
    - Same frame time
    - Adjacent pixel (edge or corner)
  - Groups-Flashes
    - First fit
    - 333 ms
    - 16.5 km
    - Boxcar clustering method
  - Flashes-Areas
    - First fit
    - ~ 180 s
    - 27,5 km
    - Boxcar method
- Group/Flash/Area Filters
  - Density
  - Glint 1
  - Contrast
  - Glint 2
  - Guilt
  - Single event
  - **Remove filtered items**

Tropical Rainfall Measuring Mission Lightning Imaging Sensor (TRMM-LIS)



- Part of the Tropical Rainfall Measuring Mission
- Launch November 1997
- Mission ended April 2015
- Event Filter List
  - Dedupe
  - Time fix
  - Ghost
  - Lollypop
  - Track
  - Blast
  - Jumper
  - **Removed filtered events**
- Geolocate Events
  - PGS Toolkit
- Clustering Parameters
  - Events-Groups
    - First fit
    - Same frame time
    - Adjacent pixel (edge or corner)
  - Groups-Flashes
    - First fit
    - 330 ms
    - 5.5 km
    - WED clustering method
  - Flashes-Areas
    - First fit
    - ~ 90 s
    - 16,5 km
    - WED method
- Group/Flash/Area Filters
  - Density
  - Glint 1
  - Contrast
  - Glint 2
  - Guilt
  - Single event
  - **Remove filtered items**

International Space Station Lightning Imaging Sensor (ISS-LIS)



- Installed on the International Space Station
- Launched February 2017
- Event Filters
  - Dedupe
  - Time fix
  - Ghost
  - Lollypop
  - Track
  - Blast
  - Jumper
  - **Filtered events removed**
- Geolocation Method
  - CATS software
- Clustering Parameters
  - Events-Groups
    - First fit
    - Same frame time
    - Adjacent pixel (edge or corner)
  - Groups-Flashes
    - First fit
    - 330 ms
    - 5.5 km
    - WED method
  - Flashes-Areas
    - First fit
    - ~ 90 s
    - 16,5 km
    - WED method Filter Groups/Flashes/Areas
- Group, Flash, and Area Filters
  - Density
  - Glint 1
  - Contrast
  - Glint 2
  - Guilt
  - Single event
  - **Remove filtered items**

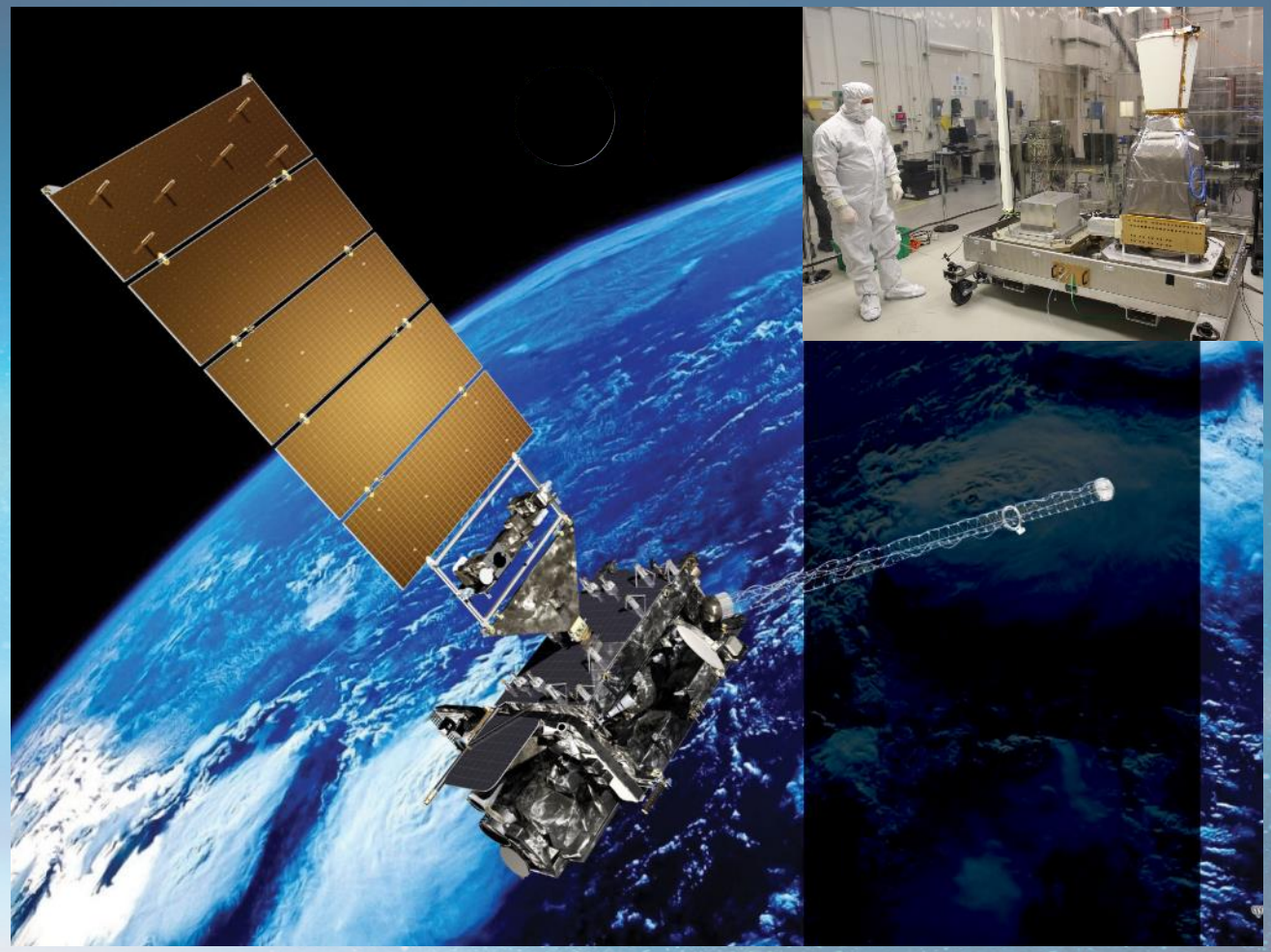
Note: Senior Review Science Panel was in support of LIS on ISS for sustained operations in FY 2021-2023 and FY 2024-2026

- **Prepare to move STP-H5 to ELC-1 site 3 when EMIT shows up (which may be a while)**
- **Prepare to cease operations whenever CLARREO-PF shows up (FY23?)**
- **NASA will be open to keeping ISS LIS going for the lifetime of the instrument**
- **If EMIT/CLARREO never show, are delayed significantly, or if another viable ISS location can be found thru the MiPROM process, then NASA would consider another Senior Review proposal in FY23 to continue operations**

CURRENT ALGORITHM (OPERATIONAL)

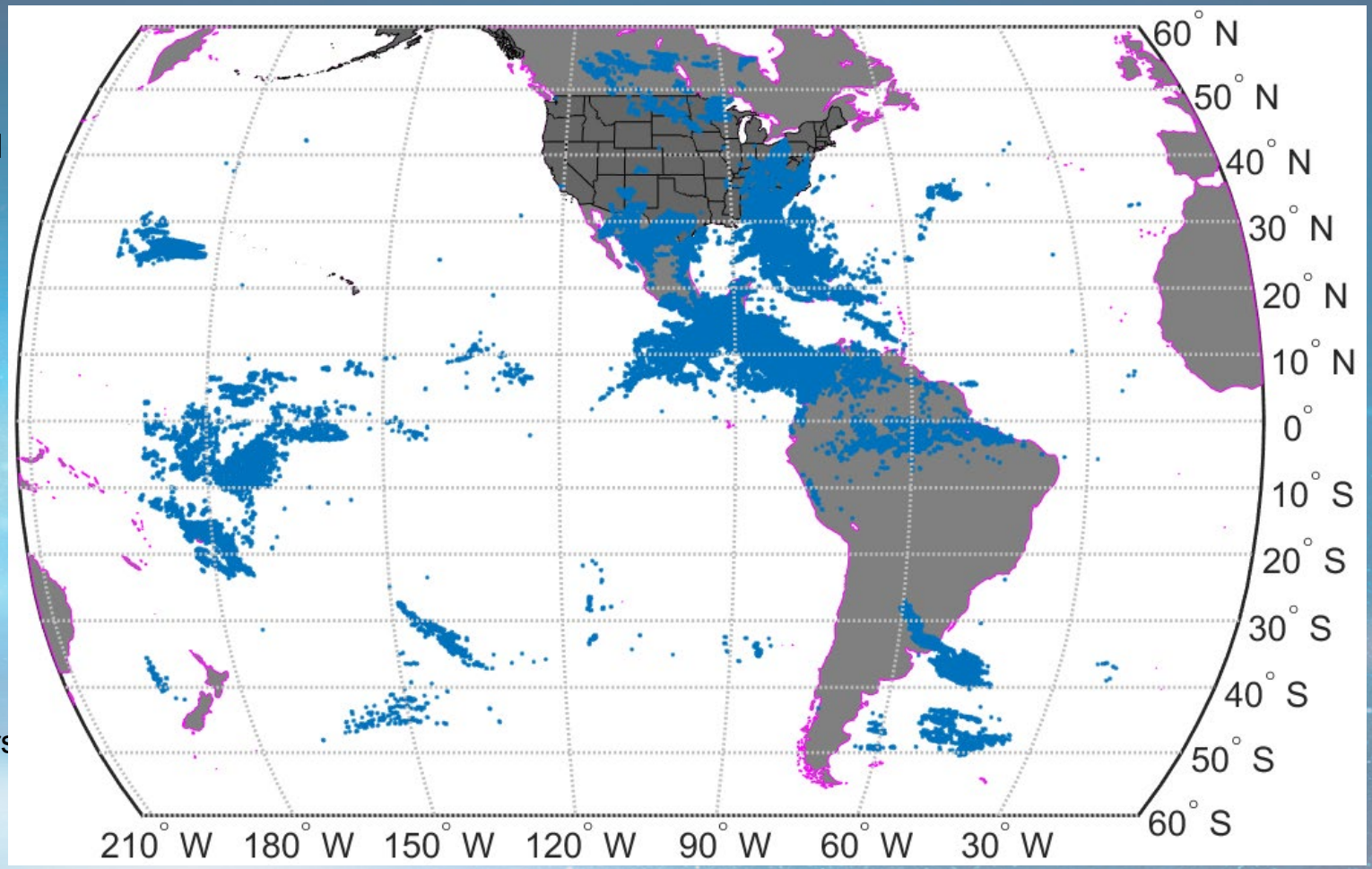
Geostationary Lightning Mapper (GLM)

- Part of the GOES-R series of Geostationary Operational Environmental Satellites
- Launched November 2016 (GLM16/GOES-East) & March 2018 (GLM17/GOES-West)
- GLM18 on GOES-T to be launched ~December 2021
- GLM19 on GOES-U to be launched ~2024



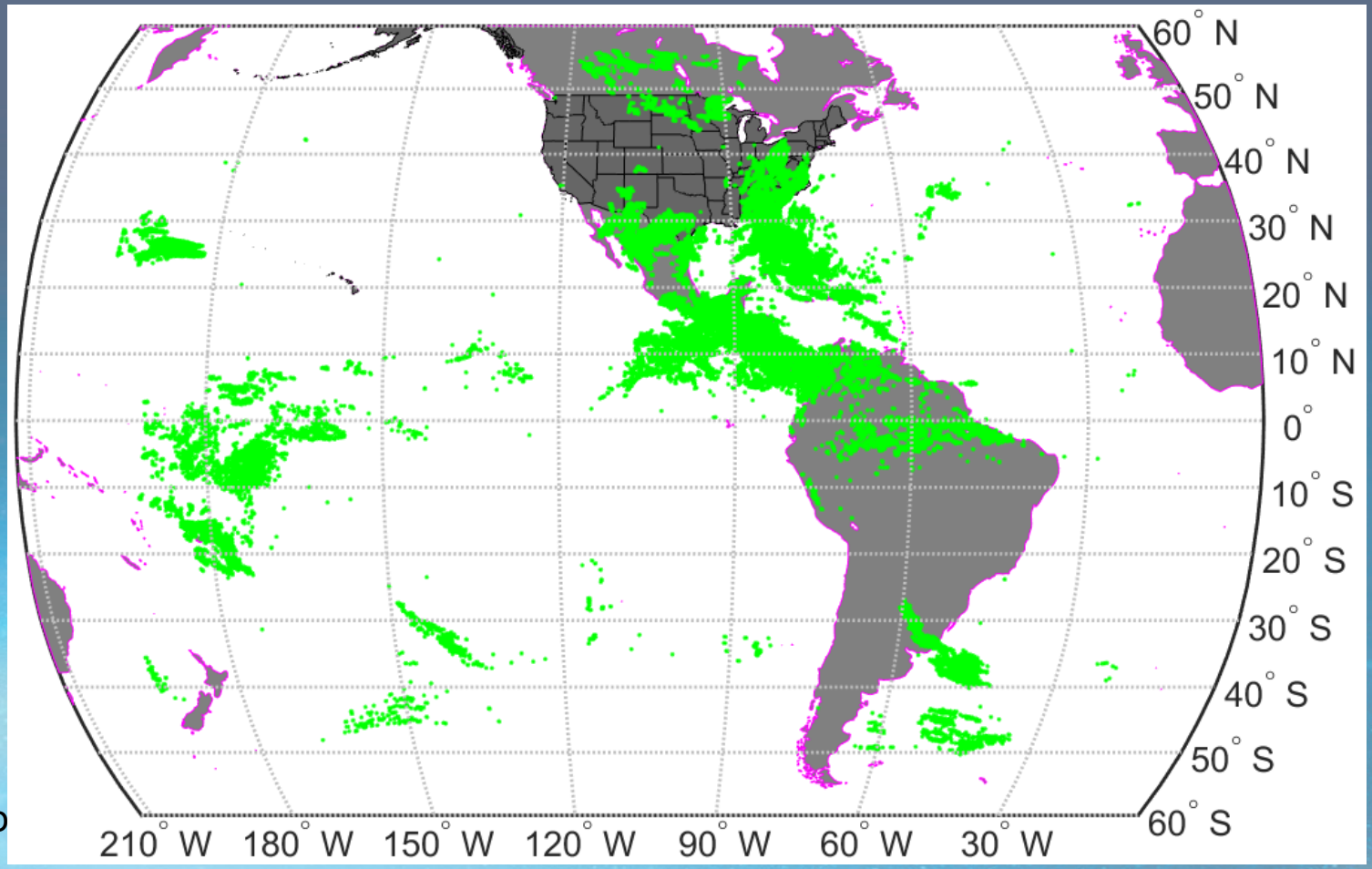
GLM L0-L1b Code

- Initial Decoding/Decompression
  - Decompresses the raw data in the CCSDS packets
  - Creates event, background, navigation, and instrument status/health streams
- Filters
  - CCD Mask
  - 2nd Level Threshold
  - CCD Transfer Noise
  - Radiation Track
  - Coherency
  - Latency
  - Contrast Leakage
  - Solar Glint
  - Blooming Filter
  - **Data flagged, not removed**
- Geolocate Method
  - LM software
  - Only geolocates events that “pass” the filters
- Sends ALL events to L2 code



GLM L1b-L2 Code

- Clustering Parameters
  - Events-Groups
    - Full fit
    - Same frame time
    - Adjacent pixel (edge or corner)
  - Groups-Flashes
    - Full fit
    - 330 ms
    - 16.5 km
    - WED method
  - No Areas
- Removes all flashes with fewer than 2 group
  - This filter is likely too harsh
  - Working on analysis to determine how to improve this filter
    - Better: Remove flashes with only 1 event?
- Outputs remaining events, groups, and flashes to NetCDF file



GLM Software/Parameter Changes (1)

- 01/13/17
  - GLM Group Energy Values are all set to the minimum value
  - GLM energy discrepancies
  - Improve GLM LCFA algorithm error-handling
  - Different units between GLM L1b and GLM L2
  - GLM Eastern RTEP mapping appears incorrect
  - GLM L2+ product metadata errors
  - Update GLM Navigation Parameters
  - Update GLM RTEP Map
- 02/17/17
  - Zero Pixels at RTEP corners in GLM Background Image
  - GLM not producing background images
- 04/24/17
  - Remove GLM Timing Artifacts
  - Change Event Filter Order to match GLM CDRL-80 Rev F
  - Implement Overshoot Filter
  - Implement Solar Glint Filter
  - Implement Crosstalk Filter
  - Update event energy computation
  - Update Block-Level Metadata
  - Update INR Implementation to GLM CDRL-46 Rev H
  - Correct event and group count variables that differ from the events, group data arrays
  - Correct GLM L2+ start/end times
  - Correct GLM LCFA file name invalid start/end date times
- 06/28/17
  - Implement GOES-16 GLM FM1 CDRL079 Rev G
  - Submitted HDF5 files for all 3 positions, and both sides of the instrument with the final PLT determined INR parameters for FM1.
  - Interim updates for event filter parameters and corrections to the radiometric calibration are also included.
  - Added latest thresholds
  - Set all A/B optical distortion coefficients to 0
  - Updated PIT to remove ambiguous coasts
  - Updated coastline ID observation start and stop times
  - Updated max solar angle
  - Updated all coherency filter parameters
  - Updated glint update period
  - Updated mask slightly (3 pixels changed)
  - Corrected errors in calibration table – no more negative calibration coefficients
  - Updated coastline ID parameters:
    - water threshold
    - midnight offset
    - initial alignment values
    - bipod coefficients,
    - earth rotation rate
    - earth rotation angle offset
    - water min
    - min coastline pixels
  - Updated the second level thresholds to correct an indexing error
  - Update primary side calibration table for GLM backgrounds

GLM Software/Parameter Changes (2)

- 07/24/17
  - GLM Eastern RTEP mapping appears incorrect
  - GLM L2+ product metadata errors
  - Update GLM Navigation Parameters
  - Update GLM RTEP Map
- 09/07/17
  - Implement GOES-16 GLM FM1 CDRL079 Rev H
  - Updated parameters for most of the event filtering parameters
  - Added data quality algorithm parameters based on the preliminary data quality product
  - Updated second level thresholds to reapply the minimum on board threshold for that channel
  - Updated overshoot filter LUT based on on-orbit data
  - Updated contrast leakage parameters to essentially turn off the filter
  - Updated glint filter parameters
  - Updated coherency filter parameters
  - Updated probability table based on on-orbit thresholds and higher amplitudes remaining after the second level threshold removes low amplitude events
  - Updated CCD frame transfer filter parameters to essentially turn off the filter
  - Added data quality parameters
  - Incorporated scaling changes into temperature conversion coefficients to mitigate focal length calculation errors that were causing nav issues
- 10/31/17
  - Banded Structure in Group Geolocation GLM L2 – Fixed “Charlie Brown” stripes in L2 groups
  - Implement data formatter burst filter
  - Fixed abnormally large group areas in the L2+ products
  - Fixed Time offset of events, groups and flashes, GLM L2+
  - GLM CALINR update to CDRL 79 Rev H
- 11/21/17
  - GLM L2 event time now has changed scale_factor = 1 millisecond
- 11/28/17
  - Radiation ‘dots’, removing single-group flashes
  - Removed duplicate events
  - Fixed GLM E-W Event Navigation Error
  - Interim fix to GLM Event Geolocation Errors
- 12/14/17
  - GLM LUT update for East
  - Update to glint filter spot amplification and contrast leakage GS parameters
  - Fix to GLM Data Burst Filter – fixed issue of crash induced empty files
- 12/14/17
- 01/10/18
  - Fix the second level threshold filter and the overshoot filter

GLM Software/Parameter Changes (3)

- 02/21/18
  - GLM LUT pre-launch update for GOES-S
- 06/19/18
  - Removed single-group flashes
  - Correct Lightning L2 event time scale_factor
  - Fix GLM L1 EFRC service crash when processing live data
- 10/15/18
  - Update GLM EFRC Algorithm to use updated CALINR format provided by GLM Flight
  - Fix GLM L2 LCFA product has 'n/a' for production_data_source
  - Use adjusted event times in Lightning L2+ product (TOF & associated)
  - Fix group and flash areas GLM L2
  - Update GLM L1b LUT for GOES East
  - Remove orphan and childless events and groups in GLM L2 & Family Links 10/29/18
  - Update GOES-16 GLM LUT to CDRL 79 Rev J
  - Updated lightning ellipsoid values (Parallax Lite)
  - New 2nd-level threshold filter w/32 levels/pixel to mitigate Bahama Bar
- 11/05/18
  - An overflow valve for when 'burst event' will cause an abnormal amount of false events to be registered in the L1b file
- 11/15/18
  - Update GOES-17 GLM LUT to CDRL079 Rev C
  - Updated the flat 2nd level threshold to mute hot pixels
- 01/18/19
- 01/28/19
  - Change GLM L2 Group and Flash Area units from km^2 to m^2
- 02/27/19
  - Second-level filter code change
  - Data burst filter code change
- 04/18/19
- 04/30/19
  - Threshold changes to mitigate noise
  - Turned back on contrast leakage filter
  - Update second-level threshold to better handle "Bahamas Bar"
  - Updated radiometric calibration to remove NaN values
  - Updated ZRDQ parameters for the data quality product
- 07/25/19
  - Minor corrections to the GLM event navigation implementation.
  - Implemented blooming filter
  - GLM metadata will now include the names of the LUTs used in production
  - GLM L2 product outputs correct yaw flip state
  - Minor corrections to the GLM event navigation
- 10/02/19
  - Updated event energy scale factor and offset
  - Corrected zero energy events/groups
- 08/25/20
- 09/03/20
- 09/21/20
- 10/27/20
  - GLM Lightning L2 Flash filter threshold code change to allow for variable number of minimum groups and events
  - Correct GOES-17 GLM L2+ Event Longitude Values west of 180°W
  - Improve how GLM Geolocate handles empty EFRC output
  - Resolves an issue where incorrect times were sometimes output for some event, group, and flash fields
- 01/29/21

Future Operational Code Improvements/Corrections

- ADR 926/928 – GLM FM1 Rev K / FM2 Rev D LUT
- ADR 1109 - For large flash areas, the area matches fill value
- ADR 549 - Eliminate GLM L1b dependency on APIDs 384 and 385
- ADR 844 - GLM L2 lightning needs _unsigned attribute on time offsets
- ADR 906 - GLM L1b – Ingest Directly Instrument Vendor CDRL079 Calibration Data Books
- ADR 1060 - Interim Solution to Facilitate GLM Gridded Data
- ADR 1140 - GLM Time Order Rule Change (Update to ADR375)
- ADR 461 - GLM L2 Data Quality Product
- ADR 645 - GLM Full-Parallax Compensation
- ADR 646 - GLM Gridded Product
- ADR 650 - Diurnal Compensation to GLM GPA Lightning Product Locations
- Feb 15-17 test of new GLM parameter set (to lower noise, increase DE, lower FAR)

GLM Code Limitations

- Algorithm latency
  - To maintain real-time data processing, code limits the size of groups and flashes
    - Groups: nominally* limited to 101 events
    - Flashes: nominally* limited to 101 groups
    - Flashes: nominally* limited to 3 seconds in duration
- Correcting/improving algorithm
  - A big change to ground processing algorithms can take ~2 years from idea to deployment
  - Very resistant to changes other than to correct bugs
    - Example: trying to get pixel ID into L2 output for over 2 years. It will likely not be in the algorithm until 2022.

*When groups are combined, the resultant group exceed 101 events. A combined a flash can exceed 101 groups, or last longer than 3 seconds. No more events can be added to the group or groups added to the flash.

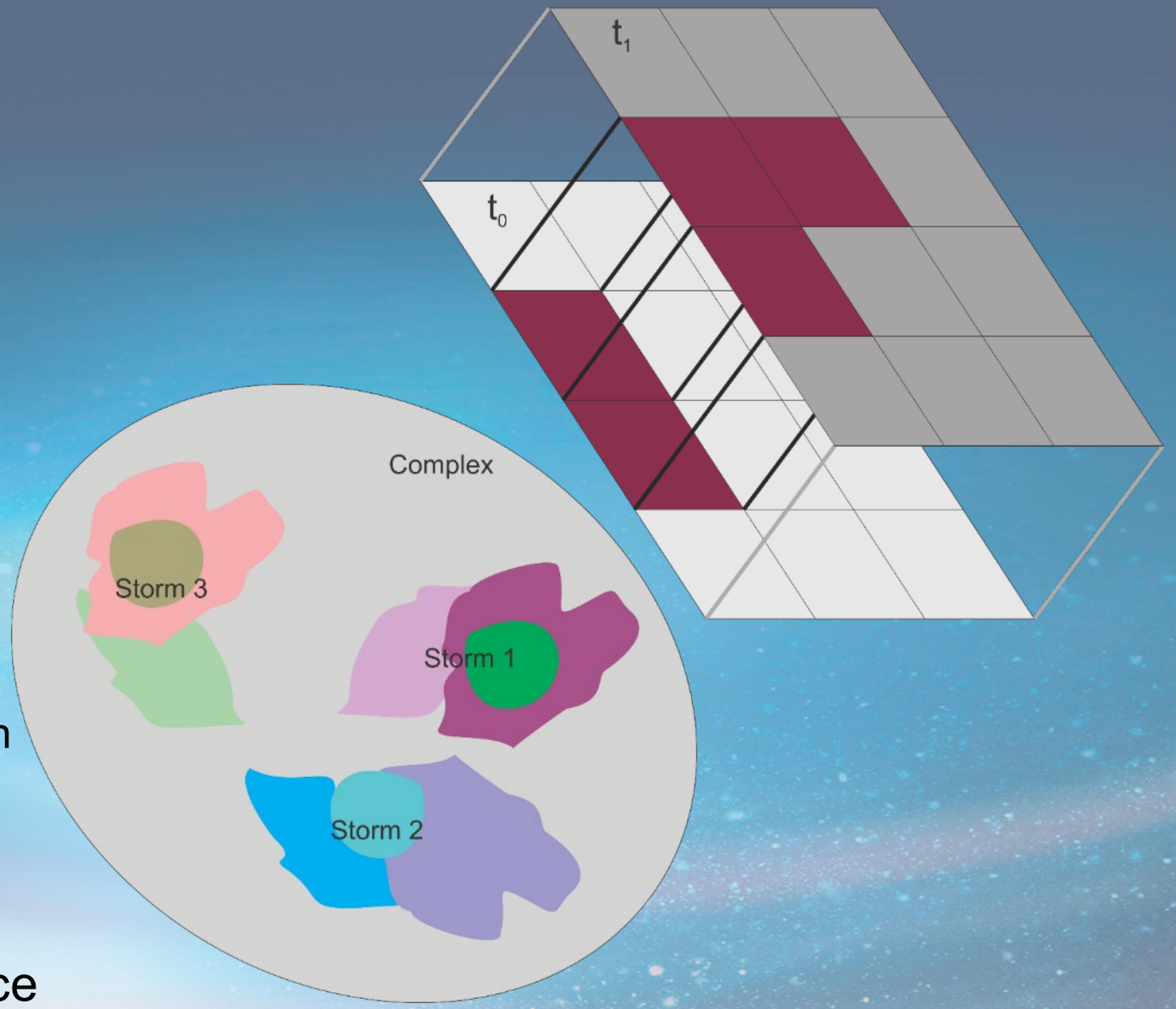
NEW NON-OPERATIONAL ALGORITHM EFFORT

Introduction to the Common Scientific Algorithm

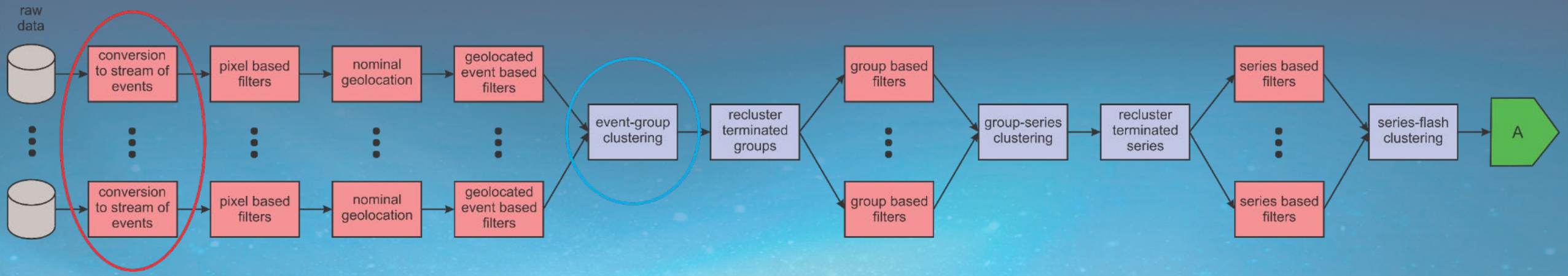
- Propose to create a Common Scientific Algorithm (CSA)
- Stand-alone code set to process orbital lightning data
- Separate from the operational or standard algorithms
- Considered “scientific”, not “operational”
  - Research oriented
  - Not obligated to be “crash-proof”
  - Not obligated to be “faster than real-time”
  - Be able to test/use experimental concepts
  - Open source
  - No use or user restrictions
  - Frequently updated
  - Use best clustering and filtering algorithms
    - Historical and/or current/future algorithms
- Algorithm would be free to use modern techniques to speed processing
  - Cloud computing
  - Parallel processing
  - Graphics card processing.
- A goal of the effort would be to make the CSA efficient enough to run in near real-time
  - Making it accessible to the operational community
  - Encourage migration of best techniques to operational codes

New Clustering Levels

- Series
  - Group that extends over time
- Storm
  - Similar to LIS/OTD Area Concept
  - Requires Flash Overlap
  - Amplitude Weighted Nominal Centroid
    - Centroid based on upper % of flash amplitudes
- Complex
  - Cluster of Storms in a Region
  - Similar to LIS/OTD Area in practice

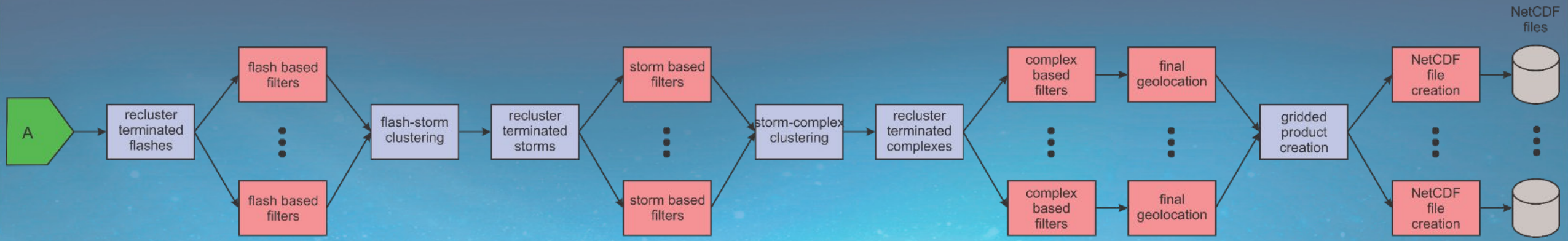


Basic Proposed Algorithm Concept (1)



- Conversion from Raw instrument data (PP)
  - Individual modules for each instrument
  - Parallel Processing Candidate - (PP)
- Pixel Based Filters (PP)
  - Only “obviously not signal” events removed
- Nominal Geolocation (PP)
  - Do not need exact location at this point in the process
- Event Geolocation Based Filters (PP)
  - Keep all but “obviously not signal” events
  - Flag, not remove events
- Clustering Event-Group
  - Including reclustering of terminated groups
- Group Based Filters (PP)
  - Keep all but “obviously not signal” groups
  - Add back any events initially marked as “not lightning” (that with added information, are now considered lightning)
- Clustering Group-Series
  - Including reclustering of terminated series
- Series Based Filters (PP)
  - Keep all but “obviously not signal” Series
  - Add back any events/groups initially marked as “not lightning” (that with added information, are now considered lightning) up-series
- Clustering Series-Flash

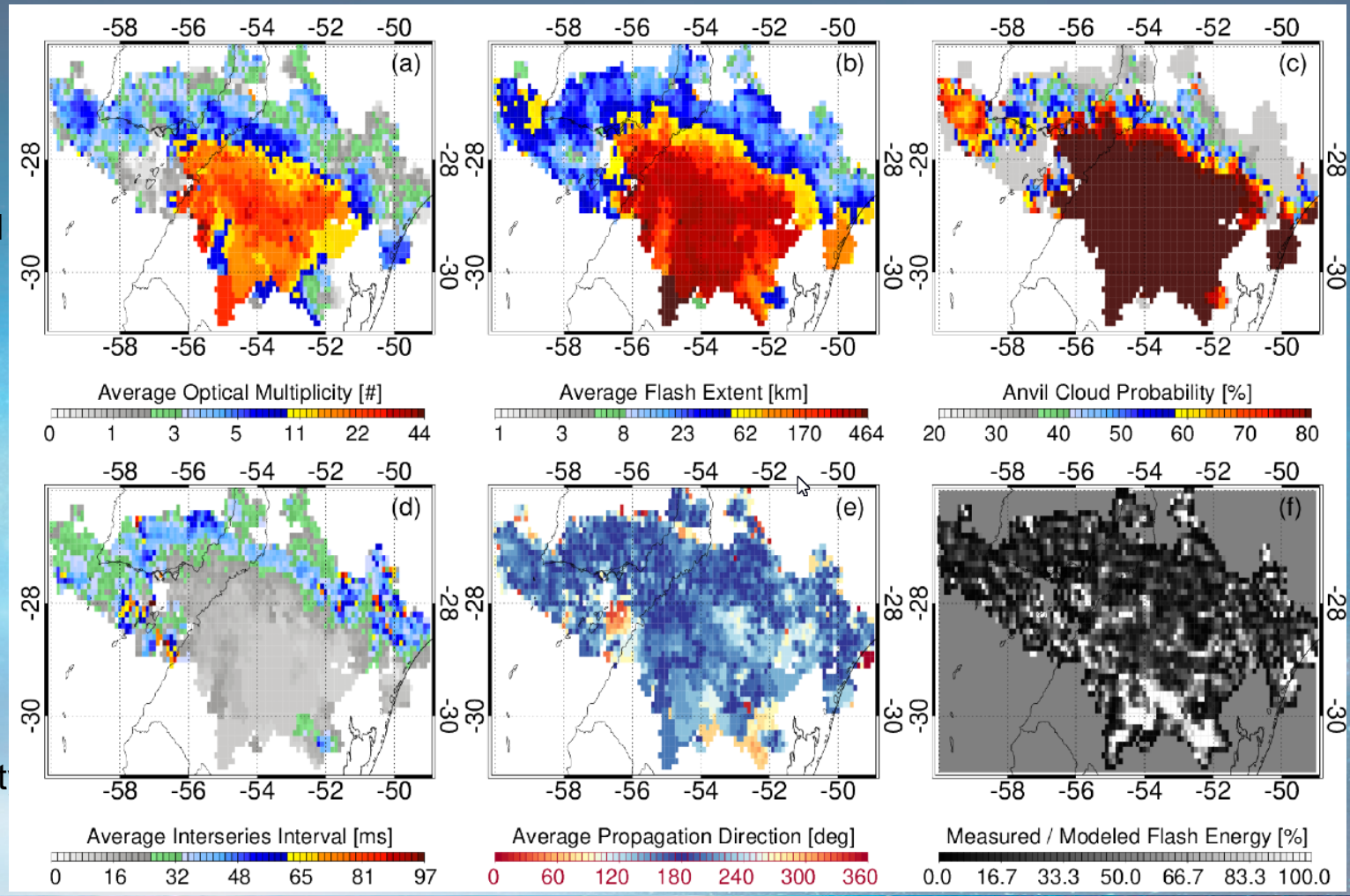
Basic Proposed Algorithm Concept (2)



- Reclustering of Terminated Flashes
- Flash Based Filters (PP)
  - Keep all but "obviously not signal" flashes
  - Add back any events/groups/series initially marked as "not lightning" (that with added information, are now considered lightning)
- Clustering Flash-Storm
  - Include reclustering of terminated storms
- Storm Based Filters (PP)
  - Keep all but "obviously not signal" storms
  - Add back any events/groups/series/flashes initially marked as "not lightning" (that with added information, are now considered lightning)
- Clustering Storm-Complex
  - Including reclustering of terminated complexes
- Complex Based Filters (PP)
  - Keep all but "obviously not signal" complexes
  - Add back any events/groups/series/flashes/storms initially marked as "not lightning" (that with added information, are now considered lightning)
- Final (precise) Geolocation (PP)
- Gridded Product Creation
- Output Results to NetCDF File (PP)
  - Gridded products included
  - Remaining non-lightning, but still "signal" objects are sent to a separate file
    - e.g., Bolide-like entities

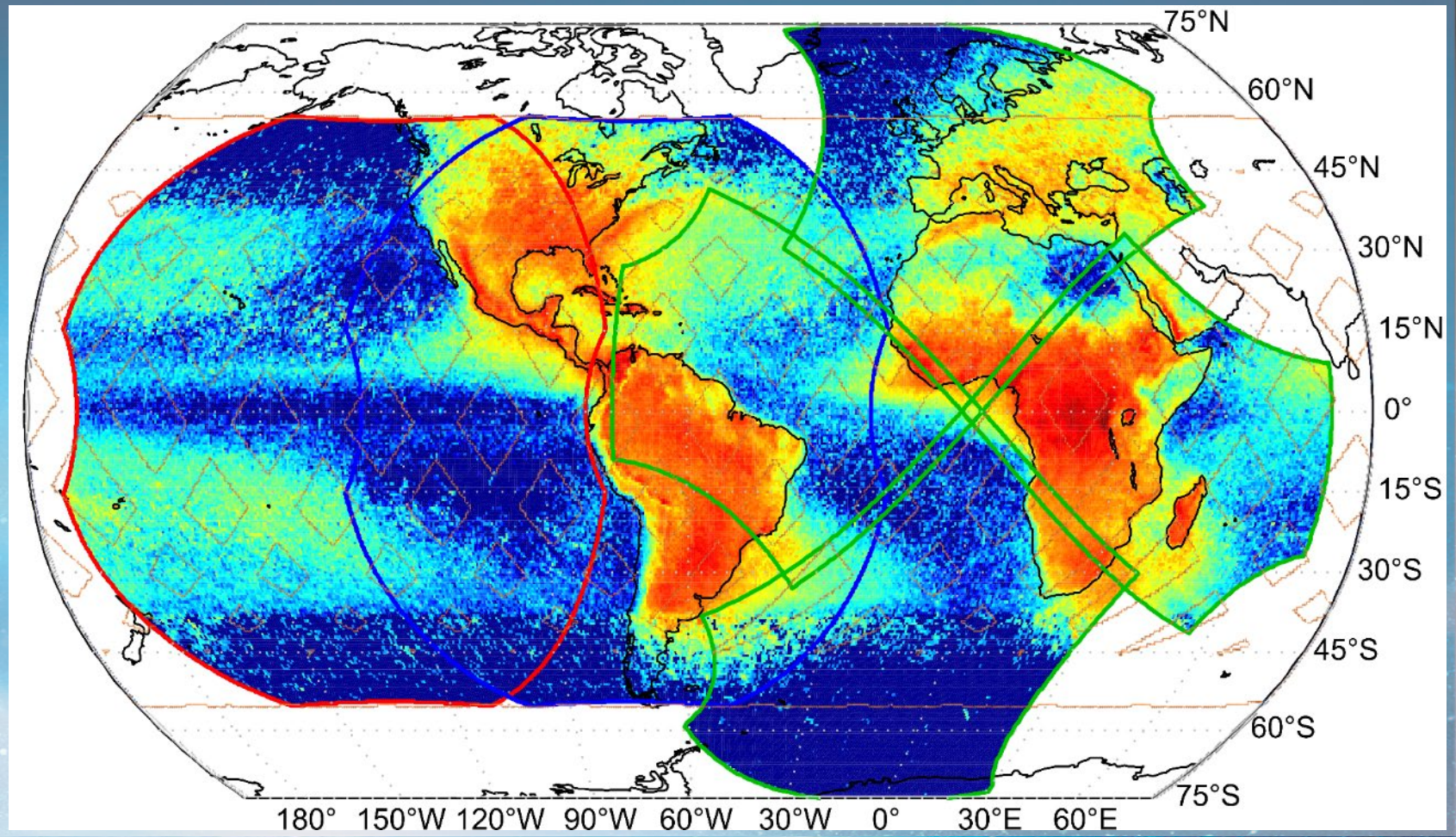
Level 3 Example Gridded Products

- Average Optical Multiplicity Per Flash
- Average Flash Extent
- Flash Extent Density
- Probability Gridpoint Is Anvil/Stratiform Cloud
- Average Interseries Interval
- Average Flash Propagation Direction
- Measured/Modeled Flash Energy
- Flash Extent Density
- Average Flash Area
- Total Energy
- Event Density
- Average Group Area
- Flash Centroid Density
- Group Centroid Density
- Accumulated Flashes
- Thunderstorm Duty
- Total Optical Energy Per Group Extent Densit
- Others…



Cross-Sensor Products Concept

- Significant Overlap Between Orbital Sensors
- Different Viewing Angles and Sensors Provide Extra Lightning Information Content
- Not Restricted to Geostationary Instruments
  - ISS-LIS
  - CLIDE
- Can Include Ground Based Lightning Detecting Systems
  - NLDN
  - ENGLN
  - ADT
  - LMA



Advantages of the Common Scientific Algorithm

- Common definition of output quantities
  - Facilitates comparisons between different sensors
- Allows for increased information content by combining different sensor results
  - Provide a single database for lightning as an essential climate variable
- Allow for rapid improvements in the algorithms (both filtering and clustering) as our understanding of lightning data improves
- Code is “open source” (no black box processing)
- Can be modified by individual users for their particular needs
- “Experimental” algorithms could be introduced and those experiments that produce very positive results can be added to the scientific code base and suggested as additions or changes to the various operational codes
- Much of the code base is already available

Action Items

- Need to refine concept
- Acquire funding
- Assemble v1.0 (current GLM algorithm)
- Invite contributions

QUESTIONS?