

Earth Mineral dust source InvesTigation (EMIT)

EMIT Greenhouse Gas Algorithms: Greenhouse Gas Point Source Mapping and Related Products

Theoretical Basis

Philip G. Brodrick, Jet Propulsion Laboratory, Pasadena, CA
Andrew K. Thorpe, Jet Propulsion Laboratory, Pasadena, CA
Claire S. Villanueva-Weeks, Jet Propulsion Laboratory, Pasadena, CA
Clayton Elder, Jet Propulsion Laboratory, Pasadena, CA
Jay Fahlen, Jet Propulsion Laboratory, Pasadena, CA
David R. Thompson, Jet Propulsion Laboratory, Pasadena, CA

Version 1.0

August 2023



Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California 91109-8099

Change Log

Version	Date	Comments
1.0	August 30, 2023	Initial Draft

Table of Contents

Coverpage	1
Table of Contents	2
1. Key Team Members	3
2. Background on the EMIT Mission and its Instrumentation	3
3. Algorithm Rationale	4
4. Algorithm Description	5
4.1 Input data	5
4.2 Theoretical description	5
4.2.1 Methane Enhancements	5
4.2.2 Plume Complex Identification	6
4.3 Practical Considerations	9
5. Output Data	9
6. Calibration, uncertainty characterization and propagation, and validation	10
6.1 Uncertainty quantification	10
6.2 Validation efforts	10
7. Constraints and Limitations	10
8. Code Repository and References	11
References	11

1. Key Team Members

A large number of individuals contributed to the development of the algorithms, methods, and implementation of the greenhouse gas mapping approach for EMIT. The primary contributors are the following:

- Philip G. Brodrick (Jet Propulsion Laboratory)
- Andrew K. Thorpe (Jet Propulsion Laboratory)
- David R. Thompson (Jet Propulsion Laboratory)
- Clayton Elder (Jet Propulsion Laboratory)
- Claire S. Villanueva-Weeks (Jet Propulsion Laboratory)
- Jay Fahlen (Jet Propulsion Laboratory)
- Red Willow Coleman (Jet Propulsion Laboratory)
- K. Dana Chadwick (Jet Propulsion Laboratory)
- Mandy Lopez (Jet Propulsion Laboratory)
- Daniel Jensen (Jet Propulsion Laboratory)
- Robert O. Green (Jet Propulsion Laboratory)

2. Background on the EMIT Mission and its Instrumentation

Mineral dust aerosols originate as soil particles lifted into the atmosphere by wind erosion. Mineral dust created by human activity makes a large contribution to the uncertainty of direct radiative forcing (RF) by anthropogenic aerosols (USGCRP and IPCC). Mineral dust is a prominent aerosol constituent around the globe. However, we have poor understanding of its direct radiative effect, partly due to uncertainties in the dust mineral composition. Dust radiative forcing is highly dependent on its mineral-specific absorption properties. The current range of iron oxide abundance in dust source models translates into a large range of values, even changing the sign of the forcing (-0.15 to 0.21 W/m²) predicted by Earth System Models (ESMs) (Li et al., 2020). The National Aeronautics and Space Administration (NASA) recently selected the Earth Mineral Dust Source Investigation (EMIT) to close this knowledge gap. EMIT was launched on July 14, 2022 to the International Space Station (ISS) to directly measure and map the soil mineral composition of critical dust-forming regions worldwide.

In addition to its primary objective described above, EMIT has demonstrated the capacity to characterize methane (CH₄) and carbon dioxide (CO₂) point source emissions by measuring gas absorption features in the shortwave infrared (Fig. 1, Thorpe et al., in press). This document breaks from the other mission Algorithm Theoretical Basis Documents (ATBDs), as the CH₄ and CO₂ products are not part of the standard series of interconnected science products. Readers should consult the LIB ATBD for the precursor products to what are used here.

The EMIT Project is part of the Earth Venture-Instrument (EV-I) Program directed by the Program Director of the NASA Earth Science Division (ESD). EMIT is comprised of a VSWIR Infrared Dyson imaging spectrometer adapted for installation on the International Space Station (ISS). EMIT measures radiance between 380 and 2500 nanometers, with an approximate 7 nm bandpass. Data are collected in a swath that is approximately 75 km wide at the equator, with an approximate ground sampling distance of 60 m.

Figure 2 shows the regions that were the primary focus of the nominal 1-year mission with a focus on mapping arid regions of the globe. The expected distribution of the EMIT revisit rates is also shown, with a median of 10 revisits per annum, and 90% of the target mask having between 3 and

15 revisits. Beyond the primary 1-year mission, any terrestrial region between ($+51.6^\circ$ and -51.6° latitude) could be targeted.

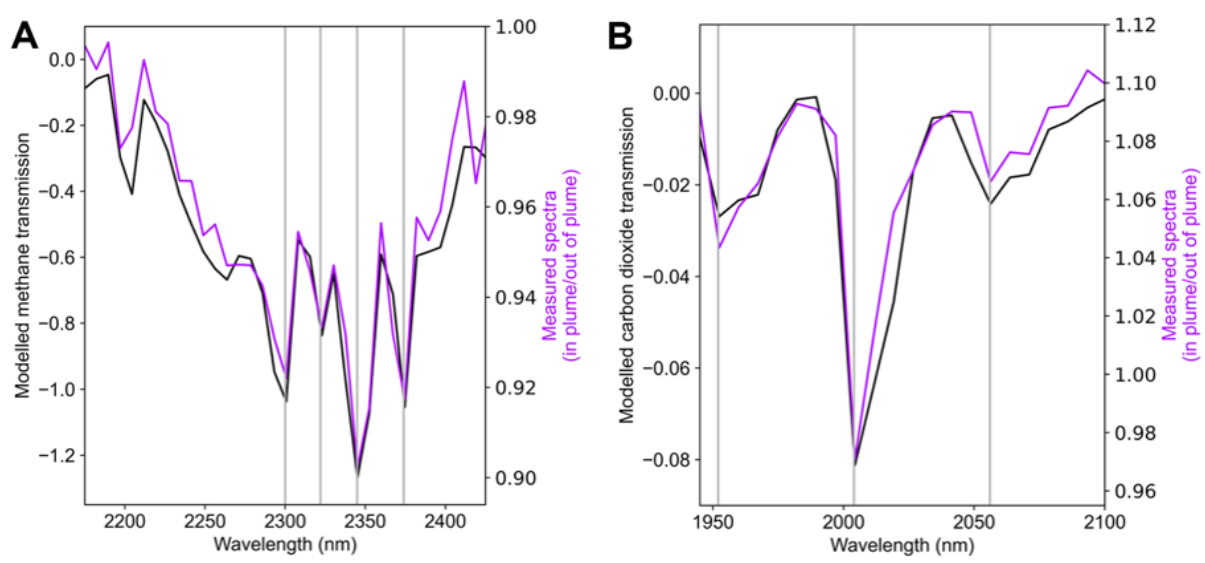


Figure 1: Examples of methane (A) and carbon dioxide (B) spectral fingerprints (purple lines) and modeled gas transmission (black line). The strongest absorption features highlighted by the vertical gray lines (Thorpe et al., in press).

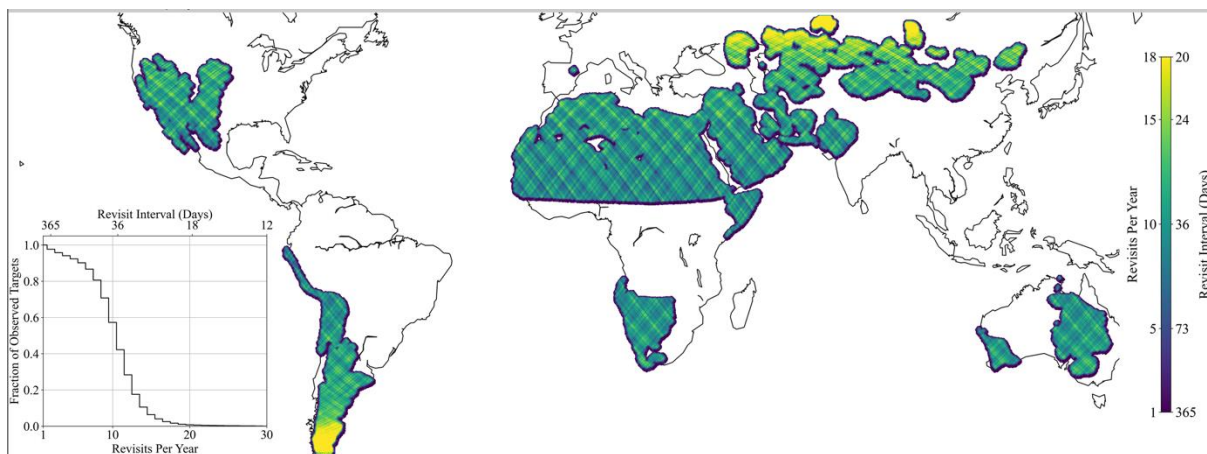


Figure 2: Total observations and average revisit interval for nominal 1-year mission (Thorpe et al., in press).

3. Algorithm Rationale

The EMIT GHG Point Source Detection approach builds on a substantial history of remote greenhouse gas detections from airborne imaging spectrometers (Thorpe et al., 2013, 2014, 2017; Thompson et al., 2015; Frankenberg et al., 2016; Duren et al., 2019; Cusworth et al., 2022). We leverage a per-column adaptive matched filter for the primary detection, due to the speed and efficacy of identifying subtle signatures. We customize the deployment in order to minimize false positives, after which a series of scientists manually review individual plumes in order to provide only instances with maximum confidence.

4. Algorithm Description

4.1 Input data

Input and output products for the EMIT CH₄ product are held as a combination of binary data cubes with detached human-readable ASCII header files, and cloud optimized GeoTIFFs (COGs; which ultimately constitute the delivery to the LP DAAC). The former adhere to the ENVI standard (accessible Aug 2023 at <https://www.nv5geospatialsoftware.com/docs/ENVIHeaderFiles.html>), and the latter adhere to the COG specification (<https://github.com/cogeotiff/cog-spec/blob/master/spec.md>, accessible Aug 2023). The header files all consist of data fields in equals-sign-separated pairs and describe the layout of the file. In the file descriptions below, n denotes the number of lines particular to the given scene, which can vary – it is normally 1280, but for some scenes at the trailing end of an orbit segment, may reach as high as 2559.

The specific input files needed for the primary GHG detection are:

1. **Measured radiance at sensor**, provided as an $n \times 1242 \times 285$ BIL interleave data cube (rows, columns, wavelength).
2. **Observation geometry**, provided as $n \times 1242 \times 11$ BIL interleave data cube (rows, columns, bands).
3. **Geographic lookup table**, provided as an orthorectified BIL interleave binary data set with 2 bands (x and y lookup indices) and a variable number of rows and columns.
4. **Mask file**, provided as an $n \times 1242 \times 7$ BIL interleave data cube (rows, columns, bands) – this is an output product of L2A with the shortname EMIT_L2A_MASK on delivery.
5. **Bandmask file**, provided as an $n \times 1242 \times 36$ BIL interleave data cube (rows, columns, bands) – this is an output product of L2A with the shortname EMIT_L2A_MASK on delivery.

4.2 Theoretical description

The procedures that follow outline a procedure for identifying methane enhancements that are clearly separable from any background noise, and which can be identified as point source emissions. As with all measurements, instrument noise and algorithmic error present some amount of false positives, which we mitigate through manual review of identified plume complexes. We describe the algorithms used here but encourage interested parties to examine the publicly available code for all details, as the precise implementation of the procedure is ultimately quite important. All algorithms used here are based heavily on existing literature (Thompson et al., 2015).

4.2.1 Methane Enhancements

EMIT’s methane enhancement product is created using an adaptive matched filter. This method searches each spatial pixel’s radiance spectrum for deviations that are characteristic of methane’s absorption spectrum. Its sensitivity is increased by using the covariance matrix estimated from the data to “whiten” or decorrelate the data during processing. There are three main components to the matched filter algorithm: the signal vector \mathbf{s} , the spectral covariance matrix \mathbf{C} and the de-meaned data $\mathbf{x} - \boldsymbol{\mu}$ for each spatial pixel. The matched filter expression is

$$MF = \frac{s^T C^{-1} (x - \mu)}{s^T C^{-1} s},$$

where μ and C are estimated from the data using the spectral radiance in each column separately. The covariance matrix is calculated using a simple shrinkage method in which the standard empirical covariance is weighted by $(1-a)$ and summed with its diagonal weighted by a , with $a = 10^{-9}$. The signal vector is defined to be the change in radiance with respect to a unit change the methane concentration length and is computed from two components. The first component is computed as a normalized finite difference derivative of the radiance at two different simulated methane quantities, given by $t = \frac{1}{L} \frac{\Delta L}{\Delta c}$, where L is computed using MODTRAN 6 with the amount of water vapor determined using the L2 retrieved water vapor quantity and Δc represents the two simulated methane concentration lengths in the MODTRAN runs used to compute ΔL . In order to better match the radiance and its calibration, the second component is simply the mean radiance, again computed column by column. Combining the two yields $s = t\mu$ evaluated as an element-wise product.

Matched filter performance is determined in part by calculating the covariance estimate using representative samples of the data. Accordingly, each spectrum in a single column is used as a sample, excluding saturated pixels, pixels flagged as clouds or surface water, and unusually bright pixels (“flares” or other hot objects). Pixels are flagged as flares using a simple radiance threshold on channel 270 (approximately 2389 nm). The threshold is the largest radiance that is likely to occur without a flame as determined using MODTRAN to model the radiance in a nadir look with the sun directly overhead having no atmospheric water vapor or methane and assuming 80% surface reflectivity. Very few materials have reflectivity exceeding 80% in this channel. Any spectra exceeding this threshold without having flares are likely to be sufficiently unusual that their exclusion from the covariance estimate will not be detrimental. These bright pixels (along with those flagged as clouds or water from the L2A mask) are excluded from the mean and covariance estimates, and also subsequently from the matched filter results, as the radiance enhancement from thermal emission can exaggerate matched filter detections.

4.2.2 *Plume Complex Identification*

The methane enhancement data contains elevated values over true methane sources, but also contains false positives induced by various surface or atmospheric characteristics. To mitigate the presence of these false positives, each scene undergoes a manual review process before being designated as a plume, where three different reviewers check each plume complex. Here, we use the term ‘plume complex’ rather than ‘plume’ because in many cases individual plumes cannot be isolated from one another given the observation geometry and spatial resolution of the instrument; a plume complex may be a single plume or composed of multiple overlapping plumes. Plume complexes may also not be true plumes – general enhancements from a given point source are also included if there is sufficient evidence that the enhancement is not a false positive, even if the enhancements lack a characteristic plume shape. We describe the manual identification and review procedure below in the order it is performed. In order for a plume complex to be reported as output, it must be identified by a primary reviewer, and pass checks by *both* reviewer 2 and reviewer 3. Reviewers may communicate with one another to improve the overall procedure, and to discuss challenging cases, but concurrence must be reached prior to publication.

The process is designed to be conservative – if a plume complex is questioned by one of the reviewers, it is excluded from delivery. Because of this conservatism, the product does not include every point source enhancement, and should not be treated as comprehensive. An internal confidence scheme is used by the reviewers strictly for their communication – any plume that makes it through the complete review process, however, is deemed high confidence.

All plume identification and review is conducted through custom layers of the Multi Mission Geospatial Information System (MMGIS), which has been configured to view matched filter and quicklook images from EMIT as they are produced by the Science Data System, and which records observations in a database for subsequent use – see Figure 3. A series of background scripts record information from reviewers and facilitate iteration on candidate plume complex imagery. The review process proceeds as follows:

Reviewer 1 – Primary Plume Complex Identification

Reviewer 1 is responsible for primary identification of candidate plume complexes. Understanding that two reviews follow, Reviewer 1 may be more permissive in the inclusion of potential plume complexes than subsequent reviews. Generally, Reviewer 1 follows a finding and flagging scheme, where any enhancement substantially elevated from the background is considered ‘found’, but may receive a ‘flag’ for a number of known issues that may lead to false positives. If a potential plume complex receives a flag (for reasons described below) it is discarded. The following sequence is used by Reviewer 1:

1. Gain confidence of the full scene
 - Assign flags for scenes overwhelmingly dominated by cloudiness, high levels of background noise, snow, and extreme topography.
2. Assess cloud cover
 - Clouds and water are masked (using masks from the L2A_MASK) during the matched filter generation, but as the masks themselves are imperfect, some amount of cloud contamination – often edges – makes it through the matched filter result. Clouds often are shown as bright in the matched filter result, and the light field can be distorted in immediate cloud proximity. Therefore Reviewer 1 should maintain a sense of cloud coverage (using scaled radiance data from EMIT as reference) within the scene, and ultimately flag any potential plume complexes that are on cloud edges or shadows.
3. Visually scan matched filter results at a coarse zoom level
 - a. Identify obvious enhancements – generally very high levels, and/or with distinct plume morphology. If no flags are identified, these are deemed ‘very high’ confidence detections.
 - b. Toggle between matched filter and EMIT radiance data (RGB), checking for clear correlations with surface features and known problem surfaces (e.g. solar panels).
 - c. When a suspected plume complex does not match the EMIT RGB layer and is not over extreme topography nor in immediate proximity to clouds, look for a plausible source like infrastructure and identify if it is a potential methane emitter. This is done using high resolution basemap imagery from multiple sources, with some external research on nearby facilities or by using google maps and street view to look for less visible viable methane emitters (trash, oil pumps, mine shafts). If no possible infrastructure is apparent, label the plume complex as ‘mystery’ for further examination. Mystery plume complexes will be accepted if

the signal is strong and fully uncorrelated with surface features, and/or if there is temporal evidence to support the plume complex (see Reviewer 2).

4. Visually scan the matched filter at a fine zoom level

Identify enhancements that still stand out from the background but are not as clear those observed from the full scene view. A confidence level of ‘high’, ‘medium’, and ‘low’ are then given based on the enhancement magnitude, enhancement size (area), and the presence of infrastructure. Steps 3b and 3c from above are repeated.

Enhancements that are tagged as ‘low’ in this confidence scheme are discarded (not reviewed by Reviewer 2 or 3) but are preserved for future review and analysis through time.

Any time a potential plume complex is identified by Reviewer 1, it is recorded into MMGIS. Additional metadata is recorded, and a global unique identifying index is assigned. The reviewer delineates the enhancement from the background, which is also subject to subsequent review. Delineation includes a small (several pixel) buffer, to demonstrate separation between the enhancement and the background.

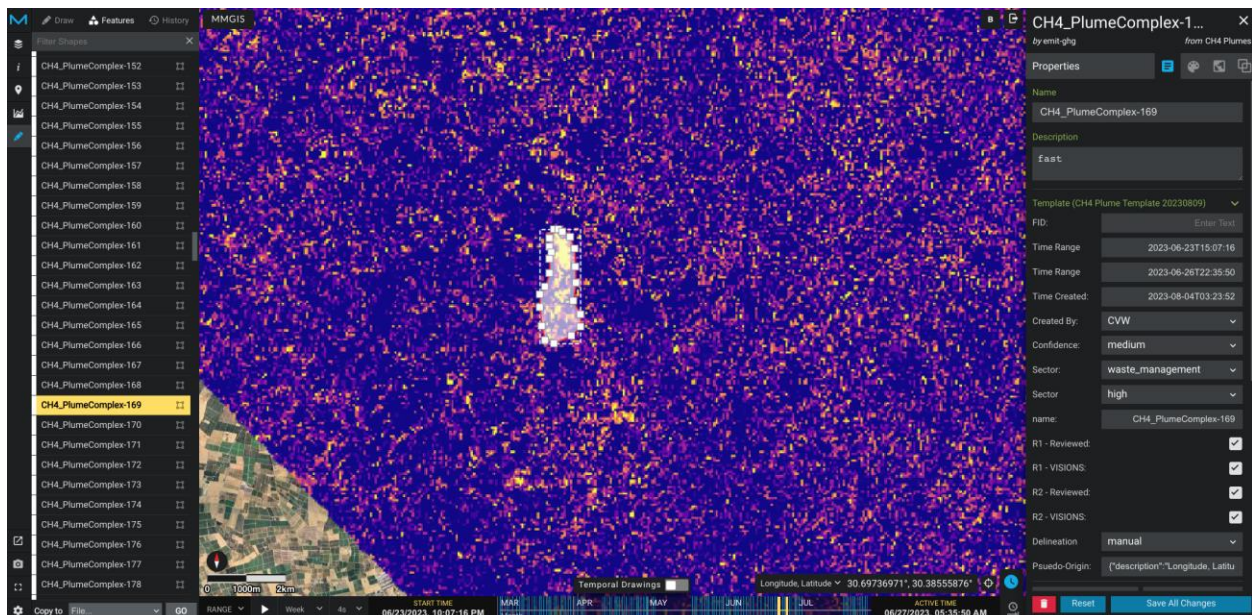


Figure 3: MMGIS interface for plume complex identification and review.

Reviewer 2 – Background Noise Consideration and Plume Shape Adjustment

Reviewer 2 examines each plume complex for quality. Two primary factors influence Reviewer 2’s decisions: 1) the background level of the scene, and 2) any temporal imagery of the same location. For the first criteria, Reviewer 2 determines if an enhancement can be reasonably separated from other bright values within the scene. If an identified plume complex sits at the same elevated rate as other noise within the scene, it is flagged and removed – notably, these enhancements may be real, but are not credibly separable from the background, and so removed in order to remain conservative in identification. Reviewer 2 also utilizes other imagery available through time as a secondary check. Spurious correlations with background materials tend to be persistent through time (given similar illumination conditions), while true plume complexes tend to drift due both to variable wind speed and direction and/or source emission rates. Areas with persistent enhancements in both shape and magnitude are very likely false positives and are removed. Areas where no plume complex is visible at one point and time but is distinctly visible in a second lend additional credibility to a complex. For plume complexes

tagged as ‘mystery’ – which do not show plausible infrastructure nearby, temporal variability or strong signatures which are substantially decorrelated from the surface are necessary for approval.

Reviewer 3 – Comprehensive Review

Prior to publishing a plume complex, a final comprehensive review is performed. Reviewer 3 first assesses the prior reviews by Reviewer 1 and 2 then performs an independent assessment of each plume complex candidate. Greenhouse gas enhancement maps are compared to EMIT RGB scenes to screen out any potential false positives associated with surface spectroscopy or atmospheric component like clouds. Enhancements are also compared with publicly available higher resolution true color imagery that provides additional spatial context. In some cases, an analysis of EMIT spectra are used to screen out surfaces that have strong spectral signatures in the shortwave infrared (for example calcium carbonate).

4.3 Practical Considerations

Computation of the adaptive matched filter is very quick relative to other steps in the EMIT SDS (e.g. L2A reflectance). The manual review process, however, is labor and time intensive. All code used for the matched filter enhancement is available at the EMIT GHG repository <https://github.com/emit-sds/emit-ghg>. The scientist review portal is based off of code available at <https://github.com/NASA-AMMOS/MMGIS/tree/development>, and interfaces with code in the EMIT GHG repository.

5. Output Data

Output data for the EMIT methane products are divided into several outputs, which are stored and delivered as a combination of COGs and GeoJSONs. As such, products are orthorectified with a variable numbers of rows and columns. Specifically the products are:

EMIT L2B Methane Enhancement Data (EMITL2BCH4ENH) and EMIT L2B Estimate Methane Plume Complexes (EMITL2BCH4PLM).

1. **Methane Enhancement Data** (EMITL2BCH4ENH), provided as a 1-band cloud optimized GeoTIFF with variable number of rows and columns, projected using WGS 84 (EPSG: 4326). This output contains the per-pixel estimated methane enhancement, with various cloud, water, and saturation masks in place to minimize false positives. Outputs are delivered on a scene-by-scene basis.
2. **Methane Plume Complexes** (EMITL2BCH4PLM), provided as a 1-band cloud optimized GeoTIFF with variable number of rows and columns, projected using WGS 84 (EPSG: 4326). This output contains the per-pixel estimated methane enhancement, cut to a specific plume complex. This product is not constrained to fall within a single scene.
3. **Methane Plume Complex Metadata** (EMITL2BCH4PLM), provided as a GeoJSON. This metadata file contains information including which scenes the plume complex came from, an outline of the plume complex, the coordinates of the maximum enhancement values, and the quantitative uncertainty of the CH₄ enhancement within the plume complex.

6. Calibration, uncertainty characterization and propagation, and validation

6.1 Uncertainty quantification

Uncertainty in the measured concentration field comes from two primary sources. First, the natural measurement noise of any physical instrument induces some variability in the measured matched filter results. We take the first noise source as a multivariate Gaussian distribution with zero mean and covariance C_{meas} . Second, background signals themselves have some variability; while in theory this “clutter” need not be Gaussian distributed, it is often effective to model them as an independent noise source with zero mean and covariance C_{bg} . These two sources are independent, so the covariances of the two noise sources add to form the total noise, $C = C_{meas} + C_{bg}$. To propagate this error through the matched filter retrieval, we recognize that the matched filter is simply a linear operator α applied to $(x - \mu)$:

$$MF = (x - \mu)^T \alpha, \quad \alpha = \frac{\{C^{-1}S\}}{S^T C^{-1}S}$$

This enables a simple analytical forward propagation of the noise distribution, transforming the noise distribution by that linear operator to yield the noise-equivalent change in methane:

$$Ne\Delta CH_4 \sim N(0, \alpha^T C \alpha)$$

This raises the question of how to estimate C . We have direct knowledge of C_{meas} from instrument radiometric models, but do not have access to C_{bg} . Moreover, the background clutter may change as a function of location and time. Fortunately, as in the construction of the matched filter itself, an unbiased estimator for the complete noise expression $Ne\Delta CH_4$ can be obtained by sample statistics of areas with no methane enhancement. Consequently, we estimate the sample covariance of the dataset outside all plume complexes within the set of scenes from which a given plume complex is identified. This empirical standard deviation indicates the uncertainty of the methane measurement throughout the scene.

In the ideal scenario, these values would be uniformly zero – in practice they tend to fall below 500 ppm m, though vary scene by scene. It is possible that unidentified plume complexes – true positives within a scene that are not ranked high enough in confidence to be determined as a plume complex – inflate this uncertainty estimate in some cases.

6.2 Validation efforts

EMIT validation activities are integral to continued demonstration of its greenhouse mapping capability. Future plans include utilizing the AVIRIS-3 instrument, an airborne equivalent of EMIT, to permit comparison of results obtained from space with EMIT. Comparison of EMIT greenhouse gas results with other satellite instruments (i.e., GHGSat, OCO-3) are also planned in addition to controlled release experiments.

7. Constraints and Limitations

No constraints or limitations are imposed on the EMIT methane products. All delivered data will have undergone quality control and should be considered valid, calibrated data up to the reported

uncertainties in input parameters. Unanticipated data corruption due to factors outside the modeling, if discovered, will be reported in peer reviewed literature and/or addenda to this ATBD.

8. Code Repository and References

All code used for the matched filter enhancement is available at the EMIT GHG repository <https://github.com/emit-sds/emit-ghg>. The scientist review portal is based on code available at <https://github.com/NASA-AMMOS/MMGIS/tree/development>, and interfaces with code in the EMIT GHG repository.

References

Thorpe, A.K., et al., Attribution of individual methane and carbon dioxide emission sources using EMIT observations from space, *Science Advances* (in press).

A. K. Thorpe, D. A. Roberts, E. S. Bradley, C. C. Funk, P. E. Dennison, I. Leifer, High resolution mapping of methane emissions from marine and terrestrial sources using a Cluster-Tuned Matched Filter technique and imaging spectrometry. *Remote Sensing of Environment*. 134, 305–318 (2013).

D. R. Thompson, I. Leifer, H. Bovensmann, M. Eastwood, M. Fladeland, C. Frankenberg, K. Gerilowski, R. O. Green, S. Kratwurst, T. Krings, B. Luna, A. K. Thorpe, Real-time remote detection and measurement for airborne imaging spectroscopy: A case study with methane. *Atmospheric Measurement Techniques*. 8, 4383–4397 (2015).

A. K. Thorpe, C. Frankenberg, D. R. Thompson, R. M. Duren, A. D. Aubrey, B. D. Bue, R. O. Green, K. Gerilowski, T. Krings, J. Borchardt, E. A. Kort, C. Sweeney, S. Conley, D. A. Roberts, P. E. Dennison, Airborne DOAS retrievals of methane, carbon dioxide, and water vapor concentrations at high spatial resolution: Application to AVIRIS-NG. *Atmospheric Measurement Techniques*. 10, 3833–3850 (2017).

A. K. Thorpe, C. Frankenberg, D. A. Roberts, Retrieval techniques for airborne imaging of methane concentrations using high spatial and moderate spectral resolution: Application to AVIRIS. *Atmospheric Measurement Techniques*. 7, 491–506 (2014).

R. M. Duren, A. K. Thorpe, K. T. Foster, T. Rafiq, F. M. Hopkins, V. Yadav, B. D. Bue, D. R. Thompson, S. Conley, N. K. Colombi, C. Frankenberg, I. B. McCubbin, M. L. Eastwood, M. Falk, J. D. Herner, B. E. Croes, R. O. Green, C. E. Miller, California's methane super-emitters. *Nature*. 575, 180–184 (2019).

C. Frankenberg, A. K. Thorpe, D. R. Thompson, G. Hulley, E. A. Kort, N. Vance, J. Borchardt, T. Krings, K. Gerilowski, C. Sweeney, S. Conley, B. D. Bue, A. D. Aubrey, S. Hook, R. O. Green, Airborne methane remote measurements reveal heavytail flux distribution in Four Corners region. *Proceedings of the National Academy of Sciences of the United States of America*. 113, 9734–9739 (2016).

D. H. Cusworth, A. K. Thorpe, A. K. Ayasse, D. Stepp, J. Heckler, G. P. Asner, C. E. Miller, V. Yadav, J. W. Chapman, M. L. Eastwood, R. O. Green, B. Hmiel, D. R. Lyon, R. M. Duren,

Strong methane point sources contribute a disproportionate fraction of total emissions across multiple basins in the United States. *Proceedings of the National Academy of Sciences*. 119, e2202338119 (2022).