The Cross-track Infrared Sounder (CrIS) NASA PCA RED Product

Joe K. Taylor[1], David. C. Tobin[1],
Graeme Martin[1], Michelle Loveless[1], Greg Quinn[1]

(1) University of Wisconsin-Madison Space Science and Engineering Center, Madison, USA

9 May 2025

25th International TOVS Study Conference (ITSC-25)
Goa, India

With acknowledgements to...

- Tim Hultberg and others involved in the hybrid PCA approach development and application for IASI.
- The CrIS NASA L1B Project, James Gleason (Co-PIs: Joe Taylor, Larrabee Strow).
- Larrabee Strow (UMBC).
- The NOAA CrIS SDR Cal Val Team, Flavio Iturbide-Sanchez (UW-SSEC lead: Dave Tobin).

Outline

- Introduction and Methodology
- Compression and Noise Reduction
- RED (Rapid Event Detection) examples
- Next Steps

Principal Component Analysis (PCA) basics

- Principal Component Analysis (PCA) is a multivariate analysis technique
- Commonly used to reduce the dimensionality of a data set with many interdependent variables
- PCA essentially performs a Singular Value Decomposition (SVD) of the covariance matrix

- A radiance spectrum $R(v)$ with N spectral channels can be written as a linear combination of scalar coefficients C_i and ***orthogonal*** Principal Components $U_i(v)$ plus the ensemble mean spectrum:

$$R(v) = \sum_{i=1}^{N} c_i \, Ui(v) \; + \; Rmean(v)$$

where C and U can be determined from Single Value Decomposition of the radiance ensemble.

- $U_i(v)$ are ordered such that $U_1(v)$ explains the most variance and $U_N(v)$ explains the least, and a reconstructed spectrum can be computed from a relatively small number of Principal Components, N' :

$$R'(v) = \sum_{i=1}^{N'} c_i \, Ui(v) \; + \; Rmean(v)$$

https://www.analyticsvidhya.com/blog/2020/12/an-end-to-end-comprehensive-guide-for-pca/

Principal Component Analysis (PCA) basics

Traditional approaches:

Independent set PCA (Global-only)

- The "training set" and "observation set" are independent
  - the training set does not include the observation set
- PCs only need to be generated and distributed once

Dependent set PCA (Local-only)

- The "training set" and "observation set" are dependent
  - the training set contains the observation set
- Efficiently captures all features in the observation set

Huang et al. (2001). Application of principal component analysis to high-resolution infrared measurement compression and retrieval. Journal of Applied Meteorology and Climatology, 40(3), 365-388.

Antonelli. (2001). Principal component analysis: A tool for processing hyperspectral infrared data. Doctoral Dissertation, The University of Wisconsin-Madison.

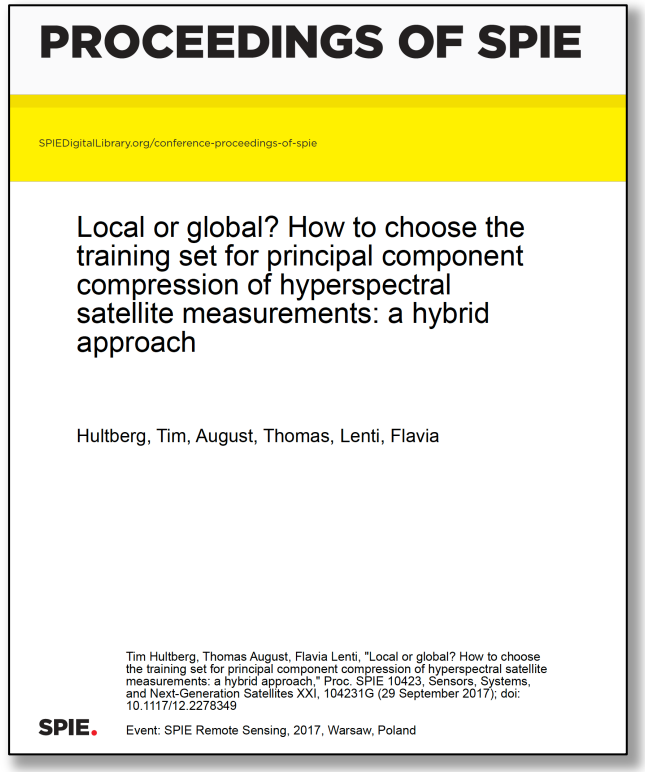
Antonelli et al. (2004). A principal component noise filter for high spectral resolution infrared measurements. Journal of Geophysical Research: Atmospheres, 109(D23).

Turner et al. (2006). Noise reduction of Atmospheric Emitted Radiance Interferometer (AERI) observations using principal component analysis. Journal of Atmospheric and Oceanic Technology, 23(9), 1223-1238.

Tobin et al (2007). Hyperspectral data noise characterization using principal component analysis: application to the atmospheric infrared sounder. Journal of Applied Remote Sensing, 1(1), 013515.

The Hybrid PCA Technique

- **The EUMETSAT Hybrid PCA technique**
  - A combination of global and local PCA approaches
  - Currently available for IASI, and will be used for IASI-NG and MTG-IRS
  - Hultberg et al. 2017 ⟶


PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Local or global? How to choose the training set for principal component compression of hyperspectral satellite measurements: a hybrid approach

Hultberg, Tim, August, Thomas, Lenti, Flavia

Tim Hultberg, Thomas August, Flavia Lenti, "Local or global? How to choose the training set for principal component compression of hyperspectral satellite measurements: a hybrid approach," Proc. SPIE 10423, Sensors, Systems, and Next-Generation Satellites XXI, 104231G (29 September 2017); doi: 10.1117/12.2278349

SPIE.

Event: SPIE Remote Sensing, 2017, Warsaw, Poland

- **Benefits**
  - Data compression and accurate reconstruction
    - ➢ CrIS data accessibility
    - ➢ Preparing for future sensors
  - Random Noise Filtering: sqrt(NPCs/Nchannels), ~73% for CrIS FSR
  - Rapid Event Detection (RED)

- **2022 EUMETSAT Conference**
  - Tobin et al. Hybrid PCA representation of CrIS data.
  - Hultberg et al. IASI hybrid principal component (PC) products becoming operational at EUMETSAT.
  - Prunet et al. Principal Component Analysis of IASI measurements for the detection of extreme atmospheric events.

- **2023 EUMETSAT Conference**
  - Taylor et al. Hybrid PCA representation of CrIS data: Progress and Rapid Event Detection

Recent: Vu Van et al. (2023). Near-real-time detection of unexpected atmospheric events using principal component analysis on the Infrared Atmospheric Sounding Interferometer (IASI) radiances. Atmospheric Measurement Techniques, 16(8), 2107-2127.

CrIS Hybrid PCA

Determine a set of global PCs, "PC_{global}"

- Based on a large representative set of spectra
- The leading $nPCs_{global}$ (150) global PCs are retained
- These global PCs are "static" and distributed to users (along with the global mean spectrum Rad_{global})

For any given granule of data

(1) Compute Coefficients for the global PCs, "**C_{global}**"
(2) Compute "Local PCs" *based on the global reconstruction residuals of this granule*, "**PC_{local}**" and the corresponding coeficients "**C_{local}**" and retain these for the leading $nPCs_{local}$ (10) local PCs (and the local mean residual spectrum Res_{local})
(3) Save C_{global}, PC_{local}, C_{local}, Res_{local} to a file for each granule

Final Reconstructed Radiances:

$$\mathbf{Rad_{Rec} = (PC_{global}\ C_{global} + Rad_{global}) + (PC_{local}\ C_{local} + Res_{local})}$$

[2223xNrec] [2223x150] [150xNrec] [2223x1] [2223 x 10] [10 x Nrec] [2223 x 1]

Global Local

CrIS FSR: ~13.7x data compression for PCA

CrIS Hybrid PCA Product (NASA L1B example, 6-minute granules)

- Standalone product that includes critical ancillary data from the L1B file in the PCA RED file (i.e., latitude, longitude, observation time, QFs, etc).
- 150 global PCs and 10 local PCs
- Scaled integer representation of PC scores combined with NetCDF integer compression
- Sensor dependent global PCs (approximately 671 million radiances per sensor)
- All three spectral bands are combined
- All FOVs are combined
- FOV independent noise normalization
- Store outlier radiances with large hybrid reconstruction residuals (limit to no more than 100 radiances per granule)
- PCA Quality Flag

Data Compression: CrIS Hybrid PCA Product (NASA L1B example, 6-minute granules)

- Sample product for NOAA-20 available via NASA GES DISC
  - based on CrIS NASA L1B Version 3
- Upcoming release for all sensors will be based on the L1B version 4 product.

- Average input file size: 168 MB
- Total input file size: 77 TB

- Average output size: 3.3 MB
- Total output size: 1.540 TB

~50x Data Compression

~13.7x from PCA

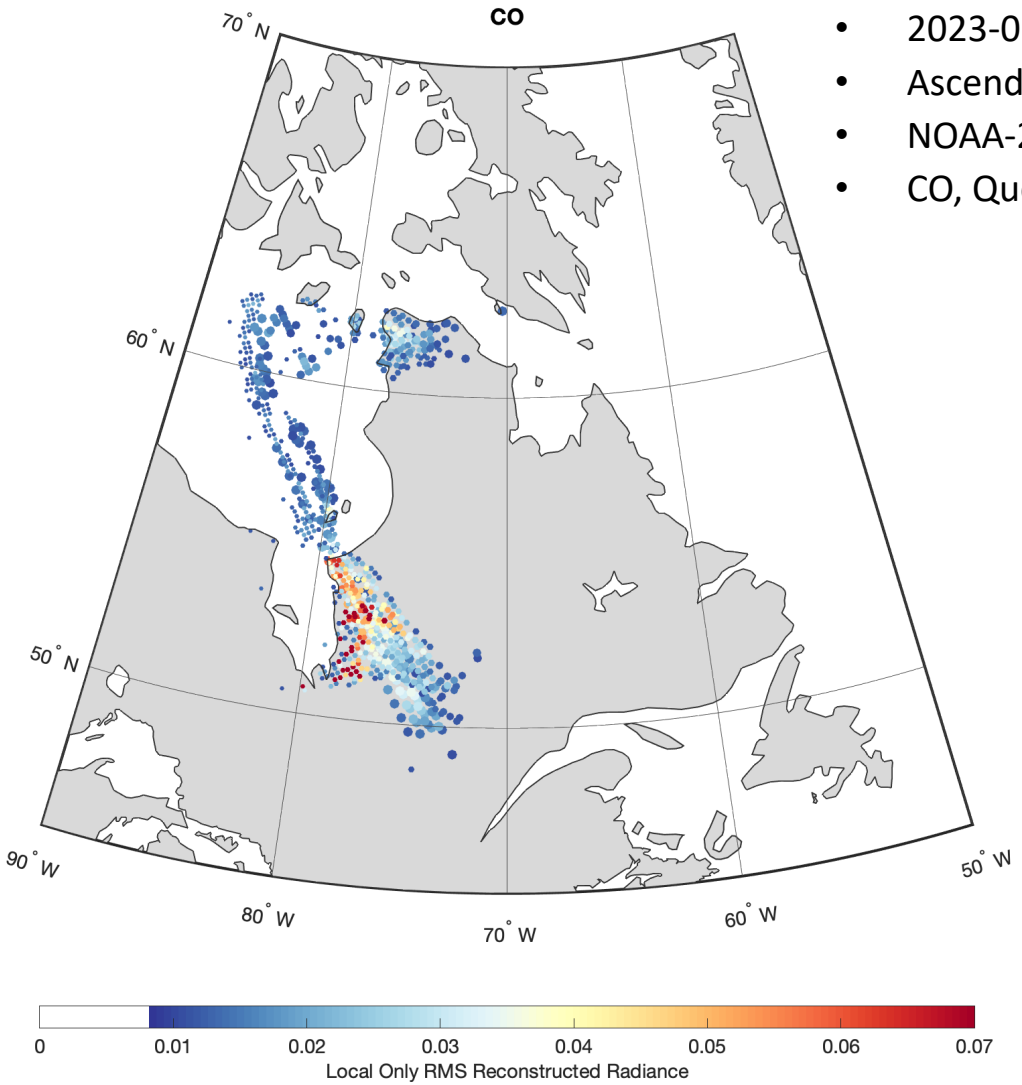
~3.6x from scaled integer rep

Rapid Event Detection: PCA RED Score Overview

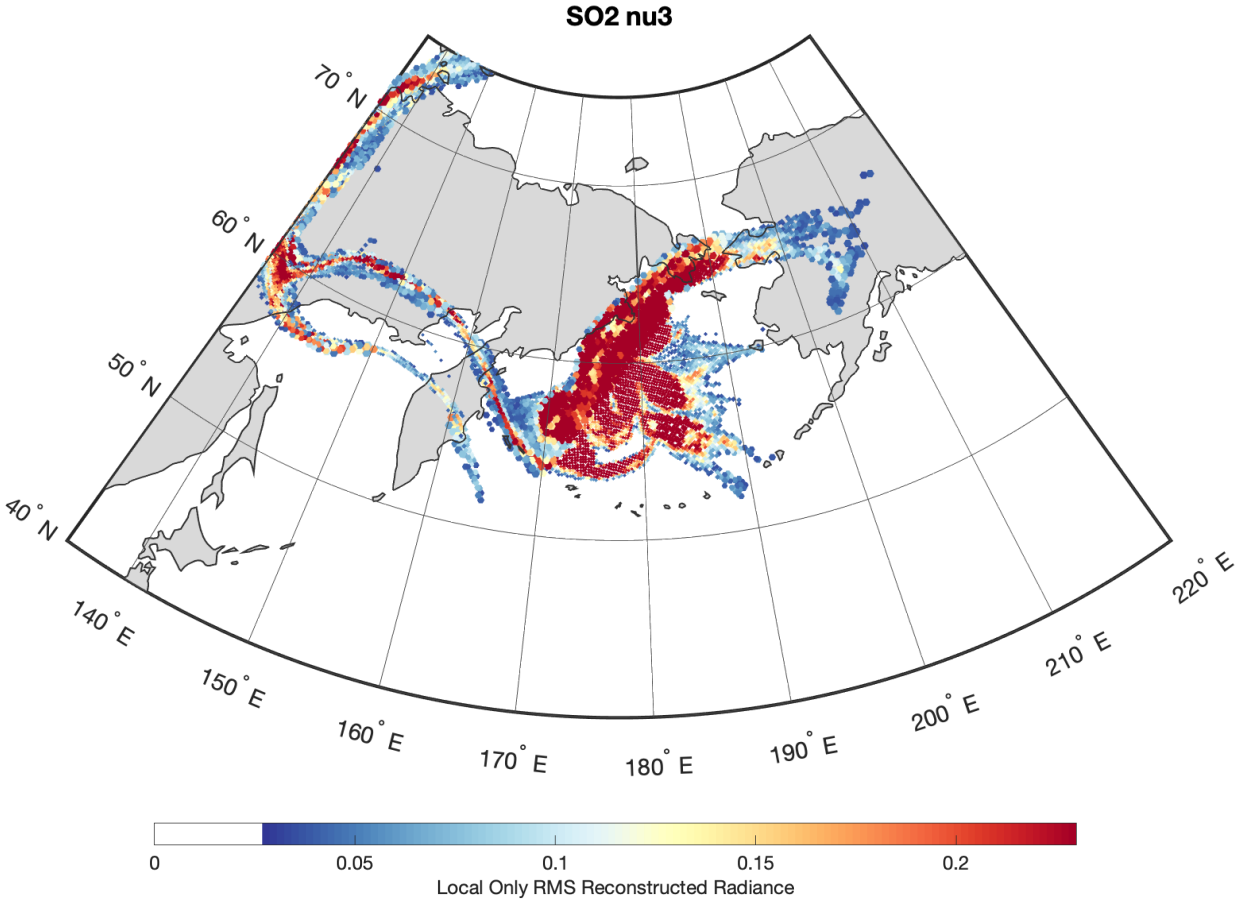
- "De-trended reconstruction scores" are generated for each observation
  - "De-trending" removes detector dependence of the reconstruction score
- 25 Spectral regions:
  - CrIS longwave, midwave, and shortwave bands,
  - Emission/absorption features (e.g., O_3, CH_4, CO, SO_2, HNO_3, N_2O, NH_3, Isoprene, PAN, Acetylene, Ethylene, Ethane, etc.)
- The local-only reconstruction score is used for Rapid Event Detection (RED)
  - PCA RED scale: 0 – 127, with scores in the 0 – 10 region indicative of nominal (nonevent) behavior, and 127 defined as the fill value for missing data.

*Adapted from Hultberg et al: 'IASI PC compression – Searching for signal in the residuals' 13886-iasi-pc-compression-searching-sinal-residuals.pdf and Hultberg, 2009, 'IASI Principal Component Compression (IASI PCC) FAQ' pdf_ipcc_faq.pdf

Examples

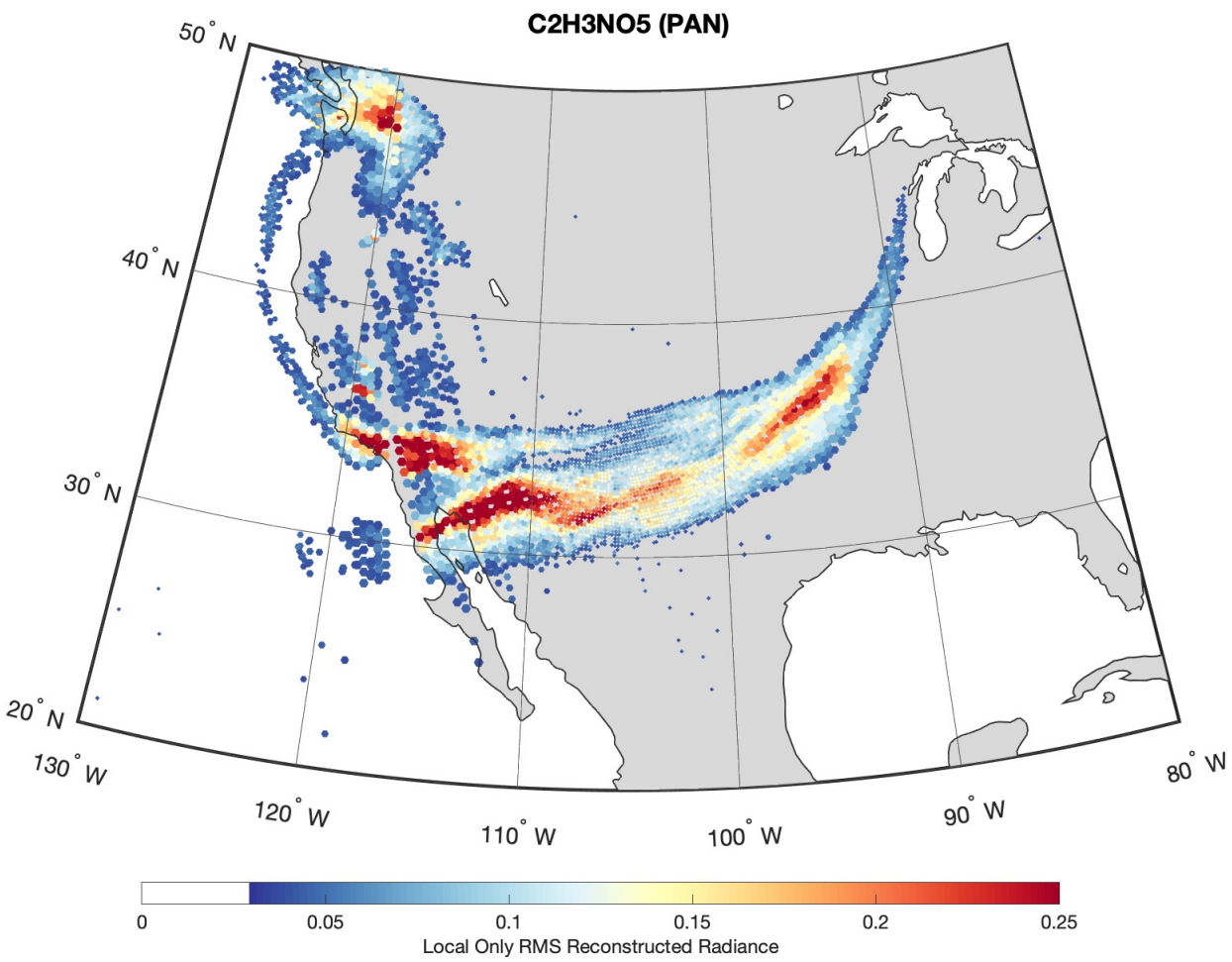


- 2023-07-14
- Ascending Overpasses
- NOAA-20 CrIS
- CO, Quebec Fires

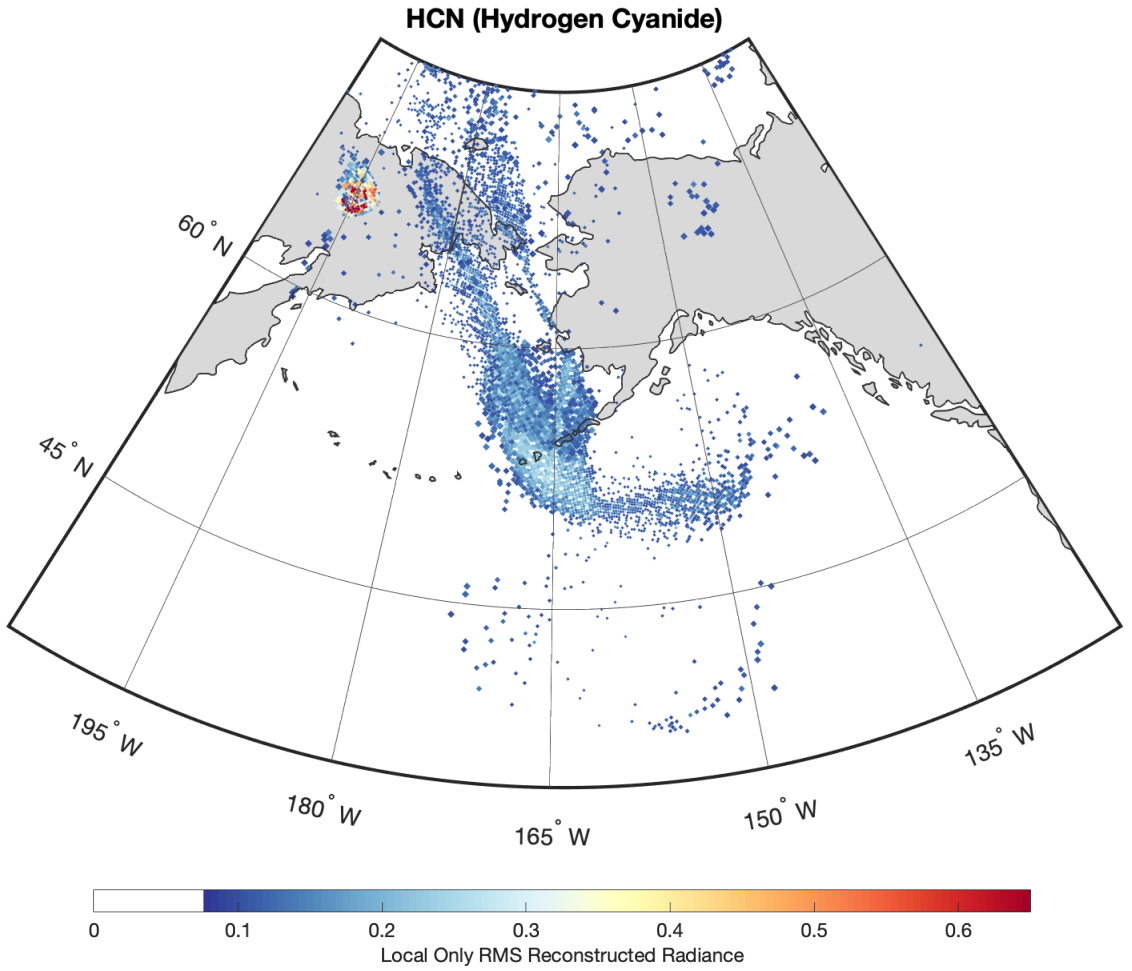


- 2019-06-25
- Descending Overpasses
- NOAA-20 CrIS
- SO2, Post Raikoke Eruption

Examples



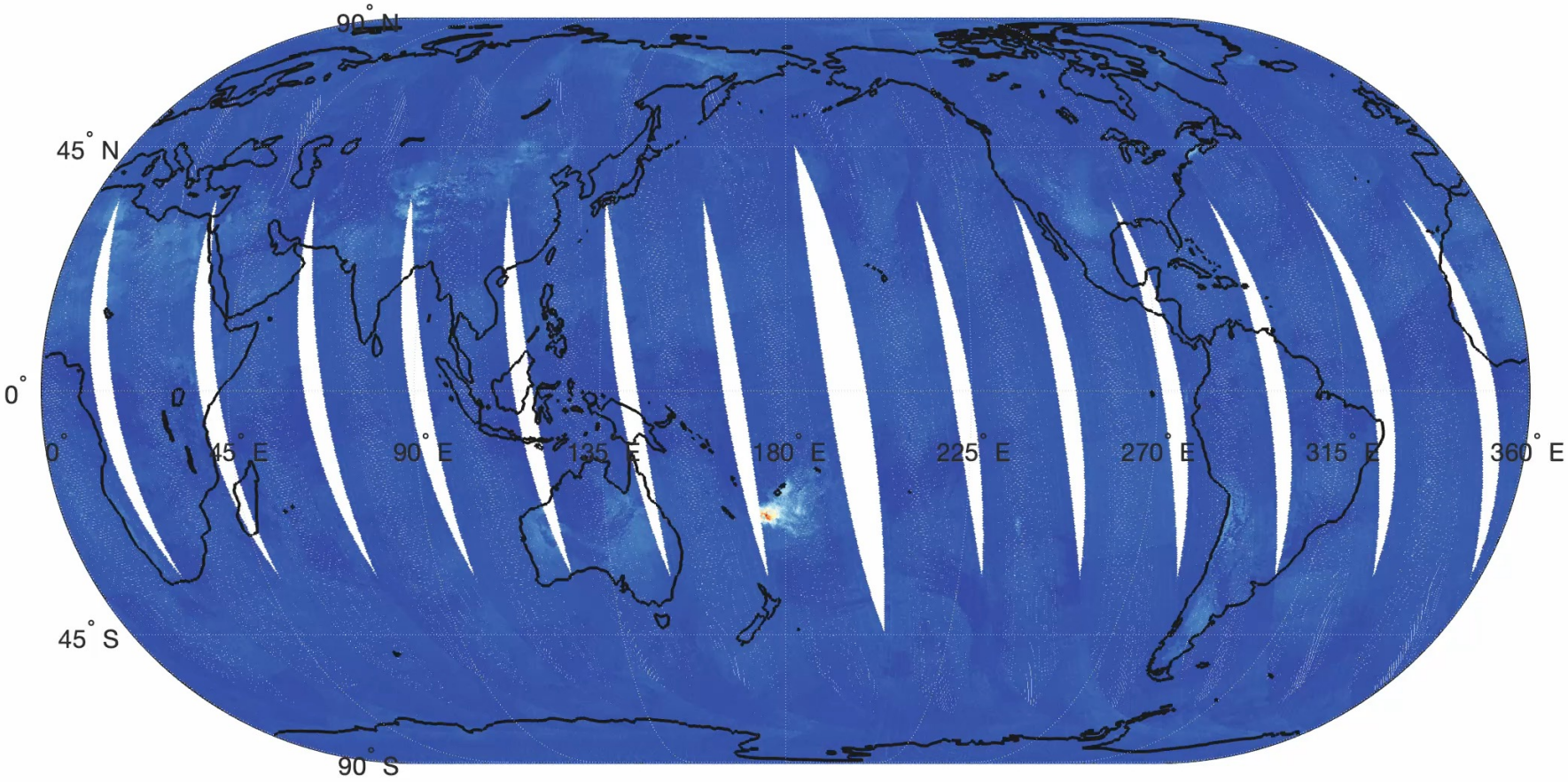
- 2020-09-12
- Ascending Overpasses
- NOAA-20 CrIS
- PAN, Western Canada and California Fires

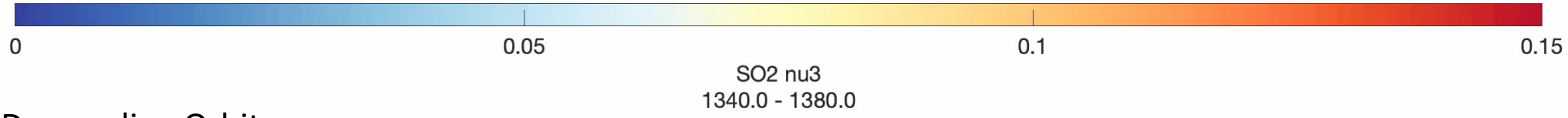


- 2023-07-17
- Descending Overpasses
- NOAA-20 CrIS
- HCN, Russian Fires (and/or Shishaldin?)

Example: Hunga Tonga (Jan 15, 2022 – Jan 25, 2022)

SO_2

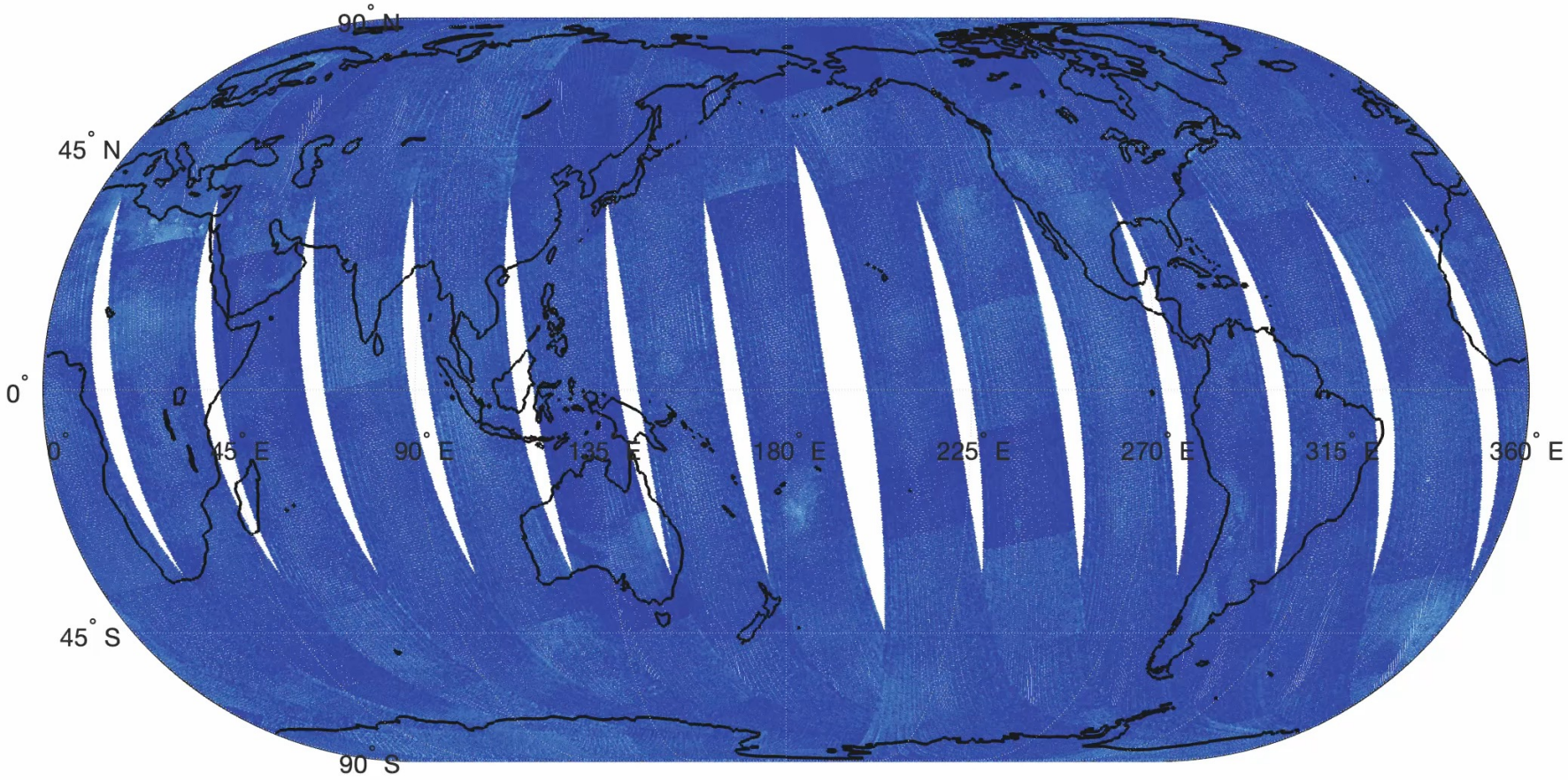


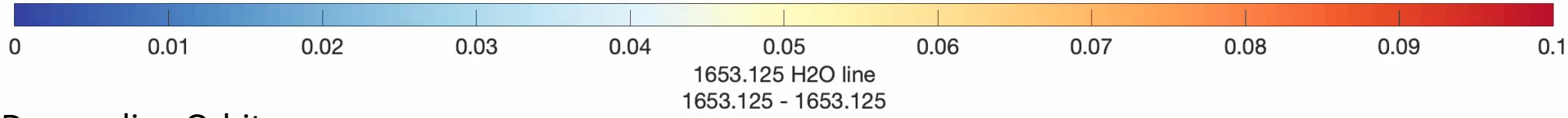


NOAA-20 CrIS
Ascending and Descending Orbits
Local only reconstructed radiances

Example: Hunga Tonga (Jan 15, 2022 – Jan 25, 2022)

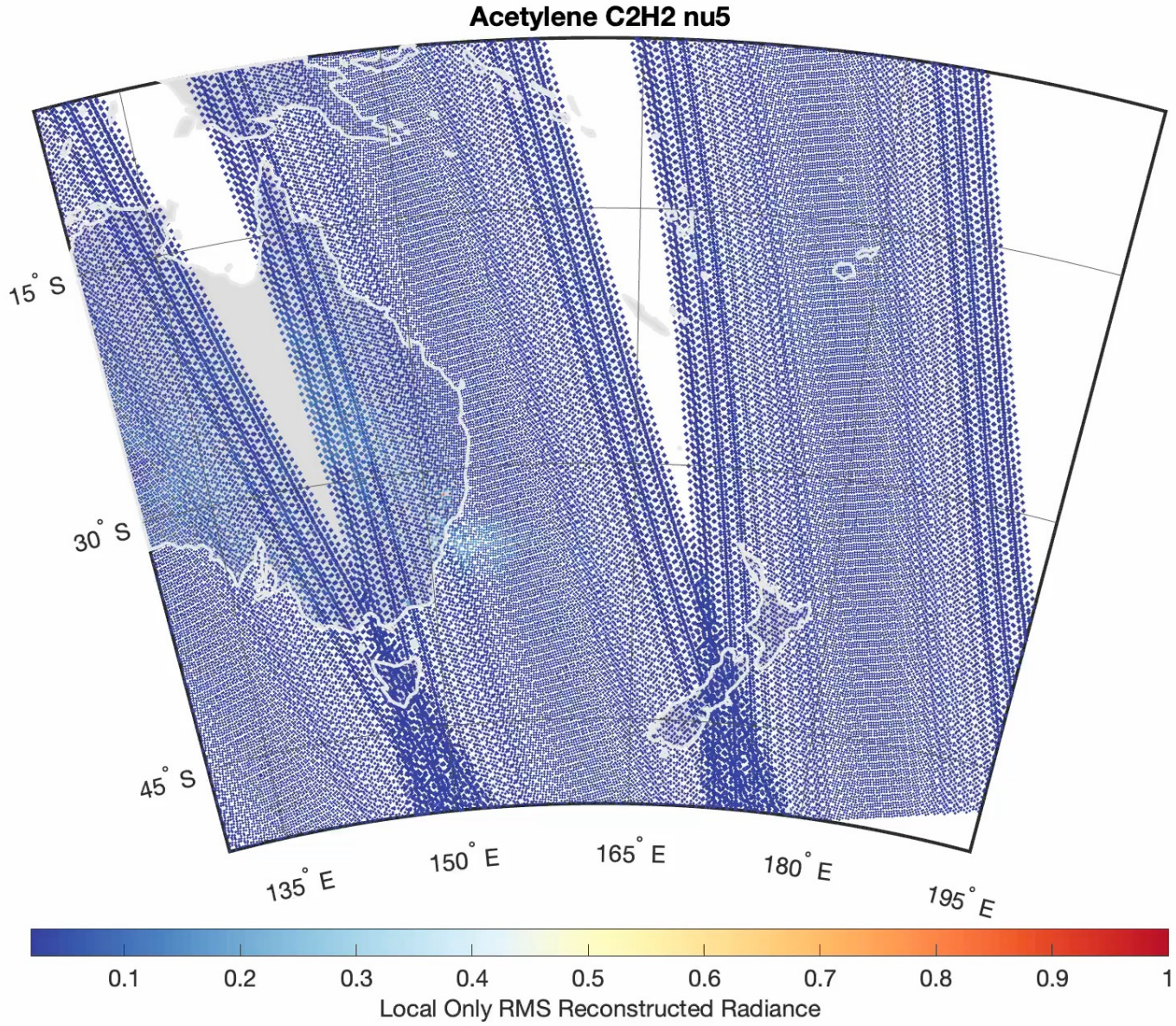
1653.125 cm^{-1}
Water Vapor





NOAA-20 CrIS
Ascending and Descending Orbits
Local only reconstructed radiances

Example: Australian Wildfires



- Dec 26, 2019 – Jan 3, 2020
- NOAA-20 CrIS
- Ascending and Descending Orbits
- Acetylene (C_2H_2)
- Local only reconstructed radiances

Summary and Next Steps

- A beta product has been generated within the CrIS NASA L1B product suite (version 3), and is available via NASA GES DISC
- The initial results are extremely promising for noise reduction, data compression, and Rapid Event Detection.
- We are currently completing a rigorous assessment of the product quality:
  - Refining the Rapid Event Detection (RED) methodology, scales, and radiance outlier thresholds.
  - Confirming the Reconstructed Radiance accuracy versus the native L1B calibrated radiances.
  - Assessing product QC and QF flags.
- Next steps:
  - Refine product file format and content (if needed).
  - Documentation.
  - Release a PCA RED product based on the CrIS NASA L1B v4 product (all CrIS sensors)
  - Create a NASA Worldview display for Rapid Event Detection (RED) and the existing IMG (collocated CrIS/VIIRS) product.

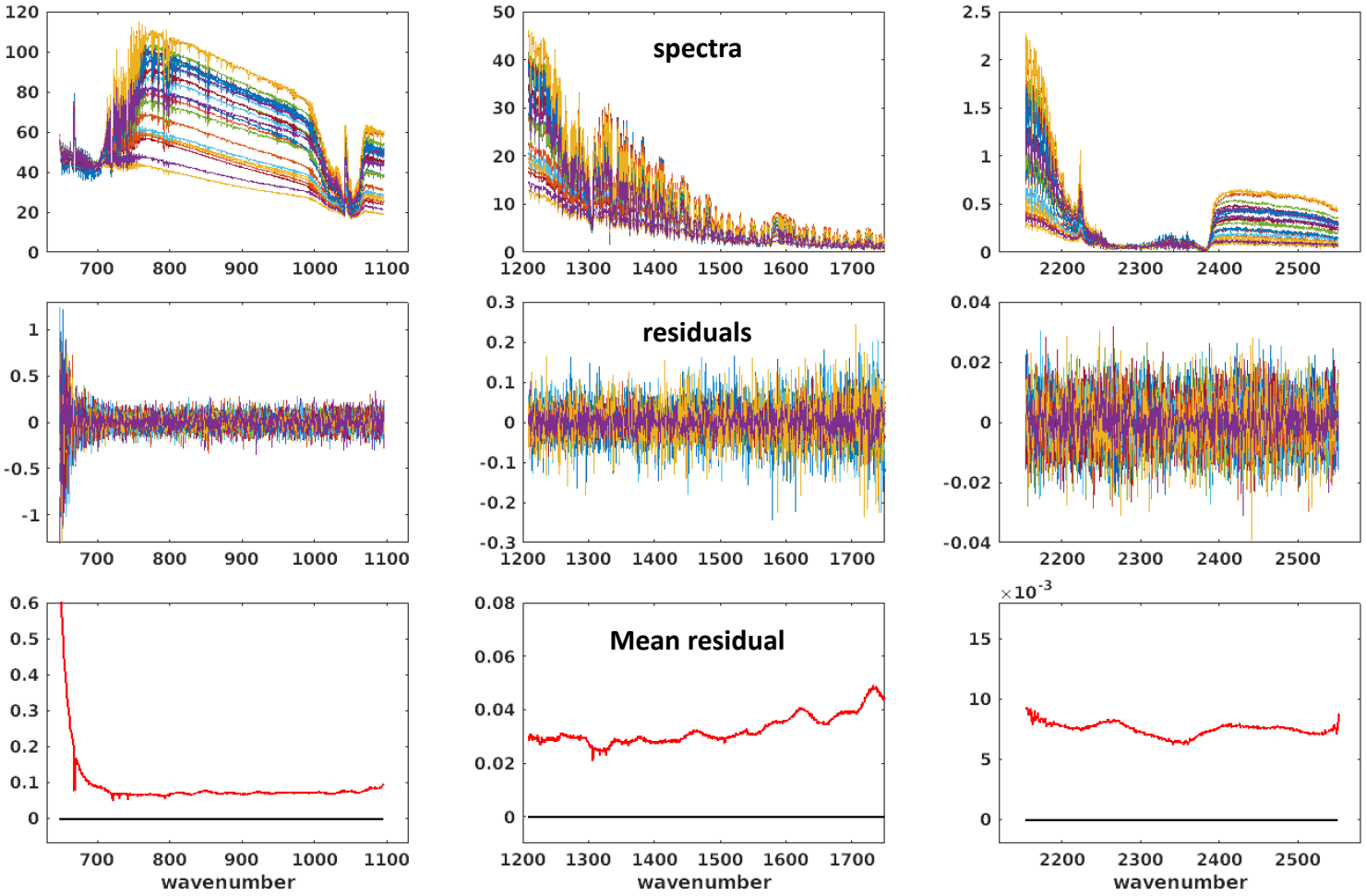
image courtesy R. Knuteson

Additional Material

Noise Reduction

- Approximately 73% reduction in random noise for CrIS FSR
  - 2223 spectral channels
  - 160 eigenvectors (150 global, 10 local)
  - sqrt($N_{PCs}/N_{channels}$)

20210413 NOAA20 radiances



STD DEV residual: looks like CrIS NESR

Example: Mount Etna, 2021-02-23, SO_2 Plume (NOAA-20)