Hybrid PCA representation of CrIS data

David Tobin[1], Joe Taylor[1], Flavio Iturbide-Sanchez[2], Liam Gumley[1], Graeme Martin[1], Michelle Loveless[1], Bob Knuteson[1], Hank Revercomb[1]

[1] Cooperative Institute for Meteorological Satellite Studies, Space Science and Engineering Center, University of Wisconsin-Madison

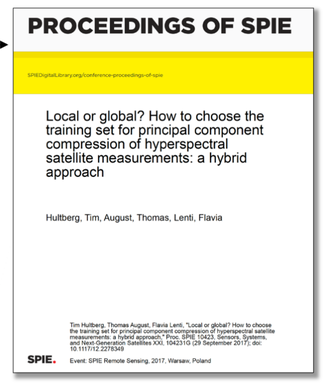
[2] NOAA, NESDIS, STAR

ITSC-24 Poster 12p.07

dave.tobin@ssec.wisc.edu

Introduction

- **The EUMETSAT Hybrid PCA technique**
  - e.g.: Hultberg et al. 2017
  - Currently available for IASI, and will be used for IASI-NG and MTG-IRS
- **This work adapts the Hybrid methodology for use with CrIS**
- **Benefits**
  - Data compression, up to ~50x
    - ➢ CrIS data accessibility
    - ➢ Preparing for future sensors
  - Random Noise Filtering
  - Rare Events Detection (RED)



CrIS Hybrid PCA Methodology

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 (and the global mean spectrum Rad_{global})

For any given granule of SCRIF 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:

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

[2223xNrec] [2223x150] [150xNrec] [2223x1] — Global; [2223 x 10] [10 x Nrec] [2223 x 1] — Local

CrIS Hybrid PCA Methodology (Cont.)

Other considerations:
- All three spectral bands combined
- All nine FOVs combined
- 150 global PCs and 10 local PCs
- Sensor independent or sensor dependent global PCs
- Noise Normalization (FOV independent)
- Scaled integer representation of PC scores combined with NetCDF integer compression
- PC representation Quality indicators
- Individual Outliers
- Local granule size
- Missing spectral bands

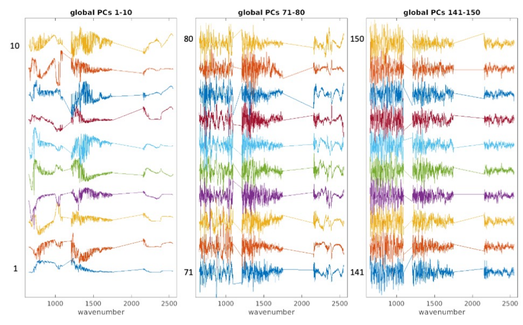
Global Training

global PCs
- Based on a large representative set of spectra
- For each sensor (SNPP and NOAA20) currently using spectra from 1 year of data (March 2018 to March 2019)
- with latitude weighting
- ~671M spectra per sensor

- PCs computed from the SVD of covariance of the radiance ensemble
- Covariance matrices of subsets (single days) of the ensemble combined into a covariance matrix for the entire ensemble following:
  - Pebay, P., 2008, Formulas for Robust, One-Pass Parallel Computation of Covariances and Arbitrary-Order Statistical Moments: Sandia Report

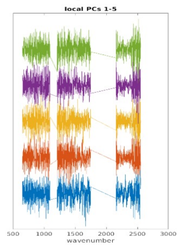
Example PCs

Global PCs



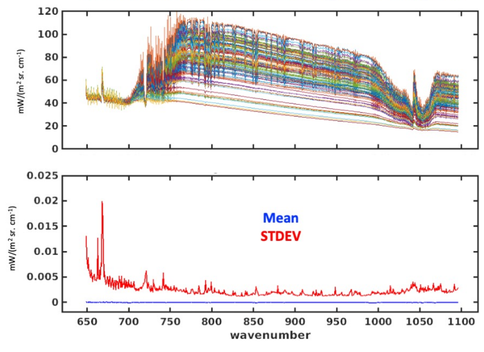
Local PCs

For a randomly selected SCRIF granule



Scaled Integer representation of PC scores

Example reconstructed radiance differences
32 Bit Float vs 16 Bit Unsigned Integers



Compression (SDR example)

Radiances from a 4 scan line CrIS granule (1080 spectra): 9,551,520 bytes

Global PCs (150 PCs, 1 time transfer):	2,653,200 bytes
Global Coefs (per granule):	648,000 bytes
Local PCs (10 PCs, per granule):	88,440 bytes
Local Coefs (per granule):	43,200 bytes

Per granule:
FSR SDR Radiances = 9,551,520 bytes
PC representation = 648,000 + 88,440 + 43,200 bytes = 779,640 bytes

➢ **PC/SDR x100 = 8.2% … PC data volume is ~8 percent of SDR radiance data volume**
➢ **SDR/PC = 12.3 … A factor of ~12 reduction in terms of data transfer expense and data storage**

And an additional compression of 2 to 4x, from scaled integer representation of the PC scores, gives **total compression of up to ~50x**

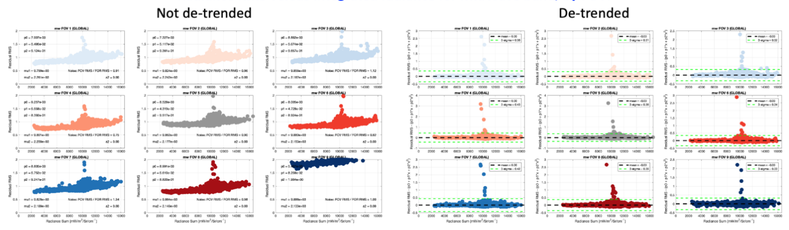
This large reduction is due to the lossy noise reduction/compression from the PCA …

PC reconstruction Quality indicators

- "De-trended" reconstruction RMS scores for each footprint and spectral band, for Global-only and Hybrid reconstructions
- Corresponding Quality Flags
- Original radiances for "Individual outlier" spectra, when identified, included in output PC files
- Useful for "rare event" identifications

Example for 20220101 Mt. Etna eruption NOAA20 data (using Global-only PC reconstruction for demonstration)
Radiance RMS versus integrated radiance for MidWave band, by FOV:

Not de-trended — De-trended

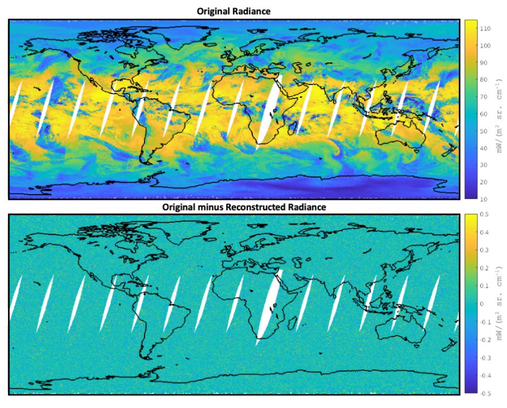


*Following Hultberg et al: 'IASI PC compression – Searching for signal in the residuals' 13886-iasi-pc-compression-searching-sinal-residuals.pdf and Hultberg, 2009, 'IASI Principal Component Compression (IASI PCC) FAQ' pdf_ipcc_faq.pdf

Random Noise Filtering

Example reconstruction residuals

900 cm^{-1} radiances
20210413 NOAA-20 data



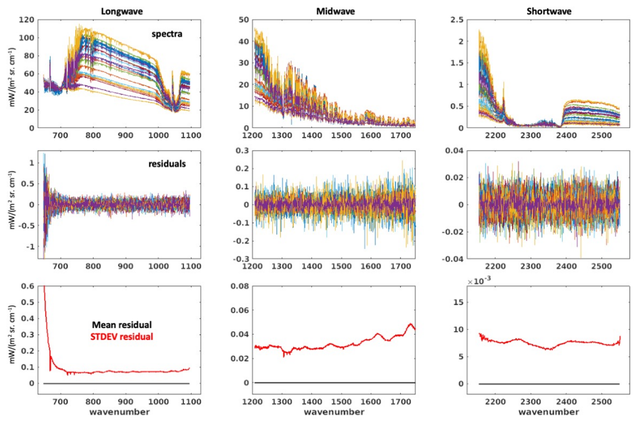
Random Noise Filtering

Example reconstruction residuals

20210413 NOAA20 radiances

~75% random Noise reduction [sqrt(160/2223)]

produces data with FOV independent random noise levels

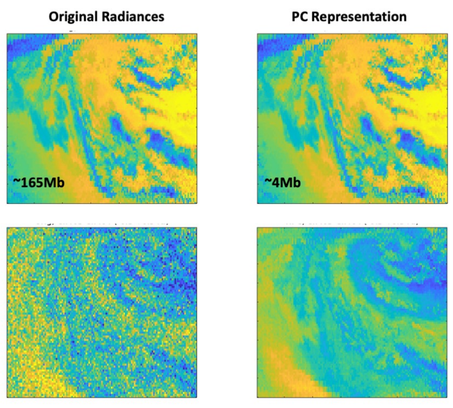


Data Compression and Random Noise Filtering

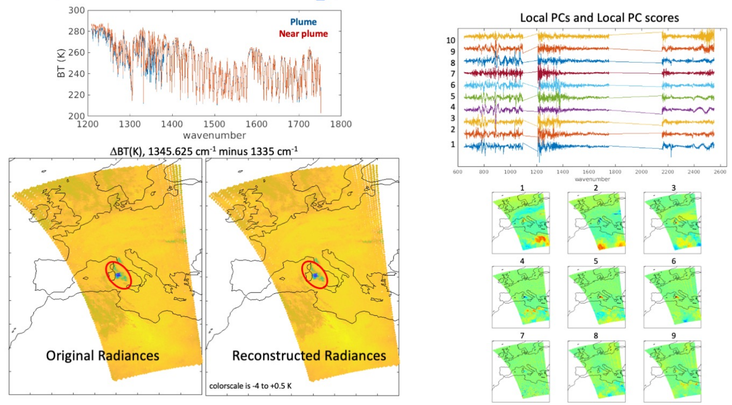
Example 40 scan line file

Image of 900 cm^{-1} radiances:
Color scale is 20 to 100 mW/(m²sr. 1/cm)

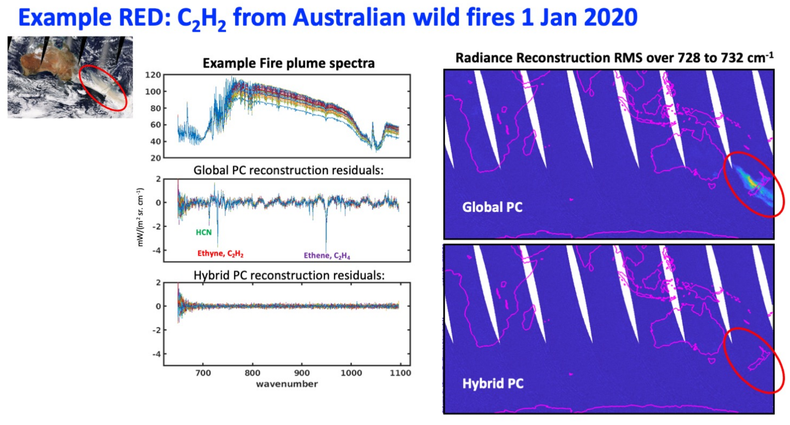
Channel difference image:
900 cm^{-1} minus 900.625 cm^{-1}:
Color scale is -0.3 to 0.3 mW/(m²sr. 1/cm)



Example RED: Mt. Etna SO_2 plume on 23 Feb 2021

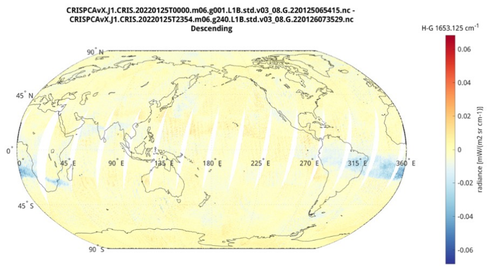


Example RED: C_2H_2 from Australian wild fires 1 Jan 2020



Example RED:

Hunga Tonga, Local contribution @ 1653.125 cm^{-1} 2022-01-25



Next Steps

- **Implement CrIS HPC software within the Community Community Satellite Processing Package (CSPP) and CrIS/VIIRS ULL processing**
  - Define product files, Documentation, provide software utilities, facilitate test users
- **Continue to assess the PC product quality**
  - A more comprehensive assessment using more data
  - Including reconstructed radiances, quality variables, and compression
  - Inputs from other groups (retrievals, NWP)
- **Implement as a prototype L1B project product, and create Worldview display for Rare Event Detection (RED), combining with existing IMG (collocated CrIS/VIIRS) product**
  - Define product files, Documentation, provide software utilities, facilitate test users