JGR Space Physics

RESEARCH ARTICLE
10.1029/2020JA027819

Special Section:
Early results from the Global-scale Observations of the Limb and Disk (GOLD) mission

Feasibility of Near-Real-Time GOLD Data Products

S. M. Codrescu[1,2], W. F. Rowland[1,2], T. M. Plummer[3], B. A. Vanier[3], T. E. Berger[4], and M. V. Codrescu[5]

[1]Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA , [2]National Centers for Environmental Information, National Oceanic and Atmospheric Administration, Boulder, CO, USA, [3]Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, CO, USA, [4]Space Weather Technology, Research, and Education Center, University of Colorado, Boulder, CO, USA, [5]Space Weather Prediction Center, National Oceanic and Atmospheric Administration, Boulder, CO, USA

Key Points:
- Near-real-time GOLD data can be useful for operational and real-time users
- Near-real-time GOLD data can be produced with a latency of 10–20 min by assuming nominal instrument pointing and satellite position
- Near-real-time GOLD data products are reasonably similar to the science products produced with definitive geolocation

Correspondence to:
W. F. Rowland,
william.rowland@noaa.gov

Citation:
Codrescu, S. M., Rowland, W. F., Plummer, T. M., Vanier, B. A., Berger, T. E., & Codrescu, M. V. (2020). Feasibility of near-real-time gold data products. *Journal of Geophysical Research: Space Physics*, *125*, e2020JA027819. https://doi.org/10.1029/2020JA027819

Received 15 JAN 2020
Accepted 2 MAY 2020
Accepted article online 13 MAY 2020

Abstract

A primary limitation for accurate specification and forecasting of the thermosphere-ionosphere (TI) system is uncertainty in the system forcing. This significantly impacts users who have operational and real-time interests in the current and future state of the TI system. Since Global-Scale Observations of the Limb and Disk (GOLD) observations are expected to provide information about both the current state and forcing of the TI system, GOLD products could be an operational asset if they could be provided in near real time. Production of GOLD data products requires knowledge of the satellite pointing and location. The current scientific processing implementation contains an operationally significant delay to await definitive location and as-flown pointing. We present the results of a demonstration low-latency processing system that assumes the nominal satellite position and pointing to produce low-latency GOLD data products. The resulting products are reasonably similar to the scientific version but are available within minutes rather than hours.

Plain Language Summary

The Global-Scale Observations of the Limb and Disk (GOLD) mission is a research mission that observes the upper atmosphere. Normally, at some point after the observations are made, scientists will analyze them. However, there are users who would benefit from near-real-time access to these data since they are expected to give a comprehensive picture of current space weather in the low Earth orbit (LEO) environment. This information would be of interest to someone who is flying a satellite in this region, for example. This article explores the possibility of processing GOLD data in near real time. The normal data processing includes a significant 2-hr delay waiting to receive the exact satellite position and pointing. This article shows that it is possible to remove most of the delay by assuming the satellite is at its normal location and pointing straight down. This assumption produces small but acceptable changes to the resulting data yet makes the data available while they are still relevant to real-time users.

1. Introduction

The Global-Scale Observations of the Limb and Disk (GOLD) mission is NASA's first commercially hosted payload. Hosted on the SES-14 spacecraft in geostationary orbit, the GOLD instrument scans the Earth disk to produce spectrographic images in ultraviolet wavelengths. These measurements are ultimately processed into gridded data products that provide information about the temperature, composition, and structure of the upper atmosphere in the thermosphere-ionosphere (TI) region. These observations are expected to provide unprecedented global-scale, continuous coverage that will help resolve spatial versus temporal dynamics of the system in response to preconditioning and system forcing (Eastes et al., 2017).

The TI system is a strongly forced system, driven by inputs from above and below. From above, it responds to changes in solar and geomagnetic forcing, while simultaneously responding to atmospheric waves and tides propagating from below. Uncertainty in this system forcing is the primary limitation for accurate specification and forecasting of the TI system (Codrescu et al., 2018). This limitation particularly impacts users who have operational and real-time interests in the current and future state of the TI system. GOLD observations are expected to provide information on both the current state and system forcing of the TI system, disambiguating spatial versus temporal effects that have not been resolved by previous observations. Because of the comprehensive and global-scale extent of the observations, GOLD data could be an operational asset if provided in near real time.

Changes to the temperature and composition of the TI system are the result of energy inputs integrated over time and therefore occur on time scales of 30 min to 1 hr (Codrescu et al., 1997; Fuller Rowell et al., 1994), meaning that a latency less than 30 min should be acceptable for operational and real-time situational awareness. The most significant technical roadblock to producing GOLD products with operationally acceptable latencies is the need to wait up to 2 hr after observations are made in order to receive as-flown spacecraft pointing and definitive ephemeris information from the satellite operator. This article presents results of a demonstration low-latency (LL) processing pipeline. The pipeline removes the 2-hr delay by using the satellite's nominal position and assuming nadir pointing. The following sections discuss the current processing pipeline for science products, the new LL processing pipeline, the impact of the LL assumption on product quality, and the latency achieved. We conclude that generation of GOLD products with near-real-time latency is feasible.

2. Methods

2.1. Science Product Processing

This paper evaluates selected dayside products from the full GOLD product suite. The GOLD science products are publicly available from the GOLD mission website. To generate nominal science (Sci) products, the GOLD mission performs processing in the following sequence (McClintock et al., 2020):

1. Level 0 (L0): The Ground Segment operated by SES collects GOLD telemetry data in 1-min files. These files are bundled and sent to the University of Colorado Laboratory for Atmospheric and Space Physics (LASP) in Boulder, CO, every 5 min for further processing.
2. Level 1A (L1A): LASP converts telemetry into 1-min files containing a time-tagged photon events list and instrument telemetry in engineering units. These files are then sent to the University of Central Florida (UCF) in Orlando, Florida, for further processing.
3. Level 1B (L1B): L1B processing is delayed approximately 2 hr until the definitive pointing and ephemeris of the SES-14 satellite are available. At that time, UCF creates 1-min files of binned data with added pointing and geolocation information.
4. Level 1C DAY (L1C DAY): For each 15-min dayside scan, UCF creates an L1C DAY file containing calibrated, geolocated data that have been further binned onto a fixed angular grid.
5. Level 2 (L2): UCF executes Level 2 algorithms to produce various Level 2 products such as TDISK and ON2.

The most significant source of latency affecting Sci product generation is the 2-hr wait for definitive ephemeris information and as-flown quaternions. The L1C DAY product contains data on a geographically fixed grid of "superpixels" across the Earth disk. L1C DAY superpixels consist of L1B pixels binned together to improve signal to noise for the resulting science product while mapping the data to a geographically fixed grid. Pointing and ephemeris data are required to geolocate L1B observations into the correct L1C DAY superpixel.

2.2. LL Processing

To achieve near-real-time results, the wait for definitive pointing and ephemeris information has to be removed. Waiting for as-flown quaternions providing definitive pointing would add up to 30 min of latency, so we assume that the instrument is pointing at nadir. Two approaches considered for satellite location were as follows: (1) Use the predicted ephemeris provided by the satellite operator, or (2) assume the satellite is always at its nominal geostationary location. Approach (2) was chosen for simplicity, robustness, and ease of implementation. Collectively, these LL assumptions allow data to be processed into a LL version of L1B, L1C, and L2 products immediately after L1A processing is complete.

A prototype LL processing pipeline was created to demonstrate the feasibility of creating near-real-time GOLD data products using the nominal pointing and station assumptions described above. The prototype pipeline is designed to run for a period of at least 6 months from October 2019 to March 2020. Results based on the first 3 months of data are analyzed in this study. Once processing is complete, the resulting LL products are uploaded to an Amazon Web Services (AWS) Simple Storage Service (S3) bucket in the cloud. Once uploaded, the product is effectively available for download by operational or real-time users.

Table 1
Metrics Used to Quantify the Feasibility of the Low-Latency Processing Pipeline

Parameter	Comparison value	Rationale
Latency	<30 min	TI system time scale
ON2	95% of pixels, <10% relative difference between LL and Sci product	GOLD Science Objectives mission requirement (Eastes et al., 2017)
TDISK	80% of pixels, <55 Kelvin difference between LL and Sci product	GOLD Science Objectives mission requirement (Eastes et al., 2017)
Geolocation	<1 L1C DAY pixel (125 km × 125 km at nadir) shift in mapping of data between LL and Sci	Likely distance shift limit for use in operational applications

The LL demonstration processing pipeline produces two L2 products: neutral temperature at approximately 160 km altitude (TDISK) and the vertical density of Oxygen O to Nitrogen N_2 ratio (ON2) products. These products were prioritized for the demonstration because of the National Oceanic and Atmospheric Administration (NOAA) Space Weather Prediction Center's (SWPC) operational interest. These simultaneous, colocated measurements could benefit upper atmospheric numerical simulation-based assimilation schemes to untangle the contributions of temperature and neutral composition to the total neutral mass density and ionosphere electron density.

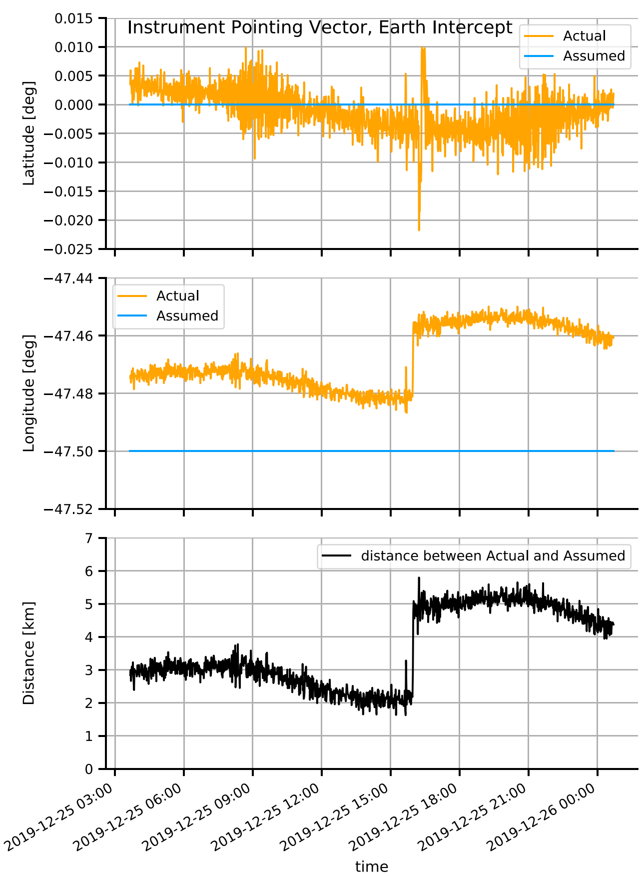


Figure 1. Top and middle show the latitude and longitude of the instrument pointing vector intersection with the Earth for *actual* and *assumed* pointing and ephemeris. The *actual* line represents as-flown pointing and definitive ephemeris as used in the Sci processing. *Assumed* position is the static geostationary orbit with nadir pointing used in the low-latency pipeline. Bottom panel shows the distance between the Earth intersection point between *actual* and *assumed*, illustrating the magnitude of error in geolocation shift introduced by the low-latency processing.

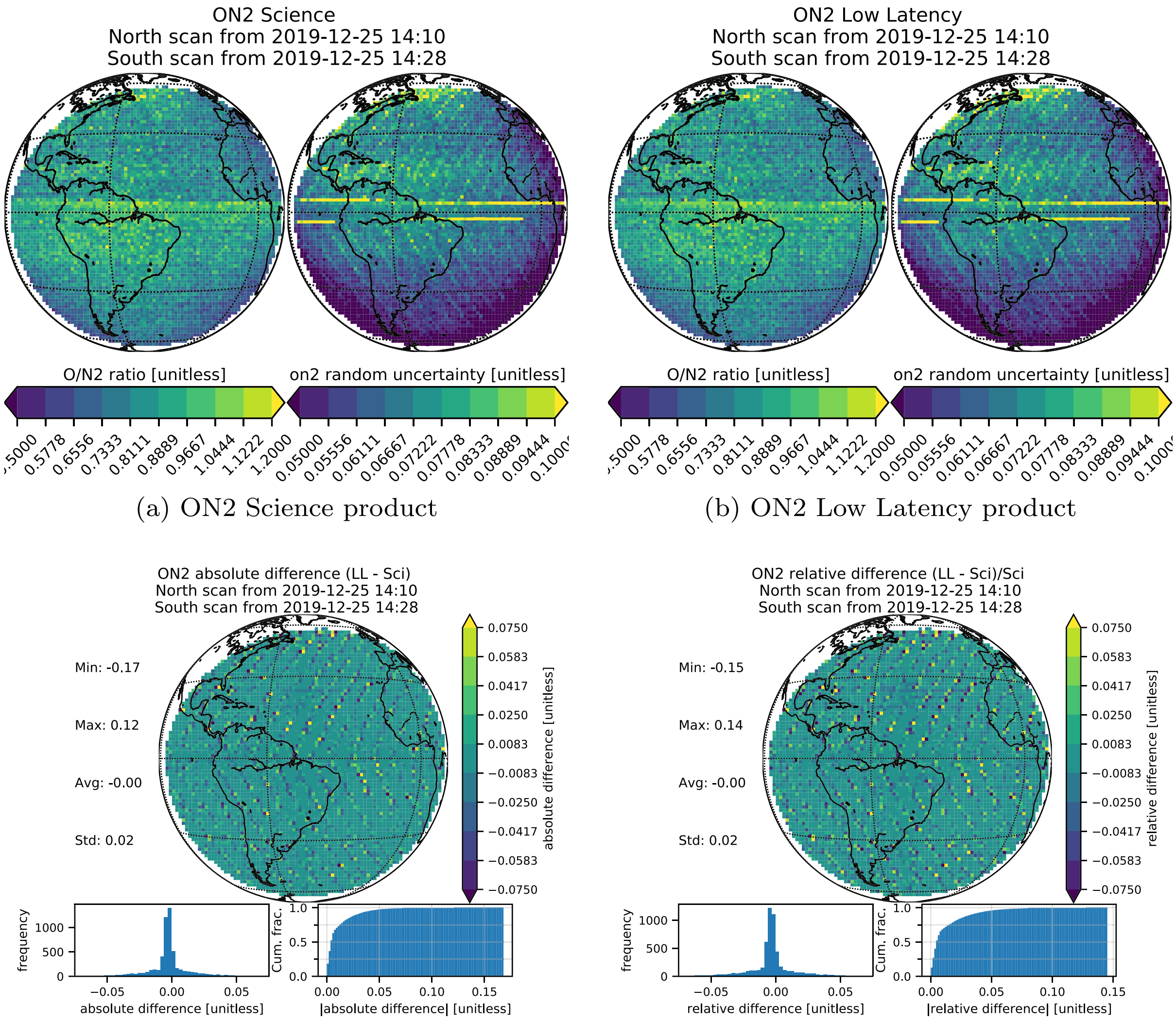
2.3. Comparison Metrics

The criteria in Table 1 are quantitative thresholds this analysis used to evaluate feasibility of the LL processing pipeline. These criteria seek to bound the maximum permissible error introduced by the pointing and ephemeris assumptions, quantify how closely LL products should resemble Sci data, and contextualize how quickly the data are needed to satisfy near-real-time uses. These criteria do not guarantee operational utility. Instead, the aim is to establish feasibility of producing near-real-time data that resemble the Sci data. If Sci data are useful, and LL data resemble Sci data, then the LL data could also be useful. Other GOLD publications are expected to highlight the utility of the data.

3. Results and Discussion

3.1. Geolocation Impacts

The impact of the LL assumptions on the geolocation of observations is shown through a comparison between the actual instrument pointing versus the LL assumed pointing in Figure 1 for 25 December 2019. The top and middle panels show the latitude and longitude of the instrument pointing vector intersection with Earth for *actual* and *assumed*. The *actual* line represents the as-flown pointing and definitive ephemeris that are used in the Sci processing. The *assumed* line gives the idealized geostationary ephemeris with nadir pointing assumed in the LL pipeline. The bottom panel shows the distance between the Earth intersection points for *actual* versus *assumed*. The bottom panel shows the magnitude of error in geolocation shift introduced by the LL processing via distance between the two points. The sudden change near 16:00 UTC is caused by a spacecraft maneuver. The LL assumptions nominally introduce less than 10 km difference between the true observation location versus the assumed location.



(a) ON2 Science product (b) ON2 Low Latency product

(c) Absolute difference (d) Relative difference

Figure 2. The ON2 product gives the ratio of the vertical column density of Oxygen O to Nitrogen N_2. (a) The science (Sci) product produced using the definitive satellite pointing and position. (b) The corresponding low-latency (LL) product produced using assumptions of nominal satellite station and nadir pointing for the same scans. (c) The absolute difference between the LL and Sci product, with a histogram of pixel differences below left and the cumulative distribution of pixel difference magnitude below right. (d) The relative difference, normalized by the Sci product, and again histogram of pixel differences and cumulative distribution of pixel difference magnitude.

3.2. Product Impacts

Impacts of the assumptions made in order to achieve LL processing were analyzed for the time period 27 September 2019 through 27 December 2019. A typical case was selected from 25 December 2019 and presented here for the ON2 product by showing a Science product (Sci) in Figure 2a, the corresponding LL product in Figure 2b, and then showing a pixel-wise absolute difference LL-Sci for Figure 2c and relative difference between the LL and Sci products (LL-Sci)/Sci in Figure 2d. Figure 3 shows the corresponding plots for the TDISK product. The difference between the Sci and LL products is generally quite subtle—usually not apparent on visual inspection of either product individually. It is only when examining the difference that a characteristic difference emerges.

To visualize a full disk image, consecutive North and South Hemisphere scans were selected. Each scan takes approximately 15 min to complete. In the region around the equator where the north and south scans overlap, the pixel values are averaged. An edge effect is visible as horizontal streaks caused by a low number

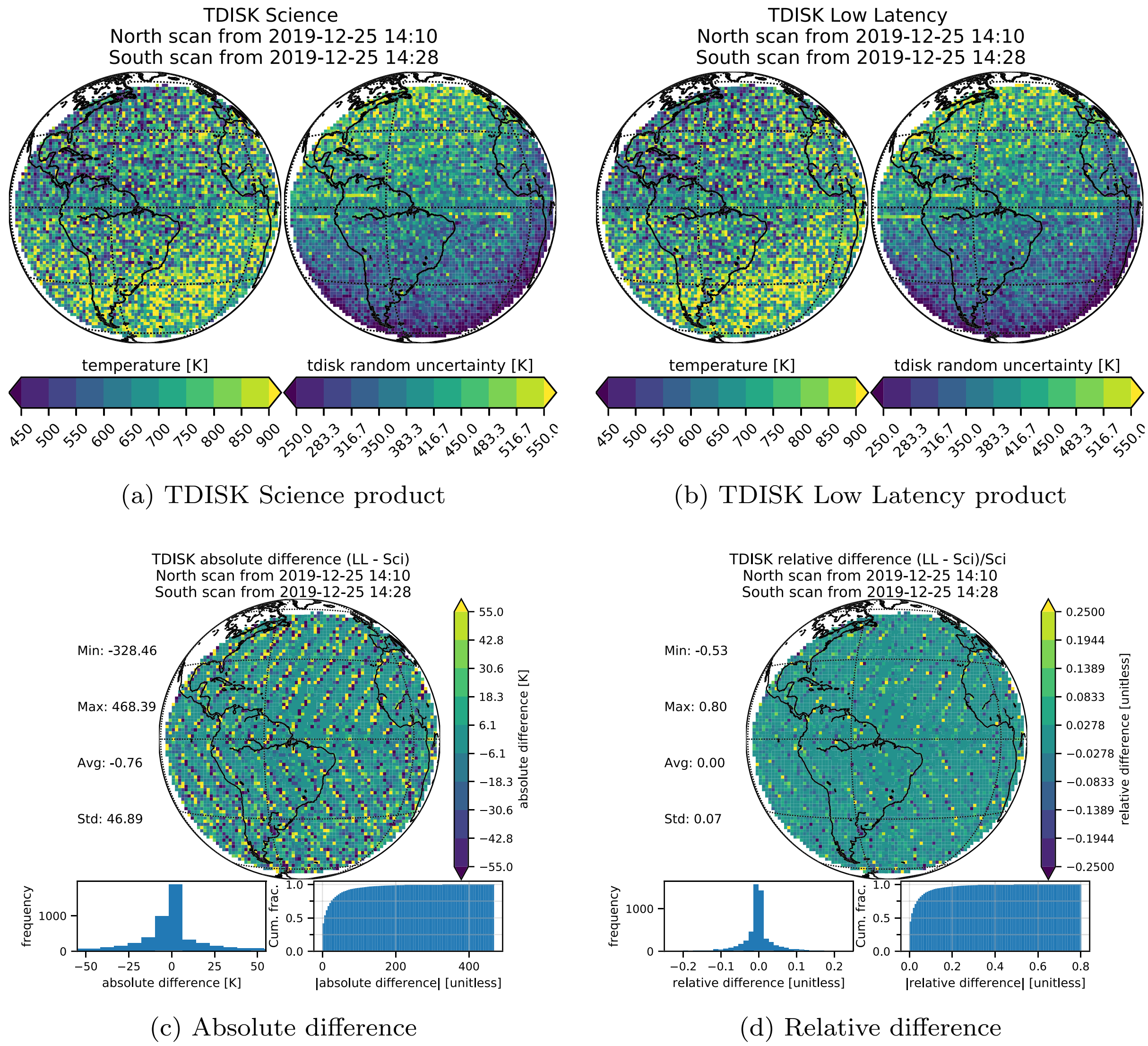


Figure 3. Same as Figure 2 but illustrating the TDISK product. TDISK gives the neutral temperature at an altitude of about 160 km.

of samples binned into these pixels as a result of the GOLD scan slit moving slightly off-parallel from the equator. For scientific use of the GOLD products, these degraded edge pixels should be discarded.

Comparisons between the LL and Sci products contain a characteristic diagonal striping pattern in the difference images. This is due to the way that the binning strategy for L1C DAY data interacts with the slightly different geolocation assumption in the LL processing versus the Sci data. L1C DAY data consist of geolocated superpixels. L1B pixels are binned into the L1C DAY fixed grid according to the observing geometry. The assumed pointing and ephemeris cause portions of the underlying observations that are binned at one location for the LL product to instead be binned into an adjacent superpixel for the corresponding Sci product (Figure 4). The L1C DAY binning implementation produces diagonal stripes of L1C DAY superpixels into which fewer L1B pixels were binned, resulting in higher uncertainty for those L1C DAY superpixels. The impact of these stripes is faintly visible in the uncertainty images of Figures 2a, 2b, 3a, and 3b. The location of these stripes of increased random error is systematically offset between LL and Sci products due to the differing geolocation, appearing as the stripes in the difference images (Figures 2c, 2d, 3c, and 3d). The TDISK product is more sensitive to noise, and therefore, the distribution of pixel difference between LL and Sci forms a wider distribution for TDISK compared to ON2. Both LL products generally resemble the corresponding Sci products. The LL ON2 product achieves more than 95% of pixels within 10% relative difference of Sci, and LL TDISK maintains more than 80% of pixels within 55 K of the Sci product. This agreement is

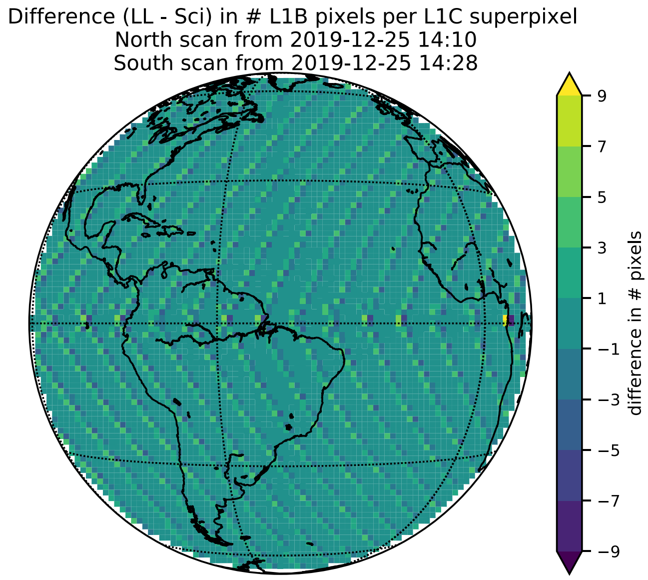


Figure 4. Difference in the number of L1B pixels binned into L1C DAY superpixel between low-latency and science products. The geolocation assumptions in the low-latency pipeline cause existing L1C DAY stripes containing fewer samples to shift position.

relatively stable throughout the day except for times near the start and end of the day when there are very few valid daytime pixels ahead or behind the terminator.

3.3. Product Arrival Latency

The product arrival latency was defined as the difference between the file last modified timestamp in Amazon S3 storage and the final pixel timestamp in the product. A histogram of this distribution is shown in Figures 5 and 6. This measure gives an accurate representation of the full latency of the system, from the time that the last observation was made until the L2 product was available to a potential user. This distribution includes in the latency from all sources, including receiving all necessary inputs, product processing latency, and latency in the delivery mechanism. Spread in the distribution results from variations in these components, such as network latency and variations in product processing latency due to computational expense scaling as a function of number of valid pixels.

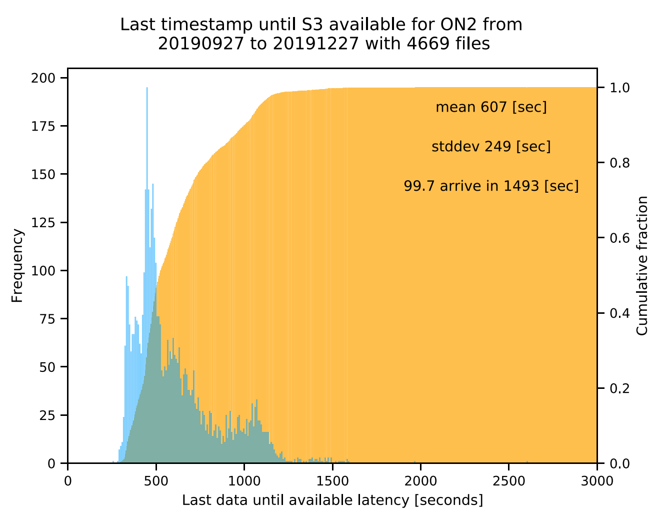


Figure 5. Latency distribution for the low-latency ON2 product. Latency measured from final pixel timestamp to last modified timestamp in cloud storage. The mean latency of 607 s corresponds to roughly 10 min.

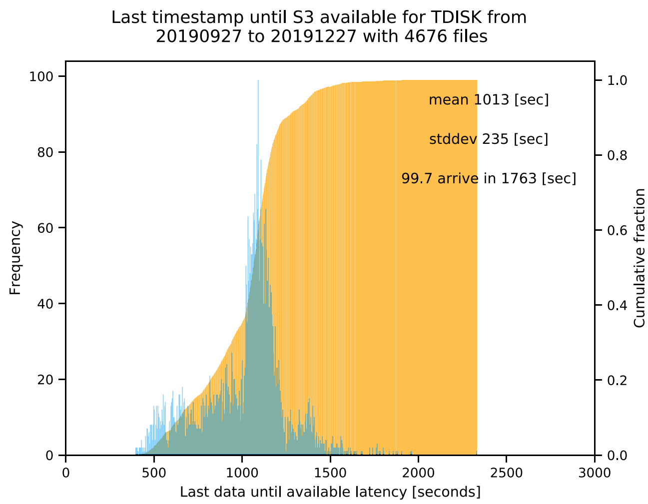


Figure 6. Latency distribution for the low-latency TDISK product. Latency measured from final pixel timestamp to last modified timestamp in cloud storage. The mean latency of 1,013 s corresponds to roughly 17 min.

Almost all products arrive in cloud storage within 20 min of the last observation they contain. The ON2 product arrives faster with a mean of approximately 10 min compared to 17 min for the TDISK product. Both ON2 and TDISK products simultaneously begin processing immediately after L1C DAY is ready; however, the computational complexity of the TDISK product is higher and therefore takes longer to complete.

3.4. Potential Future Improvements

In creation of the demonstration LL processing pipeline, a number of potential opportunities for improvement have been identified. Although SWPC believes that the current products produced by the demonstration LL processing pipeline would have value in an operational setting, these opportunities are suggested for future consideration to further improve the quality of the LL products in terms of their agreement with Sci products, and for further latency reductions.

Toward improving product quality, three changes have been identified that could increase agreement between LL and Sci products. First, in an improvement to the LL assumptions, using the predicted satellite position provided by the satellite operator, in conjunction with a short moving average of the recent pointing, is expected to better approximate the actual viewing geometry than the current static assumptions. Second, a larger L1C DAY superpixel size would improve the signal to noise and reduce the impact due to shifting of underlying L1B observations between L1C DAY superpixels. Finally, the L1C DAY binning algorithm could be modified to interpolate L1B observations overlapping multiple L1C DAY superpixels, rather than the current assignment of an entire L1B pixel based on the center of the L1B pixel. This interpolation scheme is expected to slightly improve consistency of the data across the disk in general and therefore reduce the effect of slight L1B pixel positioning differences due to real versus assumed geolocation.

The most straightforward improvement to latency would be a parallelization of the TDISK algorithm, which could bring the TDISK latency closer to that of the ON2 product. Another improvement would be to obtain a higher frequency of data delivery from the commercial satellite operator. Finally, a challenging yet significant improvement to latency would be to perform processing on chunks of data rather than full scans. The GOLD instrument takes about 15 min to perform a full hemisphere scan, which is then processed in full. Processing the scans in segments, as the data are taken, would significantly reduce the mean data arrival time but would require more serious modification to the processing routines compared to other options to reduce latency.

4. Conclusions

Approximately 3 months of available GOLD LL measurements demonstrate it is feasible to generate LL data products with latency sufficient to meet operational needs. The differences between LL and normal scientific processing are small and are not expected to compromise the use of GOLD measurements in operations. We plan to extend this study to over 6 months of GOLD measurements to further characterize impacts of the geolocation assumptions and investigate any possible seasonal effects in the difference between LL and scientific products.

Acknowledgments
The scientific (Sci) GOLD data used in this study is available from the mission webpage (https://gold.cs.ucf.edu/data/search/). Please contact the authors to establish credentials for the AWS S3 bucket containing the LL data used in this study. Thanks to LASP for implementing the LL processing system. Thanks to Computational Physics, Inc. (CPI) for supporting integration of the L2 algorithms. The CIRES portion of this work was supported by NOAA Office of Projects, Planning, and Analysis (OPPA) through Technology Maturation Program (TMP) Award 18-24, granted to NCEI in collaboration with CIRES. LASP and CPI participation was supported by a Grand Challenge grant from the Office of the Chancellor at the University of Colorado at Boulder through the Space Weather Technology, Research, and Education Center (SWx-TREC). Many thanks to the anonymous reviewers for their valuable feedback.

References

Codrescu, S. M., Codrescu, M. V., & Fedrizzi, M. (2018). An ensemble Kalman filter for the thermosphere-ionosphere. *Space Weather*, *16*, 57–68. https://doi.org/10.1002/2017SW001752

Codrescu, M. V., Fuller Rowell, T. J., & Kutiev, I. S. (1997). Modeling the F layer during specific geomagnetic storms. *Journal of Geophysical Research*, *102*(A7), 14,315–14,329. https://doi.org/10.1029/97JA00638

Eastes, R. W., McClintock, W. E., Burns, A. G., Anderson, D. N., Andersson, L., Codrescu, M., et al. (2017). The Global-Scale Observations of the Limb and Disk (GOLD) mission. *Space Science Reviews*, *212*(1), 383–408. https://doi.org/10.1007/s11214-017-0392-2

Fuller Rowell, T. J., Codrescu, M. V., Moffett, R. J., & Quegan, S. (1994). Response of the thermosphere and ionosphere to geomagnetic storms. *Journal of Geophysical Research*, *99*(A3), 3893–3914. https://doi.org/10.1029/93JA02015

McClintock, W. E., Eastes, R. W., Beland, S., Bryant, K. B., Burns, A. G., Correira, J., et al. (2020). Global Scale Measurements of the Limb and Disk (GOLD) mission implementation: 2. Observations, data pipeline and Level 1 data products. *Journal of Geophysical Research: Space Physics*, *125*, e2020JA027809. https://doi.org/10.1029/2020JA027809